Training the Archive

A Research Project on the Possibilities and Risks of Artificial Intelligence for Curatorial Practice in Museums / Ein museales Forschungsprojekt zu den Möglichkeiten und Risiken von Künstlicher Intelligenz für die kuratorische Praxis

with contributions by / mit Beiträgen von Hannes Bajohr, Nick Couldry, Elisa Giardina Papa, Adam Harvey, Mar Hicks, Mél Hogan, Moritz Ibing et al., Maya Indira Ganesh, Ulises Mejias, Matteo Pasquinelli, Gabriel Pereira, Anna Ridler, Alexa Steinbrück, Giulia Taurino & Magda Tyżlik-Carver

Training the Archive
Training the Archive
Training the Archive
Training the Archive
Training the Archive
Training the Archive
Training the Archive
Training the Archive
Training the Archive
Training the Archive
Training the Archive
Training the Archive
Training the Archive
Training the Archive
Training the Archive
Training the Archive
Training the Archive
Training the Archive
Training the Archive
Training the Archive
Training the Archive
Training the Archive

Training the Archive

Edited by / Herausgegeben von
Inke Arns, Eva Birkenstock, Dominik Bönisch & Francis Hunger
Ludwig Forum Aachen
Verlag der Buchhandlung Walther und Franz König, Cologne / Köln

Preface

Inke Arns & Eva Birkenstock

Training the Archive (2020–23) is a joint project between Ludwig Forum Aachen and HMKV Hartware Medien-KunstVerein, Dortmund, in cooperation with the Visual Computing Institute of RWTH Aachen University as a digital partner. With this publication, our joint research on the use of artificial intelligence (AI) in working with museum collections and curating exhibitions of contemporary art comes to a preliminary closure. What conclusions can be drawn after four years of reflection, production, rejection, discussion, but also criticism? On the one hand, we can look back on seven published scientific texts, four developed prototypes, ten video interviews, and an international conference with an accompanying archive exhibition at the Ludwig Forum Aachen. The inherent interdisciplinarity of the entire project enabled a continuous process of mutual negotiation, interdisciplinary communication and an intensive exchange of expertise and techniques from different cultural and research institutions. An opportunity to further develop the critical examination of digitality that is prevalent in all three institutions—not least in relation to working with collections and exhibitions—and to bring it into dialogue with our visitors and the teams involved in the project.

Prior to submitting the project application, both institutions addressed current digital transformation processes in their exhibition program as well as in their daily work. The concrete societal effects of technological 'advancements' in the form of discrimination due to algorithmic bias, exclusion through automated pattern recognition or precarious working conditions as the basis of Big Tech were already emerging at the time, and are now being debated by a broader public. *Training the Archive* was able to make an important contribution to the conscious reflection on these developments.

But why did we choose an analogue publication for the documentation of highly digital topics? To quote the media theorist Boris Groys: "A piece of writing remains stable through time; it promises trans-temporality and even immortality" (*e-flux Notes*, August 10, 2023). The book as a repository of knowledge and paper as a long-term archiving instrument were to grant our research project a temporal tenacity in the face of rapid technological change and the obsolescence with which it was concerned.

The project and this publication can be seen as an impetus to integrate the topic of AI into museums and exhibition venues on a permanent basis, and thus create a place for shared knowledge building to develop an informed approach to technology. The Kulturstiftung des Bundes' (German Federal Cultural Foundation) ambitious 'Digital Culture' program offered the ideal framework for implementing *Training the Archive* with its partners, particularly due to its long-term nature. Allowing for an unrestricted outcome and understanding any mistakes as potential were crucial to the successful implementation of our project. We hope that funding structures such as the 'Fonds Digital' will continue to provide cultural policy momentum in the future.

We would particularly like to thank the Kulturstiftung des Bundes for its generous support, our project partners, and all the people and organizations whose dedication made this publication possible: the authors, the publisher Verlag der Buchhandlung Walther und Franz König, and especially the graphic design studio Off Office, who once again developed a unique visual language and wonderful design, both for the project itself and for the book. We would also like to thank the entire project team—especially Dominik Bönisch, Francis Hunger, and all the local staff involved—for the insightful and exciting years.

With *Training the Archive,* we were able to create a trans-institutional space for the collective exploration of crucial modern-day issues, which will undoubtedly resonate in future exhibition or research projects at the participating institutions. We are grateful for this.

Vorwort

Inke Arns & Eva Birkenstock

Training the Archive (2020–23) ist ein Verbundprojekt des Ludwig Forum Aachen mit dem HMKV Hartware MedienKunstVerein, Dortmund und dem Visual Computing Institute der RWTH Aachen University als Digitalpartner. Unsere gemeinsame Forschung zum Einsatz von Künstlicher Intelligenz (KI) bei der Arbeit mit Sammlungen in Museen und beim Kuratieren von Ausstellungen für Gegenwartskunst kommt mit dieser Publikation zu einem vorläufigen Abschluss. Welches Fazit kann man nach vier Jahren des Nachdenkens, Produzierens, Verwerfens, Diskutierens, aber auch Kritisierens ziehen? Zum einen blicken wir zurück auf sieben veröffentlichte Fachtexte, vier entwickelte Prototypen, zehn Videointerviews und eine internationale Fachkonferenz mit begleitender Archivausstellung am Ludwig Forum Aachen. Die dem gesamten Projekt inhärente Interdisziplinarität ermöglichte einen kontinuierlichen Prozess des gegenseitigen Aushandelns, der disziplinübergreifenden Kommunikation sowie einen intensiven Austausch von Expertisen und Techniken unterschiedlicher Kultur- und Forschungseinrichtungen. Eine Chance, die in allen drei Institutionen vorherrschende kritische Auseinandersetzung mit Digitalität – nicht zuletzt in Bezug auf das Arbeiten mit Sammlungen und dem Ausstellungswesen – weiterzuentwickeln und in Dialog mit unseren Besucher*innen sowie mit den am Projekt beteiligten Teams zu bringen.

Beide Verbundinstitutionen verhandelten bereits im Vorfeld des Projektantrags die aktuellen digitalen Transformationsprozesse sowohl in ihrem Ausstellungsprogramm als auch in der täglichen Arbeit. Die konkreten gesellschaftlichen Auswirkungen des technischen ‚Fortschritts' in Form von Diskriminierung aufgrund algorithmischer Voreingenommenheit, Ausschlüssen durch automatisierte Mustererkennung oder prekärer Arbeitsverhältnisse als Grundlage von Big Tech bahnten sich damals bereits an und werden mittlerweile in einer breiteren Öffentlichkeit debattiert. Zu der bewussten Reflexion dieser Entwicklungen konnte *Training the Archive* einen wichtigen Beitrag leisten.

Doch warum entschieden wir uns für eine analoge Publikation bei der Dokumentation höchst digitaler Themen? Um es mit dem Medientheoretiker Boris Groys zu sagen: „A piece of writing remains stable through time; it promises trans-temporality and even immortality" (*e-flux Notes*, 10. August 2023). Das Buch als Wissensspeicher und Papier als Archivierungsinstrument sollten unserem Forschungsprojekt eine zeitliche Beharrlichkeit gegenüber dem raschen technologischen Wandel und seiner Obsoleszenz, mit denen es sich beschäftigt hat, verleihen.

Das Projekt sowie der vorliegende Band können als Anstoß verstanden werden, das Thema KI dauerhaft an Museen und in Ausstellungshäusern zu integrieren und so einen Ort für gemeinsamen Wissensaufbau zu schaffen, um einen verantwortungsvollen Umgang mit Technologien entwickeln zu können. Das ambitionierte Programm ‚Kultur Digital' der Kulturstiftung des Bundes bot gerade durch seine Langfristigkeit den idealen Rahmen, um *Training the Archive* mit seinen Partner*innen zu realisieren. Dabei eine Ergebnisoffenheit zu ermöglichen und etwaige Fehler als Potenziale zu begreifen, waren für die erfolgreiche Umsetzung unseres Projekts maßgeblich. Förderstrukturen wie der ‚Fonds Digital' wünschen wir uns auch in Zukunft als Impulse aus der Kulturpolitik.

So danken wir ganz besonders der Kulturstiftung des Bundes für die großzügige Förderung, unseren Projektpartner*innen sowie allen Personen und Instanzen, die mit ihrem unermüdlichen Einsatz diese Publikation ermöglicht haben: Allen Autor*innen, dem Verlag der Buchhandlung Walther und Franz König und besonders dem Grafikbüro Off Office, die erneut eine einzigartige Formsprache und wunderbare Gestaltung entwickelten, sowohl für das Projekt selbst als auch für die Publikation. Herzlich bedanken wir uns auch beim gesamten Projektteam – insbesondere bei Dominik Bönisch, Francis Hunger und allen involvierten Mitarbeiter*innen vor Ort – für die lehrreichen und spannenden Jahre.

Mit *Training the Archive* konnten wir einen transinstitutionellen Raum der gemeinsamen Verhandlung zentraler Themen unserer Gegenwart schaffen, der fraglos auch in zukünftigen Ausstellungs- oder Forschungsprojekten der beteiligten Häuser widerhallen wird. Dafür sind wir dankbar.

The research project *Training the Archive* aimed to apply artificial intelligence (AI) in the context of art and culture. The project placed particular emphasis on transferring current AI models to curatorial practice and questioning the extent to which emerging algorithmic systems were able to support the analysis and development of museum collections. Technology in this area developed rapidly during the project period (2020–23). Not only did machine learning methods change, but also social awareness of the topic increased significantly. The time has come, according to cultural managers Tobias Hochscherf and Martin Lätzel in their book *KI & Kultur: Chimäre oder Chance?* (*AI & Culture: Chimera or Chance?*) (2023, 9), in which people for the most part no longer understand their tools, can no longer read the data volumes. The constant technical changes in those tools available resulted in modifications to our project, which was prepared for processual alternations.

Training the Archive had the vision of creating a piece of software that, with the help of AI (whatever that might have meant at the beginning)—as the so-called 'Curator's Machine'—would partially or even fully automate the curating of art collections and exhibitions in museums. This application was not intended to follow the existing and restrictive search logics of keyword-oriented digitized archives and databases, but rather to enable the exploration of digital collections via special interfaces. Using pattern recognition could reveal connections between works of art that humans either find difficult to perceive or are able to understand only incompletely. The strengths of artificial 'intelligence' should thus become apparent when search results queried in large collections of image data no longer remained statically the same. Instead, search results would gradually adapt to the curators' personal interests based on their usage and the changing parameters.

These considerations originally referred to the digitized collection of the Ludwig Forum Aachen. There, the research question arose as to how to recognize and use thematic connections in large, unlabeled art collections in order to move beyond the concentration on canonical works or on public 'highlights.' Implementing this ambitious project in an interdisciplinary way necessitated additional project partners. The Visual Computing Institute at RWTH Aachen University contributed technical expertise, researched, and published articles on technological issues, advised the project team, and ultimately programmed the software application in iterative prototype phases. Furthermore, collaboration with HMKV Hartware MedienKunstVerein, Dortmund opened up an additional art- and media-critical perspective that resulted in extensive thesis papers—the working papers (page 106)—and the conducting of numerous video interviews with international experts. It can thus also be understood as a form of public outreach of *Training the Archive*. The aim of this working constellation was a model-like interweaving of content-critical research along with the actual development of a software prototype—The Curator's Machine—which is currently in use at the Ludwig Forum Aachen as a 'proof of concept,' and is available to all interested parties as an open-source application (page 99).

In the course of the prototyping, it became clear that a Curator's Machine would not replace the curators or automate their work but could, as a human-machine assemblage, serve as a further tool in the curatorial toolbox. This insight resulted from a study we initiated that analyzed curatorial practice in terms of software use and the conceptual procedures applied. The establishment of empirical bases met with changing technological framework conditions. At the beginning of the project, the main resource was pre-trained object recognition systems known as 'convolutional neuronal networks,' which were trained using the ImageNet image data collection. With regard to art, in particular, the object recognition of these ImageNet networks was reliable only to a limited extent. Objects were often recognized as televisions or windows, although they were in fact picture frames, so that *Training the Archive* concentrated on another aspect. Rather than trying to detect objects, the aim became to sort similar images in relation to each other and to facilitate curatorial access via visual similarities. During this process, the project team developed the first prototypes by means of which it became possible to adapt the automatically assembled image similarities by using additional annotations in such a way that these similarities could be customized and adjusted to the curatorial experts' knowledge. Initial tests were successful, but it became clear that such a modification would require a very large number of manual labels—impractical for everyday curatorial work.

During the same time period, a new pattern recognition technology, CLIP (developed by OpenAI), became publicly available. CLIP combines large language models with image processing networks into a shared, statistical text-image embedding space. Images could now be automatically addressed through text (and theoretically vice versa), an approach referred to as 'multimodal.' For *Training the Archive,* this approach meant a change in thinking. From then on, image data collections could not only be grouped automatically based on visual similarities, but also searched for and explored with the help of text queries—the 'prompts'—in particular with those keywords that did not have to be explicitly tagged for the digital images (which potentially reduces the resources required for manual data entry). This made associative searches in the datasets possible. Searches provided a (statistical) answer to queries in the form of image suggestions, e.g., for "painting about deep despair" or "drawing with red blossoms." Building on previous research on influencing pre-trained pattern

recognition networks, the Visual Computing Institute developed a so-called 'adapter' for CLIP that enabled curatorial customizations to be incorporated into the software in such a way that search sets adapted in real time. As a result, the Curator's Machine has become anything but a mere correlation-producing machine: it is a curatorial software tool for searching, filtering, and rearranging large amounts of data—in other words, a knowledge tool. At first, the software looks like a familiar search interface, with a search bar and relevant suggestions (not least thanks to our workshops with UI and UX designers). But on the canvas, users can adjust the order of the objects found, which interactively changes the text-image embedding on which the search results are based. This is modified in accordance with the curator's input behavior in order to retrieve the intentionally sought after topics of focus in the data feed. Those who find this explanation overly technical are invited to try it out for themselves (page 98).

In addition to the technological modifications mentioned above, there were also changes in the content of the project. While the research team began with an open mind about the term 'artificial intelligence,' the limitations and problems within the corresponding terminology and in terms of development economics became apparent as the project progressed. In response, *Training the Archive* increasingly distanced itself from the hype around AI and began to review its own language. Were terms like 'neural,' 'thinking,' 'learning,' 'creativity,' or 'intelligence' still appropriate? What is the function of using metaphors that humanize the subject matter in this field of research? Another instance of our distancing was related to the bias of the training data itself and the social and artistic impact this entails. In addition to the seven working papers (page 106) published on the online platform Zenodo, ten video interviews with experts that can be found on YouTube delved into questions of image production, artificial 'intelligence,' and the curatorial and artistic approach to such machines, multimodal or generative processes. A conference on November 17 and 18, 2022 at the Ludwig Forum Aachen with contributions on 'Art & Algorithms' further highlighted the current developments and debates being held (page 116). *Training the Archive,* with its various deployments, can serve as a model for digitalization projects in the field of art, as it combined an increasingly critical reflection with the innovative self-production of an open-source software tool, thus carefully creating new knowledge in overlapping fields: in computer science, critical AI studies, and curatorial studies, which in turn flowed into the discourse of art and cultural studies, digital art history, and digital humanities.

This publication completes our examination and discussion of the topic and the research project. It consists of a range of content: On the one hand, the edited transcripts of all conducted video interviews and a collection of materials. On the other hand, texts by Mél Hogan, Giulia Taurino, Moritz Ibing et al., and Hannes Bajohr were written especially for the volume and form the thematic core, which consists of four focuses.

1.) Media theorist Mél Hogan provides a 'critique of today' with the essay *Artificial Intelligence Is a Hot Mess* (page 33). The reference to 'hot mess' addresses the environmental impact of the computing and data centers that provide services for the training of large AI models, such as CLIP. It pointedly discusses the interwoven nature of Big Tech in the framework of the acute social problem of climate change, while criticizing the ideological superstructure of so-called 'longtermism' that attempts to downplay these developments. Accompanying this article are two interview transcripts: Artist and activist Adam Harvey is interviewed about his engagement with surveillance and training datasets on *Face Recognition Datasets* (page 17). Harvey has documented and discussed the often problematic scraping of training data in recent years. The second interview is entitled *How the Image Collection ImageNet Re-Constructs Reality,* and discusses the historical and functional aspects of pre-trained image recognition networks with Gabriel Pereira (page 25). Pereira and Bruno Moreschi artistically explored their notion of a 'critique of today' using the collection of the Van Abbemuseum in Eindhoven, Netherlands as an example.

2.) Addressing the second focus with the text *In Search of Boundary Objects: A Taxonomy-Based Approach to Algorithmic Co-curation in Archival Collections* (page 71), Giulia Taurino looks at the 'future of digital museums.' In her approach to the algorithm as curator, Taurino discusses a core concern of *Training the Archive,* as the different modes of human-machine co-curation are at the forefront of her argument. She explains the limitations of modern computer vision models, and elaborates on the not unproblematic relationship to the structured vocabulary of existing museum collections. This text is accompanied by the interview on *Curating Data* with Magda Tyżlik-Carver (page 55). The curator and media theorist outlines networks of relationships between humans and the non-human—algorithms, bots, software and computer infrastructures—as a future outlook on curating. In *Cultural Critique and Artificial Intelligence,* theorist, curator, and lecturer Maya Indira Ganesh discusses authoritative artistic works that have set the tone in the humanities' engagement with AI (page 63). In this work, curating art exhibitions via AI emerges as a form of knowledge production that succeeds in calling up aspects that would not be possible in academic texts alone.

As an interlude, the publication brings together materials that accompanied the research process of *Training the Archive:* sketches, screenshots and prototypes that are intended to illustrate the processual and iterative nature of the project (page 97).

3.) The third focus, 'technological state of the art,' is based on the paper *Localized Latent Updates for Fine-Tuning Vision-Language Models* by computer scientists Moritz Ibing, Isaak Lim, and Leif Kobbelt from the Visual Computing Institute at RWTH Aachen University (page 145). The authors discuss how the developed adapter can fine-tune multimodal models such as CLIP. The adapter changes the internal statistical distributions in the embedding space locally, but without destabilizing the pre-trained system, so that individual adjustments can be made in real time—a procedure that can also be used beyond the Curator's Machine. Philosopher Matteo Pasquinelli takes *A Larger Perspective on Artificial Intelligence* in his interview (page 121). He describes AI as the automation of manual, mental, visual, and organizational work and, based on this analysis, points to the diverse genealogies of so-called artificial 'intelligence.' In the interview on *Modes of Representation of AI and How to Teach It,* artist and programmer Alexa Steinbrück talks about the modes of representation of AI and what an easy-to-learn education could look like (page 129). It addresses myths and metaphors surrounding AI as well as approaches to an artistic doctrine that explores the latest technologies both critically and experimentally. Under the title *From GANs to Stable Diffusion. On Artistic Collaboration with Generative Algorithms,* artist Anna Ridler talks about her own experimental engagement with AI (page 137). Strategies of conceptual art allow her to take up current technologies without succumbing to the extractivism of image generators like Midjourney or Stable Diffusion.

4.) The final focus 'critique of tomorrow' unfolds with Hannes Bajohr's essay *Whoever Controls Language Models Controls Politics* (page 189). The author, philosopher, and literary scholar discusses the use and training of large language models, which are changing our everyday life significantly but threaten the democratic order due to the monopolistic position of a few providers. This criticism applies similarly to multimodal text-image models. Bajohr, like Mél Hogan, embeds his argument in an objection to longtermism. He sees the socialization of language models as a last resort, arguing that these should be in the hands of the self-governing public as a public good or service. Sociologists Nick Couldry and Ulises Mejias come to a similar conclusion with their thesis on *Data Colonialism* (page 165). Their interview discusses data as an abstraction of life and precisely describes how it is extracted and colonially exploited. Like Bajohr, they call for the production of the data on which AI is based to be democratic and oriented towards the public good. Artist Elisa Giardina Papa discusses *The Myth of Universality, Transparency, and Truth for What Regards Emotion in Artificial Intelligence* on page 173. She worked as a precarious clickworker herself, labeling data to detect emotions in portraits of faces. She reacted to this new, exploitative form of data work and to the pseudo-science of emotion detection in portraits with a series of artistic works. The interview *The Politics of Artificial Intelligence and Algorithmic Bias* concludes the 'critique of tomorrow' (page 181). Historian Mar Hicks relates current developments to the history of computer science. With great clarity, she discusses the gendered power structures behind computing not only as a technology, but also as a cultural technique.

Einführung

Dominik Bönisch & Francis Hunger

Das Forschungsprojekt *Training the Archive* nahm es sich zum Ziel, Künstliche Intelligenz (KI) im Kontext von Kunst und Kultur anzuwenden. Besondere Schwerpunkte lagen auf der Übertragung aktueller KI-Modelle auf die kuratorische Praxis und auf der Befragung, inwieweit die aufkommenden algorithmischen Systeme eine museale Sammlungserschließung unterstützen könnten. Im Projektzeitraum (2020–23) ist eine rasante Entwicklung der Technologie zu verzeichnen, in der sich die maschinellen Lernverfahren veränderten, aber auch die gesellschaftliche Aufmerksamkeit für das Thema deutlich anstieg. Es sei damit eine Zeit angebrochen, so meinen die Kulturmanager Tobias Hochscherf und Martin Lätzel in ihrem Buch *KI & Kultur: Chimäre oder Chance?* (2023, 9), in der die Menschen ihre Werkzeuge nicht mehr verstehen und die Datenmengen nicht mehr überblicken können. Auf die stetigen technischen Veränderungen der verfügbaren Tools reagierten wir reflexiv in der Ausgestaltung unseres Projekts, das auf prozessuale Wechsel vorbereitet war.

Die Vision von *Training the Archive* bildete eine Software, die mithilfe von KI (was auch immer das zu Beginn heißen mochte) – als sogenannte ‚Curator's Machine' – das Kuratieren von Kunstsammlungen und Ausstellungen im Museum teil- oder gar vollautomatisiert. Diese Anwendung sollte nicht den verschlagworteten und eher statischen Suchlogiken bisheriger Archive und Datenbanken folgen, sondern vielmehr über spezielle Interfaces eine Exploration von massenhaft digitalisierten Beständen komfortabler und schneller gestalten. Mittels Mustererkennung könnten Zusammenhänge und Verbindungen zwischen Kunstwerken offenbart werden, die der Mensch entweder schwer wahrnehmen oder sich nur unvollständig erschließen kann. Die Stärken der sogenannten Künstlichen ‚Intelligenz' sollten sich also zeigen, wenn in großen Bilddaten-Sammlungen Suchergebnisse nicht generisch gleichblieben. Vielmehr würden sie sich aufgrund der Nutzung sowie sich ändernder Parameter anpassen und sich somit auf die persönlichen Interessen der Kurator*innen einstellen.

Diese Überlegungen bezogen sich ursprünglich auf die digitalisierte Sammlung des Projektträgers, des Ludwig Forum Aachen. Dort stellte sich die Forschungsfrage, wie in großen, ungelabelten Kunstsammlungen thematische Zusammenhänge erkannt und genutzt werden können, und das jenseits der Konzentration auf kanonische Werke oder auf Publikums-‚Highlights'. Um das ambitionierte Vorhaben interdisziplinär umsetzen zu können, bedurfte es weiterer Projektpartner*innen. Das Visual Computing Institute der RWTH Aachen University brachte die technische Expertise ein, forschte und publizierte zu technologischen Fragestellungen, beriet das Projektteam und programmierte schließlich die Softwareanwendung in iterativen Prototypen-Phasen. Des Weiteren eröffnete die Zusammenarbeit mit dem HMKV Hartware MedienKunstVerein, Dortmund eine zusätzliche kunst- und medienkritische Perspektive, die in umfassenden Thesenpapieren – den Working Papers (Seite 108) – und in zahlreichen Videointerviews mit internationalen Expert*innen mündete und somit auch als Public Outreach von *Training the Archive* verstanden werden kann. Ziel dieser Arbeitskonstellation war eine modellhafte Verschränkung von inhaltlich-kritischer Forschung mit der konkreten Entwicklung eines Softwareprototypen – der Curator's Machine – die im Sinne eines ‚Proof of Concept' am Ludwig Forum Aachen im Einsatz ist und für alle Interessierten als Open-Source-Anwendung zur Verfügung steht (Seite 103).

Im Laufe des Prototypings wurde deutlich, dass eine Curator's Machine die Kurator*innen nicht ersetzen oder deren Arbeit automatisieren würde, sondern als Mensch-Maschine-Assemblage ein weiteres Werkzeug im kuratorischen Baukasten darstellen könnte. Zu dieser Einsicht führte eine im Projekt initiierte Studie, welche die kuratorische Praxis im Hinblick auf den Softwaregebrauch und auf angewandte konzeptionelle Verfahren analysierte. Die Erhebung empirischer Grundlagen traf auf sich ändernde technologische Rahmenbedingungen. Zu Projektbeginn waren vor allem vortrainierte Objekterkennungssysteme – sogenannte ‚Convolutional Neural Networks' – verfügbar, die anhand der Bilddatensammlung ImageNet trainiert wurden. Gerade in Bezug auf Kunst war die Objekterkennung dieser ImageNet-Netzwerke nur eingeschränkt zuverlässig – oft wurden Fernseher oder Fenster erkannt, wo es eigentlich um Bilderrahmen ging, sodass sich *Training the Archive* auf einen anderen Aspekt konzentrierte: Statt einzelne Objekte erkennen zu wollen, sollten einander ähnliche Bilder zueinander sortiert und ein kuratorischer Zugang über visuelle Ähnlichkeiten ermöglicht werden. In diesem Zuge entwickelten wir die ersten Prototypen, über die es möglich wurde, automatisch erstellte Bildähnlichkeiten durch zusätzliche Annotationen so anzupassen, dass sie personalisiert und auf das kuratorische Expert*innenwissen hin abgestimmt werden konnten. Erste Tests verliefen erfolgreich, jedoch zeigte sich, dass für eine derartige Anpassung eine sehr hohe Anzahl zusätzlicher, manueller Verknüpfungen zwischen Kunstwerken erforderlich sein würde – im kuratorischen Alltag unpraktisch.

Im gleichen Zeitraum wurde mit CLIP (von der amerikanischen Firma OpenAI entwickelt) eine neue Technologie für Mustererkennung öffentlich. Diese verbindet große Sprachmodelle mit Bilderverarbeitungsnetzwerken zu einem gemeinsamen, statistischen Text-Bild-Raum: Bilder konnten nunmehr automatisiert durch Text adressiert werden (und theoretisch umgekehrt), ein Ansatz, der als ‚multimodal' bezeichnet

wird. Für *Training the Archive* führte dieser zu einem Umdenken: Nunmehr konnten Bilddatensammlungen nicht allein anhand visueller Ähnlichkeiten automatisiert gruppiert, sondern auch mithilfe von Texteingaben – den sogenannten ‚Prompts' – durchsucht und erkundet werden. Die eingegebenen Stichwörter müssen dabei nicht vorab explizit als Beschreibungen für die Digitalisate verschlagwortet werden, was potenziell Ressourcen bei einer manuellen Erfassung einspart. Mit dem Einsatz multimodaler Modelle sind demnach assoziative Suchvorgänge in den Datenbeständen möglich geworden, die auf Anfragen wie etwa „Gemälde über tiefe Verzweiflung" oder „Zeichnung mit roten Blüten" eine Antwort in Form von statistisch sortierten Bildvorschlägen liefern.

Aufsetzend auf unserer vorhergehenden Forschung zur Beeinflussung vortrainierter Mustererkennungsnetzwerke entwickelte das Visual Computing Institute einen ‚Adapter' für CLIP, der es erlaubt, kuratorische Personalisierungen so in die Software einzubinden, dass sich die Suchmengen in Echtzeit anpassten. Im Ergebnis ist aus der Curator's Machine alles andere als eine reine Korrelation-produzierende Maschine entstanden: Sie ist ein kuratorisches Softwarewerkzeug zum Durchsuchen, Filtern und Neuanordnen großer Datenmengen – mithin ein Wissenswerkzeug. Das Interface der Software mutet dabei zunächst vertraut an, mit Suchfeld und passenden Vorschlägen (nicht zuletzt dank unserer Workshops mit UI- und UX-Designer*innen). Doch verändert die individuelle Anordnung der gefundenen Objekte, die die Nutzer*innen im Arbeitsbereich vornehmen können, interaktiv den statistischen Text-Bild-Raum, auf dem die ausgegebenen Suchergebnisse basieren. Dieser wird sinngemäß auf das Eingabeverhalten der Kurator*innen hin gekrümmt, um die intentional gesuchten Schwerpunkte auch wirklich in den eingespeisten Daten abrufen zu können. Wem das zu technisch erscheint, ist eingeladen, es selbst auszuprobieren (Seite 102).

Neben den erwähnten technologischen Wechseln kamen auch inhaltliche Veränderungen im Projekt auf. Widmete sich das Forscher*innenteam zu Beginn dem Begriff der ‚Künstlichen Intelligenz' noch unvoreingenommen, traten im Verlauf dessen Begrenzungen und die Probleme innerhalb der Terminologie und der Entwicklungsökonomie deutlich zutage. Als Reaktion distanzierte sich *Training the Archive* zunehmend von dem Hype um KI und begann, die eigene Sprache zu überprüfen. Waren Begriffe wie ‚neuronal', ‚Denken', ‚Lernen', ‚Kreativität' oder ‚Intelligenz' weiterhin angemessen? Welche Funktion kommt der Verwendung von Metaphern in diesem Forschungsbereich zu, die einen Gegenstand, ein Ding vermenschlichen sollen? Eine weitere Distanznahme bezogen wir auf den Bias der Trainingsdaten und darauf, welche gesellschaftlichen und künstlerischen Auswirkungen dieser nach sich zieht. Neben den auf der Online-Plattform Zenodo veröffentlichten sieben Working Papers (Seite 108) vertieften zusätzlich zehn Videointerviews mit Expert*innen auf YouTube die Fragen der Bildproduktion, der Künstlichen ‚Intelligenz', des kuratorischen sowie künstlerischen Umgangs mit multimodalen oder generativen Verfahren. Eine Konferenz am 17. und 18. November 2022 am Ludwig Forum Aachen mit Beiträgen zum Thema ‚Kunst & Algorithmen' machten die aktuellen Entwicklungen und geführten Debatten weiter deutlich (Seite 118). *Training the Archive* kann mit seinen verschiedenen Einsätzen als Modell für Digitalisierungsprojekte im Kunstfeld dienen, denn es vereinte eine zunehmend kritische Reflexion mit der innovativen Eigenproduktion eines Open-Source-Softwarewerkzeugs und schuf somit sorgsam neues Wissen in sich überlagernden Feldern: in der Informatik, in den Critical AI Studies und in den Curatorial Studies, die wiederum in den Diskurs der Kunst- und Kulturwissenschaften, der digitalen Kunstgeschichte bzw. der Digital Humanities hin ausstrahlten.

Die vorliegende Publikation fasst unsere fachliche Auseinandersetzung im Forschungsprozess zusammen. Sie besteht aus mehreren Inhalten: zum einen aus den gekürzten Transkripten aller geführten Videointerviews und einer Materialsammlung, zum anderen aus eigens für den Band entstandenen Texten von Mél Hogan, Giulia Taurino, Moritz Ibing et al. und Hannes Bajohr, die den thematischen Kern mit vier Schwerpunkten bilden.

1.) Die Medientheoretikerin Mél Hogan leistet eine ‚Kritik am Heute' mit ihrem Essay *Künstliche Intelligenz als ‚Hot Mess'* (Seite 43). Mit dem Verweis auf ‚Hot Mess' adressiert sie die Umweltauswirkungen jener Rechen- und Datenzentren, welche Leistungen für das Training großer KI-Modelle, wie auch CLIP, zur Verfügung stellen. Sie thematisiert eindringlich das Verwobensein von Big Tech in das akute gesellschaftliche Problem des Klimawandels und kritisiert gleichzeitig den ideologischen Überbau des ‚Longtermismus', der diese Entwicklungen herunterzuspielen versucht. Diesem Beitrag beigestellt sind zwei Interviewtranskripte: Der Künstler und Aktivist Adam Harvey wird im Gespräch über *Datensätze zur Gesichtserkennung* zu seiner Auseinandersetzung mit Überwachung und dafür vorgesehenen Trainingsdatensätzen befragt (Seite 21). Harvey hatte in den letzten Jahren die oft problematische Herkunft von Trainingsdaten dokumentiert und diskutiert. Ein weiteres Interview führt aus, *Wie die Bilddatensammlung ImageNet Wirklichkeit (re-)konstruiert,* und Gabriel Pereira erläutert die geschichtlichen und funktionalen Aspekte vortrainierter Bilderkennungsnetzwerke (Seite 29). Eine ‚Kritik am Heute' übten Pereira und sein Kollege Bruno Moreschi dabei in künstlerischer Weise mit der Arbeit *Recoding Art* (2019) am Beispiel der Sammlung des Van Abbemuseums in Eindhoven.

2.) Giulia Taurino gibt im zweiten Schwerpunkt einen Blick auf die ‚Zukunft digitaler Museen' mit dem Text *Auf der Suche nach Grenzobjekten: Ein taxonomiebasierter Ansatz zur algorithmischen Co-Kuratierung in Archivsammlungen* (Seite 83). Mit ihrer Diskussion über den Algorithmus als Kurator*in erörtert Taurino ein Kernanliegen von *Training the Archive,* denn die verschiedenen Modi der human-maschinellen Co-Kuration stehen im Vordergrund ihrer Argumentation. Sie erläutert die Grenzen moderner Computer-Vision-Modelle und arbeitet das nicht unproblematische Verhältnis zum strukturierten Vokabular bestehender Museumssammlungen heraus. Diesem Text ist das Interviewtranskript über *Das Kuratieren von Daten* mit Magda Tyżlik-Carver parallelgeführt (Seite 59). Die Kuratorin und Medientheoretikerin entwirft als Zukunftsausblick auf das Kuratieren Beziehungsgeflechte zwischen Mensch und dem Nicht-Menschlichen: Algorithmen, Bots, Software und Rechnerinfrastrukturen. Die Theoretikerin, Kuratorin und Lehrbeauftragte Maya Indira Ganesh diskutiert in ihrem Gespräch zu *Kulturkritik und Künstliche Intelligenz* maßgebliche künstlerische Arbeiten, die in der geisteswissenschaftlichen Auseinandersetzung mit KI Akzente gesetzt haben (Seite 67). Das Kuratieren von Kunstausstellungen über KI tritt hier als eine Figur der Wissensproduktion auf, der es gelingt, andere Aspekte aufzurufen, als es in akademischen Texten allein möglich ist.

Als Zwischenspiel versammelt die vorliegende Publikation Materialien, die den Forschungsprozess von *Training the Archive* begleiteten: Skizzen, Screenshots und Prototypen, welche die Prozesshaftigkeit des Vorhabens verdeutlichen sollen (Seite 97).

3.) Der dritte Schwerpunkt ‚Technologische Entwicklung' fußt auf dem Aufsatz *Lokalisierte latente Updates für die Feinabstimmung von Vision-Language-Modellen* der Informatiker Moritz Ibing, Isaak Lim und Leif Kobbelt vom Visual Computing Institute der RWTH Aachen University (Seite 155). Die Autoren diskutieren, wie der entwickelte Adapter multimodale Modelle wie CLIP feinabstimmen kann. Der Adapter ändert die internen statistischen Verteilungen im Einbettungsraum lokal, ohne dabei das vortrainierte System zu destabilisieren, sodass individuelle Anpassungen in Echtzeit ermöglicht werden – ein Verfahren, welches auch jenseits der Curator's Machine in der Machine-Learning-Community Verwendung finden kann. Der Philosoph Matteo Pasquinelli stellt in seinem Interview *Künstliche Intelligenz in breiterer Perspektive* auf (Seite 125). Er beschreibt KI als Automatisierung von manueller, geistiger, visueller und organisatorischer Arbeit und verweist ausgehend von dieser Analyse auf die vielfältigen Genealogien von sogenannter Künstlicher ‚Intelligenz'. *Zu den Repräsentationsweisen Künstlicher Intelligenz und wie eine künstlerische Lehre aussehen kann* spricht die Künstlerin und Programmiererin Alexa Steinbrück (Seite 133). Sie adressiert Mythen und Metaphern um KI und Ansätze einer künstlerischen Lehre, die die neusten Technologien kritisch und gleichzeitig experimentell erforschen. Unter dem Interview-Titel *Von GANs bis Stable Diffusion. Über die künstlerische Zusammenarbeit mit generativen Algorithmen* reflektiert die Künstlerin Anna Ridler ihre ganz eigene experimentelle Auseinandersetzung mit KI (Seite 141). Strategien der Konzeptkunst erlauben es ihr, aktuelle Technologien aufzugreifen, ohne dem Extraktivismus von Bildgeneratoren wie Midjourney oder Stable Diffusion zu erliegen.

4.) Der letzte Schwerpunkt ‚Kritik am Morgen' entfaltet sich mit Hannes Bajohrs Text *Wer die Sprachmodelle beherrscht, beherrscht auch die Politik* (Seite 197). Der Schriftsteller, Philosoph und Literaturwissenschaftler bespricht den Einsatz und das Training großer Sprachmodelle, welche nicht nur unseren Alltag maßgeblich verändern, sondern auch durch die Monopolstellung weniger Anbieter*innen die demokratische Ordnung bedrohen. Diese Kritik trifft in ähnlicher Weise für multimodale Text-Bild-Modelle zu. Bajohr nimmt mit seiner Argumentation – wie zuvor Mél Hogan – Anstoß am Longtermismus. Als letztes Mittel sieht er die Vergesellschaftung der Sprachmodelle und führt aus, dass diese als öffentliches Gut oder Dienstleistung in den Händen der selbstverwalteten Öffentlichkeit liegen sollten. Zu einer ähnlichen Einsicht kommen die Soziologen Nick Couldry und Ulises Mejias mit ihrer These des *Datenkolonialismus* (Seite 169). Ihr Interview diskutiert Daten als Abstraktion des Lebens und beschreibt präzise, wie diese extrahiert und kolonial ausgebeutet werden. Wie Bajohr fordern sie eine demokratische und gemeinwohlorientierte Produktion jener Daten, welche der KI zugrunde liegen. Die Künstlerin Elisa Giardina Papa diskutiert ab Seite 177 ihren Beitrag *Emotionen und der Mythos von Universalität, Transparenz und Wahrheit durch Künstliche Intelligenz*. Sie arbeitete selbst als prekäre Clickworkerin und labelte Daten zur Erkennung von Emotionen in Porträts. Auf diese neue, ausbeuterische Form der Datenarbeit und auf die Pseudo-Wissenschaft der Emotionsdetektion reagierte sie mit einer Reihe künstlerischer Arbeiten. Das Videointerview *Künstliche Intelligenz und algorithmischer Bias als politisches Feld* schließt die ‚Kritik am Morgen' ab (Seite 185). Die Historikerin Mar Hicks setzt aktuelle Entwicklungen mit der Geschichte der Informatik in Beziehung. Mit großer Klarheit erörtert sie die geschlechtsspezifischen Machtstrukturen, die hinter dem Computing nicht nur als Technologie, sondern auch als Kulturtechnik stehen.

In Conversation with
Adam Harvey on “Face Recognition Datasets”

The interview discusses Adam Harvey’s critical examination of machine learning training datasets. He states: “I don’t think it’s possible to destroy the complete existing face detection computer vision infrastructure, nor do I want to. But it is possible to limit its growth and limit its dangerous potential.”

Transcript of the interview with Adam Harvey, conducted by Francis Hunger on 2021–05–14.

Hunger: Hi Adam. Large-scale image collections emerged around the mid-2010s for training artificial weighted networks for pattern recognition. Most known and critically discussed is ImageNet, which is used for object detection and classification. Your project *MegaPixels* (2017–20) does not look into object detection, but into face recognition and the respective training datasets. Before talking about *MegaPixels,* could you tell me a bit more about your preceding project called *CV Dazzle* (2010), and how *MegaPixels* emerged from that?

Harvey: *MegaPixels,* which is now called *Exposing.ai* (2021), is a progression from *CV Dazzle,* which is from eleven years ago now. The connection is slightly complex, but if you allow me to rant for a while, I think I can explain it. It's that *CV Dazzle* or Computer Vision Dazzle, with 'Dazzle' referring to a World War One camouflage style, was developed as a counter-surveillance project, with dual objectives of proving that computer vision camouflage is possible and that the camouflage can be done in an aesthetic way. It's a functional aesthetic. There was an additional discussion here about how camouflage is an evolving strategy and the goal of *CV Dazzle* is to minimize one's appearance in the way you're being observed, analyzed, or recognized by biometric analysis software. *CV Dazzle* targeted face detection. And face detection is a form of object detection where the face is the object. And if you can break face detection then you break face recognition. So, the goal of *CV Dazzle* is simply to break that image analysis pipeline at the easiest point. And I did that by targeting a very specific algorithm called the Viola Jones Algorithm and Haar Cascade. *CV Dazzle* worked by targeting this very specific algorithm that was deployed on a multitude of surveillance infrastructure devices. As I was posting more examples of *CV Dazzle,* they were showing up in scientific research papers. And researchers were looking at ways to make sure face recognition worked when people were wearing a lot of makeup, and then an image from *CV Dazzle* was included in that research paper. Now this was really interesting to me, because I did it primarily as an art project, as a hacker project to find the vulnerability, exploit that, and then promote it in the hopes that people would adopt this strategy and adapt it to their own means. After the Viola Jones Haar Cascade algorithm a rather new technique emerged: the 'neural networks.' What you see happening is, every time that you try to attack a neural network, you're not really attacking it. An adversarial exploit is not so much an adversarial maneuver as a protagonistic contribution to the network. The strategy from about 2015 onwards, and therefore from *CV Dazzle,* flipped—and so did my thinking. *MegaPixels* is an evolution of the *CV Dazzle* project, but looks at attacking the supply chains, instead of attacking the algorithm directly. If you are able to limit the amount of data that's going into a face detection or a face recognition algorithm, you limit the power to the point that if you don't have any data, you don't have an algorithm and you can't do face recognition. If you begin to remove that data you begin to remove the capability of neural networks. So, my intention with *MegaPixels* is to try to thwart the growth. I don't think it's possible to destroy the complete existing face detection computer vision infrastructure, nor do I want to. But it is possible to limit its growth and limit its dangerous potential.

Hunger: The Brainwash Café dataset and the MS-CELEB dataset were some of the first datasets which you and Jules LaPlace researched for *MegaPixels*. So, how did you come across these datasets and what is so special about them?

Harvey: It coincides with ImageNet that researchers began to adapt a strategy that was first popularized to the dataset called Labeled Faces in the Wild (LFW), which was released in 2007. At that time the economics of collecting data changed. You didn't have to pay people, you didn't have to offer a consent form, because there was so much data and so few rules about collecting it that researchers realized they could go online and scrape as much data as they wanted. There are a lot of problems with that, because the data is not clean, it's biased, it's not representational. Primarily what I'm looking at with *MegaPixels* is not specifically those bias issues of the data. Instead, I'm looking at the origins and the endpoints. The origins are often online. But as I was reading more [scientific] papers, I found some shocking types of data acquisition. Researchers became so comfortable with the idea of scraping and collecting data online that it normalized the lack of conventional research principles, for instance, that you would typically require a consent. And researchers engaged in a type of data collection irrational exuberance where everything became data and everything was an opportunity to collect a dataset; there are several examples that I have written about on the *MegaPixels* and now the *Exposing.ai* site. For example, that Brainwash dataset is called Brainwash, because it refers to a café where people did laundry. And it had a webcam—that was a popular thing to do for a while. But when they put up a webcam, they never had the idea that those webcam images would be scraped by a professor at Stanford, packaged into a dataset for face detection, published online, downloaded by the National University of Defense Technology in China, and used for military applications. Well, that is a clear transgression of one's expectations of privacy, even in a public place. And that was one that I highlighted and that had a strong response. The creator took the dataset down, but without saying anything about it. Often there isn't an admission of guilt. The Duke MTMC dataset was quite different, because there was no scraping of the Internet. The researchers set up cameras on campus and then recorded students going to class, coming from their dormitory. You can see in the video that there are very small signs on the ground that people glance at and don't understand what's going on and keep walking. And nobody changes their behavior in the videos. So I watched them all. There are thousands of people in the videos and I didn't see anyone turn around and walk the other way. Yet, the researchers seemed to think this approach qualified as informed consent. And the main issue in a lot of the datasets is that there was no informed consent. Researchers were generally so excited about their research project that they forgot that academic research typically requires some level of informed consent. So, the Duke MTMC dataset was shared online and then it was used in various forms of military research projects and eventually the Duke MTMC dataset became the most widely used and highly cited dataset for 'person re-identification.' And person re-identification is what is used to track people throughout a city with a network of surveillance cameras. This dataset had a very direct impact on research growth and surveillance technology. The author then took it down and, the reason I bring it up is because he did offer an apology—and he was the only one.

Hunger: Could you point out a little bit more, what's your take on why at this point scientific ethics is not sufficiently effective?

Harvey: As I mentioned, there is a missing component. It seems to be the elephant in the room. The researchers get quiet when you start talking about data consent and the origins of data. Where does the data come from and who gave permission? *Exposing.ai* is about coming to terms with the reality of artificial intelligence and that you can't develop it without data. You can model it, you can create fake data or synthetic data and this works to a certain extent, but it's not a perfect replacement, especially for very demanding tasks like face recognition. If you peel back some of the technical layers of what face recognition is doing, it's still working like the Eigenface algorithm in that way: when your face is run through the layers of a neural network you're being compared to the neural network's memory of every face that the network has looked at before. And if you don't have those faces, then you can't be compared to anything.

Hunger: Adam, it seems to me that the machine learning and computer vision research community is pretty hermetic, in the sense that they do not recognize—or only to a really limited extent—the research that is done from humanity's perspective or from a critical AI perspective.

Harvey: For quite a few years public scrutiny was largely missing for academic researchers. And it is important. Often academic researchers are misunderstood as operating in a completely benevolent academic environment. This is not true. A lot of the researchers are funded by large corporations, Google, Facebook, car manufacturing companies, Yahoo, even defense agencies have sponsored quite a bit of the research. If you look back at the original dataset, the largest one, basically the fountainhead from which flow many of the face datasets, called Yahoo Flickr Creative Commons 100 Million—this was actually created with the involvement of researchers from Lab41, a name that not many people have heard of. Lab41 is a research subsidiary group of the Central Intelligence Agency, the CIA. And this dataset was created along with researchers at a laboratory that is part of the United States national security network. So it's not too hard to see what the intentions are behind a lot of these datasets. When people say 'academic,' they think of idyllic campuses in universities with students doing work for the good of humanity. I would like to think this is also true. But the reality is, a lot of the datasets and the research papers that go along with it have joint authorship, and the funding is from a range of groups from commercial, to defense, to scientific. Some of them are completely okay and in the public interest. But as you drill into the papers you realize that it's not okay to say: "This is an academic paper, therefore there should be nothing wrong with it." You have to look to the acknowledgements and the funding section to determine who is really behind the paper and whose research it is to begin with.

Hunger: To come to an end, let me provoke you a bit with the observation that your practice has more and more transformed from an artistic practice towards an activist one. Where do you see yourself?

Harvey: I find both of those terms not a good fit, because they're often used against me. And an example is the dataset called UnConstrained College Students. The dataset sounds like a party, but it's a dataset of faces that were captured by a professor using a long range telephoto or basically a spy camera pointed across campus and photographing students on their way to class or to lunch, which was in fact funded by US military, defense and intelligence agencies. The makers of this dataset shot back at me, because my research was an art project and that was not valid to them, it wasn't in their arena. They only talked to other academics, scientists and 'legitimate' researchers. I like artistic practice and I like activist practice, but they are often not seen by outsiders as legitimate. So it depends on where I am, whether I identify with those terms or not. But there is a lot of room for this convergence in artistic research. And people in the academic community should give more legitimacy to artistic research, because in some cases I've seen artistic research that is much better than academic research. Often artistic research points out the problems of academic research and that's one reason why people try to dismiss it. But, if someone doesn't like your project, they'll try to find anything to dismiss you. Through the project, I have actually met quite a few researchers who are open and realized that it is a problem and they don't want to be on the wrong side of history. So it's important to have these slightly adversarial projects to provoke people to make legitimate change.

Scan the QR code below to watch the full video interview online.

Im Gespräch mit
Adam Harvey über „Datensätze zur Gesichtserkennung“

Im Interview geht es um Adam Harveys kritische Untersuchung von Trainingsdatensätzen des maschinellen Lernens. Er sagt dazu: „Ich glaube nicht, dass es möglich ist, die gesamte bestehende Infrastruktur der Gesichtserkennung zu zerstören, und das will ich auch nicht. Aber ich möchte ihr Wachstum begrenzen und ihr gefährliches Potenzial einschränken.“

Transkript des Interviews mit Adam Harvey, geführt von Francis Hunger am 14–05–2021.

Hunger: Seit Mitte der 2010er Jahre sind umfassende Bildsammlungen entstanden, um künstliche, gewichtete Netzwerke für die Mustererkennung zu trainieren. Am bekanntesten und kritisch diskutiert ist ImageNet, das zur Objekterkennung und -klassifizierung verwendet wird. Dein Projekt *MegaPixels* (2017–20) befasst sich nicht mit der Objekterkennung, sondern mit der Gesichtserkennung und den entsprechenden Trainingsdatensätzen. Könntest du, wenn du über *MegaPixels* sprichst, ein wenig mehr über dessen Vorgängerprojekt *CV Dazzle* (2010) erzählen und wie *MegaPixels* daraus hervorgegangen ist?

Harvey: *MegaPixels,* das jetzt *Exposing.ai* (2021) heißt, ist eine Weiterentwicklung von *CV Dazzle,* das nun schon elf Jahre alt ist. *CV Dazzle* oder Computer Vision Dazzle – ‚Dazzle' bezieht sich auf eine Tarntechnik aus dem Ersten Weltkrieg – wurde als Projekt zur Abwehr von Überwachungsmaßnahmen entwickelt und sollte zum einen beweisen, dass Camouflage gegen Computer Vision möglich ist, und zum anderen, dass die Tarnung auf ästhetische Weise erfolgen kann. Es ist eine funktionale Ästhetik. Es gab hier eine zusätzliche Diskussion darüber, dass Tarnung eine sich entwickelnde Strategie ist, und das Ziel von *CV Dazzle* ist es, das eigene Erscheinungsbild in Bezug auf die Beobachtung, Analyse und Erkennung durch biometrische Analysesoftware zu minimieren. *CV Dazzle* war auf die Gesichtserkennung ausgerichtet. Und Gesichtserkennung ist eine Form der Objekterkennung, bei der das Gesicht das Objekt ist. Und wenn man die Detektierung des Gesichtes stört, dann stört man auch dessen Erkennung. Das Ziel von *CV Dazzle* besteht also darin, die Bildanalyse-Pipeline an der einfachsten Stelle zu unterbrechen. Und das habe ich erreicht, indem ich einen ganz bestimmten Algorithmus, den Viola-Jones-Haar-Cascade-Algorithmus, ins Visier genommen habe. Und *CV Dazzle* funktionierte deswegen, weil ein spezifischer Algorithmus auf einer Vielzahl von Überwachungsgeräten eingesetzt wurde. Und als ich immer mehr Beispiele für *CV Dazzle* veröffentlichte, stellte ich fest, dass sie dann in anderen Forschungsarbeiten auftauchten. Wissenschaftler*innen untersuchten, wie man sicherstellen kann, dass die Gesichtserkennung funktioniert, wenn Menschen viel Make-up tragen, und dann wurde ein Bild von *CV Dazzle* in diese Forschungsarbeit aufgenommen. Das war interessant für mich, denn ich habe es in erster Linie als Kunstprojekt gemacht, als Hackerprojekt, um eine Schwachstelle zu finden, sie auszunutzen und danach zu verbreiten, in der Hoffnung, dass die Leute diese Strategie übernehmen und für ihre eigenen Zwecke anpassen würden. Nach dem Viola-Jones-Haar-Cascade-Algorithmus entwickelte sich eine neue Technologie: sogenannte ‚neuronale Netzwerke'. Jedes Mal, wenn man versucht, ein neuronales Netzwerk anzugreifen, greift man es leider nicht wirklich an. Ein adversarialer Hack ist weniger ein gegnerisches Manöver als vielmehr ein protagonistischer Beitrag zum Netzwerk. Die seit 2015 eingesetzte Strategie, beginnend mit *CV Dazzle,* kippte, und mein Denken änderte sich entsprechend. *MegaPixels* ist eine Weiterentwicklung von *CV Dazzle,* aber mit dem Ziel, die Lieferketten anzugreifen, anstatt den Algorithmus direkt anzugreifen. Wenn es gelingt, die Datenmenge zu begrenzen, die in einen Gesichtserkennungsalgorithmus einfließt, kann man die Leistung so weit einschränken, dass man ohne Daten keinen Algorithmus hat und keine Gesichtserkennung durchführen kann. Indem man anfängt, die Daten zu entfernen, entfernt man auch die Fähigkeiten neuronaler Netzwerke. Meine Absicht bei diesem Projekt *MegaPixels* ist also, das Wachstum zu bremsen. Ich glaube nicht, dass es möglich ist, die gesamte bestehende Infrastruktur der Gesichtserkennung zu zerstören, und das will ich auch nicht. Aber ich möchte ihr Wachstum begrenzen und ihr gefährliches Potenzial einschränken.

Hunger: Der Brainwash-Café-Datensatz und der MS-CELEB-Datensatz waren einige der ersten Datensätze, die du und Jules LaPlace für das *MegaPixels*-Projekt recherchiert habt. Wie bist du auf diese Datensätze gestoßen und was ist das Besondere an ihnen?

Harvey: Es fällt mit ImageNet zusammen, dass Forscher*innen eine Strategie übernahmen, die zuerst für den Datensatz Labeled Faces in the Wild (LFW) bekannt wurde, der im Jahr 2007 veröffentlicht worden ist. Zu dieser Zeit änderten sich die Ökonomien der Datenerhebung. Man musste die Leute nicht mehr bezahlen, man brauchte keine Einverständniserklärung abzugeben, denn es gab so viele Daten und so wenige Regeln für deren Erhebung, dass die Forscher*innen erkannten, dass man online gehen und so viele Daten abgreifen konnte, wie man wollte. Das bringt eine Menge Probleme mit sich, denn die Daten sind nicht sauber, sie sind durch Bias verzerrt, sie sind nicht repräsentativ. Bei *MegaPixels* geht es mir nicht in erster Linie um den Bias in den gesammelten Daten, sondern ich betrachtete die Herkünfte und die Endpunkte. Die Ursprünge sind oft online. Aber als ich weitere [wissenschaftliche] Aufsätze las, fand ich einige schockierende Arten der Datenerfassung. Ich denke, das liegt daran, dass sich die Forscher*innen so sehr an die Idee gewöhnt haben, Daten online zu sammeln und zu erfassen, dass das Fehlen der üblichen Forschungsprinzipien, zu denen normalerweise auch eine Einwilligung zählt, normalisiert wurde. Die Forscher*innen haben sich auf einen irrationalen Überschwang bei der Datenerfassung eingelassen, bei dem alles zu Daten wurde und alles eine Gelegenheit war, einen Datensatz zu sammeln. Es gibt mehrere Beispiele, über die ich auf der *MegaPixels*- und jetzt der *Exposing.ai*-Website geschrieben habe. Der Brainwash-Datensatz heißt übrigens Brainwash, weil es sich um ein Café handelt, in dem die Leute Wäsche waschen. Und das Café hatte eine Webcam, die eine Zeit lang sehr üblich waren. Aber wenn man eine Webcam aufstellt, hat man keine Ahnung, dass diese Webcam-Bilder von einem Professor in Stanford ausgewertet, in einen Datensatz für Gesichtserkennung aufgenommen, online publiziert, von der Nationalen Universität für Verteidigungstechnologie in China heruntergeladen und für militärische Zwecke verwendet wurden. Das ist also ein ganz klarer Verstoß gegen die Anforderungen an Privatsphäre, selbst an einem öffentlichen Ort. Darauf habe ich hingewiesen, und es folgte eine deutliche Reaktion. Der Ersteller hat den Datensatz offline gestellt, ohne jedoch etwas dazu zu sagen. Oft gibt es kein Schuldeingeständnis. Eine ziemliche Ausnahme ist der Duke-MTMC-Datensatz. Dieser Datensatz war ganz anders, denn er wurde nicht aus dem Internet erhoben. Die Wissenschaftler*innen stellten Kameras auf dem Campus auf und zeichneten Studierende auf, die zum Unterricht gingen oder aus ihrem Wohnheim kamen. Man kann in den Videos sehen, dass es kleine Hinweise auf dem Boden gibt, auf die die Leute schauen, aber nicht verstehen, was los ist, und weitergehen. Und niemand ändert sein Verhalten in den Videos. Ich habe mir alle angesehen. In den Videos sind Tausende von Menschen zu sehen, und ich habe niemanden gesehen, der sich um 180 Grad gedreht hätte und in die andere Richtung gegangen wäre. Dennoch schienen die Wissenschaftler*innen zu glauben, dass dies als aktive Einwilligung ausreichte. Und das Hauptproblem bei vielen Datensätzen ist, dass es aber überhaupt keine aktive Einwilligung gab. Die Wissenschaftler*innen waren im Allgemeinen so von ihrem Forschungsprojekt begeistert, dass sie vergaßen, dass akademische Forschung in der Regel aktive Zustimmung erfordert. Ähnliches geschah mit dem Duke-MTMC-Datensatz: Er wurde online verbreitet und dann in verschiedenen militärischen Forschungsprojekten verwendet, und schließlich wurde der Duke-MTMC-Datensatz zum meistgenutzten und am häufigsten zitierten Datensatz für die ‚Re-Identifizierung' von Personen. Und diese Re-Identifizierung von Personen wird verwendet, um Menschen in einer Stadt zu verfolgen. Man hat ein Netz von Überwachungskameras in der ganzen Stadt. Dieser Datensatz hatte also einen direkten Einfluss auf die Entwicklung der Forschung und der Überwachungstechnologie. Der Autor hat den Aufsatz dann aus dem Internet genommen, und ich erwähne ihn deshalb, weil er sich entschuldigt hat, und er war der Einzige.

Hunger: Könntest du ein bisschen mehr darauf eingehen, warum aus deiner Sicht an diesem Punkt eine Wissenschaftsethik nicht ausreichend greift?

Harvey: Wie ich bereits erwähnt habe, gibt es eine fehlende Komponente, über die nicht gern gesprochen wird. Die Wissenschaftler*innen werden still, wenn man über die Zustimmung zu Daten und deren Herkunft sprechen will. Woher kommen die Daten und wer hat sie freigegeben? Und beim Projekt *Exposing.ai* geht es wirklich darum, sich mit der Realität der Künstlichen Intelligenz auseinanderzusetzen und dass man sie nicht ohne Daten entwickeln kann. Man kann sie modellieren, man kann gefälschte Daten oder synthetische Daten erstellen, und das funktioniert bis zu einem gewissen Grad, aber es ist kein perfekter Ersatz, vor allem nicht für sehr anspruchsvolle Aufgaben wie die Gesichtserkennung. Wenn man einige der technischen Schichten der Gesichtserkennung entfernt, funktioniert sie immer noch wie der Eigenface-Algorithmus: Wenn dein Gesicht durch die Schichten eines neuronalen Netzwerks läuft, wirst du mit den im neuronalen Netzwerk gespeicherten Gesichtern verglichen, die es zuvor betrachtet hat. Und wenn man diese Gesichter nicht hat, dann kann man ganz einfach mit nichts verglichen werden.

Hunger: Ich habe den Eindruck, dass die Forschungsgemeinschaft im Bereich des maschinellen Lernens und der Computer Vision ziemlich hermetisch ist, in dem Sinne, dass sie die Forschung, die aus einer gesellschaftswissenschaftlichen oder kritischen KI-Perspektive betrieben wird, nur in sehr geringem Maße anerkennt.

Harvey: Die akademischen Forscher*innen wurden einige Jahre lang kaum von der Öffentlichkeit kontrolliert. Und das ist wichtig: Es ist ein häufiges Missverständnis, dass Akademiker*innen in einem völlig gutartigen akademischen Umfeld arbeiten. Das stimmt so nicht. Viele der Forscher*innen werden von großen Unternehmen finanziert, von Google, Facebook, Automobilhersteller*innen, Yahoo, sogar das Militär hat die Forschung gesponsert. Wenn man sich jenen ursprünglichen Datensatz ansieht, den größten, der im Grunde die Quelle für viele der Gesichtsdatensätze darstellt, nämlich Yahoo Flickr Creative Commons 100 Million, so wurde dieser unter Beteiligung von Forscher*innen des Lab41 erstellt, einem Namen, von dem nur wenige Leute gehört haben. Lab41 ist eine Forschungsuntergruppe der Central Intelligence Agency, der CIA. Und dieser Datensatz wurde also zusammen mit Forscher*innen eines Labors erstellt, das Teil des nationalen Sicherheitsnetzes der Vereinigten Staaten ist. Es ist nicht allzu schwer zu erkennen, welche Absichten hinter vielen dieser Datensätze stehen. Und wenn die Leute ‚akademisch' sagen, denken sie zuerst an einen idyllischen Universitätscampus mit Studierenden, die zum Wohle der Menschheit arbeiten. Ich würde gern glauben, dass dem so ist, doch die Realität sieht so aus, dass viele der Datensätze und die dazugehörigen Forschungsarbeiten von mehreren Autor*innen gemeinsam verfasst und von verschiedenen Gruppen aus der Wirtschaft, dem Militär und der Wissenschaft finanziert werden. Einige davon sind völlig in Ordnung und im öffentlichen Interesse. Aber wenn man sich mit den wissenschaftlichen Aufsätzen befasst, stellt man fest, dass es nicht in Ordnung ist, zu sagen: „Dies ist ein akademischer Aufsatz, also sollte daran nichts auszusetzen sein." Man muss sich die Danksagungen und den Abschnitt über die Finanzierung ansehen, um herauszufinden, wer wirklich hinter dem Aufsatz steht und wessen Forschung es ist.

Hunger: Abschließend möchte ich ein wenig mit dem Statement provozieren, dass sich deine Praxis immer mehr von einer künstlerischen zu einer aktivistischen entwickelt hat. Inwiefern siehst du das so?

Harvey: Ich finde, dass diese beiden Begriffe nicht gut passen, weil sie oft gegen mich verwendet werden. Ein Beispiel ist der Datensatz UnConstrained College Students. Der Datensatz hört sich nach einer Party an, aber es ist ein Datensatz von Gesichtern, die von einem Professor aufgenommen wurden, der ein Teleobjektiv mit großer Reichweite, im Grunde eine Spionagekamera, benutzte, die auf den Campus gerichtet war und Studierende auf ihrem Weg zum Unterricht oder zum Mittagessen fotografierte, was vom US-Militär, den Verteidigungsbehörden und den Geheimdiensten finanziert wurde. Die Ersteller*innen dieses Datensatzes reagierten negativ auf mich, weil meine Recherche ein Kunstprojekt war und das für sie nichtig ist, weil ich nicht aus ihrem beruflichen Feld stammte. Sie sprachen nur mit anderen Akademiker*innen, Wissenschaftler*innen und ‚seriösen' Forscher*innen. Ich mag also künstlerische und aktivistische Praktiken, aber sie werden von Außenstehenden oft nicht als legitim angesehen. Es hängt also davon ab, wo ich mich befinde, ob ich mich mit diesen Begriffen identifizieren kann oder nicht. Es gibt eine Menge Raum für dieses Wechselspiel der künstlerischen Forschung. Und die Menschen in der akademischen Gemeinschaft sollten der künstlerischen Forschung mehr Legitimität einräumen, denn in einigen Fällen habe ich künstlerische Forschung gesehen, die viel besser ist als akademische. Künstlerische Forschung weist oft auf Pro-

bleme in der akademischen Forschung hin, und das ist ein Grund, warum diese Leute sie ignorieren wollen. Und wenn ihnen dein Projekt nicht gefällt, werden sie alles versuchen, um dich zu ignorieren. Aber durch das Projekt habe ich auch Forscher*innen kennengelernt, die offen sind und erkannt haben, dass dies ein Problem ist, und die nicht auf der falschen Seite der Geschichte stehen wollen. Daher halte ich es für wichtig, diese leicht kontroversen Projekte durchzuführen, um Menschen dazu zu bewegen, legitime Veränderungen vorzunehmen.

Den QR-Code scannen, um das vollständige Videointerview anzusehen.

In Conversation with Gabriel Pereira on “How the Image Collection ImageNet Re-Constructs Reality”

The interview explores historical and functional aspects of the ImageNet image data collection. Pereira investigated this aspect of Visual Computing in collaboration with his research and artist colleague Bruno Moreschi, using the collection of the Van Abbemuseum, Eindhoven as an example.

Transcript of the interview with Gabriel Pereira, conducted by Francis Hunger on 2022–01–25.

Hunger: ImageNet is one of the most important image data collections for the training of so-called 'artificial neural networks,' or better, 'weighted' networks. Could you introduce what ImageNet is, what it does, and where it comes from?

Pereira: Yeah, ImageNet is really an important dataset. It is a collection of images that have been organized and labeled. There are over 14 million images and it comes from the work of Fei-Fei Li and her team at Stanford University. When I say that ImageNet is a dataset, that might not do it justice, because it has been so influential that it has grown to be more than just a dataset and it is almost one of the most important things that happened to computer vision over the past 15 years, since its creation. It has been really one of the reasons why computer vision has become so popular, because it gave a large amount of structured data to train computer vision algorithms and it went on to be used for many challenges. In these challenges, different researchers and computer scientists were trying to create algorithms that performed to a certain degree of efficiency. They used the ImageNet dataset to perform benchmarks with those algorithms. And so it really powered the growth of the field. It is hard, sometimes, for us researchers to say 'determined,' but ImageNet really influenced what the field has done over the past 15 years or so.

Hunger: Could you describe a little more how ImageNet interacts with computer vision algorithms?

Pereira: In the case of ImageNet, we have to go back to the beginning. It is a large dataset, a collection of images, and they first have to be collected. In this case, a lot of those 14 million images of ImageNet come from the Internet itself. These are images that people have produced, that they have taken of their families and their houses, and so on, and that they have uploaded to places like Flickr where they didn't give it a copyright condition. So, what the researchers of ImageNet did was to scrape the Internet of those images and create a large pool of images. So it already comes from there. And then, what those researchers did was having to organize them according to a taxonomy. They had to categorize those images according to what they are. You need to say—for all of those categories—that this is a table or that is a window or that is a cell phone. Fei-Fei Li decided to use WordNet, which is a system that was created in the mid-80s with funding from the DARPA [Defense Advanced Research Projects Agency]. It already comes from a military background. And WordNet is basically a taxonomy. It divides the world into 175,000 items, so-called 'synsets.' It defines what the things are and how these things relate. So, a window is part of a building, for example, and so on. And it gives structure to the dataset. Then the next aspect of this is that someone has to sit there and decide that this is an image of a table or this an image of a window. And computers can't do that, because it is very difficult to know what is a table, and what is a window. Fei-Fei Li and her team decided to use, and this was a very important thing that ImageNet did, crowdwork, or microwork, from Amazon Mechanical Turk. Amazon Mechanical Turk is a platform created by Amazon where you can create microtasks, for instance, ask a person if this is a window or that is a table. And thousands and thousands of workers ended up working for this and they aren't recognized in the dataset and they weren't paid well at all, they weren't listened to, in the making of the dataset. Finally, we have the use, how those systems are used to do those kinds of image recognition that we might see in our day-to-day life. Basically, those datasets can be downloaded for free, and people can use it to train those algorithms in those categories that were defined, that were adopted from WordNet. And the work of Nanna Thylstrup is really fantastic on this, how those images continue to exist there and are used to do very specific things, but sometimes, they are out of place, they might even be toxic, the way that those images have been appropriated and the way that they are used in this system. We don't have very much moderation. There is one more thing which is really important: this system already defines what the problem is and how to solve it. You can't go back and add more categories to WordNet. All the objects that exist in the world are already defined, and what ImageNet does very well is ignore its own limitations, ignore its own situated way of 'seeing' the world.

Hunger: Currently, we are seeing a lot of bold promises in the tech-bubble, I would say, overconfident promises about AI and what it can do and solve—is this a strategy? I mean with a certain level of knowledge you must realize that these promises can't be kept.

Pereira: Many of the people who are working in AI probably don't believe in AI in the same way that they tell the public that they do. However, there is something interesting in the way that those forms of discourse also construct what comes to exist. They also construct which kinds of projects are done and they also shape what sorts of applications we see of AI in the world. To think of one example, in all of those commercial computer vision systems they pretend that they can see everything—and of course they can't. But to some extent, this imaginary of what computer vision is and what computer vision can do, is also part of the fight that the companies have to face to also justify their existence in the world and justify the power they exert in society. That is a really important thing for us—as researchers and artists—to go against. To reshape the imaginary that people in society have of those systems. I already mentioned the workers that are creating those computer vision systems, cataloging those images and that they get very little money for doing this kind of work. And in this case, there is something that is ignored in ImageNet in the way that the media talks about it, and I think we need to recover it.

Hunger: You've worked on visual experiments with ImageNet on the example of art museum collections. I really liked how you compared modern art to objects from the IKEA catalogue. What were your most significant findings?

Pereira: So, this is the work of *Recoding Art* (2019), which I conducted together with the researcher and artist Bruno Moreschi, who is a very dear collaborator of mine. We worked with the Van Abbemuseum in the Netherlands and we had two exhibitions in the collection analyzed by computer vision from six computer vision services or libraries. And of those computer vision services or libraries, many of them were commercial and private and we don't exactly know which kinds of datasets they use. This includes Google Cloud Vision and Amazon Rekognition. They may use datasets from many different places, but of course all of those datasets have been shaped by ImageNet, because it is the most influential dataset. What we noticed when we started, using computer vision to analyze those artworks, was that there were many errors or unexpected results. But we understood those errors to be very revealing of both the art system and the computer vision systems that we were using. Part of this is because originally, computer vision was not trained on art. What we did then was to look at those unexpected results and first, they are very revealing of the way that computer vision has been trained to see and of the ways of seeing that they embody. It is a very commercial way of seeing, it is a product-focused way of seeing, to the extent that you show Marcel Duchamp's *Fountain* from 1917, one of the most important works of contemporary or modern art, and it reads it as an urinal, which is of course exactly what Duchamp was critiquing. So, computer vision misses sort of the given subtext or other kinds of understandings about this image, and it deceives us to some extent. We understand some of the origins of those datasets that are training those computer vision systems. So, those images come from day-to-day lives of people in their homes, or they come from this kind of collection where windows or tables or even cell phones are really prominent. And on the other hand, those glitches, they are revealing of the art system, they are somehow poetic. The art system is a very hierarchical system where people are not listened to about what they think about the artworks. Very often, the only people who matter in the museum are the artists and the curators and nobody else can be part of the conversation about what the artwork is doing. It's very top-down. But, using computer vision to look at those artworks differently provides, in my opinion, a very creative potential for looking at the art museum and those artworks with fresh sets of eyes, to the extent that suddenly we can see things that we were unable to see before, make connections that we couldn't before. And it can lead visitors, for example, to be part of this conversation in ways that they weren't invited to do before. What happens when you look at Duchamp's *Fountain* as a mere urinal again? What happens when you look at an artwork and you see a table, or you see something from everyday life, or you see it just as a cell phone? What kind of interesting connections emerge when we listen to those unexpected results?

Hunger: Could you provide a few more examples of what the machine detected in those artwork collections?

Pereira: For example, the amount of artworks that get read as 'cats,' as if there is a cat in the image when in reality actually there isn't one, but suddenly we're seeing all of those cats in images that are reminding us of this everyday imagery of our lives. Other than that, for example, many times the artwork gets read as a table or as a window, which is interesting, because it might relate to the shape of the artwork, reminding us somehow of a window—it's rectangular, it has a frame around it. And that's really interesting, because suddenly we're seeing the artwork, but we're focusing on what supports it, not on the artwork itself. Other than that, and this is a bit more problematic, very often when there are women in those artworks, they get read as 'racy,' they get read as if there is something that is provocative or pornographic in that image, even when those women are not naked. So, there is something about the gendered way of seeing of those computer vision systems. They don't work as well as we were told they do. They are not as efficient, they are not as neutral, and they are not inevitable.

Scan the QR code below to watch the full video interview online.

Im Gespräch mit
Gabriel Pereira über „Wie die Bilddatensammlung ImageNet Wirklichkeit (re-)konstruiert“

Das Interview erkundet geschichtliche und funktionale Aspekte der Bilddatensammlung ImageNet. Diese untersuchte Pereira in Zusammenarbeit mit seinem Forscher- und Künstlerkollegen Bruno Moreschi am Beispiel der Sammlung des Van Abbemuseums, Eindhoven.

Transkript des Interviews mit Gabriel Pereira, geführt von Francis Hunger am 25–01–2022.

Hunger: ImageNet ist eine der wichtigsten Bilddatensammlungen für das Training sogenannter neuronaler Netze, oder besser gesagt, gewichteter Netze der Künstlichen ‚Intelligenz'. Könntest du kurz erklären, was ImageNet ist, was es überhaupt macht und woher es kommt?

Pereira: ImageNet ist wirklich ein sehr wichtiger Datensatz. Es ist eine Sammlung von Bildern, die sortiert und gelabelt wurden. Es gibt über 14 Millionen Bilder und es ist das Ergebnis der Arbeit von Fei-Fei Li und ihrem Team an der Stanford University. Wenn ich sage, dass ImageNet ein Datensatz ist, wird das der Sache vielleicht nicht ganz gerecht, denn es war so einflussreich, dass es mehr als nur ein Datensatz ist. Es ist eines der wichtigsten Dinge, die in den letzten 15 Jahren im Bereich computerbasierten Sehens und Computer Vision passiert sind. ImageNet ist einer der Gründe, warum Computer Vision so populär wurde, da es eine große Menge an strukturierten Daten für das Training von Computer-Vision-Algorithmen verfügbar machte. Es wurde für viele Wettbewerbe verwendet. In diesen Wettbewerben versuchten verschiedene Forscher*innen und Informatiker*innen, Algorithmen zu entwickeln, die einen bestimmten Grad an Effizienz erreichen, und sie nutzten den ImageNet-Datensatz, um Benchmarks mit diesen Algorithmen zu setzen. Auf diese Weise wurde das Wachstum dieser Disziplin wirklich vorangetrieben. Es fällt vielleicht schwer zu sagen, ImageNet habe die Forschung ‚bestimmt', aber ImageNet hat die Entwicklung des Fachgebiets in den letzten 15 Jahren seit seiner Gründung wirklich stark beeinflusst.

Hunger: Gabriel, könntest du etwas genauer beschreiben, wie ImageNet mit Computer-Vision-Algorithmen interagiert?

Pereira: Im Fall von ImageNet müssen wir an den Anfang zurückgehen. Es handelt sich um einen großen Datensatz von Bildern und diese müssen zunächst gesammelt werden. In diesem Fall stammen viele der 14 Millionen Bilder, die Teil von ImageNet sind, aus dem Internet. Es handelt sich um Bilder, die Menschen von ihren Familien, ihren Häusern usw. gemacht haben, und die sie auf Plattformen wie Flickr hochgeladen haben, wo sie nicht mit einem Copyright versehen wurden. Die ImageNet-Forscher*innen haben also das Internet nach solchen Bildern durchforstet und einen großen Pool von Bildern angelegt. Und dann mussten die Forscher*innen sie nach einer Taxonomie organisieren. Sie mussten diese Bilder kategorisieren, um zu bestimmen, was sie sind. Man muss genau sagen, dass dieses ein Tisch ist, oder jenes ein Fenster oder Mobiltelefon. Dazu mussten sie einer Taxonomie folgen, einem System, das die Wörter in strenge Kategorien einteilt. Fei-Fei Li entschied sich für WordNet, ein System, das Mitte der 1980er Jahre mit finanzieller Unterstützung der DARPA [Defense Advanced Research Projects Agency] entwickelt wurde. Es hat also bereits einen sehr militärischen Hintergrund. Es unterteilt die Welt in 175.000 Dinge, sogenannte ‚Synsets'. Es definiert, was die Dinge sind und wie sie zusammenhängen. So ist ein Fenster Teil eines Gebäudes usw. Das gibt dem Datensatz eine Struktur. Der nächste Aspekt ist, dass sich jemand hinsetzen und entscheiden muss, dass dies ein Bild eines Tisches oder ein Bild eines Fensters ist. Und das können Computer nicht, denn es ist sehr schwierig, zu wissen, was ein Tisch und was ein Fenster ist. Fei-Fei Li und ihr Team beschlossen daher, und das war ein sehr wichtiger Aspekt von ImageNet, Crowdwork oder Mikroarbeit von Amazon Mechanical Turk zu nutzen. Amazon Mechanical Turk ist eine von Amazon entwickelte Plattform, auf der man kleinste Aufgaben erstellen kann, also eine Person fragen kann, ob dies ein Fenster oder ein Tisch ist. Tausende und Abertausende von Arbeiter*innen haben dafür gearbeitet, ohne dass sie in dem Datensatz als Autor*innen sichtbar werden und ohne, dass sie dafür gut bezahlt wurden, ohne, dass man sie wahrgenommen hat. Und schließlich geht es um die Verwendung dieser Systeme für die Bilderkennung, wie wir sie im täglichen Leben sehen können. Im Grunde genommen können diese Datensätze kostenlos heruntergeladen werden, und die Leute können sie nutzen, um die Algorithmen mit den Kategorien zu trainieren, die definiert und von WordNet übernommen wurden. Die Arbeit von Nanna Thylstrup ist in dieser Hinsicht wirklich fantastisch. Sie zeigt, wie dort Bilder weiterhin existieren und für ganz bestimmte Dinge verwendet werden, und manchmal passen sie nicht richtig, sie können sogar toxisch sein, aufgrund dessen wie man sich die Bilder angeeignet hat und wie sie in diesem System verwendet werden. Es wird dort wenig moderiert. Es gibt noch eine weitere Sache, die wirklich wichtig ist: Das System definiert bereits, was das Problem ist, und wie man es lösen kann. Man kann nicht zurückgehen und weitere Kategorien zu WordNet hinzufügen. Es ist bereits definiert, was alle Objekte sind, die in der Welt existieren. Und was ImageNet sehr gut kann, ist, seine eigenen Beschränkungen zu ignorieren, seine eigene spezielle, sehr situierte Art, die Welt zu ‚sehen'.

Hunger: Momentan sehen wir in der Tech-Bubble eine Menge Versprechen, ich würde sagen: übermäßig zuversichtliche Versprechen in Bezug auf KI und die Leistungen, die sie erbringen soll – handelt es sich dabei um eine Strategie? Ich meine, mit einem gewissen Kenntnisstand muss man doch erkennen, dass diese großen Versprechen nicht eingehalten werden können.

Pereira: Viele derjenigen, die im Bereich der Künstlichen Intelligenz arbeiten, glauben wahrscheinlich nicht in dem Maße an KI, wie sie es der Öffentlichkeit erzählen. Interessant ist jedoch die Art und Weise, in der diese Formen des Diskurses konstruieren, was im Entstehen begriffen ist. Sie bestimmen auch, welche Art von Projekten durchgeführt werden, und sie bestimmen, welche Art von Anwendungen wir in der Welt der KI sehen. Ein Beispiel: Bei all diesen kommerziellen Computer-Vision-Systemen wird so getan, als könnten sie alles sehen, aber das können sie natürlich nicht. Aber in gewisser Weise ist diese Vorstellung davon, was Computer Vision ist und was Computer Vision kann, auch Teil des Kampfes, dem sich die Unternehmen stellen müssen, um ihre Existenz in der Welt und die Macht, die sie in der Gesellschaft ausüben, zu rechtfertigen. Und ich denke, dass es für uns als Forscher*innen und Künstler*innen wirklich wichtig ist, dagegen vorzugehen. Wir müssen die Vorstellung, die die Menschen in der Gesellschaft von diesen Systemen haben, korrigieren. Ich habe die Arbeiter*innen erwähnt, die diese Computer-Vision-Systeme entwickeln, diese Bilder katalogisieren und die nur sehr wenig Geld für diese Art von Arbeit bekommen. In diesem Fall gibt es etwas, das bei ImageNet

in der Art und Weise, wie die Medien darüber sprechen, ignoriert wird. Und ich denke, wir müssen das sichtbar machen.

Hunger: Bruno Moreschi und du, ihr habt an visuellen Experimenten mit ImageNet am Beispiel von Kunstmuseumssammlungen gearbeitet. Mir hat sehr gut gefallen, wie ihr moderne Kunst mit Objekten aus dem IKEA-Katalog verglichen habt. Was waren eure wichtigsten Erkenntnisse?

Pereira: Das ist die Arbeit *Recoding Art,* die ich 2019 zusammen mit dem Forscher und Künstler Bruno Moreschi – den ich sehr schätze – durchgeführt habe. Wir haben mit dem Van Abbemuseum in den Niederlanden kooperiert und zwei Ausstellungen aus der Sammlung mithilfe der Computer Vision von sechs Computer-Vision-Diensten oder -Bibliotheken analysiert. Und von diesen Computer-Vision-Diensten waren viele kommerziell und privat, und wir wissen also nicht genau, welche Arten von Datensätzen sie verwenden. Das umfasst Google Cloud Vision und Amazon Rekognition. Sie können Datensätze verschiedenster Herkunft verwenden, aber natürlich sind alle diese Datensätze durch ImageNet geprägt, denn es ist der einflussreichste Datensatz. Bei der Analyse von Kunstwerken mithilfe von Computer Vision stellten wir fest, dass es viele Fehler oder unerwartete Ergebnisse gab. Diese Fehler haben wir als sehr aufschlussreich, sowohl für das Kunstsystem als auch für die von uns verwendeten Computer-Vision-Systeme, verstanden. Das liegt zum Teil daran, dass Kunst zum Training von der Computer Vision ursprünglich nicht verwendet wurde. Wir haben uns dann diese unerwarteten Ergebnisse angeschaut, und sie sind sehr aufschlussreich hinsichtlich der Art, wie Computer-Vision-Systeme trainiert wurden, und für die Art des Sehens, die sie verkörpern. Es ist eine sehr kommerzielle Art des Sehens, eine sehr produktorientierte Art des Sehens, die so weit geht, dass man Marcel Duchamps *Fountain* von 1917, eines der wichtigsten Werke der zeitgenössischen oder modernen Kunst, als Urinal interpretiert, was genau das ist, was der Künstler kritisierte. Es fehlen also eine Art Subtext oder andere Arten des Verständnisses dieses Bildes und es betrügt uns in gewisser Weise. Wir verstehen einige der Ursprünge dieser Datensätze, mit denen diese Computer-Vision-Systeme trainiert werden. Diese Bilder stammen aus dem täglichen Leben der Menschen, aus ihren Wohnungen, oder sie stammen aus Sammlungen, in denen Fenster, Tische oder Handys im Vordergrund stehen. Auf der anderen Seite sind eben diese Störungen, die für das Kunstsystem bezeichnend sind, irgendwie poetisch. Das Kunstsystem ist ein sehr hierarchisches System, in dem v. a. den Besucher*innen nicht zugehört wird, was sie über die Kunstwerke denken. Oft sind die einzigen, die im Museum etwas zu sagen haben, die Künstler*innen und die Kurator*innen, und niemand sonst kann sich an der Diskussion über die Wirkung der Kunstwerke beteiligen. Es ist sehr top-down. Der Einsatz von Computer Vision, um diese Kunstwerke anders zu betrachten, bietet meines Erachtens ein sehr kreatives Potenzial, um das Kunstmuseum und die Kunstwerke mit neuen Augen betrachten, und zwar in einem Maße, dass wir plötzlich Dinge sehen können, die wir vorher nicht sahen, und Verbindungen herstellen können, die wir vorher nicht wahrnehmen konnten. Und das kann dazu führen, dass Besucher*innen an diesem Gespräch teilnehmen, zu dem sie vorher nicht eingeladen waren. Was passiert, wenn man Duchamps *Fountain* noch einmal als Urinal betrachtet? Was passiert, wenn man ein Kunstwerk betrachtet und einen Tisch sieht oder etwas aus dem täglichen Leben oder ein Handy? Welche interessanten Verbindungen ergeben sich, wenn wir diese unerwarteten Ergebnisse sehen?

Hunger: Könntest du noch ein paar weitere Beispiele nennen, was die Maschine in diesen Kunstsammlungen detektiert hat?

Pereira: Zum Beispiel die vielen Kunstwerke, die als ‚Katzen' gelesen werden, als ob es eine Katze auf dem Bild gäbe, obwohl es in Wirklichkeit gar keine gibt, aber plötzlich sehen wir alle diese Katzen in Bildern, die uns an diese alltägliche Symbolik des täglichen Lebens erinnern. Außerdem wird das Kunstwerk oft als Tisch oder als Fenster gelesen, was interessant ist, weil es vielleicht mit der Form des Kunstwerks zusammenhängt, das uns irgendwie an ein Fenster erinnert – es ist rechteckig, es hat einen Rahmen. Und das ist wirklich interessant, denn plötzlich sehen wir das Kunstwerk, aber wir konzentrieren uns auf das, was es zusammenhält, nicht auf das Kunstwerk selbst. Abgesehen davon, und das ist etwas problematischer, werden diese Kunstwerke, wenn sie Frauen zeigen, sehr oft als ‚aufreizend' gelesen, als etwas Provokantes oder Pornografisches, selbst wenn die Frauen nicht nackt sind. Es geht also um die geschlechtsspezifische Sichtweise dieser Computer-Vision-Systeme. Letztendlich funktionieren diese nicht so gut, wie sonst behauptet wird. Sie sind nicht so effizient, sie sind nicht neutral und sie sind auch nicht unvermeidlich.

Den QR-Code scannen, um das vollständige Videointerview anzusehen.

Artificial Intelligence Is a Hot Mess

Mél Hogan

Liar, liar, the world's on fire / Whatcha gonna do when it all burns down /
Fire, fire burning higher / Still got time to turn it all around—Dolly Parton

This essay outlines the underlying philosophies generating the 'AI' hype and explains how the terminology itself has become a marketing tactic that functions in part to cover up the very real and urgent problems 'AI' generates and exacerbates. Specifically, this essay draws attention to how 'AI' impacts the environment, and is utterly unsustainable technologically and ideologically on an already burning, dried-out, and depleted planet—a planet drained socially, politically, and environmentally.

On March 29, 2023, Emily M. Bender tweeted a thread critiquing the Future of Life Institute's (2023) Open Letter calling to "Pause Giant AI Experiments." As Bender makes clear, the Institute is a longtermist operation, which essentially means that as an organization they value—and worry about—threats that could lead to the extinction of human potential in whatever amorphous shape and format 'potential' may take in the far future. As critics point out, however, longtermists worry about the abstract at the expense of real, on-the-ground, current issues confronting (biological) humans and all living things on earth (Torres 2021). In general, longtermists are white, wealthy men who are hyper-invested in digital technologies—both in financial terms and in terms of their own personal sense of worth—and who value the idea of future humans (usually as digital humans) living in simulated virtual worlds powered by the vast energy of outer space, while also, ironically, decrying the rogue 'AI' technologies that they are helping fund and create. These imagined threats of rogue 'AI' are too numerous to outline here, but range from deeply entrenched algorithmic biases that ruin people's lives, to mass disinformation and deep faking breaking a shared reality, causing economic, social, and political unrest, to more speculative scenarios like rogue 'Artificial Intelligence' shutting down critical infrastructure like water or energy, or automating weapons that target and kill. While these scenarios are cause for some concern, they overlook the fact that the 'AI' imagined useful to longtermist realizations exacerbate climate change and other social and political problems in the here and now (Klein 2023).

While longtermists lack sophistication in their thinking and feeling, they hold a lot of decision-making power when it comes to how 'AI' might be deployed, and to what ends. The term 'AI' in itself does a lot of the work of misleading the public as to what it is; 'artificial intelligence' implies, among other things, the ability to acquire a kind of agency, in a way that the 'large language model' or 'machine learning' does not, or does much less evocatively. Stefano Quintarelli (2019) has made the case that if 'AI' were understood for what it is, as "Systematic Approaches to Learning Algorithms and Machine Inferences (SALAMI)," the public would be very differently oriented to the technology. Similarly, Francis Hunger (2023) proposes that we replace the deceiving terminology of 'AI' because of how it anthropomorphizes technology and fuels the hype. For now, 'AI' is gaining considerable ground in the public imaginary, as a technology that will revolutionize the world. What 'AI' will actually one day become, however, is always deployed by the industry as a promise to be harnessed and as a threat to be mitigated. They pay for this version of 'AI,' as hype, to be peddled and taken up by mainstream media. They invest in their own fantasies and embroil the rest of us in them. This means that, despite their lack of real expertise or care, they disproportionately influence the shape of the future. The 'AI' future imagined by longtermists is, by any account, ridiculous and violent, but in a world driven by endless capitalist exploitation, it will be legitimized by all sorts of things under the guise of how it sustains 'human potential,' as longtermist Toby Ord (2020) put it in his book *The Precipice*—i.e., the inherent inevitability of 'AI' as scientific progress, and, in turn, the accepted notion that Big Tech (Birch 2022) knows what's best about future-making.

There's an entire psychology to longtermism that remains to be fully unpacked, especially as to how it intersects and overlaps with transhumanism, eugenics and pronatalism, effective altruism, accelerationism, and other right-wing ideologies held by tech CEOs. Some of these movements, and their origins, have been thoughtfully picked apart by writer/scholars such as Malcolm Harris (2023) and Émile P. Torres (2023), among others. Torres (2021) in particular addresses the multi-billion-dollar longtermist movement, and warns that "the point is that longtermism might be one of the most influential ideologies that few people outside of elite universities and Silicon Valley have ever heard about," and adds—importantly—that longtermism "is not equivalent to 'caring about the long term' or 'valuing the wellbeing of future generations.'" Their language is murky, but their idea is this: embodied humans do not contain inherent value. It's 'AI's' ability to convert humanity into a kind of digital potential that is the longtermist's dream. This is why this preamble is very important to all discussions of 'AI;' it reveals it as a tool and tactic with very specific politics—now grouped as TESCREAL (transhumanism, extropianism, singularitarianism, cosmism, rationalism, effective altruism, and longtermism).* To be clear, the impact that the wealth of longtermists (or TESCREAL more generally) could have on tackling climate change, for example, would be tremendous in terms of the improvement on the quality of life of humans, plants and animals. (Here, we might think of Elon Musk or Sam Bankman-Fried, for example.) But, for them, there's no glory in salvaging the planet—rather, they create new problems that can be solved with their new 'AI' toys.

This is why critics like Bender (2023) rightly anticipate that 'AI' developments, deployments, and regulation by longtermists are a 'hot mess.' The *Merriam-Webster* dictionary (n.d.) defines "hot mess" as "something in a state of extreme disorder or disarray" and—at the same time—"attractive and sexy." In these past few months especially, the media hype surrounding 'AI,' as well as the onslaught of criticism against the overpromises made by the 'AI' industry and its investors, reflect the sexy disarray and lure of the topic on all sides. Describing 'AI' as a 'hot mess' is also perfect for how it all at once signals resignation or dismay and—inadvertently—relates the inevitable environmental fallout, precisely because of the intermingling of longtermist ideology (Torres 2021), the concentration of compute power within Big Tech (Birch 2022), and the climate collapse that provides the ever-emerging context where these projects take hold. 'Hot mess,' in short, draws attention to the literal heat of collapse for 'AI.' As Nicole Starosielski (2014, 2506) explained almost a decade ago, "Heat exchanges are not confined to communications systems, but move across and through infrastructure, ecologies, and bodies," of which 'AI' is only the latest iteration and possibly the greatest—or final—threat. To be clear, while catastrophic for the planet in its current instantiations, the threat is not so much the technology itself, which could be harnessed differently, scaled down, or used with accountability, as it is the thinking and feeling of those at the helm of 'AI' projects and Big Tech companies, seemingly unable to build collectively, or work relationally. The failure to think relationally means that 'AI' is destructive, exploitative, and extractivist.

The environmental impacts of 'AI'

Recent studies on the environmental impacts of training and using models for 'AI' deployments have revealed them to be highly energy- and water-intensive, and extremely high in carbon dioxide emissions (Luccioni 2023a; Zhang 2023). While already exorbitant, the numbers in these studies have not yet fully accounted for emissions associated with the computers, hard drives, or any other equipment or material infrastructure manufactured and used for the models' training, nor for the ways in which models are repeated for training and will only be scaled up exponentially in the future (Zhang 2023). Of note on this point, Bender and her collaborators (2021), including the now infamously fired from Google 'AI' ethicists Timnit Gebru and Margaret Mitchell, explained in their widely-cited article *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* that 'AI' models are trained and retrained many times over during research and development, and that estimates should be viewed as minimums. To date, few studies make clear the scale of this retraining.

In 2019, Emma Strubell and her collaborators demonstrated that the process of training a single deep learning natural language processing model could lead to more than 600,000 pounds of carbon dioxide emissions. In other words, as a headline reviewing the paper put it, "training a single AI model can emit as much carbon as

* See Troy (2023); Gebru (2023).

five cars in their lifetimes" (and that includes manufacturing the car) (Hao 2019). To illustrate these impacts, they explain in *Gizmodo* that the energy used for "AI models like OpenAI's GPT-3, which powers the world-famous ChatGPT, could power an average American's home for hundreds of years," a finding based on a 2023 report by the Stanford Institute for Human-Centered Artificial Intelligence (HAI) (DeGeurin 2023). And as Walid Saad (2023), an Associate Professor of Electrical and Computer Engineering at Virginia Tech, recently explained on *Fox5 Washington DC,* 'AI' uses incredible resources that are already in short supply and unsustainable. 'AI' amplifies climate change.

However, like most climate issues, it remains difficult for those critical of industrial 'AI' to convey the scale of these impacts to a public that neither knows how to buy into these discussions nor fully comprehends just what 'AI' encompasses. With these—often startling— numbers in hand, however, computer scientists can begin to raise awareness about the 'AI' industry (Amazon, OpenAI, Google, NVIDIA, Baidu, Microsoft, Intel, Meta, etc.), or, more specifically, about industrial 'AI.' For an episode of *The Data Fix* podcast, I spoke with Sasha Luccioni (2023b), who works on the ethical and societal impacts of machine learning models and datasets; she made it clear that the goal of measuring impacts in this way was to eventually create standards for the industry. There is a sense that the industry can be rehabilitated or at least regulated from the outside, and that measuring impacts is at least one important step towards putting policy into practice.

Sustainable 'AI'

In 2021, Aimme van Wynsberghe wrote *Sustainable AI: AI for sustainability and the sustainability of AI,* in which she outlines a clear trajectory for 'AI' ethics in three waves. It's worth noting that these are more or less the same waves identified in critical media studies and other disciplines, perhaps with slightly different crests. The first wave was sensing a general threat of 'AI' from a Terminator- or M3GAN-like autonomous robot, or more utopic views of uploading our minds to the cloud—a fear of AGI (artificial general intelligence) without yet using the term. That first wave was speculative and about 'AI's' potential. The second wave looked at more practical issues, such as surveillance, privacy concerns, black-boxed algorithms, and the 'explainability' of algorithms, the biases in datasets and search engines, and the lack of fair representation in Big Data, to the promises of facial recognition and sentiment analysis and other things rendered quantified. The third wave, she argues, is "one that confronts the environmental disaster of our time head-on and actively seeks to engage academics, policy makers, AI developers and the general public with the environmental impact of AI" (van Wynsberghe 2021, 213).

In the article, van Wynsberghe (2021) makes a compelling case for doing 'AI' work 'responsibly.' Like Luccione, Strubell, and others, she offers a set of solutions, such as using carbon measures and reporting them as matter of public policy (we should note that this has been tried with data centers and water; though it has failed thus far, it remains a good idea) (Hölzle 2022), using smaller datasets when possible, and using data centers closer to the source of modeling and experimentation, and at times of day when energy is in lower demand, and so on (Fagerström 2023). These are all very practical, implementable ideas. She goes into more detail about these workarounds that are especially important to 'AI' engineers and scientists, those working directly with the technology. One such industry response that has already emerged, as recently pointed out to me by Fenwick McKelvey (2023a), co-director of Concordia University's Applied 'AI' Institute, is the idea of 'Frugal AI:' "a technique that promises the use of less data and less compute power while guaranteeing robustness within the intended field of use for a given AI model" (Larsson 2022).

For those of us on the outside, suffice it to say that there are more or less polluting ways to work with industrial 'AI,' but solutions revolve around maintaining the work—'AI' as an inevitability. Even as the Future of Life Institute's Open Letter surfaced in March 2023, asking that 'AI' research be paused for six months, the intention is simply gloom-hype deployed to draw more attention to 'AI' without any real plausibility of pausing their work or addressing current risks to biases, privacy, or content ownership.

As Meredith Whittaker (2021, 51) writes,

> *This is a perilous moment. Private computational systems marketed as artificial intelligence (AI) are threading through our public life and institutions, concentrating industrial power, compounding marginalization, and quietly shaping access to resources and information. In considering how to tackle this onslaught of industrial AI, we must first recognize that the 'advances' in AI celebrated over the past decade were not due to fundamental scientific breakthroughs in AI techniques. They were and are primarily the product of significantly concentrated data and compute resources that reside in the hands of a few large tech corporations.*

Taking actions to remedy this within one context, Emily Tucker (2022), the Executive Director of the Center on Privacy & Technology, a think tank at Georgetown Law that focuses on the impacts of surveillance policy on marginalized people, writes: "Starting today, the Privacy Center will stop using the terms 'artificial intelligence,' 'AI,' and 'machine learning' in our work to expose and mitigate the harms of digital technologies in the lives of individuals and communities." Here she echoes Whittaker's argument that 'AI' is not only a technological innovation but also a foil—through language and discourse—that distracts us from naming the actual mechanisms and corporate players involved in 'AI' deployments, and those that should be held to account for what they do.

With these warnings, ultimately, we have to decide if we want to make something that is socially and politically corrupt more 'environmentally friendly.' Or if we want to deal with actual social problems in possibly non-technological ways first—through activism, community-building, education and policy work (Fenwick 2023b).

Looking through any tech company report on their environmental practices, we can find claims about sustainability and the importance of protecting ecosystems. On the surface, it appears that Big Tech is taking on the task of making water and global energy more sustainable for the environment's sake, to fight climate change. But problems always crop up alongside these claims. Many of these tech companies have noted that they're paying off other companies and consumers to give them 'green credits' for their coal electricity usage. And this conceals the fact that the vast majority of computer energy use comes from coal-powered manufacturing, and that the coal-powered Internet makes up at least 40% of the energy source (Epstein 2016). Big Tech also unwittingly follows its own uneasy logic as an industry. For example, it'll be all too happy to provide you an 'AI' toothbrush or 'AI' refrigerator under claims that it is more efficient and 'smart'—but it won't account for what it calls 'externalities' like all the e-waste, planned obsolescence, or increased consumerism that results from those so-called innovations. 'AI' as an industry has no inherent capacity to overthrow the interests that are exploiting people and nature: it is much more likely to exacerbate environmental collapse while offering us products that are more 'environmentally friendly.' One is about replacing objects and services, the other is about anchoring ideologies and politics—a hot mess!

'AI' Data Centers

Data centers are the underlying data storage infrastructure of the Internet. Data centers also centralize computational power to process data. With the recent turn to 'AI' by all major Big Tech companies, the infrastructure to support data has also changed. 'AI' requires more power, more water, more mineral and material support (Hogan et al. 2022). While some components of 'AI' might render tasks in the data center more efficient and cut emissions by optimizing energy consumption, predicting equipment failures, and automating maintenance tasks, as journalist Matt Hay (2023) puts it, "AI might hurt the environment before it begins to help the climate crisis." The efficiency of the data centers themselves, while notable, are a foil against addressing the very real, very urgent concerns that computational thinking and doing exacerbate.

With the significant growth of data generated by 'AI' in terms of its applications, data centers have grown in size to accommodate increasing demands for storage and processing power, to process data, make predictions, synthesize and analyze data, and for managing cyberattacks as security risks grow with 'AI.' (Examples of companies hosting 'AI' on their own servers in their privately-owned data centers include Google's DeepMind;** Amazon's SageMaker;*** Microsoft's Project Natick;**** and Alibaba's ET Brain,***** among many others). These types of corporate initiatives have led to the construction of large-scale data centers that can house hundreds of thousands of servers, as well as the development of edge data centers to bring computing power closer to users. Because of the heat generated by servers, liquid cooling systems are now generally used, and they are much more efficient and environmentally friendly air-cooling systems. While 30% of data centers worldwide remain in the US (Daigle 2021), there's a push to move data warehouses to places like the near Arctic of Scandinavia (Vincent 2016),****** or underwater in Europe (Cellan-Jones 2020), so that the cooling is part of the land- and/or water-scape. Overall, the industry's goal is to enable the ever-increasing demands of computing, not to rethink or curtail them. That is why illustrating the scale of these investments is worthwhile.

It's difficult to know at any given point which data center is used mostly for 'AI' operations as opposed to more traditional data tasks, but we can assume that the world's largest are in some ways implicated with the 'AI' turn, and fueling it. The Meta Fort Worth Data Center in Texas spans over 2.5 million square feet and is designed to provide high-performance computing and storage services for the company's own 'AI' workloads (DPR Construction 2023). A close second is Switch's (2023) SuperNAP in Las Vegas, which spans 2 million square feet and is used by various online retailers for their 'AI' workloads. Similarly, both the Microsoft Quincy (USA) Data Centre and Digital Realty's Mega Data Centre in Singapore—each of which spans around 1.5 million square feet—are used for high-performance computing by companies like IBM and Yahoo for their 'AI' workloads. In Asia, Alibaba's A100 Data Centre in Zhangbei spans over 1 million square feet and is built specifically for 'AI' and high-performance computing (Feifei 2022). Many more data centers are being built or revamped to accommodate the growing demands of 'AI,' and all tout their green energy plans and planet-saving strategies, as the concept of 'sustainability' is well served by corporate ideals of development and innovation.

Environmental 'AI'

Across academic disciplines, 'sustainability' is a contested word. It carries with it implicit and often unacknowledged deep philosophical claims that are entangled with all kinds of problematic power relations (Lally 2021). This moment of enduring climate catastrophe is important because it's not a moment from which we could or should argue for sustainability. It's not enough (Parton 2023). We cannot ask of future generations to do as we have done in the past hundred-plus years; our idea of sustainability, it turns out, hinges on growth by way of destruction, colonialism, extractivism, and exploitation. It's a failed idea. Simply put, sustainability is no longer feasible in the anthropocene. When we consider sustainability through extractivism to be a failed approach, a settler-colonial one, one that leads to ongoing environmental devastation—what can/must we imagine instead? How can this position of living under the threat of climate collapse, not to mention a pandemic, at the time of an industrial 'AI' revolution which requires tremendous computational power, be reconciled? Do these feel like contradictions or necessary parallel movements? With such massive investments made in prediction technologies, whose or what logics are we creating, and following? How does an environment seeped in the promise of ubiquitous computation warp and limit our imagination in thinking about the future?

The observation here is that we are not neutral observers and thinkers and feelers, uninfluenced or unmoved by the environment which we inhabit. This relationship to our world and ideas we form around 'AI' are important if we consider humans to be complex amalgams of embodied expectations, ideologies, desires, and fears (Crawford 2022, 18–19) but also, as many indigenous scholars like Zoe Todd, Leanne Simpson, and Jason Lewis, among others, have argued for so long, understanding humans (and non-humans for that matter) as inseparable from the environment. In short, settler-colonial science and technology sees humans as custodians of nature, where nature is the domain of humans, and the environment as a natural resource to be managed and extracted towards goals of human progress, which are evidenced by science and technology. In broad strokes, Indigenous ways of knowing and being offer a different starting point from settler-colonial visions for 'AI.' Those of us invested in settler-colonial ideals, or privileged by longstanding settler-colonial formations, have inadvertently bought into the data dream as the techno fix.

In this way, 'AI,' as settler-colonial science promises, reveals patterns and hard truths by way of deep learning that will in turn offer us predictions to learn and live from. Never mind that those predictions come largely from us, ultimately, the large datasets we've inadvertently been generating as Internet users. The settler conditions of being for 'AI' to work this way are to separate humans from environment, but also humans from embodiment. Even as we imagine 'AI' for the goal of stewardship and conservation and repair to, for example, make agriculture or city living more efficient, we are not transforming the conditions that have led us here and keep us here. We are limiting the radical potential of environmental politics. While capitalism frames technological innovation in terms of progress, we know that 'AI' is used for mining, oil exploitation, and deforestation—so it's important to note that 'AI' deployments are also not inherently motivated towards climate mitigation or repair, no matter how urgent the situation has become. It depends on what gets funded and deployed.

Recent predictions have been made using 'AI' that we will hit the 1.5-degree Celsius threshold in the next 10–12 years. We are also promised that 'natural' disasters and climate change can be usefully tracked by 'AI.' There are a lot of very impressive 'AI' tools for environmental management, for mapping, conserving and rewilding the planet. 'AI' can also help with military, surveillance, and medical breakthroughs and various other applications in the sciences. 'AI' cars—which are in some ways the face of autonomous 'AI'—come at a great carbon cost because of the underlying technology that requires real-time sensing, camera-based computer vision, seamless data streams, and so on. Aspirations of the autonomous car are also a perfect example of how as an object, 'AI' can offer a solution—though usually towards technocratically efficient ends—but is literally fueled by data in highly costly ways.

That's why 'AI' in itself is difficult to classify as simply good or bad, compelling or troubling. Often, they are both at the same time. However, either way, these are still not sustainable projects in our current instantiation of industrial 'AI'—where endless data streams make more demands on other resources. Thus, the outcome of a particular 'AI'—good, bad, or weird, whatever its value—doesn't mean it can be run in an ecology of technology, which has a technocratic bias, that sees business operating at all costs, a bias that is deeply entrenched in the material infrastructures that shape it, and in turn, that shape possible futures, or even how they can be re-imagined. Once we are comfortably wrapped in a blanket of 'AI,' our ability to think or feel beyond it will be seriously limited.

Conclusion

Instead of looking at these projects for their potential for the environment or as something that can be managed, I'd argue that more time needs to be dedicated to thinking about how 'AI' futurities extend the settler-colonial project of at best custodianship and at worst endless extraction—over nature rather than relationally, or as "kinship," as Jason Lewis and his team (2018), and others, write about it. Resistance might also include feminist standpoint theory—theories from the margins and that questions power—and that undermine the cloak of scientific authority that 'AI' wears to propel its vision forward as objective and logical. In other words, feminist and decolonial critiques of 'AI' are our best bets for bringing us back to a collective and connected mindset.

** See https://www.deepmind.com/ or *Announcing Google DeepMind* (Hassabis 2023).
*** See https://aws.amazon.com/sagemaker/.
**** See https://natick.research.microsoft.com/.
***** See *Et Brain: Exploring New Uses for Data and AI* (Alibaba Cloud 2018).
****** See *Social Media Hyperscale Data Center* (DPR Construction 2012).

These ‘AI’ settler-colonial futures are evidenced in obvious ways by, for example, the way ChatGPT used Kenyan workers to clean up its data (largely for use in the Global North). But there are other ways in which these politics—shrouded in the idea of autonomy and intelligence—cover up the magic of ‘AI’ with a magic trick. None of these critiques of ‘AI’ are new—for decades, scholars have been telling us that technologies like ‘AI’ reflect a value system—in this case, the values of the owners of the means of production in particular, who seem sheltered from the unpredictabilities of the world and averse to the complexities of being a human collective on this planet. Possibly the best illustration of this is when Tesla faked its self-driving car capabilities in a promotional demo video. In a testimonial, a Tesla worker said, “The intent of the video was not to accurately portray what was available for customers in 2016. It was to portray what was possible to build into the system” (Elluswamy 2023). But Elon Musk, the company’s CEO, promoted the video at the time, tweeting that Tesla vehicles require “no human input at all” to drive through urban streets to highways and eventually to find a parking spot (ibidem). I could list hundreds of projects like this—but the point I want to end on is twofold: one is that the hype of ‘AI’ tech functions through a perpetual promise about the future, and that that future—if drawn from datasets of the past—can hardly be predictive of anything more than what has been possible through masculinist, settler-colonial science and engineering. Two, all of these promises are stored in huge data warehouses, powered by mostly coal, some renewables, faulty ideas of carbon offset, and no real accountability for the climate crisis which is now a condition of living with ‘AI.’

‘AI’ is a hot mess. It is hyped and critiqued, all day and night, on social media by scholars, artists, journalists, and industry experts. The hype is the typical tech bluster that most of us are used to by now. But the critiques feel heavier than the usual responses to tech hype; we sense despair, sarcasm, anger and vigilance. Maroussia Lévesque (2023) tweeted the concept of “degenerative AI” to explain the “decay and deterioration of discourse through averaged out good enough guesstimates.” The idea is that what is truly generative must be fresh, tended to, grown in conditions that favor intended outcomes. So, if climate change is not the problem directly being solved by ‘AI’ or any other technology—especially those that harm the environment—then they are part of the problem. Nothing, at this juncture, can justify its existence without centering the planet’s flourishing.

As I hope to have shown in this short piece, sustainability ideals are far from sufficient for dealing with the energy-, water- and land-hungry data centers that power and propel industrial ‘AI’ and its heated future fantasies. Sustainability is in itself a settler-colonial project that fails to recognize that nature is not—and never was—a resource to be extracted. Thinking about ‘AI’ as ‘environmental’ (being) rather than as ‘sustainable’ (doing) can help mitigate the harms underway, and turn our attention back to non-technological approaches towards more just and equitable life on Earth.

Bibliography

Alibaba Cloud. 2018. “Et Brain: Exploring New Uses for Data and AI.” *Alibaba Cloud Community*, April 26, 2018. https://www.alibabacloud.com/blog/et-brain-exploring-new-uses-for-data-and-ai_585388.

Bender, Emily M. (@emilymbender). 2023. “That AI Letter … Here‘s a Quick Rundown,” Twitter Thread, March 29, 2023, 5:36–6:24 AM. https://twitter.com/emilymbender/status/1640920936600997889.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” In *FAccT ’21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922.

Birch, Kean. 2022. “Big Tech, Little Oversight.” *Policy Options*, February 23, 2022. https://policyoptions.irpp.org/magazines/february-2022/big-tech-little-oversight.

Cellan-Jones, Rory. 2020. “Microsoft’s Underwater Data Centre Resurfaces after Two Years.” *BBC.com*, September 14, 2020. https://www.bbc.com/news/technology-54146718.

Crawford, Kate. 2022. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press.

Daigle, Brian. 2021. “Data Centers Around the World: A Quick Look.” *United States International Trade Commission Executive Briefings on Trade*, May 2021. https://www.usitc.gov/publications/332/executive_briefings/ebot_data_centers_around_the_world.pdf.

DeGeurin, Mack. 2023. “Move Aside, Crypto. AI Could Be the Next Climate Disaster.” *Gizmodo*, April 3, 2023. https://gizmodo.com/chatgpt-ai-openai-carbon-emissions-stanford-report-1850288635.

DPR Construction. 2012. “Social Media Hyperscale Data Center.” Accessed May 22, 2023. https://web.archive.org/web/20120618131955/http://www.dpr.com/projects/sweden-data-center.

———. 2023. “Meta Fort Worth Data Center.” Accessed May 22, 2023. https://www.dpr.com/projects/facebook-fort-worth-data-center.

Elluswamy, Ashok. 2023. “Tesla Engineer Testified that Promotional Self-Driving Video Was Staged.” *CBC News*, January 18, 2023. https://www.cbc.ca/news/business/tesla-deposition-self-driving-claim-1.6717564.

Epstein, Alex. 2016. “The Truth About Apple’s ‘100% Renewable’ Energy Usage.” *Forbes.com*, January 8, 2016. https://www.forbes.com/sites/alexepstein/2016/01/08/the-truth-about-apples-100-renewable-energy-usage/?sh=4be7a151189c.

Fagerström, Ola. 2023. “Fact Check: Ola Fagerström and Microsoft’s Surface Emissions Estimator.” *Environmental Variables* (podcast by Asim Hussain), April 26, 2023. Audio, 45:25. https://podcastaddict.com/environment-variables/episode/156784079.

Feifei, Fan. 2022. “Alibaba Sets Up Largest Intelligent Computing Center.” *ChinaDaily.com*, September 1, 2022. https://www.chinadaily.com.cn/a/202209/01/WS631004b6a310fd2b29e75575.html.

Fenwick, McKelvey. 2023a. “Imaginaries.” *The Data Fix* (podcast by Mél Hogan), May 30, 2023. Audio, 51:08. https://shows.acast.com/the-data-fix/episodes/imaginaries-with-fenwick-mckelvey.

———. 2023b. “Let’s Base AI Debates on Reality, not Extreme Fears About the Future.” *The Conversation*, April 3, 2023. https://theconversation.com/lets-base-ai-debates-on-reality-not-extreme-fears-about-the-future-203030.

Future of Life Institute. 2023. “Pause Giant AI Experiments: An Open Letter.” Last modified March 22, 2023. https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

Gebru, Timnit (@timnitGebru). 2023. “TESCREAL Bundle of Ideologies,” Twitter Tweet, February 26, 2023, 10:18 AM. https://twitter.com/timnitGebru/status/1629893565731184640.

Hao, Karen. 2019. “Training a Single AI Model Can Emit As Much Carbon As Five Cars In Their Lifetimes.” *MIT Technology Review*, June 6, 2019. https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes.

Harris, Malcolm. 2023. *Palo Alto: A History of California, Capitalism, and the World*. Boston, MA: Little, Brown and Company.

Hassabis, Demis. 2023. “Announcing Google DeepMind.” *Google DeepMind*, April 20, 2023. https://www.deepmind.com/blog/announcing-google-deepmind.

Hay, Matt. 2023. “AI Might Hurt the Environment Before It Begins to Help the Climate Crisis.” *Evening Standard*, April 26, 2023. https://www.standard.co.uk/comment/comment/ai-chatgpt-damage-hurt-environment-help-climate-crisis-b1076895.html#comments-area.

Hogan, Mél, Dustin Edwards, and Zane Griffin Cooper. 2022. “5 Things About Critical Data Center Studies.” *Commonplace*, October 31, 2022. https://doi.org/10.21428/6ffd8432.af5934aa.

Hölzle, Urs. 2022. “Our Commitment to Climate-Conscious Data Center Cooling.” *The Keyword* (Google), November 21, 2022. https://blog.google/outreach-initiatives/sustainability/our-commitment-to-climate-conscious-data-center-cooling.

Hunger, Francis. 2023. “Unhype Artificial ‘Intelligence’! A Proposal to Replace the Deceiving Terminology of AI.” *Zenodo*, April 12, 2023. https://doi.org/10.5281/zenodo.7524493.

Klein, Naomi. 2023. “AI Machines Aren’t ‘Hallucinating’. But Their Makers Are.” *The Guardian*, May 8, 2023. https://www.theguardian.com/commentisfree/2023/may/08/ai-machines-hallucinating-naomi-klein.

Lally, Róisín, ed. 2021. *Sustainability in the Anthropocene: Philosophical Essays On Renewable Technologies*. Lanham, MD: Lexington Books.

Larsson, Simone. 2022. “Frugal AI: Value at Scale Without Breaking the Bank.” *data iku*, April 22, 2022. https://blog.dataiku.com/frugal-ai-value-at-scale-without-breaking-the-bank.

Lévesque, Maroussia (@Maroussia_____). 2023. “Claiming the Term ‘Degenerative AI’,” Twitter Tweet, May 1, 2023, 3:33 PM. https://twitter.com/Maroussia_____/status/1653150553583525888.

Lewis, Jason Edward, Noelani Arista, Archer Pechawis, and Suzanne Kite. 2018.“Making Kin with the Machines.” *Journal of Design and Science*, July 16, 2018. https://doi.org/10.21428/bfafd97b.

Luccioni, Sasha. 2023a. “The Mounting Human and Environment Costs of Generative AI.” *ArsTechnica*, April 12, 2023. https://arstechnica.com/gadgets/2023/04/generative-ai-is-cool-but-lets-not-forget-its-human-and-environmental-costs.

———. 2023b. “Greener.” *The Data Fix* (podcast by Mél Hogan), March 20, 2023. Audio, 49:57. https://www.thedatafix.net/episodes/009.

Merriam-Webster Dictionary. n.d. “Hot Mess.” Accessed May 22, 2023. https://www.merriam-webster.com/dictionary/hot%20mess.

Ord, Toby. 2020. *The Precipice: Existential Risk and the Future of Humanity*. New York, NY: Hachette Books.

Parton, Dolly. 2023. “Dolly Parton—World on Fire (Official Audio),” YouTube, May 11, 2023. Video, 4:28. https://www.youtube.com/watch?v=MLIGxNZeW78.

Quintarelli, Stefano. 2019. “Let’s Forget the Term AI. Let’s Call Them Systematic Approaches to Learning Algorithms and Machine Inferences (SALAMI).” *Quinta’s Weblog*, November 24, 2019. https://blog.quintarelli.it/2019/11/lets-forget-the-term-ai-lets-call-them-systematic-approaches-to-learning-algorithms-and-machine-inferences-salami.

Saad, Walid. 2023. “AI and its effects on the environment.” *FOX 5 Washington DC* (interview), April 12, 2023. Video, 4:58. https://www.fox5dc.com/video/1206605.

Starosielski, Nicole. 2014. “The Materiality of Media Heat.” *International Journal of Communication* 8, no. 1: 2504–08. https://ijoc.org/index.php/ijoc/article/download/3298/1268.

Torres, Émile P. 2021. “Against Longtermism.” *Aeon*, October 19, 2021. https://aeon.co/essays/why-longtermism-is-the-worlds-most-dangerous-secular-credo.

———. 2023. “Risk.” *The Data Fix* (podcast by Mél Hogan), May 10, 2023. Audio, 54:42. https://www.thedatafix.net/episodes/014.

Strubell, Emma, Ananya Ganesh, and Andrew McCallum. 2019. “Energy and Policy Considerations for Deep Learning in NLP.” *The 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, arXiv, June 5, 2019. https://arxiv.org/abs/1906.02243#.

Switch. 2023. “About Switch.” Accessed May 16, 2023. https://www.switch.com/about.

Troy, Dave. 2023. “The Wide Angle: Understanding TESCREAL—the Weird Ideologies Behind Silicon Valley’s Rightward Turn.” *The Washington Spectator*, May 1, 2023. https://washingtonspectator.org/understanding-tescreal-silicon-valleys-rightward-turn.

Tucker, Emily. 2022. “Artifice and Intelligence.” *Center on Privacy & Technology* (Georgetown Law), March 8, 2022. https://medium.com/center-on-privacy-technology/artifice-and-intelligence%C2%B9-f00da128d3cd.

Vincent, James. 2016. “Mark Zuckerberg Shares Picture from Facebook’s Cold, Cold Data Center.” *The Verge*, September 29, 2016. https://www.theverge.com/2016/9/29/13103982/facebook-arctic-data-center-sweden-photos.

Whittaker, Meredith. 2021. “The Steep Cost of Capture.” *Interactions* 28, no. 6 (November–December): 50–55. https://doi.org/10.1145/3488666.

Wynsberghe, Aimee van. 2021. “Sustainable AI: AI for Sustainability and the Sustainability of AI.” *AI and Ethics* 1, no. 3 (February): 213–18. https://doi.org/10.1007/s43681-021-00043-6.

Zhang, Sharon. 2023. “Report on ChatGPT Model’s Emissions Offers Rare Glimpse of AI’s Climate Impacts.” *Truthout*, April 3, 2023. https://truthout.org/articles/report-on-chatgpt-models-emissions-offers-rare-glimpse-of-ais-climate-impacts.

Künstliche Intelligenz als ‚Hot Mess‘

Mél Hogan

Liar, liar, the world's on fire / Whatcha gonna do when it all burns down /
Fire, fire burning higher / Still got time to turn it all around – Dolly Parton

Dieser Essay skizziert die dem ‚KI'-Hype zugrunde liegenden Philosophien und erklärt, wie die Terminologie selbst zu einer Marketingtaktik wird, die z. T. dazu dient, die sehr realen und drängenden Probleme zu verschleiern, die ‚KI' hervorruft und verschärft. Insbesondere macht dieser Aufsatz darauf aufmerksam, wie ‚KI' sich auf die Umwelt auswirkt und auf einem bereits brennenden, ausgetrockneten und erschöpften Planeten – einem Planeten, der sozial, politisch und ökologisch ausgelaugt ist – technologisch und ideologisch völlig unhaltbar ist.

Am 29. März 2023 twitterte Emily M. Bender einen Beitrag, in dem sie den Offenen Brief des Future of Life Institute (2023) kritisierte, in dem dazu aufgerufen wird, „die gigantischen KI-Experimente zu stoppen". Wie Bender klarstellt, verfolgt das Institut den Ansatz des ‚longtermism' (dt.: Longtermismus), was im Wesentlichen bedeutet, dass es sich als Organisation um Bedrohungen sorgt, die zum Aussterben menschlichen Potenzials führen könnten, in welcher amorphen Form und in welchem Format auch immer dieses ‚Potenzial' in ferner Zukunft auftreten mag. Kritiker*innen weisen jedoch darauf hin, dass sich die sogenannten ‚Longtermists' um Abstrakta sorgen und dabei die tatsächlichen Probleme, mit denen (biologische) Menschen und alle Lebewesen auf der Erde konfrontiert sind, außer Acht lassen (Torres 2021). Im Allgemeinen handelt es sich bei den Longtermists um weiße, wohlhabende Männer, die stark in digitale Technologien investiert haben – sowohl in finanzieller Hinsicht als auch in Bezug auf ihr persönliches Selbstwertgefühl – und die die Vorstellung zukünftiger Menschen (in der Regel als digitale Menschen) pflegen, die in simulierten virtuellen Welten leben, welche von der gewaltigen Energie des Weltraums angetrieben werden, während sie – ironischerweise – gleichzeitig die bösartigen ‚KI'-Technologien verurteilen, die sie mitfinanzieren und entwickeln. Diese imaginierten Bedrohungen sind zu zahlreich, um sie hier aufzuzählen, doch sie reichen von tief verwurzelten algorithmischen Verzerrungen, die das Leben der Menschen ruinieren, über massenhafte Desinformation und tiefgreifende Fälschungen, die eine gemeinsame Realität zerstören und wirtschaftliche, soziale und politische Unruhen verursachen, bis hin zu eher spekulativen Szenarien, in denen eine bösartige ‚Künstliche Intelligenz' kritische Infrastrukturen wie die Wasserversorgung oder die Energieversorgung lahmlegt oder tödliche Waffen automatisiert. Diese Szenarien sind zwar besorgniserregend, übersehen aber die Tatsache, dass jene ‚KI', die für die Verwirklichung des Longtermismus als nutzbringend imaginiert wird, bereits heute den Klimawandel und andere soziale und politische Probleme verschärft (Klein 2023).

Während das longtermistische Denken und Fühlen unausgereift ist, verfügen dessen Akteur*innen über große Entscheidungsmacht, wenn es darum geht, wie und zu welchem Zweck ‚KI' eingesetzt werden könnte. Der Begriff ‚KI' trägt viel dazu bei, die Öffentlichkeit in die Irre zu führen; ‚Künstliche Intelligenz' impliziert unter anderem die Fähigkeit, über eine gewisse Handlungsfähigkeit zu verfügen, und zwar dergestalt, wie es das ‚große Sprachmodell' oder ‚maschinelles Lernen' nicht oder deutlich weniger eindrücklich tun. Stefano Quintarelli (2019) hat dafür plädiert, dass die Öffentlichkeit ganz anders auf die Technologie reagieren würde, wenn ‚KI' als das verstanden würde, was sie ist, nämlich als „Systematic Approaches to Learning Algorithms and Machine Inferences (SALAMI)". In ähnlicher Weise schlägt Francis Hunger (2023) vor, die irreführende Terminologie der ‚KI' zu ersetzen, da sie die Technologie anthropomorphisiert und den Hype anheizt. Derzeit gewinnt ‚KI' in der öffentlichen Imagination erheblich an Boden als eine Technologie, die die Welt revolutionieren wird. Was ‚KI' jedoch eines Tages tatsächlich sein wird, wird von der Industrie gleichzeitig als Versprechen dargestellt, das es zu nutzen gilt, und als Bedrohung, die es zu entschärfen gilt. Sie zahlt dafür, dass diese Version von ‚KI' als Hype zirkuliert und von den Mainstream-Medien aufgegriffen wird. Sie investiert in ihre eigenen Fantasien und verwickelt alle anderen in diese Erzählungen. Das bedeutet, dass die Industrie die Gestaltung der Zukunft unverhältnismäßig stark beeinflusst, obwohl es ihr an echtem Fachwissen und Sorgfalt mangelt. Die ‚KI'-Zukunft, die sich die Longtermists ausma-

len, ist in jeder Hinsicht irrwitzig und brutal. Aber in einer Welt, die von endloser kapitalistischer Ausbeutung angetrieben wird, wird ‚KI' durch alles Mögliche legitimiert, unter dem Deckmantel, dass sie das „menschliche Potenzial" aufrechterhalte, wie es der Longtermist Toby Ord (2020) in seinem Buch *The Precipice* ausdrückt – und zwar als inhärente Unvermeidbarkeit von ‚KI' als wissenschaftlichem Fortschritt und, im Gegenzug, als die Vorstellung, dass „Big Tech" (Birch 2022) weiß, was das Beste für die Zukunftsgestaltung ist.

Es gibt eine ganze Psychologie des Longtermismus, die es erst noch zu Verstehen gilt, insbesondere im Hinblick darauf, wie sie sich mit Transhumanismus, Eugenik und Pronatalismus, effektivem Altruismus, Akzelerationismus und anderen rechtsgerichteten Ideologien von Tech-CEOs überschneidet und überlappt. Einige dieser Bewegungen und ihre Ursprünge wurden unter anderem von Autor*innen/Wissenschaftler*innen wie Malcolm Harris (2023) und Émile P. Torres (2023) analysiert. Torres (2021) befasst sich insbesondere mit der Multimilliarden-Dollar-Bewegung der Longtermists und warnt: „Der Punkt ist, dass der Longtermismus eine der äußerst einflussreichen Ideologien werden könnte, von der jedoch nur wenige Menschen außerhalb der Eliteuniversitäten und des Silicon Valley je gehört haben", und fügt hinzu – und das ist wichtig –, dass Longtermismus „nicht gleichbedeutend ist mit ‚sich um die Langfristigkeit kümmern' oder ‚das Wohlergehen künftiger Generationen schätzen'". Ihre Sprache ist schwammig, aber ihre Idee ist die folgende: Der Mensch als körperliches Wesen habe keinen inhärenten Wert. Vielmehr sei es die Fähigkeit der ‚KI', die Menschheit in eine Art digitales Potenzial zu verwandeln, so der Traum der Longtermists. Aus diesem Grund ist diese Vorbemerkung für alle Diskussionen über ‚KI' sehr wichtig; sie enthüllt sie als ein Werkzeug und eine Taktik mit einer sehr spezifischen Politik – die inzwischen unter dem Acronym TESCREAL (Transhumanismus, Extropianismus, Singularitarismus, Kosmismus, Rationalismus, Effektiver Altruismus und Longtermismus) zusammengefasst wird.* Um es klar zu sagen: Der Einfluss, den der Reichtum der Longtermists (oder allgemeiner von TESCREAL) beispielsweise auf die Bekämpfung des Klimawandels haben könnte, wäre im Hinblick auf die Verbesserung der Lebensqualität von Menschen, Pflanzen und Tieren gewaltig. Man denke hier z. B. an Elon Musk oder Sam Bankman-Fried. Aber für sie ist die Rettung des Planeten nichts Glorreiches – vielmehr erzeugen sie neue Probleme, die sie mit ihren neuen ‚KI'-Spielzeugen lösen können.

Deshalb antizipieren Kritiker*innen wie Bender (2023) zu Recht, dass Entwicklungen, Verwendungen und Regulierungen von ‚KI' durch Longtermists zu einer ‚Hot Mess' führen. Das *Merriam-Webster* Wörterbuch (o. J.) definiert „hot mess" als „etwas, das sich in einem Zustand extremer Unordnung oder Unübersichtlichkeit" befinde und gleichzeitig „attraktiv und sexy" sei. Gerade in den letzten Monaten spiegelt der Medienhype um ‚KI', aber auch der Sturm der Kritik an den überzogenen Versprechungen der ‚KI'-Industrie und ihrer Investor*innen die sexy Unordnung und Verlockungen des Themas auf allen Seiten wider. ‚KI' als ‚Hot Mess' zu bezeichnen, ist auch deshalb so treffend, weil es gleichzeitig Resignation oder Bestürzung signalisiert und – ungewollt – auf die unausweichlichen ökologischen Folgen verweist, und zwar gerade wegen der Verstrickung in longtermistisch orientierte Ideologie (Torres 2021), der Konzentration von Rechenkapazitäten innerhalb von Big Tech (Birch 2022) und dem Klimakollaps, der einen neuen Kontext liefert, in dem sich diese Projekte entfalten. Kurz gesagt, ‚Hot Mess' lenkt die Aufmerksamkeit auf die buchstäbliche Hitze des Zusammenbruchs der ‚KI'. Wie Nicole Starosielski (2014, 2506) vor fast einem Jahrzehnt erklärte, „ist der Wärmeaustausch nicht auf Kommunikationssysteme beschränkt, sondern bewegt sich über und durch Infrastrukturen, Ökologien und Körper", von denen ‚KI' nur die jüngste Iteration und möglicherweise die größte – oder letzte – Bedrohung darstellt. Um es klar zu sagen: Die Bedrohung liegt nicht so sehr in der Technologie selbst – die anders genutzt, kleiner skaliert oder verantwortungsvoller eingesetzt werden könnte – als vielmehr im Denken und Fühlen derjenigen, die an der Spitze von ‚KI'-Projekten und Big-Tech-Unternehmen stehen und anscheinend nicht in der Lage sind, kollektiv oder relational zu agieren. Das Versäumnis, relational zu denken, bedeutet, dass ‚KI' zerstörerisch, ausbeuterisch und extraktivistisch ist.

* Dazu Troy (2023); Gebru (2023).

Jüngste Studien über die Umweltauswirkungen des Trainings und der Nutzung von ‚KI'-Modellen haben gezeigt, dass diese sehr energie- und wasserintensiv sind und extrem hohe Kohlendioxidemissionen verursachen (Luccioni 2023a; Zhang 2023). Die Zahlen in diesen Studien sind zwar bereits exorbitant, berücksichtigen aber noch nicht in vollem Umfang die Emissionen, die mit den Computern, Festplatten oder anderen Geräten oder der materiellen Infrastruktur verbunden sind, die für das Training des Modells hergestellt und verwendet werden, und auch nicht die Art und Weise, in der Modelle wiederholt trainiert und in Zukunft auch noch exponentiell gesteigert werden (Zhang 2023). Bemerkenswert ist in diesem Zusammenhang, dass Bender und ihre Mitarbeiter*innen (2021) – darunter die inzwischen von Google entlassenen ‚KI'-Ethikerinnen Timnit Gebru und Margaret Mitchell – in ihrem viel zitierten Artikel *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* geschrieben haben, dass ‚KI'-Modelle während der Forschung und Entwicklung viele Male trainiert und umtrainiert werden und dass Schätzungen als Mindestwerte angesehen werden sollten. Bislang gibt es nur wenige Studien, die das Ausmaß des wiederholten Trainings deutlich machen.

Im Jahr 2019 wiesen Emma Strubell und ihre Mitarbeiter*innen nach, dass das Training eines einzigen Deep-Learning-Modells zur Verarbeitung natürlicher Sprache zu mehr als 270 Tonnen an Kohlendioxidemissionen führen könnte. Mit anderen Worten: „Das Training eines einzigen KI-Modells kann so viel Kohlendioxid ausstoßen wie fünf Autos in ihrem Lebenszyklus" (einschließlich der Herstellung des Autos) (Hao 2019). In *Gizmodo* wird zur Veranschaulichung dieser Auswirkungen erklärt, dass die Energie, die für „KI-Modelle wie GPT-3 von OpenAI, das dem weltberühmten ChatGPT zugrunde liegt, verwendet wird, das Haus eines durchschnittlichen Amerikaners Hunderte von Jahren lang mit Strom versorgen könnte", eine Feststellung, die auf einem Bericht des Stanford Institute for Human-Centered Artificial Intelligence (HAI) aus dem Jahr 2023 beruht (DeGeurin 2023). Und wie Walid Saad (2023), außerordentlicher Professor für Elektro- und Computertechnik an der Virginia Tech University, kürzlich auf *Fox5 Washington DC* erklärte, verbraucht ‚KI' massive Ressourcen, die bereits knapp und nicht nachhaltig sind. ‚KI' verstärkt den Klimawandel.

Wie bei den meisten Klimaproblemen ist es für Kritiker*innen der industriellen ‚KI' jedoch schwierig, das Ausmaß dieser Auswirkungen einer Öffentlichkeit zu vermitteln, die weder weiß, wie sie sich in diese Diskussionen einbringen kann, noch voll und ganz begreift, was ‚KI' alles umfasst. Mit diesen oft verblüffenden Zahlen können Informatiker*innen jedoch das Bewusstsein für die ‚KI'-Industrie (Amazon, OpenAI, Google, NVIDIA, Baidu, Microsoft, Intel, Meta usw.) oder speziell für die industrielle ‚KI' schärfen. Für eine Folge des Podcasts *The Data Fix* sprach ich mit Sasha Luccioni (2023b), die sich mit den ethischen und gesellschaftlichen Auswirkungen von Modellen und Datensätzen des maschinellen Lernens befasst. Sie machte deutlich, dass die Aufzeichnung von Auswirkungen dazu dienen solle, Standards für die Branche zu schaffen. Es bestehe das Gefühl, dass die Branche rehabilitiert oder zumindest von außen reguliert werden könne, und dass die Messung der Auswirkungen zumindest ein wichtiger Schritt zur Umsetzung politischer Entscheidungen in die Praxis sei.

Nachhaltige ‚KI'

Im Jahr 2021 schrieb Aimee van Wynsberghe *Sustainable AI: AI for sustainability and the sustainability of AI,* in dem sie in drei Wellen skizziert, wie sich ‚KI'-Ethik entwickelt hat. Es ist erwähnenswert, dass es sich dabei mehr oder weniger um dieselben Wellen handelt, die in den kritischen Medienwissenschaften und anderen Disziplinen identifiziert wurden, vielleicht in leicht unterschiedlichen Ausformungen. Die erste Welle zeichnete sich durch die Wahrnehmung einer allgemeinen Bedrohung durch ‚KI' in Form eines Terminator- oder M3GAN-ähnlichen autonomen Roboters oder einer eher utopischen Vorstellung vom Hochladen unserer Gedanken in die Cloud aus – die Furcht vor einer ‚Allgemeinen Künstlichen Intelligenz' (engl.: AGI – ‚Artifical General Intelligence'), ohne sie jedoch so zu nennen. Diese erste Welle war spekulativ

und beschäftigte sich mit dem Potenzial von ‚KI'. Die zweite Welle befasste sich mit praktischeren Themen wie Überwachung, Datenschutz, Blackbox-Algorithmen oder der ‚Erklärbarkeit' von Algorithmen, den Biases in Datensätzen und Suchmaschinen und der mangelnden fairen Darstellung in Big Data bis hin zu den Versprechungen von Gesichtserkennung und Stimmungsanalyse und anderen quantifizierbaren Dingen. Die dritte Welle, so van Wynsberghe, sei „eine, die sich der Umweltkatastrophe unserer Zeit frontal stellt und aktiv versucht, Akademiker*innen, politische Entscheidungsträger*innen, KI-Entwickler*innen und die breite Öffentlichkeit mit den Umweltauswirkungen der KI zu konfrontieren" (van Wynsberghe 2021, 213).

In diesem Artikel liefert van Wynsberghe (2021) überzeugende Argumente für eine ‚bewusste' Arbeit im Bereich der ‚KI' und stellt wie Luccione, Strubell und andere eine Reihe von Lösungen vor, z. B. den Einsatz von Kohlenstoffmessungen und die Berichterstattung darüber als Aufgabe öffentlicher Politik (wir sollten anmerken, dass dies bereits bei Datenzentren und Wasser versucht wurde, gescheitert ist und dennoch eine gute Idee bleibt) (Hölzle 2022), die Verwendung kleinerer Datensätze, wenn möglich, und die Nutzung ortsnaher Rechenzentren, die näher am Ausgangsort der Modellierung und der Experimente liegen, sowie zu Tageszeiten, in denen weniger Energie benötigt wird, und so weiter (Fagerström 2023). Alles sehr praktische, umsetzbare Ideen. Sie geht näher auf diese Abhilfemöglichkeiten ein, die vor allem für ‚KI'-Ingenieur*innen und -Wissenschaftler*innen wichtig sind, also für diejenigen, die direkt mit dieser Technologie arbeiten. Wie mir Fenwick McKelvey (2023a), Co-Direktor des Instituts für angewandte ‚KI' der Concordia University, kürzlich mitteilte, gibt es darauf bereits aus der Industrie heraus eine erste Reaktion: die Idee der sparsamen ‚KI' unter dem Stichwort ‚Frugal AI': „eine Technik, die die Verwendung von weniger Daten und weniger Rechenleistung verspricht und gleichzeitig die Robustheit eines bestimmten KI-Modells innerhalb des vorgesehenen Anwendungsbereichs garantiert" (Larsson 2022).

Für uns Außenstehende genügt es, festzustellen, dass es mehr oder weniger umweltschädliche Wege gibt, mit industrieller ‚KI' zu arbeiten, doch die Lösungsvorschläge drehen sich letztlich darum, die Arbeit aufrechtzuerhalten – ‚KI' als eine Unvermeidlichkeit. Selbst der im März 2023 veröffentlichte Offene Brief des Future of Life Institute (2023), in dem gefordert wird, die ‚KI'-Forschung für sechs Monate zu unterbrechen, ist nichts weiter als ein Depri-Hype, der dazu dient, mehr Aufmerksamkeit auf ‚KI' zu lenken, ohne wirklich plausible Vorschläge, ihre Arbeit zu unterbrechen oder die aktuellen Risiken für Biases, Datenschutz oder Urheberrechte anzugehen.

Wie Meredith Whittaker (2021, 51) schreibt:

> *Dies ist ein gefährlicher Moment. Proprietäre Softwaresysteme, die als künstliche Intelligenz (KI) vermarktet werden, durchdringen unser öffentliches Leben und unsere Institutionen, konzentrieren industrielle Macht, verstärken die Marginalisierung und formen unmerkbar den Zugang zu Ressourcen und Informationen. Wenn wir darüber nachdenken, wie wir diesen Ansturm der industriellen KI bewältigen können, müssen wir zunächst erkennen, dass die im letzten Jahrzehnt gefeierten ‚Fortschritte' in der KI nicht auf grundlegende wissenschaftliche Durchbrüche bei den KI-Techniken zurückzuführen sind. Sie waren und sind in erster Linie das Ergebnis einer erheblichen Konzentration von Daten- und Rechenressourcen, die sich in den Händen einiger weniger großer Technologiekonzerne befinden.*

Emily Tucker (2022), geschäftsführende Direktorin des Center on Privacy & Technology, ein Think Tank am Georgetown Law, der sich auf die Auswirkungen der Überwachungspolitik auf marginalisierte Menschen konzentriert, erklärte: „Ab heute wird das Privacy Center die Begriffe ‚Künstliche Intelligenz', ‚KI' und ‚maschinelles Lernen' in seiner Arbeit nicht mehr verwenden, um die Schäden digitaler Technologien für das Leben von Einzelpersonen und Gemeinschaften aufzudecken und zu mindern." Hier greift sie Whittakers Argument auf, dass ‚KI' nicht nur eine technologische Innovation sei, sondern auch eine diskursive und sprachliche Folie, die uns von den tatsächlichen Mechanismen und Unternehmensakteur*innen ablenke, die am Einsatz von ‚KI' beteiligt seien, und von denjenigen, die für ihr Tun zur Rechenschaft

gezogen werden sollten. Angesichts dieser Warnungen sollten wir entscheiden, ob wir etwas, das sozial und politisch korrupt ist, ‚umweltfreundlicher' machen wollen. Oder ob wir tatsächliche soziale Probleme zuerst auf möglicherweise nicht-technologische Weise angehen wollen – durch Gemeinschaftsbildung, Aktivismus, Bildung und politische Arbeit (Fenwick 2023b).

Schaut man sich die Berichte von Technologieunternehmen über ihre Umweltpraktiken an, so findet man Behauptungen über Nachhaltigkeit und die Bedeutung des Schutzes von Ökosystemen. Oberflächlich betrachtet sieht es so aus, als ob Big Tech sich der Aufgabe stellt, Wasser und globale Energie um der Umwelt willen nachhaltiger zu gestalten, um den Klimawandel zu bekämpfen. Aber im Zuge dieser Behauptungen tauchen immer wieder Probleme auf. Viele dieser Technologieunternehmen mussten feststellen, dass sie andere Unternehmen und Verbraucher*innen für den Erhalt ‚grüner Credits' bezahlen, die den Kohlestromverbrauch ausgleichen sollen. Dies verschleiert die Tatsache, dass der überwiegende Teil des Energieverbrauchs von Computern aus der Kohleverstromung stammt und dass das mit Kohle betriebene Internet mindestens 40 % des Energieverbrauchs ausmacht (Epstein 2016). Unbewusst folgt Big Tech zudem seiner eigenen problematischen Industrielogik. Zum Beispiel bietet Big Tech nur allzu gern eine ‚KI'-Zahnbürste oder einen ‚KI'-Kühlschrank an und behauptet, dass diese effizienter und ‚smarter' seien – aber sogenannte ‚externe Effekte' bleiben unberücksichtigt, z. B. all der Elektroschrott, die geplante Obsoleszenz oder der gesteigerte Konsum, der sich aus diesen angeblichen Innovationen ergibt. Die ‚KI'-Branche hat kein inhärentes Interesse, jene Antriebskräfte zu überwinden, die zur Ausbeutung von Mensch und Natur führen – es ist wahrscheinlicher, dass sie den ökologischen Kollaps verschlimmert, während sie uns Produkte anbietet, die ‚umweltfreundlicher' sind. Es geht also zum einen um das Ersetzen von Gegenständen und Dienstleistungen und zum anderen um die Verankerung von Ideologien und Politik – welche ‚Hot Mess'!

‚KI'-Rechenzentren

Rechenzentren sind die dem Internet zugrunde liegende Infrastruktur zur Datenspeicherung. In Rechenzentren wird auch die Rechenleistung zur Verarbeitung von Daten zentralisiert. Mit der jüngsten Hinwendung aller großen Technologieunternehmen zur ‚KI' hat sich auch die Dateninfrastruktur verändert. ‚KI' erfordert mehr Energie, mehr Wasser, mehr Industriemineralien und mehr Kapital (Hogan et al. 2022). Während möglicherweise einige Komponenten der ‚KI' die Aufgaben im Rechenzentrum effizienter machen und die Emissionen reduzieren könnten, indem sie den Energieverbrauch optimieren, Geräteausfälle vorhersagen und Wartungsaufgaben automatisieren, könnte, wie der Journalist Matt Hay (2023) es ausdrückt, „KI der Umwelt schaden, bevor sie beginnt, in der Klimakrise zu helfen." Die Effizienz der Rechenzentren selbst ist zwar beachtlich, verdeckt aber die sehr realen, sehr dringenden Probleme, die durch das rechnergestützte Denken und Handeln verschlimmert werden.

Mit dem beträchtlichen Wachstum der Daten, die durch ‚KI'-Anwendungen erzeugt werden, sind die Rechenzentren größer geworden, um den steigenden Bedarf an Speicher- und Verarbeitungsleistung zu decken, um Daten zu verarbeiten, Vorhersagen zu machen, Daten zu synthetisieren und zu analysieren und um Cyberangriffe zu bewältigen, da die Sicherheitsrisiken mit ‚KI' zunehmen (Beispiele für Unternehmen, die ‚KI' auf eigenen Servern in ihren Rechenzentren hosten, sind unter anderem Googles DeepMind[**], Amazons SageMaker[***], Microsofts Project Natick[****] und Alibabas ET Brain[*****]). Derartige Unternehmensinitiativen haben zum Bau großer Rechenzentren geführt, die Hunderttausende von Servern beherbergen können, sowie zur Entwicklung von Edge-Datenzentren, um die Rechenleistung näher an die Nutzer*innen zu bringen. Aufgrund der von Servern erzeugten Wärme werden heute in der Regel Flüssigkeitskühlsysteme verwendet, denn sie sind wesentlich effizienter und umweltfreundlicher als Luftkühlsysteme. Zwar befinden sich nach wie vor 30 % der Rechenzentren weltweit in den USA (Daigle 2021), aber es gibt Bestrebungen, Rechenzentren

** Siehe https://www.deepmind.com/ oder vgl. *Announcing Google DeepMind* (Hassabis 2023).
*** Siehe https://aws.amazon.com/sagemaker/.
**** Siehe https://natick.research.microsoft.com/.
***** Vgl. *Et Brain: Exploring New Uses for Data and AI* (Alibaba Cloud 2018).

an Orte wie Skandinavien in Arktisnähe (Vincent 2016)[******] oder unter Wasser in Europa (Cellan-Jones 2020) zu verlegen, sodass die Kühlung Teil der Erd- oder Wasserlandschaft ist. Insgesamt besteht das Ziel der Industrie darin, die ständig steigenden Anforderungen an die Datenverarbeitung zu befriedigen und nicht, sie zu überdenken oder einzuschränken. Aus diesem Grund lohnt es sich, das Ausmaß dieser Investitionen zu veranschaulichen.

Es ist schwer zu sagen, welches Rechenzentrum zu einem bestimmten Zeitpunkt hauptsächlich für ‚KI'-Operationen verwendet wird und nicht für andere, traditionellere Rechenaufgaben. Aber wir können davon ausgehen, dass die größten Rechenzentren der Welt an der ‚KI'-Wende mitbeteiligt sind und sie vorantreiben. Das Meta Fort Worth Data Center in Texas erstreckt sich über mehr als 230.000 qm und ist darauf ausgelegt, Hochleistungsrechen- und Speicherdienste für unternehmenseigene ‚KI'-Aufgaben bereitzustellen (DPR Construction 2023). Dicht gefolgt von SuperNAP der Firma Switch (2023) in Las Vegas, das sich über 185.000 qm erstreckt und von verschiedenen Online-Händler*innen für ihre ‚KI'-Workloads genutzt wird. In ähnlicher Weise werden sowohl das Microsoft Quincy Data Centre (USA) als auch das Mega Data Centre von Digital Reality in Singapur mit einer Fläche von jeweils rund 140.000 qm durch Unternehmen wie IBM und Yahoo für ihre ‚KI'-Arbeit genutzt. In Asien erstreckt sich das A100 Data Centre von Alibaba in Zhangbei über mehr als 93.000 qm und wurde speziell für ‚KI'- und High-Performance-Computing gebaut (Feifei 2022). Viele weitere Rechenzentren werden derzeit errichtet oder umgestaltet, um den wachsenden Anforderungen von ‚KI' gerecht zu werden, und alle werben mit ihren Plänen für grüne Energie und Strategien zur Schonung des Planeten, da das Konzept der ‚Nachhaltigkeit' gut zu den unternehmerischen Entwicklungs- und Innovationsidealen passt.

‚KI' für die Umwelt

In allen akademischen Disziplinen ist ‚Nachhaltigkeit' ein umstrittenes Wort. Es beinhaltet implizite und oft uneingestandene tiefgreifende philosophische Behauptungen, die mit allen Arten von problematischen Machtverhältnissen verwoben sind (Lally 2021). Die in der heutigen Zeit anhaltende Klimakatastrophe ist wichtig, weil dies kein Zeitpunkt mehr ist, an dem wir um Nachhaltigkeit bitten sollten. Das allein reicht nicht aus (Parton 2023). Wir können von künftigen Generationen nicht verlangen, dass sie so handeln, wie wir es in den letzten mehr als einhundert Jahren getan haben; unsere Vorstellung von Nachhaltigkeit, so stellt sich heraus, beruht auf Wachstum durch Zerstörung, Kolonialismus, Extraktivismus und Ausbeutung. Sie ist eine gescheiterte Idee. Einfach ausgedrückt: Nachhaltigkeit ist im Anthropozän nicht länger ausreichend. Wenn wir Nachhaltigkeit durch Extraktivismus für einen gescheiterten Ansatz halten, einen kolonialen Ansatz, der zu fortlaufender Umweltzerstörung führt, was können und müssen wir uns stattdessen vorstellen? Wie lässt sich diese Situation – unter der Bedrohung durch einen Klimakollaps zu leben, ganz zu schweigen durch eine Pandemie – mit einer industriellen ‚KI'-Revolution, die enorme Rechenleistung erfordert, vereinbaren? Ist das ein Widerspruch oder eine notwendige Parallelbewegung? Angesichts massiver Investitionen in Vorhersage-Technologien: Wessen und welcher Logik folgen wir und welche Logik erzeugen wir? Wie formt und limitiert ein Umfeld, das vom Versprechen einer allgegenwärtigen Datenverarbeitung durchdrungen ist, unsere Vorstellung von der Zukunft?

Es geht darum, dass wir in unseren Beobachtungen, unserem Denken und Fühlen nicht neutral sind und von der Umgebung, in der wir leben, nicht unbeeinflusst oder unbewegt bleiben. Diese Beziehung zu unserer Welt und die Vorstellungen, die wir uns von ‚KI' machen, sind wichtig, wenn wir den Menschen als komplexes Amalgam aus verkörperten Erwartungen, Ideologien, Begehren und Ängsten betrachten (Crawford 2022, 18–19), aber auch, wie viele indigene Wissenschaftler*innen wie Zoe Todd, Leanne Simpson und Jason Lewis u. a. seit langem argumentieren, wenn wir den Menschen (und auch Nicht-Menschen) als untrennbar von der Umwelt verstehen. Kurz gesagt, die siedler-koloniale Wissenschaft und Technologie sieht den Menschen als Hüter*in der Natur, wobei die Natur die Domäne des Menschen ist, und die Um-

****** Vgl. *Social Media Hyperscale Data Center* (DPR Construction 2012).

welt als eine natürliche Ressource angesehen wird, die mit dem Ziel des menschlichen Fortschritts, evidenzbasiert durch Wissenschaft und Technologie, verwaltet und extrahiert werden kann. Indigene Wissens- und Seinsweisen bieten – grob gesagt – einen anderen Ausgangspunkt als siedler-koloniale Visionen für ‚KI'. Diejenigen von uns, die in siedler-koloniale Ideale investiert haben oder durch langjährige siedler-koloniale Formationen privilegiert sind, haben ungewollt dem Traum von ‚Daten als schnelle Lösung' Glauben geschenkt.

Wie es siedler-koloniale Wissenschaft verspricht, enthüllt ‚KI' Muster und harte Wahrheiten mittels Deep-Learning, welche uns wiederum Vorhersagen anbieten, auf deren Basis wir leben und lernen. Dabei spielt es keine Rolle, dass diese Vorhersagen größtenteils von uns stammen, und zwar aus den großen Datensätzen, die wir als Internetnutzer*innen unbeabsichtigt erzeugt haben. Die (kolonialen) Siedler-Voraussetzungen dafür, dass ‚KI' auf diese Weise funktioniert, bestehen darin, Menschen von der Umwelt, aber auch, Menschen von ihrer Körperlichkeit zu trennen. Selbst wenn wir uns vorstellen, dass ‚KI' das Ziel verfolgt, die Umwelt zu schützen und zu reparieren, um beispielsweise die Landwirtschaft oder das Leben in der Stadt effizienter zu gestalten, verändern wir nicht die Grundlagen, die uns hierhin geführt haben und die uns hier halten. Wir schränken das radikale Potenzial der Umweltpolitik ein. Während im Kapitalismus technologische Innovationen als Fortschritt dargestellt werden, wissen wir, dass ‚KI' für den Bergbau, die Erdölförderung und die Abholzung von Wäldern eingesetzt wird – daher ist es wichtig, festzustellen, dass der Einsatz von ‚KI' nicht inhärent für den Klimaschutz oder zur Klimareparatur intendiert ist, auch wenn die Situation sehr dringlich geworden ist. Es hängt vielmehr davon ab, was genau finanziert und eingesetzt wird.

Kürzlich wurde mithilfe von ‚KI' vorhergesagt, dass wir in den nächsten 10 bis 12 Jahren die Schwelle von 1,5 Grad Celsius erreichen werden. Man verspricht uns auch, dass ‚natürliche' Katastrophen und der Klimawandel mithilfe von ‚KI' sinnvoll gemessen und nachvollzogen werden können. Es gibt viele sehr beeindruckende ‚KI'-Tools für Umweltmanagement, für Kartierung, den Schutz und die Wiederbelebung des Planeten. ‚KI' kann auch beim Militär, bei der Überwachung, bei medizinischen Erkenntnissen und bei verschiedenen anderen Anwendungen in der Wissenschaft helfen. ‚KI'-Autos – die in gewisser Weise das Gesicht autonomer ‚KI' sind – sind mit hohen Kohlenstoffkosten verbunden, da die zugrundeliegende Technologie Echtzeitsensorik, kamerabasierte Computer Vision und nahtlose Datenströme usw. erfordert. Die Bestrebungen in Bezug auf das autonome Auto sind auch ein perfektes Beispiel dafür, wie ‚KI' objektifiziert eine Lösung bieten kann – wenn auch in der Regel zu technokratisch effizienten Zwecken –, aber durch Daten auf höchst kostspielige Weise buchstäblich angeheizt wird.

Deshalb ist ‚KI' an sich schwer als gut oder schlecht, überzeugend oder beunruhigend zu klassifizieren. Oft ist sie mehreres gleichzeitig. Wie dem auch sei, in der gegenwärtigen Form der industriellen ‚KI' handelt es sich noch immer nicht um nachhaltige Projekte, weil endlose Datenströme immer weitere Ressourcen beanspruchen. Ob gut, schlecht oder seltsam, unabhängig von ihrem Wert – die Ergebnisse, die eine spezifische ‚KI' liefert, rechtfertigen nicht ihren Einsatz in einem technologischen Umfeld, welches dadurch geprägt ist, dass Unternehmen um jeden Preis produzieren müssen – eine Prägung, die tief in die maßgeblichen Infrastrukturen eingeschrieben ist und die wiederum mögliche Zukunftsszenarien und unsere Vorstellungen von Zukunft präkonfiguriert. Wenn wir es uns erst einmal unter einer Decke aus ‚KI' bequem gemacht haben, wird unsere Fähigkeit, darüber hinaus zu denken oder zu fühlen, stark eingeschränkt sein.

Schlussfolgerung

Anstatt diese Projekte im Hinblick auf ihr Potenzial für die Umwelt oder als etwas, das gehandhabt werden kann, zu betrachten, sollte mehr Zeit darauf verwendet werden, darüber nachzudenken, wie ‚KI'-Zukunftsszenarien das koloniale Projekt der Siedler*innen ausweitet, das im besten Fall auf Vormundschaft und im schlimmsten Fall auf

endlose Ausbeutung abzielt – und zwar über die Natur verfügend und nicht im Sinne von „Kinship" (dt. in etwa: Seelenverwandtschaft), wie Jason Lewis et al. (2018) dazu schreiben. Zum Widerstand könnte auch Theorie aus feministischer Perspektive gehören – Theorien, die von den Rändern ausgehen und die Macht infrage stellen – und die den Mantel der wissenschaftlichen Autorität untergraben, den die ‚KI' trägt, um ihre Vision als objektiv und logisch zu propagieren. Mit anderen Worten: Feministische und dekoloniale Kritiken sind unsere beste Chance, uns zu einer kollektiven und vernetzten Denkweise zurückzubringen.

Diese siedler-kolonialen Zukünfte von ‚KI' zeigen sich beispielsweise in der Art und Weise, wie ChatGPT kenianische Arbeiter*innen zur Bereinigung seiner Daten einsetzt (die dann größtenteils im Globalen Norden verwendet werden). Aber es gibt noch andere Wege, auf denen diese Politik – umhüllt von der Idee der Autonomie und der Intelligenz – die Magie der ‚KI' mit einem Zaubertrick verschleiert. Keine dieser Kritiken an ‚KI' ist neu – seit Jahrzehnten weisen uns Wissenschaftler*innen darauf hin, dass Technologien wie ‚KI' ein bestimmtes Wertesystem widerspiegeln – in diesem Fall vor allem die Werte der Eigentümer*innen der Produktionsmittel, die vor den Unwägbarkeiten der Welt geschützt zu sein scheinen und die von der Komplexität des Daseins als menschlichem Kollektiv auf diesem Planeten nichts wissen wollen. Das vielleicht beste Beispiel dafür ist ein Werbevideo, in dem Tesla die Fähigkeiten seines selbstfahrenden Autos vortäuschte. In einer Stellungnahme sagte ein Tesla-Mitarbeiter: „Die Absicht des Videos war nicht, genau darzustellen, was den Kunden im Jahr 2016 zur Verfügung stand. Es ging darum, darzustellen, was in das System eingebaut werden könnte" (Elluswamy 2023). Elon Musk, der Vorstandsvorsitzende des Unternehmens, warb damals für das Video und twitterte, dass Tesla-Fahrzeuge „keinerlei menschliches Zutun" benötigten, um durch städtische Straßen auf Autobahnen zu fahren und schließlich einen Parkplatz zu finden (ebenda). Ich könnte Hunderte solcher Projekte aufzählen, aber ich möchte mit zwei Punkten enden: Erstens, dass der Hype der ‚KI'-Technologie durch ein ständiges Versprechen über die Zukunft funktioniert, und dass diese Zukunft – wenn sie aus Datensätzen der Vergangenheit gezogen wird – kaum mehr vorhersagen kann als das, was maskulinistische, siedler-koloniale Wissenschaft und Technik ermöglicht hat. Zweitens sind all diese Versprechungen in riesigen Datenzentren gespeichert, die größtenteils mit Kohle, ein paar erneuerbaren Energien und fehlerhaften Versprechungen vom Kohlendioxidausgleich betrieben werden und keine wirkliche Verantwortung für die Klimakrise übernehmen, die nun das Leben mit ‚KI' bestimmt.

‚KI' ist eine ‚Hot Mess'. Sie wird Tag und Nacht in den sozialen Medien von Wissenschaftler*innen, Künstler*innen, Journalist*innen und Branchenexpert*innen angepriesen und kritisiert. Der Hype ist ein typisches technisches Getöse, an das die meisten von uns inzwischen gewöhnt sind. Aber die Kritiken fühlen sich heftiger an als die üblichen Reaktionen auf den Tech-Hype; wir spüren Verzweiflung, Sarkasmus, Wut und Wachsamkeit. Maroussia Lévesque (2023) twitterte das Konzept der „degenerativen KI", um den „Verfall und die Verschlechterung des Diskurses durch an Mittelwerten orientierten, gerade so akzeptablen, Schätzungen" zu erklären. Die Idee ist, dass das, was wirklich generativ ist, frisch sein und gepflegt werden muss und unter Bedingungen wächst, die die beabsichtigten Ergebnisse begünstigen. Wenn also der Klimawandel nicht das Problem ist, das durch ‚KI' oder andere Technologien – insbesondere solche, die der Umwelt schaden – direkt gelöst werden soll, dann sind sie Teil des Problems. Nichts kann zum jetzigen Zeitpunkt seine Existenz rechtfertigen, ohne das Gedeihen des Planeten in den Mittelpunkt zu stellen.

Wie ich in diesem kurzen Beitrag hoffentlich gezeigt habe, reichen die Ideale der Nachhaltigkeit bei weitem nicht aus, um mit den energie-, wasser- und flächenhungrigen Rechenzentren fertig zu werden, die die industrielle ‚KI' und ihre aufgeheizten Zukunftsphantasien an- und vorantreiben. Nachhaltigkeit ist an sich schon ein siedlungskoloniales Projekt, das verkennt, dass die Natur keine Ressource ist – und auch nie war –, die man ausbeuten kann. ‚KI' als ‚ökologisch' (Sein) und nicht als ‚nachhaltig' (Tun) zu betrachten, kann dazu beitragen, die laufenden Schäden zu mindern und unsere Aufmerksamkeit wieder auf nichttechnologische Ansätze für ein gerechteres Leben auf der Erde zu lenken.

Literaturverzeichnis

Alibaba Cloud. 2018. „Et Brain: Exploring New Uses for Data and AI“. *Alibaba Cloud Community*, 26. April 2018. https://www.alibabacloud.com/blog/et-brain-exploring-new-uses-for-data-and-ai_585388.

Bender, Emily M. (@emilymbender). 2023. „That AI Letter … Here's a Quick Rundown“, Twitter Thread, 29. März 2023, 5:36–6:24 Uhr. https://twitter.com/emilymbender/status/1640920936600997889.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major und Shmargaret Shmitchell. 2021. „On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?“. In *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922.

Birch, Kean. 2022. „Big Tech, Little Oversight“. *Policy Options*, 23. Februar 2022. https://policyoptions.irpp.org/magazines/february-2022/big-tech-little-oversight.

Cellan-Jones, Rory. 2020. „Microsoft's Underwater Data Centre Resurfaces after Two Years“. *BBC.com,* 14. September 2020. https://www.bbc.com/news/technology-54146718.

Crawford, Kate. 2022. *Atlas of AI: Power, Politics and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale, University Press.

Daigle, Brian. 2021. „Data Centers Around the World: A Quick Look“. *United States International Trade Commission Executive Briefings on Trade*, Mai 2021. https://www.usitc.gov/publications/332/executive_briefings/ebot_data_centers_around_the_world.pdf.

DeGeurin, Mack. 2023. „Move Aside, Crypto. AI Could Be The Next Climate Disaster“. *Gizmodo*, 3. April 2023. https://gizmodo.com/chatgpt-ai-openai-carbon-emissions-stanford-report-1850288635.

DPR Construction. 2012. „Social Media Hyperscale Data Center“. Aufgerufen am 22. Mai 2023. https://web.archive.org/web/20120618131955/http://www.dpr.com/projects/sweden-data-center.

———. 2023. „Meta Fort Worth Data Center“. Aufgerufen am 22. Mai 2023. https://www.dpr.com/projects/facebook-fort-worth-data-center.

Elluswamy, Ashok. 2023. „Tesla Engineer Testified that Promotional Self-Driving Video Was Staged“. *CBC News*, 18. Januar 2023. https://www.cbc.ca/news/business/tesla-deposition-self-driving-claim-1.6717564.

Epstein, Alex. 2016. „The Truth About Apple's '100% Renewable' Energy Usage“. *Forbes.com*, 8. Januar 2016. https://www.forbes.com/sites/alexepstein/2016/01/08/the-truth-about-apples-100-renewable-energy-usage/?sh=4be7a151189c.

Fagerström, Ola. 2023. „Fact Check: Ola Fagerström and Microsoft's Surface Emissions Estimator“. *Environmental Variables* (Podcast von Asim Hussain), 26. April 2023. Audio, 45:25. https://podcastaddict.com/environment-variables/episode/156784079.

Feifei, Fan. 2022. „Alibaba Sets Up Largest Intelligent Computing Center“. *ChinaDaily.com*, 1. September 2022. https://www.chinadaily.com.cn/a/202209/01/WS631004b6a310fd2b29e75575.html.

Fenwick, McKelvey. 2023a. „Imaginaries“. *The Data Fix (*Podcast v. Mél Hogan), 30. Mai 2023. Audio, 51:08. https://shows.acast.com/the-data-fix/episodes/imaginaries-with-fenwick-mckelvey.

———. 2023b. „Let's Base AI Debates on Reality, not Extreme Fears About the Future“. *The Conversation*, 3. April 2023. https://theconversation.com/lets-base-ai-debates-on-reality-not-extreme-fears-about-the-future-203030.

Future of Life Institute. 2023. „Pause Giant AI Experiments: An Open Letter“. Letzte Änderung 22. März 2023. https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

Gebru, Timnit (@timnitGebru). 2023. „TESCREAL Bundle of Ideologies“, Twitter Tweet, 26. Februar 2023, 10:18 Uhr. https://twitter.com/timnitGebru/status/1629893565731184640.

Hao, Karen. 2019. „Training As a Single AI Model Can Emit As Much Carbon As Five Cars In Their Lifetimes“. *MIT Technology Review*, 6. Juni 2019. https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes.

Harris, Malcolm. 2023. *Palo Alto: A History of California, Capitalism, and the World*. Boston, MA: Little, Brown and Company.

Hassabis, Demis. 2023. „Announcing Google DeepMind“. *Google DeepMind*, 20. April 2023, https://www.deepmind.com/blog/announcing-google-deepmind.

Hay, Matt. 2023. „AI Might Hurt The Environment Before It Begins To Help The Climate Crisis“. *Evening Standard*, 26. April 2023. https://www.standard.co.uk/comment/comment/ai-chatgpt-damage-hurt-environment-help-climate-crisis-b1076895.html#comments-area.

Hogan, Mél, Dustin Edwards und Zane Griffin Cooper. 2022. „5 Things About Critical Data Center Studies“. *Commonplace*, 31. Oktober 2022. https://doi.org/10.21428/6ffd8432.af5934aa.

Hölzle, Urs. 2022. „Our Commitment to Climate-Conscious Data Center Cooling“. *The Keyword* (Google), 21. November 2022. https://blog.google/outreach-initiatives/sustainability/our-commitment-to-climate-conscious-data-center-cooling.

Hunger, Francis. 2023. „Unhype Artificial 'Intelligence'! A Proposal to Replace the Deceiving Terminology of AI“. *Zenodo*, 12. April 2023. https://doi.org/10.5281/zenodo.7524493.

Klein, Naomi. 2023. „AI Machines Aren't 'Hallucinating'. But Their Makers Are“. *The Guardian*, 8. Mai 2023. https://www.theguardian.com/commentisfree/2023/may/08/ai-machines-hallucinating-naomi-klein.

Lally, Róisín, Hrsg. 2021. *Sustainability in the Anthropocene: Philosophical Essays on Renewable Technologies*. Lanham, MD: Lexington Books.

Larsson, Simone. 2022. „Frugal AI: Value at Scale Without Breaking the Bank“. *data iku*, 22. April 2022. https://blog.dataiku.com/frugal-ai-value-at-scale-without-breaking-the-bank.

Lévesque, Maroussia (@Maroussia_____). 2023. „Claiming the Term ‚Degenerative AI'“, Twitter Tweet, 1. Mai 2023, 15:33 Uhr. https://twitter.com/Maroussia_____/status/1653150553583525888.

Lewis, Jason Edward, Noelani Arista, Archer Pechawis und Suzanne Kite. 2018. „Making Kin with the Machines“. *Journal of Design and Science*, 16. Juli 2018. https://doi.org/10.21428/bfafd97b.

Luccioni, Sasha. 2023a. „The Mounting Human and Environment Costs of Generative AI“. *ArsTechnica*, 12. April 2023. https://arstechnica.com/gadgets/2023/04/generative-ai-is-cool-but-lets-not-forget-its-human-and-environmental-costs.

———. 2023b. „Greener“. *The Data Fix (*Podcast v. Mél Hogan), 20. März 2023. Audio, 49:57. https://www.thedatafix.net/episodes/009.

Merriam-Webster Dictionary. o. J. „Hot Mess“. Aufgerufen am 22. Mai 2023. https://www.merriam-webster.com/dictionary/hot%20mess.

Ord, Toby. 2020. *The Precipice: Existential Risk and the Future of Humanity*. New York, NY: Hachette Books.

Parton, Dolly. 2023. „Dolly Parton – World on Fire (Official Audio)“, YouTube, 11. Mai 2023. Video, 4:28. https://www.youtube.com/watch?v=MLIGxNZeW78.

Quintarelli, Stefano. 2019. „Let's Forget the Term AI. Let's Call Them Systematic Approaches to Learning Algorithms and Machine Inferences (SALAMI)“. *Quinta's Weblog*, 24. November 2019. https://blog.quintarelli.it/2019/11/lets-forget-the-term-ai-lets-call-them-systematic-approaches-to-learning-algorithms-and-machine-inferences-salami.

Saad, Walid. 2023. „AI and its effects on the environment“. *FOX 5 Washington DC* (Interview), 12. April 2023. Video, 4:58. https://www.fox5dc.com/video/1206605.

Starosielski, Nicole. 2014. „The Materiality of Media Heat“. *International Journal of Communication* 8, Nr. 1: 2504–8. https://ijoc.org/index.php/ijoc/article/download/3298/1268.

Torres, Émile P. 2021. „Against Longtermism“. *Aeon*, 19. Oktober 2021. https://aeon.co/essays/why-longtermism-is-the-worlds-most-dangerous-secular-credo.

———. 2023. „Risk“. *The Data Fix (*Podcast v. Mél Hogan), 10. Mai 2023. Audio, 54:42. https://www.thedatafix.net/episodes/014.

Strubell, Emma, Ananya Ganesh und Andrew McCallum. 2019. „Energy and Policy Considerations for Deep Learning in NLP“. *The 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, arXiv, 5. Juni 2019. https://arxiv.org/abs/1906.02243#.

Switch. 2023. „About Switch“. Aufgerufen am 16. Mai 2023. https://www.switch.com/about.

Troy, Dave. 2023. „The Wide Angle: Understanding TESCREAL—the Weird Ideologies Behind Silicon Valley's Rightward Turn“. *The Washington Spectator*, 1. Mai 2023. https://washingtonspectator.org/understanding-tescreal-silicon-valleys-rightward-turn.

Tucker, Emily. 2022. „Artifice and Intelligence“. *Center on Privacy & Technology* (Georgetown Law), 8. März 2022. https://medium.com/center-on-privacy-technology/artifice-and-intelligence%C2%B9-f00da128d3cd.

Vincent, James. 2016. „Mark Zuckerberg Shares Picture from Facebook's Cold, Cold Data Center“. *The Verge*, 29. September 2016. https://www.theverge.com/2016/9/29/13103982/facebook-arctic-data-center-sweden-photos.

Whittaker, Meredith. 2021. „The Steep Cost of Capture“. *Interactions* 28, Nr. 6 (November–December): 50–55. https://doi.org/10.1145/3488666.

Wynsberghe, Aimee van. 2021. „Sustainable AI: AI for Sustainability and the Sustainability of AI“. *AI and Ethics* 1, Nr. 3 (Februar): 213–18. https://doi.org/10.1007/s43681-021-00043-6.

Zhang, Sharon. 2023. „Report on ChatGPT Model's Emissions Offers Rare Glimpse of AI's Climate Impacts“. *Truthout*, 3. April 2023. https://truthout.org/articles/report-on-chatgpt-models-emissions-offers-rare-glimpse-of-ais-climate-impacts.

In Conversation with Magda Tyżlik-Carver on “Curating Data”

Magda Tyżlik-Carver researches relational networks of relationships between humans and the non-human. In the interview, she relates these notions to concepts of the post-human as curatorial entities, that is, to algorithms, bots, software, and computing infrastructure.

Transcript of the interview with Magda Tyżlik-Carver, conducted by Francis Hunger on 2021–04–22.

Hunger: In regard to curation you pointed out a genealogy on which your notion of 'post-human curation' builds. Could you describe this concept a little bit more for us?

Tyżlik-Carver: It starts with the figure of the curator, the next figure is the figure of curating, and then there is the figure of the curatorial. I needed to introduce these three figures in order to move to the fourth figure, which is this figure of a post-human curating. And these three figures are really helpful, because they point to the moments in the history of curating where something changed in the practice. So, if we start with 'curators' as these persons who emerged, then they emerged with museums and collections. And their job really was to look after these collections. But their job was also to collect, to get to know the objects in the collections, what they are, to trace their origin, keep inventory. And that's why often when we talk about curators, we're talking about a sort of guard or caretaker. And then 'curating,' which is this next figure and refers to a moment when it starts to define a set of practices, when it's not just about collections, but also how to exhibit them, how to put them known to the public, how to contextualize them, also how to organize them in an exhibition space or how to place them in relation to each other. And of course, these practices already existed in museums, but especially in the mid-century with the rise of curators as well as curatorial approaches, which were much more idiosyncratic and also developed to accommodate new art objects. These were often conceptual objects in minimalist art, or even the material objects, as described by Lucy Lippard. So, these are all different examples of how curating started to expand beyond the traditional spaces, where this caring for objects happened. The third figure is the figure of 'the curatorial,' which is proposed as a philosophy of curating by Jean-Paul Martinon. This philosophy of curating is really a result of a period, lasting two decades or longer, where curatorial graduate degrees proliferated, and where art biennials also helped to increase the importance of the independent curator within the art world. And so what is interesting here is that the curatorial refers to this field of knowledge and practice and the curating becomes not just about putting objects in space, but also generating knowledge around that, not only about the objects, but also how these objects are related in the exhibition context, how they are contextualized, and so on. Now we can introduce the fourth figure, which is 'post-human curating,' and this figure complicates this idea of curating that is done just by people and that is just about artworks or other cultural objects and that only takes place in art or cultural institutions. So, that's quite important and so in the context of my research into post-human curating, on the one hand we can say that machines and software and platforms become part of curating and automation is increasingly an element of curatorial practice. And we are aware of algorithmic curating and content curating where curating is shared and distributed across human and non-human elements especially on [social media] platforms. And then on the other hand, I also think of curating where users and not just curators curate the content and work within the institutions of platforms and blogs, and so on.

Hunger: You've already mentioned the notion of data curation. One could say it's not just that everyone is an artist today, but everyone is a curator—a data curator in a way. In your opinion, what would be the ideal education for a data curator, compared to an art curator?

Tyżlik-Carver: For me, what is important is not actually looking for a difference between the art curator's and the data curator's educations, but rather to think about what are possible pedagogies that perhaps could bring together the art and data elements. And in the same way where art is this thing that allows us to sort of poke at things, not take them for granted and unpack them at the same time and maybe take apart and deconstruct and reconstruct them in a slightly different way. And it would be helpful to also think about how data curating is a problem that is considered solely part of computer science and data science. This is a massive issue, because we are so implicated in it, there is a need for more literacy, but also not just to know what's happening and how to process data, but also how to actually intervene in that. How to be like an artist, how to unpack it, how to poke at it—and first and most importantly—how not to take it for granted, because that is a massive issue at the moment.

Hunger: Let's sharpen the notion of the 'post-human' a little bit more at this point. When we refer to algorithms, to pattern recognition, to databases, to networks, we refer usually to machine automation. Is that already a reference to the non-human? Because in the end, it's still humans who have set all those non-human agents in motion.

Tyżlik-Carver: Post-humanism helps to think beyond this dichotomy between human and non-human, man and woman, and so on. What we have to face again and again is the fact that it is not so straightforward to dismantle these relations. Algorithms, databases, automation, pattern recognition are not naturally occurring phenomena, but results of material practices of certain histories and other elements that condition how these tools are deployed. How are they used, who is using them, and on whom? So, once we start to question this, we can see that—with the work of a number of people who have revealed the bias that exists in different algorithmic models or in databases, and so on—what actually happens is that this Humanism with a capital 'H' still exists. It is now especially visible in the form of white supremacy, which can be observed in systematic injustice that is executed through facial recognition algorithms, for example, that especially struggle with recognizing or differentiating between faces that are Black. Or in transphobic algorithms that have also bias in the gender recognition software often used by corporations. So, these are really interesting examples. And my reference to post-human and to post-humanism goes through feminism and a materialist understanding which is also a historical one. And in Rosi Braidotti's definition of post-humanism—one of the ways in which she defines post-humanism—is that it's this historical moment that marks the end of the opposition between humanism and anti-humanism. Namely, between this Humanism with a capital 'H,' the Humanism of the Enlightenment, and anti-humanism, the rejection of 'human' as a category. It is more focused on tracing frameworks or looking for more affirmative models and real alternatives. And this is helpful, because it introduces this idea that this critique also needs to be followed by action, where we are obviously suggesting different ways out of the madness that we see around us, which a lot of the time is so destructive and depressing. What are these imaginative powers that could help us in thinking about the different models and different ways of working with algorithms, and that all data has to be extractivist? What does it mean that we extract and do we always have to act in an extractivist way? Do we need to build certain models at all? So, these are the questions that we need to ask, and post-humanism, or the way in which Braidotti and Haraway and others work with it, this new materialist approach, is really opening that up. And not to completely remove the human from the picture, not at all, but actually to re-center and look at the relationality that exists in these systems, that it's not just us, but it's us in relation to others and other things. When we say 'us,' who do we mean?

Hunger: Thank you. In your recent paper, *Curating Data Infrastructures of Control and Affect ... and Possible Beyonds* (2021), you stated, and I'm quoting from this paper: "The practice of curating data is also an epistemological practice that needs interventions to consider futures, but also account for the past. This can be done by asking where data come from. The task in curating data is to reclaim their traceability and to account for their lineage."

Tyżlik-Carver: This definitely builds on a scholarship that is de-colonial and works with other practices and specifically looks at de-colonial museums, where we can already trace this complex and difficult past that we have to deal with, as curators or as Europeans who are implicated in this history in a particular way. And we can actually learn so much from that work that has been done in relation to physical objects that have been taken and stolen and moved to Europe, or bodies that have been taken and stolen and moved across the Atlantic. Actually, what's happening today with data practices is based on similar values of extractions and taking away without asking for permission and without negotiating positions of that taking away. The question is: what is 'data' in that sense? When we think about it, not just in terms of 'data is this set of numbers organized in tables,' for example, but we actually think about 'what is the process of taking that data?' It is not a given, we have to take it. And so, what does it mean to take that data? And when I talk about accounting for the lineage, it is really to face the truth of the violence of taking data. Also with subjects like Big Data. This subject exists in our imaginary so strongly that in a sense it becomes a body itself, and a body with life. It has its own life. So, in the article I actually refer to a number of examples, but one that I discuss a little bit deeper is an artwork by Sissel Marie Tonn and Jonathan Reus, two artists based in the Netherlands. In that work they use seismic data from the Groningen area in the Netherlands, which is an area that has been affected by man-made earthquakes that can be traced back to the mining industry which was there, and translate this gathered data into a bodily experience through touch and through movement, sound and contraction. So, what they do is that they make these vests that can be worn on the body which make your body shake. When you enter, you can imagine that it's almost like a database that is spatially distributed and where the data is brought from this mechanical or automated way back into the human environment, where it is possible to make it more affective and bring it back to enter human life, not just computational life. This is also going back to my previous point of thinking 'why is data just within certain disciplines'—computer science and data science—that see data in one particular way? Whereas, if we look at the examples of work, including yours and a number of others, it's just such a rich proposition to see and imagine what data is. And that is so important to good life! If we think of data just as numbers in tables, we end up with continuous bias and that is impossible to solve, simply because we just have to face the fact that we are subjects that are situated and have our opinions. And it's very difficult to get rid of that and pretend that databases don't carry all this baggage with them. Whereas, in examples like these artistic projects that use data, and with a lot of empathy, it opens up a possibility of how we can actually work with data.

Scan the QR code below to watch the full video interview online.

Im Gespräch mit
Magda Tyżlik-Carver über „Kuratieren von Daten“

Magda Tyżlik-Carver erforscht relationale Beziehungsnetze zwischen Menschen und dem Nicht-Menschlichen. Sie setzt im Interview diese Begriffe in Beziehung zu Konzepten des Posthumanen als kuratorische Entitäten, d. h. zu Algorithmen, Bots, Software und Computerinfrastrukturen.

Transkript des Interviews mit Magda Tyżlik-Carver, geführt von Francis Hunger am 22–04–2021.

Hunger: In Bezug auf das Kuratieren hast du auf eine Genealogie verwiesen, auf die dein Begriff des ‚Post-Human Curating' aufbaut. Könntest du dieses Konzept etwas ausführlicher für uns beschreiben?

Tyżlik-Carver: Es beginnt mit der Figur des / der Kurator*in, dann ist die nächste Figur die des Kuratierens, und dann gibt es die Figur des Kuratorischen. Und ich musste diese drei Figuren einführen, um zu der vierten Figur zu kommen, welches die Figur des posthumanen Kuratierens ist. Ich denke, dass diese drei Figuren wirklich hilfreich sind, weil sie auf die Momente in der Geschichte des Kuratierens hinweisen, in denen sich etwas in der Praxis verändert hat. Wenn wir also mit ‚Kurator*innen' als Personen beginnen, dann treten sie in Museen und Sammlungen auf. Und ihre Aufgabe war es, sich um die Sammlungen zu kümmern. Ihre Aufgabe war es auch, zu sammeln, die Objekte in den Sammlungen einzuordnen, zu wissen, worum es sich handelt, die Herkunft zurückzuverfolgen und den Bestand zu inventarisieren. Und das ist der Grund, warum wir oft, wenn wir über Kurator*innen sprechen, von einer Art Bewacher*in oder Betreuer*in sprechen. Und als nächstes: ‚Kuratieren', das ist die nächste Figur, die sich auf einen Moment bezieht, in welchem wir nicht mehr nur über Kurator*innen sprechen, sondern wenn ‚Kuratieren' eine Reihe von Praktiken beschreibt, wenn es nicht nur um Sammlungen geht, sondern auch darum, wie man sie ausstellt, wie man sie der Öffentlichkeit zugänglich macht, wie man sie kontextualisiert, auch wie man Objekte in einem Ausstellungsraum organisiert oder wie man sie in Beziehung zueinander setzt. Und natürlich gab es diese Praktiken bereits in den Museen, aber ich denke, sie traten dann besonders in der Mitte des 20. Jahrhunderts auf, mit dem Aufkommen von Kurator*innen und kuratorischen Ansätzen, die viel eigenwilliger und auch ausgefeilter waren, um eine neue Art von Kunstobjekten einzubinden. Das waren oft konzeptuelle Objekte, z. B. minimalistische Kunst, oder auch die materiellen Objekte, wie sie Lucy Lippard beschrieben hat. Das sind verschiedene Beispiele dafür, wie das Kuratieren begann und sich über die traditionellen kuratorischen Räume hinaus ausdehnte, in denen das Kuratieren und die Bewahrung der Objekte ursprünglich stattfand. Die dritte Figur ist die Figur des ‚Kuratorischen', die als eine Art Philosophie des Kuratierens von Jean-Paul Martinon vorgeschlagen wird. Diese Philosophie des Kuratierens ist das Ergebnis eines Prozesses von zwei Jahrzehnten oder länger, in dem nach und nach Kurator*innen mit Universitätsabschlüssen auftraten, und in dessen Zuge Kunstbiennalen dazu beitrugen, die Bedeutung der unabhängigen Kurator*innen innerhalb der Kunstwelt zu steigern. Interessant ist hier, dass sich das Kuratorische auf dieses Feld des Wissens und der Praxis bezieht und dass es beim Kuratieren nicht mehr nur darum geht, Objekte im Raum zu platzieren, sondern auch darum, kontextuelles Wissen zu generieren, nicht nur über die Objekte, sondern auch darüber, wie diese Objekte im Ausstellungskontext in Beziehung zueinander stehen, wie sie kontextualisiert werden und so weiter. Nun führen wir die vierte Figur ein, das ‚posthumane Kuratieren', welche die bisherige Idee des Kuratierens verkompliziert, die nur von Menschen gemacht wird, sich nur um Kunstwerke oder andere kulturelle Objekte dreht, und die nur in Kunst- oder Kulturinstitutionen stattfindet. Das ist wichtig, und im Kontext meiner Forschung zum posthumanen Kuratieren ist es wirklich interessant, dass einerseits Maschinen, Software und Plattformen Teil des Kuratierens werden. Es zeigt sich, wie Automatisierung zunehmend ein Element der kuratorischen Praxis wird, und wir wissen um das algorithmische Kuratieren und Content-Kuratieren besonders auf Online-Plattformen und ähnlichem, welches sich über menschliche und nicht-menschliche Elemente verteilt. Und andererseits ist auch über Kuratieren nachzudenken, wo Nutzer*innen und nicht nur Kurator*innen den Inhalt kuratieren und sozusagen innerhalb von Institutionen wie Online-Plattformen und Blogs arbeiten.

Hunger: Du hast bereits den Begriff der Datenkuration erwähnt. Man könnte zugespitzt sagen, heute ist nicht mehr jede*r ein*e Künstler*in, sondern jede*r ist ein*e Kurator*in – ein*e Datenkurator*in. Was wäre deiner Meinung nach die ideale Ausbildung eines/einer Datenkurator*in, verglichen mit dem/der Kunstkurator*in?

Tyżlik-Carver: Für mich ist der Unterschied zwischen der Ausbildung eines/einer Kunstkurator*in und der Ausbildung eines/einer Datenkurator*in nicht so wichtig. Mir geht es eher darum, darüber nachzudenken, was mögliche pädagogische Methoden wären, welche die Kunst- und Datenelemente zusammenbringen könnten. Und zwar auf die gleiche Weise, wie Kunst es uns erlaubt, die Dinge zu hinterfragen, sie nicht als selbstverständlich zu betrachten und gleichzeitig zu analysieren und vielleicht einen Teil zu nehmen und zu dekonstruieren und auf eine etwas andere Art neu zu rekonstruieren. Und man könnte darüber nachdenken, wie Datenkuration stärker inkludiert werden könnte, weil es ein Problem ist, dass Datenkuration allein als Teil der Informatik und Data Sciences betrachtet wird. Das ist ein massives Problem. Ich glaube, weil wir so sehr darin verwickelt sind, benötigen wir mehr Bildungskompetenz in Bezug auf Daten. Es ist wichtig, nicht nur zu wissen, wie Daten verarbeitet werden, sondern auch, wie man tatsächlich interveniert, wie man in eine Künstler*innenrolle schlüpft und das Geschehen analysiert, wie man es ausprobiert und vor allem, wie man es nicht einfach als gegeben hinnimmt, denn ich glaube, das ist im Moment ein großes Problem.

Hunger: Lass uns an dieser Stelle den Begriff des ‚Posthumanen' noch etwas präzisieren. Wenn wir von Algorithmen, von Mustererkennung, von Datenbanken, von Netzwerken sprechen, beziehen wir uns meist auf maschinelle Automatisierung. Ist das schon der Verweis auf das Nicht-Humane? Weil es am Ende immer noch Menschen sind, die all diese nicht-menschlichen Akteur*innen in Bewegung gesetzt haben.

Tyżlik-Carver: Der Posthumanismus hilft dabei, über diese Dichotomien von Mensch und Nicht-Menschlichem, Mann und Frau, und weitere hinauszudenken. Wir müssen immer wieder feststellen, dass es nicht so einfach ist, diese Art von Beziehungen zu demontieren, also dass es die Algorithmen, Datenbanken, Automatisierungen und Mustererkennung gibt, und dass sie keine natürlich vorkommenden Phänomene sind, sondern Ergebnisse materieller Praktiken, bestimmter geschichtlicher Entwicklungen und anderer Elemente, die beeinflussen, wie diese Werkzeuge eingesetzt werden. Wie werden sie benutzt, wer benutzt sie, an wem? Wer entwirft sie und für wen? Wenn wir das hinterfragen, werden die Folgen klarer. Daraus folgt, dass die Leute, die aufgezeigt haben, wie Bias in verschiedenen algorithmischen Modellen oder in Datenbanken vorherrschen, verdeutlichen, dass es diesen Humanismus mit dem großen ‚H' immer noch gibt. Er ist jetzt besonders sichtbar in Form einer Vorherrschaft der Weißen, die man als systematische Ungerechtigkeit beobachten kann, zum Beispiel anhand von Gesichtserkennungsalgorithmen, die sich besonders schwertun, Gesichter zu erkennen oder zu unterscheiden, wenn eine Person Schwarz ist. Oder in transphoben Algorithmen, die einen Bias in Bezug auf die Erkennung des Geschlechts in der Software haben, die oft von Unternehmen verwendet wird. Das sind wirklich sehr interessante Beispiele, und ich denke, mein Bezug zum Posthumanen und Posthumanismus ist beeinflusst durch den Feminismus und ein materialistisches Verständnis, welches auch eine historische Dimension hat. Und in Rosi Braidottis Definition des Posthumanismus bezieht sie sich darauf, dass der Posthumanismus diese Art von historischem Moment ist, der das Ende der Opposition zwischen Humanismus und Anti-Humanismus markiert. Also sozusagen zwischen diesem Humanismus mit großem ‚H', also dem Humanismus der Aufklärung, und dem Anti-Humanismus, also der Ablehnung des Menschen als Kategorie. Sie sagt, dass es stärker darum geht, die Rahmungen nachzuvollziehen oder nach affirmativen Modellen zu suchen und nach wirklichen Alternativen. Das ist hilfreich, weil es die Idee einführt, dass dieser Kritik eben auch ein Vorschlag folgen muss, wie wir einen anderen Weg einschlagen aus dem Wahnsinn, den wir um uns herum sehen, der so häufig so zerstörerisch und deprimierend ist. Worin liegt die Vorstellungskraft, die uns helfen könnte, über verschiedene Modelle und verschiedene Möglichkeiten der Arbeit mit Algorithmen nachzudenken, oder darüber, dass alle Daten extraktivistisch sein müssen? Was bedeutet es, dass wir extrahieren und müssen wir immer extraktivistisch agieren? Müssen wir überhaupt bestimmte Modelle bauen? Das sind die Fragen, die wir stellen müssen, und ich denke, dass der Posthumanismus oder die Art und Weise, wie Braidotti und Haraway und andere damit umgehen, dass dieser neue materialistische Ansatz diese Fragestellungen wirklich neu aufrollt. Und zwar nicht, um die Menschen komplett aus dem Bild zu entfernen, ganz und gar nicht, sondern um die Relationalität, die in diesen Systemen existiert, neu zu zentrieren und um zu betrachten, dass es nicht nur um uns geht, sondern um uns in Beziehung zu Anderen und zu anderen Dingen. Wenn wir ‚wir' sagen, wen meinen wir dann?

Hunger: In deinem Text *Curating Data Infrastructures of Control and Affect … and Possible Beyonds* (2021) hast du festgestellt, ich zitiere: „Die Praxis des Kuratierens von Daten ist auch eine epistemologische Praxis, die Eingriffe zur Einbeziehung der Zukunft und der Vergangenheit erfordert. Dies kann durch die Frage geschehen, woher Daten kommen. Aufgabe des Kuratierens von Daten ist es, ihre Rückverfolgbarkeit einzufordern und ihre Herkunft zu berücksichtigen."

Tyżlik-Carver: Dies baut auf einer dekolonialen Forschung auf, die andere Praktiken einsetzt und speziell dekoloniale Museen in den Blick nimmt, wo wir diese komplexe und schwierige Vergangenheit nachverfolgen können, mit der wir uns auseinandersetzen müssen, gerade als Kurator*innen oder als Europäer*innen, die auf besondere Weise in diese Geschichte verwickelt sind. Und ich denke, wir können tatsächlich so viel von dieser Arbeit lernen, die in Bezug auf die physischen Objekte geleistet wurde, die gestohlen und nach Europa gebracht wurden, oder in Bezug auf die Körper, die gestohlen und über den Atlantik geschifft wurden. Das, was heute mit Datenpraktiken geschieht, basiert auf ganz ähnlichen Wertvorstellungen der Extraktion und des Wegnehmens, ohne um Erlaubnis zu bitten, und ohne die Bedingungen des Wegnehmens auszuhandeln. Wenn wir also darüber nachdenken, so folgt daraus die grundlegende Frage: Was sind ‚Daten'? Es zeigt sich, dass Daten nicht nur Zahlen sind, die in Tabellen organisiert werden. Tatsächlich denken wir stattdessen darüber nach, was der Prozess ist, um diese Daten zu extrahieren. Sie sind nicht gegeben, wir müssen Daten erheben. Was bedeutet es also, diese Daten zu entnehmen? Wenn ich vorschlage, dass ich die Herkünfte der Daten berücksichtige, dann bedeutet das, sich der Wahrheit zu stellen, dass das Extrahieren von Daten ein gewalttätiger Akt ist. Aber auch bei solchen Themen wie Big Data. Dieses Thema existiert so stark in unserem Imaginären, dass es in gewissem Sinne selbst zu einem Körper wird und zu einer Art Körper mit Leben. Es führt ein Eigenleben. In dem Artikel gehe ich auf eine Reihe von Beispielen ein, aber eines, das ich etwas ausführlicher bespreche, ist die Arbeit von Sissel Marie Tonn and Jonathan Reus, zwei Künstler*innen aus den Niederlanden. Sie arbeiten mit seismischen Daten aus der Region Groningen in den Niederlanden, einem Gebiet, das von menschengemachten Erdbeben betroffen war, die auf die dortige Bergbauindustrie zurückzuführen sind, und übersetzen diese Daten in eine körperliche Erfahrung, durch Berührung, Bewegung und Klang. Was sie also tun, ist, dass sie diese Westen herstellen, die man am Körper tragen kann und die den Körper zum Zittern bringen, und es gibt auch eine Art Sound, der damit einhergeht, und das passiert in dem Areal, wo alles auf eine bestimmte Weise installiert ist. Man kann es sich fast wie eine Datenbank vorstellen, die räumlich verteilt ist, und dass die Daten aus dieser maschinellen und automatisierten Sichtweise zurück in die menschliche Umgebung gebracht werden, wo es möglich ist, sie affektiver zu machen und sie zurück in das menschliche Leben zu führen und nicht nur in das rechnerische Leben. Das ist auch ein Rückverweis auf mein vorheriges Nachdenken darüber, warum Daten nur innerhalb bestimmter Disziplinen, der Informatik und den Data Sciences, welche wiederum Daten auf eine bestimmte Art und Weise betrachten, eine Rolle spielen. Wenn wir uns die Beispiele anderer Arbeiten ansehen, einschließlich deiner und einer Reihe anderer, gibt es vielfältige Vorschläge, um sich vorzustellen, was Daten sind. Und das ist so wichtig für ein gutes Leben! Wenn wir uns Daten nur als Zahlen in Tabellen vorstellen, landen wir bei einem kontinuierlichen Bias, und das lässt sich nicht auflösen, weil wir einfach der Tatsache ins Auge sehen müssen, dass wir Subjekte sind, die situiert sind, und die ihre Meinungen haben. Es ist sehr schwierig, das loszuwerden und so zu tun, als ob Datenbanken nicht all diesen Ballast mit sich tragen würden. Bei künstlerischen Projekten hingegen, die Daten verwenden, und zwar mit viel Einfühlungsvermögen, öffnet dies eine Möglichkeit, wie wir tatsächlich mit Daten arbeiten können.

Den QR-Code scannen, um das vollständige Videointerview anzusehen.

In Conversation with
Maya Indira Ganesh on “Cultural Critique and Artificial Intelligence”

Indira Ganesh discusses how artistic works have influenced the critique of AI in recent years. The interview further investigates the differences between academic text production and art exhibition curating, and how they reflect on our world and the development of AI technologies.

Transcript of the interview with Maya Indira Ganesh, conducted by Francis Hunger on 2023–06–13.

Hunger: I would like to revisit several discussions around artificial intelligence critique of the past few years. How and when did you become aware of it? What was the first discussion about AI where you felt: this needs to be investigated?

Indira Ganesh: When I was studying the emergence of ethics in the driverless car context, I found this separation between understandings of ethics and politics between myself studying the driverless car and somebody else studying recommender systems. I was studying the social and cultural shaping of ethics and autonomy in the driverless car, and I was describing this technology in terms of its imaginaries, its data infrastructures and as a 20th century media that was transforming. I found that in tracing the shaping of autonomous vehicle ethics, there were different sorts of communities, for instance, there was the machine ethics crowd who wanted to formalize and operationalize ethics within computational systems, and then there were those people who were talking about ethics, but they were also trying to dismantle Big Tech. And I kept finding this sort of separation between ethics and politics. There were people who wanted AI, but better, and there were those who wanted no AI at all, and there were those who had no interest in the digital and were only interested in social institutions and later had started thinking about algorithmic technologies. Then there were those who were concerned with the planetary, but weren't thinking at all about the digital. I found these separations and exclusions and distinctions interesting in terms of how people were approaching AI and what they were asking of it. Maybe a year or so ago, I brought some friends together to write something about AI as media, and because we all come from media studies and cultural studies backgrounds and I wanted to investigate what it means for this rather niche, rather new field of media studies. What do we say to AI as media? And these discussions were really interesting, they were all on Zoom, they were during the pandemic and it ended up being a short paper that Johannes Bruder and I presented at a conference this year, called *The Functions of Criticism* (University of Cambridge). And in that, we were really asking through those conversations with friends this question of what is critique for and we arrived at this position of 'lethargic critique.' And we're recognizing that this Big Tech critique is a very active, very committed, very passionate encounter with Big Technology, with law and regulation. This is work that I've done in the past, before becoming an academic, when I was working in civil society organisations, notable places like Tactical Tech. But I think what Johannes and I were trying to think about was more around the conditions of life and how we live with AI as media now. We are very influenced by Tung-Hui Hu's recent book *Digital Lethargy: Dispatches from the Age of Disconnection* (2022). And we also built on Eve Sedgwick's idea of this "paranoid and reparative reading" (1997), in which paranoia is this very passionate and involved engagement that is very invested in the present and in the future and we were thinking about artworks and these positionalities that don't care. What does it mean to not care or not have the ability to care, not have the means to care? When you are not powerful in society and your words are not heard, what critique is possible when you are within the system?

Hunger: I would like to revisit with you the emergence of this discussion. I remember Safiya Umoja Noble's book *Algorithms of Oppression* (2018), which was about search engines, not so much about artificial intelligence in a more narrow sense, and then the essay by Kate Crawford and Trevor Paglen *Excavating AI* (2019). What did these texts mean for you?

Indira Ganesh: I like sort of situating our discussion in [artistic] work that's been really significant and pivotal. *Excavating AI* and *Anatomy of AI* (2018) by Kate Crawford and Vladan Joler—what I really like about them is they are quite distinct methodological critiques, and ways of interrogating these technologies from within the technology. So going into the databases, understanding how they are architected and organized, understanding their histories and making you feel, and this is the case of *Excavating AI* telling the history of ImageNet and where the [training] images come from. And I think that this is a really great way to do critique. And similarly, *Anatomy of AI* uses the very distinct metaphor of a pipeline and a supply chain, what people like to call 'the value chain of AI.' And [the authors] show you how things are actually architected and organized within the system. What I like about it, that it's also visual, it's accessible media. It engages a much wider audience. Probably one of the most standard pieces in this way is *Mega-Pixels* (2017–20) by Adam Harvey and Jules LaPlace, who work on facial recognition technologies and systems. It's again interesting, because that's a very high level of confidence and skill in working with AI algorithmic technologies and computer vision systems, and I think that's not something that everybody can do. But I think that's so important, because they become these places from which all of us then can say: "Look, it is possible to actually develop this critique in these ways." And this is what I actually love about this work. And the last thing is quite a different piece of work, because it's an artistic work. I mean some of these are all on that border of research and art and critical academic scholarship, but there is this other piece that I really like by Zach Blas, called *Facial Weaponization Suite* (2012–14). It's a very material object that sits on a wall and it's made of very real material like silicon and it's created on this understanding of surveillance, and the question: "What does a 'gay' face even look like?" There were some artworks that were so eccentric and clever and thoughtful in how they were not big grand gestures, but they are still so powerful and incisive and insightful. There is the very famous one by Mimi Onuoha called *Library of Missing Datasets* (2016). And it's so simple. It's a filing cabinet and it's just data no one ever thinks to asks about. And it's that negative space that Onuoha is gesturing to. But [she] still makes it really powerful and compelling. I love that piece very much. And then there was another really hilarious one by Surya Mattu and Tega Brain, which was called *Unfit Bits* (2015) and it was just about a Fitbit, attached to a dog or to a bicycle wheel, and it was about hacking systems that want your data to be able to then target you to make predictions about, segment, and profile you. Today we might look back on those works which are six or seven or eight years old and say: "Oh that's about data privacy. And that's not about AI." And I feel like it's important to bring these things constantly into the connection, into the discussion with AI, because data is the fuel of AI. We have to keep making these connections, and if I want to curate something about AI, I would bring some of these older works in that make us stop and say: "Well, what does this have to do with AI?" Even asking that question is vitally important.

Hunger: Looking into the future, it seems as if we as critical 'AI' researchers are more or less obliged to talk about so-called 'general artificial intelligence' and I would actually like to avoid this topic, since I deem it a pseudo-discussion of a very unlikely scenario. Instead of investing our time in this phantasm, where do you think could the critical 'AI' discourse continue and be most fruitful?

Indira Ganesh: I think AI will become quite boring and other things will happen, other manifestations. It's a really lively research topic and area, but there are also many parts to it, like data infrastructures, algorithmic technologies, automated decision making, natural language processing, vision technologies. I would love to read about what they're actually for and with time people will adopt and do things with them. Some of them will be absolutely awful and terrible and we should never have them. And maybe there'll be some interesting uses. But I would like to read about the process of how people do that and also how they reject technologies and say: 'They don't work.' Or, you know, sort of a critical AI discourse that marries recent histories of technology critique with the present and the future. So, there is a great book called *Voices in the Code* (2022) by David G. Robinson, which is about the history of how doctors and transplant surgeons made decisions about who would get a kidney. And because the economy of getting a kidney, keeping it on ice, storing it and then transplanting it to another person is very much about who's deserving and who's at risk. And this technology, like many others now, are being subject to algorithmic systems. And *Voices in the Code* isn't about health care. It's about human biases and flaws and it's about how algorithmic systems reshape and force us to relook at why do we want to give a kidney to this person and not that person? What are the metrics by which we assess the risk and assess who is more deserving? And why do we believe that algorithmic technologies will make this system better and fairer and more transparent and where are they not. So, I want to read more of these kinds of things that are about the worlds that we live in and that we inhabit, and about AI discourse that is about art and technology. Again, coming back to Tega Brain's work, which is fantastic: Tega Brain and Sam Lavigne have *Solar Protocol* (2021–23) or their project *Perfect Sleep* (2021–22) about sleep pods that are sort of curious and cute and deeply technical and allow us to look around at the world and say: "Oh, this is what it's like to actually live in this future!" I'm convinced we're already in the future. Or Winnie Soon's work *Unerasable Characters* (2019–22) which is all about censorship and the new forms of technology from machine learning, but also earlier kinds of technology. So, the works that situate us where we are and allow us to appreciate and slowly, slowly come to terms with where they already are, is what I think makes for a critical AI discourse that I'm interested in.

Scan the QR code below to watch the full video interview online.

Im Gespräch mit
Maya Indira Ganesh über „Kulturkritik und Künstliche Intelligenz“

Indira Ganesh diskutiert im Interview, wie künstlerische Arbeiten die Kritik an KI in den vergangenen Jahren beeinflusst haben. Dabei kommen die Unterschiede zwischen akademischer Textproduktion und dem Kuratieren von Kunstausstellungen zur Sprache und wie diese unsere Welt sowie die Entwicklung von KI-Technologien reflektieren.

Transkript des Interviews mit Maya Indira Ganesh, geführt von Francis Hunger am 13–06–2023.

Hunger: Ich möchte auf einige kritische Diskussionen der letzten Jahre über Künstliche ‚Intelligenz' kommen. Wie und wann bist du auf dieses Thema aufmerksam geworden? Was war die erste Diskussion über KI, bei der du das Gefühl hattest, dass diese untersucht werden muss?

Indira Ganesh: Als ich die Fragestellung der Ethik im Zusammenhang mit fahrer*innenlosen PKWs untersuchte, stellte ich fest, dass es eine Trennung gibt zwischen dem Verständnis von Ethik und Politik, wenn ich mich z. B. mit dem autonomen Fahrzeug beschäftige, oder jemand anderes das mit Empfehlungssystemen tut. Ich untersuchte die soziale und kulturelle Gestaltung von Ethik und Autonomie im fahrer*innenlosen PKW und beschrieb diese Technologie im Hinblick auf ihre Vorstellungen, ihre Dateninfrastrukturen und als ein Medium des 20. Jahrhunderts, das sich im Wandel befindet. Als ich die Entwicklung der Ethik autonomer Fahrzeuge erforschte, zeigte sich, dass es verschiedene Arten von Communities gab, z. B. die Gruppe der Maschinenethiker*innen, die die Ethik innerhalb von Computersystemen formalisieren und operationalisieren wollte, und dann gab es die Leute, die über Ethik sprachen, aber auch versuchten, Big Tech zu demontieren. Und immer wieder stieß ich auf die Trennung von Ethik und Politik. Es gab Leute, die wollten KI, aber besser, und es gab solche, die überhaupt keine KI wollten, und es gab solche, die kein Interesse am Digitalen hatten und sich nur für soziale Institutionen interessierten und erst später anfingen, über algorithmische Technologien nachzudenken. Dann gab es diejenigen, die sich um den Planeten sorgten, aber nicht über das Digitale nachdachten. Ich fand diese Trennungen, Ausschlüsse und Unterscheidungen interessant in Bezug darauf, wie die Menschen sich der KI näherten und was sie von ihr erwarteten. Vielleicht vor einem Jahr, habe ich einige Freund*innen zusammengebracht, um etwas über KI als Medium zu schreiben. Und weil wir alle aus der Medienwissenschaft oder den Kulturwissenschaften kommen, wollte ich untersuchen, was das für dieses eher nischenhafte, eher neue Feld der Medienwissenschaft bedeutet. Was sagen wir zu KI als Medium? Und diese Diskussionen waren wirklich interessant, sie fanden alle auf Zoom statt, denn sie fanden während der Pandemie statt, und es endete mit einem kurzen Aufsatz, den Johannes Bruder und ich dieses Jahr auf einer Konferenz mit dem Titel *The Functions of Criticism* (Universität Cambridge) präsentierten. Darin haben wir uns in Gesprächen mit Freund*innen die Frage gestellt, wozu Kritik gut ist, und wir kamen zu der Position der ‚lethargischen Kritik'. Wir haben herausgearbeitet, dass diese Big-Tech-Kritik eine sehr aktive, sehr engagierte, sehr leidenschaftliche Auseinandersetzung mit Big Technology, mit Recht und Regulierung ist. Solch eine Arbeit habe ich in der Vergangenheit gemacht, bevor ich Akademikerin wurde, als ich für zivilgesellschaftliche Organisationen gearbeitet habe, und zwar an angesehenen Orten wie Tactical Tech. Aber ich glaube, was Johannes und ich versucht haben, war mehr über die Lebensbedingungen und die Art und Weise, wie wir jetzt mit KI als Medium leben, nachzudenken. Wir sind sehr beeinflusst von Tung-Hui Hus kürzlich erschienenem Buch *Digital Lethargy: Dispatches from the Age of Disconnection* (2022). Wir haben auch auf Eve Sedgwicks Idee einer „paranoiden und reparativen Lesart" (1997) aufgebaut, in der Paranoia ein leidenschaftliches und nahegehendes Engagement ist, das sehr in die Gegenwart und die Zukunft gerichtet ist. Und wir dachten über Kunstwerke und persönliche Standpunkte nach, die unbedeutend sind. Was bedeutet es, keine Rolle zu spielen oder nicht die Fähigkeit zu haben, eine Rolle zu spielen, nicht die Mittel zu haben, eine Rolle zu spielen? Wenn man in der Gesellschaft nicht mächtig ist und nicht gehört wird, welche Kritik ist dann möglich, wenn man sich im System befindet?

Hunger: Ich möchte mit dir noch einmal auf den Anfang oder die Entstehung dieser Diskussion zurückkommen. Ich erinnere mich an Safiya Umoja Nobles Buch *Algorithms of Oppression* (2018), in dem es um Suchmaschinen ging, nicht so sehr um Künstliche Intelligenz im engeren Sinne, und dann an den Essay von Kate Crawford und Trevor Paglen *Excavating AI* (2019). Was haben diese Texte für dich bedeutet?

Indira Ganesh: Ich möchte unsere Diskussion mithilfe einiger [künstlerischer] Arbeiten verorten, die wirklich wichtig und entscheidend waren. *Excavating AI* und *Anatomy of AI* (2018) von Kate Crawford und Vladan Joler gefallen mir sehr gut, weil sie ganz unterschiedliche methodologische Kritiken und Möglichkeiten darstellen, diese Technologien von innen heraus zu befragen. Es geht also darum, die Datenbanken zu untersuchen, zu verstehen, wie sie aufgebaut und organisiert sind, ihre Geschichte zu verstehen. Und das ist der Fall bei *Excavating AI,* welches die Geschichte von ImageNet erzählt, woher die [Trainings-]Bilder stammen. Und ich denke, dass dies eine wirklich großartige Möglichkeit ist, Kritik zu üben. Ganz ähnlich verwendet *Anatomy of AI* eine sehr eigene Metapher der Pipeline und der Lieferketten, die als ‚Wertschöpfungskette der KI' bezeichnet werden. Und die Autor*innen zeigen auf, wie die Dinge innerhalb dieses Systems tatsächlich aufgebaut und organisiert sind. Was mir daran gefällt, ist, dass es auch visuell ist, es ist ein zugängliches Medium. Es spricht ein viel breiteres Publikum an. Ich würde sagen, dass *MegaPixels* (2017–20) von Adam Harvey und Jules LaPlace, die sich mit Gesichtserkennungstechnologien und -systemen beschäftigen, wahrscheinlich eines der Standardwerke in dieser Hinsicht ist. Das ist wiederum interessant, denn da gibt es ein sehr hohes Maß an Selbstvertrauen und Können bei der Arbeit mit algorithmischen KI-Technologien und Computer-Vision-Systemen. Und ich denke, das ist eine Fähigkeit, die nicht viele haben. Aber das ist so wichtig, weil sie zu diesen Orten werden, von denen aus wir sagen können: „Schaut mal, es ist möglich, diese Kritik auf solch eine Weise zu entwickeln." Und das ist es, was ich an dieser Arbeit so liebe. Mein letzter Punkt ist eine ganz andere Arbeit, weil es eine künstlerische Arbeit ist. Ich meine, einige dieser Arbeiten bewegen sich an der Grenze zwischen Forschung und Kunst und kritischer akademischer Wissenschaft, aber diese andere Arbeit von Zach Blas mit dem Titel *Facial Weaponization Suite* (2012–14) gefällt mir sehr. Es ist ein sehr materielles Objekt, das an einer Wand hängt und aus einem sehr realen Material wie Silikon besteht. Es wurde auf der Grundlage dieses Verständnisses von Überwachung geschaffen und stellt dabei die Frage: „Wie sieht ein ‚schwules' Gesicht aus?" Es gab einige Kunstwerke, die so exzentrisch und klug und durchdacht waren, dass sie keine großen Gesten darstellten, aber dennoch so kraftvoll, prägnant und aufschlussreich waren, dass sie einige dieser Spannungen des algorithmischen Lebens zusammenbrachten. Da ist zum Beispiel das sehr berühmte Werk von Mimi Onuoha *Library of Missing Datasets* (2016). Und es ist so einfach. Es ist ein Aktenschrank und es geht um Daten, nach denen niemand fragt. Und es ist dieser negative Raum, auf den Onuoha mit dieser Geste hinweist. Aber [sie] macht es trotzdem sehr kraftvoll und fesselnd. Ich liebe dieses Kunstwerk. Und dann gab es noch eine andere wirklich lustige Arbeit von Surya Mattu und Tega Brain mit dem Titel *Unfit Bits* (2015), die sich um Fitbit-Tracker drehte, die an einem Hundehalsband oder am Rad eines Fahrrades befestigt waren. In der Arbeit ging es darum, Systeme zu hacken, die deine Daten haben wollen, um dich dann gezielt anzusprechen, um Vorhersagen über dich zu machen, dich zu segmentieren und Profile von dir zu erstellen. Heute blicken wir vielleicht auf diese sechs, sieben oder acht Jahre alten Arbeiten zurück und sagen: „Oh, da geht es um Datenschutz. Und da geht es nicht um KI". Ich denke, es ist wichtig, diese Dinge immer wieder mit KI in Verbindung zu bringen, denn Daten sind der Treibstoff für KI. Wir müssen diese Verbindungen immer wieder herstellen, und wenn ich etwas zum Thema KI kuratiere, würde ich einige dieser älteren Arbeiten einfügen, die uns dazu bringen, innezuhalten und zu fragen: „Was hat das mit KI zu tun?" Allein schon diese Frage zu stellen, ist von entscheidender Bedeutung.

Hunger: Wenn wir in die Zukunft blicken, scheint es, als ob wir als kritische ‚KI'-Forscher*innen mehr oder weniger gezwungen sind, über eine sogenannte ‚Allgemeine Künstliche Intelligenz' zu sprechen, und ich würde dieses Thema eigentlich gerne vermeiden, da ich es für eine Pseudo-Diskussion über ein sehr unwahrscheinliches Szenario halte. Anstatt Zeit in dieses Phantasma zu investieren, wie könnte deiner Meinung nach der kritische ‚KI'-Diskurs fortgeführt werden und am fruchtbarsten sein?

Indira Ganesh: Ich denke, dass KI ziemlich langweilig werden wird und andere Dinge passieren, und sich andere Erscheinungsformen entwickeln werden. Es ist ein sehr lebendiges Forschungsthema und -gebiet, aber es gibt auch viele Teile davon, z. B. Dateninfrastrukturen, algorithmische Technologien, automatisierte Entscheidungsfindung, Verarbeitung natürlicher Sprache und Bildverarbeitungstechnologien. Ich würde gerne darüber lesen, wozu sie eigentlich gut sind. Und mit der Zeit werden die Menschen sie übernehmen und Dinge mit ihnen tun, wobei einige von ihnen absolut schrecklich und furchtbar sein werden. Und wir sollten sie niemals haben, und ich denke, wir wissen bereits, was einige dieser Dinge sind. Vielleicht gibt es auch einige interessante Verwendungsmöglichkeiten. Aber ich würde gerne über den Prozess lesen, wie die Menschen das machen und auch darüber, wie sie Technologien ablehnen und sagen: „Sie funktionieren nicht". Oder eine Art kritischer KI-Diskurs, der die jüngere Geschichte der Technologiekritik mit der Gegenwart und der Zukunft verbindet. Es gibt ein großartiges Buch mit dem Titel *Voices in the Code* (2022) von David G. Robinson, in dem es darum geht, wie Ärzt*innen und Transplantationschirurg*innen Entscheidungen darüber treffen, wer eine Niere bekommt. Und weil die Ökonomie, eine Niere zu bekommen, sie auf Eis zu legen, sie zu lagern und dann einer anderen Person zu transplantieren, sehr stark davon abhängt, wer es ‚verdient' und wer ‚mehr' gefährdet ist. Und bei dieser Technologie, wie bei vielen anderen auch, geht es nicht um Gesundheitsfürsorge. Es geht um menschliche Voreingenommenheit und Fehler, und es geht darum, wie algorithmische Systeme uns umgestalten und uns zwingen, neu zu überlegen, warum wir dieser Person eine Niere geben wollen und jener nicht. Nach welchen Maßstäben bewerten wir das Risiko und beurteilen, wer sie mehr verdient? Und warum glauben wir, dass algorithmische Technologien dieses System besser, gerechter und transparenter machen werden, und wo sind sie es nicht? Ich möchte also mehr von diesen Dingen lesen, die sich mit den Welten beschäftigen, in denen wir leben und die wir bewohnen. Und über KI-Diskurse, die sich mit Kunst und Technologie befassen. Um noch einmal auf die Arbeit von Tega Brain zurückzukommen, die fantastisch ist: Tega Brain und Sam Lavigne haben *Solar Protocol* (2021–23) oder ihr Projekt *Perfect Sleep* (2021–22) über Schlafkapseln, die irgendwie seltsam und niedlich und zutiefst technisch sind und uns erlauben, die Welt zu betrachten und zu sagen: „Oh, so sieht es aus, wenn man tatsächlich in dieser Zukunft lebt!" Ich bin davon überzeugt, dass wir bereits in der Zukunft leben. Oder die Arbeit *Unerasable Characters* (2019–22) von Winnie Soon, in der es um Zensur und die neuen Formen des maschinellen Lernens geht, aber auch um frühere Arten von Technologie. Die Werke, die uns dort verorten, wo wir stehen, und uns erlauben, zu erkennen und uns langsam damit abzufinden, wo sie sich bereits befinden, machen meiner Meinung nach einen kritischen KI-Diskurs aus, an dem ich interessiert bin.

Den QR-Code scannen, um das vollständige Videointerview anzusehen.

In Search of Boundary Objects: A Taxonomy-Based Approach to Algorithmic Co-curation in Archival Collections

Giulia Taurino

From early card catalogs and physical indexes arranged in file cabinets, to modern computerized systems and searchable databases, archival practices in museums and libraries have relied on taxonomic logics that facilitate information retrieval and obtain knowledge from local repositories. Traditionally, archivists, librarians, and curators have been responsible for documenting, organizing, and curating metadata related to each artifact in a given archive. In recent years, advancements in artificial intelligence (AI) have offered unprecedented support to the curatorial and analytical process that turns archival data into accessible, interconnected, and intelligible information. This chapter tackles the notion of 'algorithm as curator' (Hendricks 2017) in archival settings, building upon the work of scholars who have theorized a co-operative framework for human-algorithmic practice (Dekker and Tedone 2019; Tedone 2019). Having outlined a few definitions of algorithmic curation, this chapter will focus on the limits of state-of-the-art computer vision models and discuss a co-curatorial approach to training machine learning models based on historical datasets and structured vocabularies from heritage collections.

Knowledge curation and information retrieval are primary concerns of museums, libraries, archives and other institutional spaces that host vast repositories. The access to heritage records, both physical and digital, can be driven by queries for single entities (e.g., a specific painting) or by the search for assemblages and coherent clusters of objects (e.g., all artworks from a certain artist). On occasion, archival records are made accessible to the public through curated exhibitions that provide guided experiences and help visitors locate each artifact in its historical, geographical, or thematic context. In museums, artworks' labels, summaries and descriptions allow visitors to get a glimpse of the archive as they walk through art displays. The archival narratives we find in galleries are the result of a lengthy process that involves researching, cataloging and labeling objects. At their minimum, metadata and keywords represent the infrastructure of knowledge in cultural heritage institutions.

Similarly, whether in the form of unstructured data disseminated across the web, or as labeled records stored and indexed in digital repositories, the increasing mass of networked content is in constant need of categorization to support the large-scale retrieval process that makes information accessible to Internet users. The pace of online content production requires alternative solutions to manual classification that can make tasks like text and image categorization faster and, in some cases, more accurate. Among others, the development of machine learning techniques for information extraction and content curation presents a unique opportunity to automate the management of digital content.

While online content classification is still to some extent outsourced to teams of crowd workers in charge of tagging and annotating digital items, automated content filtering constitutes the main curatorial strategy of most content-hosting platforms relying on recommender systems. Moreover, major search engines routinely implement algorithmic indexing for content ranking, prioritization and personalization. The whole of these practices across online media ecosystems has been described as 'curation by code' (Morris 2015) or 'machine learning-based curation' (Heuer 2020), with reference to the use of computer programming and algorithms for the organization of content on the Internet. Widely found in museology and library science to address the ensemble of recovery, care, preservation, and exhibition practices for the long-term support of heritage or natural resource collections, the metaphor of curatorship in AI is used to broadly describe the sorting of digital content. Curation algorithms are commonly deployed on the web to select and arrange information spanning from journalism newsfeeds to textual data, user-generated images, music tracks, and audiovisual material. Scholars have taken different stances on algorithmic curation. The urge to find keywords to illustrate new media dynamics resulted in the adoption of the term 'algorithmic curation' for the analysis of Internet culture, from social networks to content aggregators and streaming platforms. For instance, a theoretical framework for understanding web curation (Davis 2016) distinguishes human curation online between

productive practices (i.e., performative production of content) and consumptive practices (i.e., users' selection of content for media consumption). These domains co-exist within the boundaries of network curation and the curatorial code (ibid., 6). Network curation is an underlying effect of social interaction and collective decision-making that may affect individual choices. It can be implemented algorithmically in recommendation engines through collaborative filtering models that arrange content based on both personal preferences and similarity among users. Finally, forms of machine-based curation are defined as the 'curatorial code,' the entanglement of platform architectures and algorithmic operations that partake in the ordering of content online (ibidem). This framework draws upon Hogan's exhibitional approach (2010), which compares online sites to exhibition spaces where users submit their own artifacts (data) held in storehouses (databases) and mediated by algorithms that decide not only which items may be visible, but also how they should be ordered and displayed. At the turn of the twenty-first century, archivists, artists, and institutions started blending with this exhibitional aspect of Internet culture and established their presence online by creating their own databases, websites, and social media profiles in a collective curatorial movement. At the same time, beyond acting as gatekeepers of archival collections, curators have taken on overseeing the presentation, authenticity, and attributed value of records. Nagler and del Pesco (2011) point at this phenomenon as part of a broader transition in the role of art curators from archival caretakers and exhibition-makers (Obrist 2014, 25) to taste-makers, which accompanied the introduction of algorithmic systems for information filtering. As a consequence of this shift, the term 'curator' has acquired a blurred definition, often being associated with multiple identities: reminiscent of traditional academic titles, as a 'scientist,' 'scholar' or 'custodian' (Horie 1986), and, in online jargon, disguised as a 'node' (Graham et al. 2010), a 'software,' an 'algorithm.' Among others, artist Anne-Marie Schleiner adopted the term "net curators" to reference the conflating of users—"cultural producers (artists) and consumers, readers and writers, and information filters and collectors" (Schleiner 2003, 1)—into a mass of curatorial agents on the web.

Algorithmic curation in platform environments, however, differs substantially from archival contexts in both scope and application. The purpose of museum archiving is far from profit-driven dynamics that regulate the staging of personalized experiences and online media exposure by means of algorithmic technologies. In fact, prior to entering the realm of Internet studies, the term curation was discussed in relation to 'digital curation' (Yakel 2007) and its 'data management lifecycle' (Higgins 2008). Grounded in archival and information science, these studies return to the etymological origins of curation (Lat.: *cūrāre,* Eng.: 'to take care of'), and are primarily concerned with attending to records' acquisition, storage, maintenance, and preservation, as well as with the overall process of classification and dissemination of knowledge. At a time when digital curation was at the center of the debate in heritage institutions, the act of curating through computer programming emerged from the domain of media practice. In 2008, Joasia Krysa published a thesis on "software curating"—i.e., software as curatorial object and curatorial platform—, "making a direct reference to software art and collapsing firm distinctions between curating, programming and artistic practice" (Krysa 2008, 21). Her work resulted in *kurator,* "a proposal for an experimental, permutational software application capable of curating exhibitions" (Krysa 2014) through a combination of human and computer interventions. This project stands out as one of the precursors of contemporary algorithmic art experiments that tinker with the idea of AI as an active participant in the curatorial process.

More recently, in the attempt to relocate the roots of curatorial work in the context of algorithmic culture, Cairns and Birchall (2013) reflected in *Curating the Digital World* on the role of museums—and curators—as cultural filters for the abundance of content available online. As the very act of curating is being redefined outside of the physical spaces of archives and exhibition galleries, "a museum now needs to know as much about algorithms as art history" (ibidem). To further problematize this duality, Hendricks (2017) provided a comprehensive overview of the notion of algorithm as curator in museums, touching on several topics—namely, the interplay between

computer/curator, database/narrative, digital reproduction/value assessment. Her essay is a final response to our uncertainties: algorithmic curation is indeed changing the way we collect, store, access, consume and interpret culture. By 2019, the conversation on algorithmic curation has become well-established in the domains of archival and media studies as well as art history. Academic papers on 'networked co-curation' (Dekker and Tedone 2019) and 'human-algorithmic curation' (Tedone 2019) invite us to expand our view of the socio-technical specificities of algorithmic filtering and introduce the concept of curating with or against the machine. The curatorial function is explored in the interaction with AI, conferring curatorial agency to both human and non-human actors involved in the co-creative process of selecting and organizing information. Editors, designers, archivists, community curators and algorithms all partake in modeling the act of curatorship. As Dekker and Tedone (2019, 4) outline, "to come to terms with the new role of the curator, and perhaps to distinguish it from the contemporary art curators, new concepts emerged such as 'immaterial curating' (Krysa 2006), 'distributed curating' (Krysa 2008; 2013), 'computer-aided curating' (Grubinger 2006), and 'post-human curating' (Tyżlik-Carver 2016; 2017)." Beyond specific definitions, it is understood that networked images inevitably undergo some form of algorithmic co-curation at the platform level, where users and algorithms alike contribute to sort content. However, this interdisciplinary debate tends to overlook the epistemological and pragmatic foundations of the curatorial work before images are shared online. At its origins, the process of curating collections relies on the creation of taxonomic systems for the correct classification and interpretation of records. The acquisition of knowledge around the object curated often starts from a catalog (e.g., a list of items) and a taxonomic representation (e.g., a list of categories) that undergo a process of ordering or grouping. Even when the curation results in an arbitrary selection of items, a category in its own right—'miscellanea'—is created to account for the lack of information, metadata, or artistic interconnection. To explore the relation between algorithmic co-curation and the taxonomic logics of AI, the following paragraphs take a closer look into the current state of machine vision, a discipline that marked yet another shift in the field of AI for cultural heritage management and preservation.

As the debate on co-curatorial practices branches out into archival science, museology and media studies, machine learning researchers in the field of computer science have worked intensively on advancing AI applications in computer vision. Computer vision is a sub-field of AI that focuses on training machines to detect, interpret, and derive information from digital images. In recent years, this domain underwent a rapid evolution in what concerns the understanding, development and application of deep learning methods based on Convolutional Neural Networks for feature extraction, achieving promising results when performing a wide range of tasks, including object recognition and image classification. Computer vision models can be used in optical character recognition for the transcription of handwritten or typewritten textual data, in content-based image retrieval to enrich metadata and facilitate queries, in image processing for restoration purposes via noise and blur reduction or structural and textural inpainting. This acquired aptitude of computing machines generates new algorithmic curatorial affordances: not only are computers able to select and curate archival material, but they are also capable of inferring textual information from images, labeling objects, or recovering degraded, damaged records and corrupted visual data. For domains that rely extensively on visual culture, such as the arts and media, AI's optical faculty expands the outreach of hybrid curatorial intelligence into the entire archival pipeline, including the processes of data archeology, restoration and classification. In addition to ordering content in hierarchical queues and organized streams, machines are now equipped with the processing ability of defining objects and developing taxonomic systems.

The algorithmic act of identifying, describing, and naming visual objects starts from a training dataset: an initial catalog consisting of a collection of samples and exemplary records used to teach machine learning models to make inferences based

on visual cues. At its present stage, the algorithmic catalog is expansive. Most recent benchmark datasets—i.e., datasets used both for training and testing—reflect a tendency towards generalization and adaptability in a non-task-specific way. Deep learning marked a "shift towards homogenization: rather than having bespoke feature engineering pipelines for each application, the same deep neural network architecture could be used for many applications" (Bommasani et al. 2022, 4). For the past ten years, this level of generalization has been achieved by using supervised learning, a subfield of machine learning that leverages on labeled datasets for training algorithms to make predictions. One of the most influential datasets in computer vision is ImageNet. ImageNet is a large-scale ontology of annotated images built upon the hierarchical structure of WordNet (Fellbaum 1998), a comprehensive lexical database that reconstructs the semantic and lexical relations of the English language. This image dataset was announced as a "critical resource for developing advanced, large-scale content-based image search and image understanding algorithms, as well as for providing critical training and benchmarking data for such algorithms" (Deng et al. 2009, 1).

For each 'synset' (synonym set: groups of words related to a single concept) in WordNet, ImageNet aimed at providing 500–1,000 images (ibidem), totaling, at its completion, almost 15 million images related to nouns and 22 thousand categories (Crawford and Paglen 2021). This operation was meant to support the process of transferring into machines those same cognitive behaviors that we find in the human brain, where learning is linked to recurrent observation and real-world experience. Automated interpretation of images based on statistical surveys of training sets, however, comes with challenges and limitations. For instance, most of the images in ImageNet are sourced from the Internet as born-digital records via web-scraping, while taxonomic relationships are based on an English database and crowd annotation, making the object classes, taxonomy, and labels problematic. In their archeology of datasets, Crawford and Paglen conclude that canonical "training sets of labeled images that are ubiquitous in contemporary computer vision and AI are built on a foundation of unsubstantiated and unstable epistemological and metaphysical assumptions about the nature of images, labels, categorization, and representation" (ibidem). Moreover, as Koch et al. remark, "dataset audits have revealed concerning biases that have direct implications for algorithmic bias and harms […]. Problematic categorical schemas have been identified in popular image datasets, including poorly-formulated categories and the inclusion of derogatory and offensive labels" (Koch et al. 2021, 1-2).

Other than raising a series of ethical questions on the social impacts of computer vision, training machines to perform tasks related to visual processing is a computationally demanding endeavor. To limit both the social and computational costs of existing benchmark datasets, AI researchers are exploring self-supervised learning, a method that consists in pre-training models on unannotated data. Scholars at Stanford University have pointed at this trend as "a paradigm shift with the rise of models (e.g., BERT, DALL-E, GPT-3) trained on broad data (generally using self-supervision at scale) that can be adapted to a wide range of downstream tasks. We call these models foundation models […]" (Bommasani et al. 2022, 1). In a blog post on foundation models published in 2022, IBM describes them as "the future of AI": "flexible, reusable AI models that can be applied to just about any domain or industry task," "trained on large, unlabeled datasets and fine-tuned for an array of applications" (Murphy 2022). Yet, these models cannot escape the risks of reproducing defects at scale, even when datasets are unlabeled. To control the surge of automated errors in self-supervised learning, foundation models require artisanal care in the curation of the data selected for the training phase.

No matter how expansive the catalog is or how thorough the categorical schemas, looking at millions of images does not make up for the missing data in digital repositories. Academic studies have repeatedly shown that algorithmic bias is a direct symptom of biased input datasets—be they labeled or unlabeled—providing only a partial view of the world and perpetuating discrimination. Pereira and Moreschi (2021, 1) exposed these shortcomings by testing commercial image-recognition models on artworks and unveiling AI's "capitalist and product-focused reading of the world." For

years, machine learning applications have prioritized the optimization of model architectures over data curation (Jakubik et al. 2022). From an archival perspective, many popular training datasets are fundamentally un-curated. This has led researchers and practitioners to urge a change from model-centric to data-centric approaches that emphasize "the systematic design and engineering of data" (ibid., 3). Among others, Jo and Gebru (2020) solicited the adoption of archival strategies to sociocultural datasets that leverage on ethical and inclusive data collection practices, standardized documentation and maintenance plans. In the context of data-centric AI, archival datasets represent a critical asset. Not only do they offer practical frameworks for data collection standards, but they can also provide curated datasets with structured vocabularies and cataloging references for guiding machines to take a historical attitude towards heritage records.

A Historical Attitude

Taking a historical attitude in machine learning requires distinct curation choices that prioritize the inclusion of archival collections over web-scraped data. When the ImageNet project was first presented in 2009, digitization efforts in museums were still at early stages. Even when machine-readable archival material became available, issues related to accessibility and lack of funding, including quality of data, copyright attribution, IP regulations (Fiorucci et al. 2020), posed challenges to the use of heritage records in machine learning. In the meantime, major tech companies such as Microsoft and Google have produced their own datasets, with thousands of images showing everyday objects, buildings, humans, animals, natural scenes. Most of these datasets are still dominated by images found online. The result is that the knowledge passed through to AI often lacks historical depth. As Groys (2016) observes, "the Internet is organized in a less historicist way than traditional libraries and museums. The most interesting aspect of the Internet as an archive is precisely the possibilities for decontextualization and recontextualization through the operations of cut and paste that the Internet offers its users." Although enticing, the possibility of taking things out of their historical context can produce controversial outcomes. History is the ground on which we organize, recover and retrieve knowledge. In Foucault's words, history itself "organizes the document, divides it up, distributes it, orders it, arranges it in levels, establishes series, distinguishes between what is relevant and what is not, discovers elements, defines units, describes relations" (Foucault 1972, 6–7).

The Internet's tendency to disaggregate data in hyper-personalized experiences over multiple platforms has fostered harmful phenomena, such as the spreading of misinformation and disinformation, and the creation of information silos. Attempts to re-aggregate and anthologize online content (Doueihi 2011; Taurino 2020) via automated algorithmic curation and user-generated queues led to unstandardized curatorial practices, undermining the idea of networked encyclopedic knowledge the Internet was originally meant to represent. A computer vision model fed with a plethora of web-sourced images risks replicating a dynamic where curation is just a synonym for mechanical, unregulated content classification. The advancement of archival digitization programs in GLAM* institutions introduces new opportunities to refine AI's visual syntax and educate curatorial algorithms to recognize art, historical documents and heritage records, while avoiding fragmented cataloging practices. Observing machine learning as an archival science (Taurino and Smith 2022), we propose in this chapter a framework for training machine learning models that builds upon curated archival datasets and a taxonomy-based approach to algorithmic co-curation. The term algorithmic co-curation is used here to define the set of co-curatorial activities that intervene at the level of dataset design and model training. In this framework, cultural heritage collections are deployed in computer vision with respect to two aspects: 1.) improving qualitative domain knowledge in AI models (training the algorithm), and 2.) improving human-algorithmic collaboration and computer-aided curatorial work (training the archive).

At the scale of human cognition, once the brain acquires the basic faculty to discern objects, a constant retraining enables us to perfect the ability of interpreting visual cues and gain different levels of expertise. It is an ongoing process that involves

* The acronym GLAM stands for galleries, libraries, archives and museums, and is used to refer to cultural heritage institutions that aim at knowledge preservation and dissemination.

everyday learning experiences with real-world objects, mediated forms of knowledge, individual and collective memories. Contemporary pre-trained AI models can achieve relatively high confidence scores on a broad set of visual tasks, with fairly accurate general knowledge inferences. They are likely, however, to underperform or return errors when it comes to domain-specific knowledge, such as interpreting artworks and historical records. As foundation models gain momentum, one solution to the shortcomings of AI is to co-curate machine learning models by using heritage datasets either in the pre-training process or in the phase of re-training for fine-tuning on specialized tasks. For instance, Garcia et al. (2020) propose to use the SemArt dataset—a collection of fine-art painting images paired with a list of attributes and artistic commentary (Garcia and Vogiatzis 2018)—to train a baseline model (VIKING) specifically designed to extract knowledge from images of paintings for semantic art understanding and question-answering. Their study demonstrates that VIKING outperforms existing computer vision models when it comes to the analysis of artworks (Garcia et al. 2020, 14). Exposing algorithms to field-specific cultural records and more nuanced abstract categories has proven to aid AI systems to replicate expert skills. Researchers have described this type of knowledge as "qualitative domain knowledge which requires intensive engagement and interpretation by humans with sophisticated skills" (Deng et al. 2020, 3). Domain knowledge supports models that are more statistically rigorous and avoid interpretability issues by maintaining human-understandable inputs and outputs (ibid., 4). Feeding machine intelligence with specialized knowledge helps complement computers' outstanding ability for information retention with the necessary human expertise that generated entire fields such as cultural studies or art history. For this purpose, historical documentation is fundamental to fill the lexical and visual gaps found in 'non-specific' machine learning models. Yet, a generic archival dataset as a whole might just not be enough. We need to identify archival collections in museums or libraries that can function as boundary objects (Star and Griesemer 1989), meaning symbolic—both abstract and concrete—references that algorithms can use to support forms of co-operation with human actors. Taking a museum of zoology as a case study, Star and Griesemer describe four types of boundary objects that can sustain interdisciplinary cooperation between field experts: repositories, ideal types, coincident boundaries, and standardized forms (ibid., 411). In particular, repositories can take "the form of a set of modular things. These are things that might be individually removed without collapsing or changing the structure of a whole. A library, for example, or a collection of case studies […] is a repository" (Star 2010, 603).

As we argue in this chapter, identifying boundary objects within large-scale heritage collections constitutes a strategic point of departure for collecting training datasets that can help machine learning models to answer domain-related questions. Large, structured archival repositories comprise historical repertoires in the form of iterative and representative records, making them a great pool from which one can draw relevant samples for learning specific tasks. Moreover, the modular structure of curated digital archives enables selecting either a whole repository (e.g., all items in a contemporary art collection) or sub-clusters of interest (e.g., a folder containing artworks created by a single artist). Folders may follow different principles of organization, depending on the features of the archival material we seek to highlight. For example, in art collections, folders can be categorized by geographic location, time period, or art movement, or based on temporal, spatial, or stylistic proximity. The same principles and time-space boundaries applied to photo-journalism archives might end up generating a miscellanea of images, covering separate events and topics. A folder structure in a repository can help redistributing records into distinct clusters contingent to the scope of the task demanded from the machine learning model. For instance, if we want to train a computer vision model to answer questions about what type of event is reported in specific photo-journalism coverage, we might need to curate folders that include visual records with common subjects and themes, rather than dividing them by location, date or style.

The organizational mindset of structured archives favors the transfer of knowledge between human and algorithmic curator from the dataset design to the training phase. While the design of the dataset can be clearly imagined as a direct act of

cooperation between humans and machines, where archivists or researchers curate catalogs, repositories, and folder systems that can turn into relevant sources of visual data for training sets, when it comes to the actual phase of machine learning, the cooperative process might take different paths depending on the size of the dataset. When it comes to small heritage archives, one option is to take a taxonomy-based approach that not only supports the use of 'expert' datasets, but also provides a framework for semantic interpretation. Among others, Michaud et al. (2018) discuss content-based image retrieval methods in application to small cultural heritage collections. They compare the performance of different algorithmic models on both 'generic' image datasets (like ImageNet) and 'expert' datasets (i.e., ROMANE 1K, Coin Collection Online Catalogue). In their paper, they present an "unsupervised framework, which aims at combining automatically the information from various local descriptors" (ibid., 161). They notably propose a set of procedures for features extraction based on the Bags of Visual Words model, a pre-deep learning method that treats image features as words (key points associated to descriptors) and uses clustering to generate a high-dimensional visual vocabulary. Their study shows that, when applied to cultural heritage collections, the Bags of Visual Words model yields more accurate results than Convolutional Neural Networks models (ibid., 169).

Other studies suggest that the use of domain-specific vocabularies, word embeddings, and taxonomic labels might enhance the performance of image analysis, classification and retrieval. In a paper published in 2019, Garcia et al. introduce two methods for obtaining context-aware embeddings, one based on visual elements and one based on non-visual attributes defined in a knowledge graph. Their research demonstrates that "context-aware embeddings are beneficial in many automatic art analysis problems, improving art classification accuracy by up to a 7.3% with respect to classification baselines" (Garcia et al. 2019, 8). Moreover, a study published in 2021 proposes to "incorporate coarse taxonomic labels to train image classifiers in fine-grained domains" (Su and Maji 2021, 1) by using semi-supervised learning (i.e., a combination of labeled data and unlabeled data) to train a computer vision model on a biology dataset. The results indicate that the implementation of taxonomy-aware user input produces improvements in fine-grain image classification (ibid., 10). While taxonomies in arts and culture don't always follow a rigid hierarchy of classes and orders, taxonomic logics can still be successfully adopted to transfer human domain-specific knowledge to algorithmic systems. Among other resources, controlled vocabularies provide a knowledge organization system based on indexing schemes that have been designed and maintained by experts to reduce lexical ambiguity and ensure terminological consistency. In library science, these lists of terms and phrases represent a standard reference for organizing content in catalogs, annotating images and labeling archival material with metadata.

Similar to other classification systems, controlled vocabularies have drawbacks: they need to be constantly updated to avoid outdated terminology; when used for information retrieval, precision comes at the risk of exhaustivity; the design of these systems requires ethical considerations on lexical choices. Nevertheless, controlled vocabularies, along with other categorical 'schemata' that are used for systematic knowledge organization, can be effectively adopted in supervised and semi-supervised vocabulary-informed learning to improve accuracy in computer vision models trained for object detection and image categorization (Fu and Sigal 2016). With their own existing limitations, controlled vocabularies and classification schemes can favor the implementation of human curation when working with pre-trained machine learning models, by providing a guideline for evaluating the label distribution, prioritizing sets of labels produced by pre-trained models over others or excluding entire sets of labels that are deemed not relevant for a domain-specific application.

Conclusions

In the previous sections we have visited several perspectives on how archival and curatorial work might intersect with algorithmic practices in application to heritage archives. Although many archival repositories are still in the process of digitizing analog

records, some historical collections have already come to comprise millions of images, making them good candidates for producing training datasets in AI. Moreover, rigorous practices of curatorship are widely established in the field of art history and museum studies, and heritage datasets are often accompanied by thorough documentation. This is the case for the Getty Research Institute, which created the Cultural Objects Name Authority (CONA), a structured vocabulary for compiling metadata about visual arts and cultural heritage. Additionally, the Getty Vocabulary Program maintains the Categories for the Description of Works of Art (CDWA), a list of best practices for cataloging artworks, which is incorporated in Cataloging Cultural Objects (CCO), a manual for describing heritage records. Whether in the form of repositories, folders, catalogs, or vocabularies, reasoning in terms of boundary objects might be the key for defining practices of algorithmic co-curation that allow researchers and experts from different disciplines to cooperate in the design of AI able to make field-specific inferences. While most common image recognition models are trained via web scraping on a random selection of objects, training the archive always implies a principle of human curation and organization of the dataset at the source, which is then transferred to machine vision models. In a way, the process of training the archive starts from training the algorithm. This chapter has provided an overview of how a joint (both human- and machine-based) training translates into practices of algorithmic co-curation. Having outlined early definitions of algorithmic curation in relation to Internet culture and computational art practices, we advanced an approach to algorithmic co-curation grounded in archival and information science. In this framework, algorithms are called to cooperate with researchers, archivists and curators during the phase of model design and training, as much as human curators are invited to cooperate with algorithms during networked co-curatorial practices. Heritage collections were discussed as the preferential settings in which this collaborative work should take place, at the intersection between algorithms (i.e., teaching machines to learn from labeled or unlabeled archival records) and human curators (i.e., teaching archivists, curators, librarians, researchers to work with machine learning models and evaluate machine-generated metadata). Furthermore, in order to establish the level of human intervention necessary in the training process, we advanced a taxonomy-based approach to algorithmic co-curation that takes into account the specificities of each heritage collection and the queries that might be used for information retrieval.

As highlighted earlier in this paper, current state-of-the-art computer vision models are based on several problematic assumptions made at the level of dataset design. Canonical training sets like ImageNet source images from the Internet, with limited attention for the definition of boundary objects that can sustain collaborative work and support historically-grounded knowledge. In these large-scale image datasets, human curation is often replaced by automated data collection and, when present, takes the form of generic crowd annotation, thus understating the importance of field-specific expertise. Finally, benchmark datasets deployed at scale for model training and validation by a variety of machine learning communities pose issues for institutional diversity. Scholars have found a "concentration across the field on datasets that have been introduced by researchers situated within a small number of elite institutions" (Koch et al. 2021, 1). This can be for several reasons, including lack of funding and resources for the creation of more institutionally diverse training sets, issues of accessibility to heritage data due to copyright or privacy regulations, or challenges with digitization that prevent archival material from being machine-readable.

Using historical archives as training sets can help us overcome some of these challenges and address the failures of AI models based on large-scale web-scraped datasets, while also enabling new directions in machine learning applications. On the one hand, sourcing datasets from heritage archives adjusts the focus of AI towards the importance of field-specific documentation and taxonomic understanding, making historical data and cultural memory the backbone of machine learning models. On the other hand, this approach to algorithmic co-curation promotes a cooperative model of machine-assistance based on domain expertise. As such, it diversifies training datasets not only in terms of specialized repositories hinged on localized knowledge, but also

in terms of institutional roots. Similarly, archival and curatorial work can benefit from the use of AI to automate part of the archiving and management process. In particular, with regard to metadata extraction (e.g., using OCR for transcribing handwritten manuscripts or image recognition for labeling objects), machine learning models have proven to be an important resource for improving the quality of content classification, indexing, and retrieval in archival settings (Colavizza et al. 2022).

In this scenario, we reintroduced the notion of algorithmic co-curation not only as a theoretical framework that can be used to describe instances where machines curate content over the Internet in more or less hybrid and interactive ways. Rather, we proposed to utilize it as a methodological approach to information organization and extraction that relies on both human and machine intelligence to preserve collective knowledge and memory. With this updated definition, the paper aims to shift scholarly attention from the concept of the algorithm as curator to the very practice of curating the algorithm. That is, rather than focusing on redefining the role of the curator in the context of algorithmic culture, it calls for establishing a praxis of algorithmic design that fits into pre-established curatorial traditions. In this attempt, we suggest that developing machine learning models that can be successfully integrated into existing archival workflows will contribute to the creation of AI-enabled curatorial practices that go beyond standardized tasks based on superordinate categories to embrace forms of specialized algorithmic co-curation that leverage fine-grained analysis and domain knowledge.

Bibliography

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein et al. 2022. "On the Opportunities and Risks of Foundation Models." arXiv. http://arxiv.org/abs/2108.07258.

Cairns, Susan, and Danny Birchall. 2013. "Curating the Digital World: Past Preconceptions, Present Problems, Possible Futures." In *Museums and the Web,* edited by Nancy Proctor and Rich Cherry. Silver Spring, MD: Museums and the Web. https://mw2013.museumsandtheweb.com/paper/curating-the-digital-world-past-preconceptions-present-problems-possible-futures.

Colavizza, Giovanni, Tobias Blanke, Charles Jeurgens, and Julia Noordegraaf. 2022. "Archives and AI: An Overview of Current Debates and Future Perspectives." *Journal on Computing and Cultural Heritage* 15 (1): 1–15. https://doi.org/10.1145/3479010.

Crawford, Kate, and Trevor Paglen. 2021. "Excavating AI: The Politics of Images in Machine Learning Training Sets." *AI & SOCIETY* 36 (June): 1105–16. https://doi.org/10.1007/s00146-021-01162-8.

Davis, Jenny L. 2016. "Curation: A Theoretical Treatment." *Information, Communication & Society* 20 (5): 770–83. https://doi.org/10.1080/1369118X.2016.1203972.

Dekker, Annet, and Gaia Tedone. 2019. "Networked Co-Curation: An Exploration of the Socio-Technical Specificities of Online Curation." *Arts* 8 (3), 86. https://doi.org/10.3390/arts8030086.

Deng, Changyu, Xunbi Ji, Colton Rainey, Jianyu Zhang, and Wei Lu. 2020. "Integrating Machine Learning with Human Knowledge." *IScience* 23 (11): 101656. https://doi.org/10.1016/j.isci.2020.101656.

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. "ImageNet: A Large-Scale Hierarchical Image Database." In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–55. IEEE. https://doi.org/10.1109/CVPR.2009.5206848.

Doueihi, Milad. 2011. *Digital Cultures*. American ed. Cambridge, MA: Harvard University Press.

Fellbaum, Christiane, ed. 1998. *WordNet: An Electronic Lexical Database*. Preface by George Miller. Cambridge, MA: The MIT Press.

Fiorucci, Marco, Marina Khoroshiltseva, Massimiliano Pontil, Arianna Traviglia, Alessio Del Bue, and Stuart James. 2020. "Machine Learning for Cultural Heritage: A Survey." *Pattern Recognition Letters* 133 (May): 102–8. https://doi.org/10.1016/j.patrec.2020.02.017.

Foucault, Michel. 1972. *The Archaeology of Knowledge*. New York: Pantheon Books.

Fu, Yanwei, and Leonid Sigal. 2016. "Semi-Supervised Vocabulary-Informed Learning." arXiv. http://arxiv.org/abs/1604.07093.

Garcia, Noa, Benjamin Renoust, and Yuta Nakashima. 2019. "Context-Aware Embeddings for Automatic Art Analysis." In *Proceedings of the 2019 International Conference on Multimedia Retrieval*, 25–33. https://doi.org/10.1145/3323873.3325028.

Garcia, Noa, and George Vogiatzis. 2018. "How to Read Paintings: Semantic Art Understanding with Multi-Modal Retrieval." arXiv. http://arxiv.org/abs/1810.09617.

Garcia, Noa, Chentao Ye, Zihua Liu, Qingtao Hu, Mayu Otani, Chenhui Chu, Yuta Nakashima, and Teruko Mitamura. 2020. "A Dataset and Baselines for Visual Question Answering on Art." arXiv. http://arxiv.org/abs/2008.12520.

Graham, Beryl, Sarah Cook, and Steve Dietz. 2010. *Rethinking Curating: Art after New Media*. Leonardo Books. Cambridge: The MIT Press.

Groys, Boris. 2016. "The Truth of Art." *e-flux* 71 (March). https://www.e-flux.com/journal/71/60513/the-truth-of-art.

Grubinger, Eva. 2006. "C@C: Computer-Aided Curating (1993–1995)." In *Curating Immateriality: The Work of the Curator in the Age of Network Systems (DATA Browser 3)*, edited by Joasia Krysa, 107–15. New York: Autonomedia.

Hendricks, Manique. 2017. "The Algorithm as Curator: In Search of a Non-Narrated Collection Presentation." *Stedelijk Studies Journal* 1. https://doi.org/10.54533/StedStud.vol005.art11.

Heuer, Hendrik. 2020. Users and Machine Learning-Based Curation Systems. University of Bremen: Dissertation. https://doi.org/10.26092/ELIB/241.

Higgins, Sarah. 2008. "The DCC Curation Lifecycle Model." *International Journal of Digital Curation* 3 (1): 134–40. https://doi.org/10.2218/ijdc.v3i1.48.

Hogan, Bernie. 2010. "The Presentation of Self in the Age of Social Media: Distinguishing Performances and Exhibitions Online." *Bulletin of Science, Technology & Society* 30 (6): 377–86. https://doi.org/10.1177/0270467610385893.

Horie, Charles V. 1986. "Who Is a Curator?" *International Journal of Museum Management and Curatorship* 5 (3): 267–72. https://doi.org/10.1080/09647778609515029.

Jakubik, Johannes, Michael Vössing, Niklas Kühl, Jannis Walk, and Gerhard Satzger. 2022. "Data-Centric Artificial Intelligence." arXiv. http://arxiv.org/abs/2212.11854.

Jo, Eun Seo, and Timnit Gebru. 2020. "Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning." In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 306–16. https://doi.org/10.1145/3351095.3372829.

Koch, Bernard, Emily Denton, Alex Hanna, and Jacob G. Foster. 2021. "Reduced, Reused and Recycled: The Life of a Dataset in Machine Learning Research." arXiv. http://arxiv.org/abs/2112.01716.

Krysa, Joasia, ed. 2006. *Curating Immateriality: The Work of the Curator in the Age of Network Systems*. New York: Autonomedia Press.

———. 2008. *Software Curating: The Politics of Curating in/as (an) Open System(s)*. University of Plymouth: Dissertation. https://doi.org/10.24382/4984.

———. 2013. "Some Questions on Curating as (Public) Interface to the Art Market." *APRJA* (archive), April 12, 2013. https://www.academia.edu/30947090/Some_Questions_on_Curating_as_Public_Interface_to_the_Art_Market.

———. 2014. "Kurator: a Proposal for an Experimental, Permutational Software Application Capable of Curating Exhibitions." In *Networks*, edited by Lars Bang Larsen, 118–21. Documents of Contemporary Art. Cambridge, MA: The MIT Press.

Murphy. 2022. "What Are Foundation Models?" *IBM Research* (blog), May 9, 2022. https://research.ibm.com/blog/what-are-foundation-models.

Michaud, Dorian, Thierry Urruty, Philippe Carré, and François Lecellier. 2018. "Adaptive Features Selection for Expert Datasets: A Cultural Heritage Application." *Signal Processing: Image Communication* 67 (September): 161–70. https://doi.org/10.1016/j.image.2018.06.011.

Morris, Jeremy Wade. 2015. "Curation by Code: Infomediaries and the Data Mining of Taste." *European Journal of Cultural Studies* 18 (4–5): 446–63. https://doi.org/10.1177/1367549415577387.

Nagler, Christian, and Joseph del Pesco. 2011. "Curating in the Time of Algorithms." *Fillip* 15 (Autumn): 52–61. https://www.academia.edu/15258682/Curating_in_the_Time_of_Algorithms.

Obrist, Hans Ulrich. 2014. *Ways of Curating*. First American edition. New York: Faber and Faber.

Pereira, Gabriel, and Bruno Moreschi. 2021. "Artificial Intelligence and Institutional Critique 2.0: Unexpected Ways of Seeing with Computer Vision." *AI & SOCIETY* 36 (4): 1201–23. https://doi.org/10.1007/s00146-020-01059-y.

Schleiner, Anne-Marie. 2003. "Fluidities and Oppositions among Curators, Filter Feeders and Future Artists." *intelligent agent* 3, no. 1 (March). http://www.intelligentagent.com/archive/v03[1].01.curation.schleiner.pdf.

Star, Susan Leigh. 2010. "This Is Not a Boundary Object: Reflections on the Origin of a Concept." *Science, Technology, & Human Values* 35 (5): 601–17. https://doi.org/10.1177/0162243910377624.

Star, Susan Leigh, and James R. Griesemer. 1989. "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39." *Social Studies of Science* 19 (3): 387–420. https://doi.org/10.1177/030631289019003001.

Su, Jong-Chyi, and Subhransu Maji. 2021. "Semi-Supervised Learning with Taxonomic Labels." arXiv. http://arxiv.org/abs/2111.11595.

Taurino, Giulia. 2020. "Redefining the Anthology: Forms and Affordances in Digital Culture." https://doi.org/10.6092/UNIBO/AMSDOTTORATO/9365.

Taurino, Giulia, and David Smith. 2022. "Machine Learning as an Archival Science: Narratives behind Artificial Intelligence, Cultural Data, and Archival Remediation." In *NeurIPS AI Cultures Workshop*. https://ai-cultures.github.io/papers/machine_learning_as_an_archiva.pdf.

Tedone, Gaia. 2019. "Human-Algorithmic Curation: Curating with or against the Algorithm." In *2019 Conference on Computation, Communication, Aesthetics & X*, 125–39. Milan, Italy: xCoAx. https://2019.xcoax.org/xCoAx2019.pdf.

Tyżlik-Carver, Magda. 2016. Curating in/as Commons. Posthuman Curating and Computational Cultures. Aarhus University: Dissertation.

———. 2017. "| Curator | Curating | the Curatorial | Not-Just-Art Curating: A Genealogy of Posthuman Curating." *Springerin*, no. 1. https://www.springerin.at/en/2017/1/kuratorin-kuratieren-das-kuratorische-nicht-nur-kunst-kuratieren/.

Yakel, Elizabeth. 2007. "Digital Curation." *OCLC Systems & Services: International Digital Library Perspectives* 23 (4): 335–40. https://doi.org/10.1108/10650750710831466.

Auf der Suche nach Grenzobjekten: Ein taxonomiebasierter Ansatz zur algorith-mischen Co-Kuratierung in Archivsammlungen

Giulia Taurino

Von frühen Zettelkatalogen und physischen Registern in Aktenschränken bis hin zu modernen computergestützten Systemen und durchsuchbaren Datenbanken haben sich Archivierungspraktiken in Museen und Bibliotheken auf taxonomische Logiken gestützt, die das Auffinden von Informationen erleichtern und Wissen aus lokalen Beständen vermitteln. Traditionell waren Archivar*innen, Bibliothekar*innen und Kurator*innen für die Dokumentation, Organisation und Kuratierung von Metadaten zu jedem Artefakt in einem bestimmten Archiv zuständig. In den letzten Jahren hat Fortschritt in der Künstlichen Intelligenz (KI) den kuratorischen und analytischen Prozess, der Archivdaten in zugängliche, vernetzte und verständliche Informationen umwandelt, in nie dagewesener Weise unterstützt. Dieses Kapitel befasst sich mit dem Begriff des ‚Algorithmus als Kurator' (Hendricks 2017) in archivarischen Umgebungen und baut auf der Arbeit von Wissenschaftler*innen auf, die einen kooperativen Rahmen für die menschlich-algorithmische Praxis theoretisiert haben (Dekker und Tedone 2019; Tedone 2019). Nachdem einige Definitionen der algorithmischen Kuratierung skizziert wurden, konzentriert sich das Kapitel auf die Grenzen moderner Computer-Vision-Modelle und erörtert einen co-kuratorischen Ansatz für das Training von maschinellen Lernmodellen auf der Grundlage historischer Datensätze und strukturierter Vokabulare aus historischen Sammlungen.

Die Kuratierung von Wissen und das Abrufen von Informationen sind Hauptanliegen von Museen, Bibliotheken, Archiven und anderen institutionellen Einrichtungen, die über umfangreiche Bestände verfügen. Der Zugang zu physischen und digitalen Archivalien kann durch Anfragen nach einzelnen Objekten (z. B. einem bestimmten Gemälde) oder durch die Suche nach Zusammenstellungen und kohärenten Gruppen von Objekten (z. B. alle Kunstwerke bestimmter Künstler*innen) erfolgen. Gelegentlich werden Archivalien der Öffentlichkeit durch kuratierte Ausstellungen zugänglich gemacht, die den Besucher*innen helfen, jedes Artefakt in seinem historischen, geografischen und thematischen Kontext zu verorten. In Museen ermöglichen Schilder, Zusammenfassungen und Beschreibungen von Kunstwerken den Besucher*innen einen Einblick in das Archiv, während sie durch die Kunstausstellungen gehen. Die archivarischen Erzählungen, die wir in Galerien finden, sind das Ergebnis eines langwierigen Prozesses, der die Erforschung, Katalogisierung und Beschriftung von Objekten umfasst. Metadaten und Schlüsselwörter sind das Minimum der Wissensinfrastruktur in Institutionen des kulturellen Erbes.

Unabhängig davon, ob es sich um unstrukturierte Daten handelt, die über das Internet verbreitet werden, oder um gelabelte Datensätze, die in digitalen zentralen Ablagen gespeichert und indiziert werden, muss die zunehmende Masse an vernetzten Inhalten ständig kategorisiert werden, um den groß angelegten Sammelprozess zu unterstützen, der den Internetnutzer*innen Informationen zugänglich macht. Das Tempo der Produktion von Online-Inhalten erfordert alternative Lösungen zur manuellen Klassifizierung, die Aufgaben wie die Kategorisierung von Texten und Bildern schneller und in einigen Fällen auch genauer machen. Unter anderem bietet die Entwicklung maschineller Lerntechniken für die Informationsextraktion und die Aufbereitung von Inhalten eine einzigartige Möglichkeit, die Verwaltung digitaler Inhalte zu automatisieren.

Während die Klassifizierung von Online-Inhalten teilweise immer noch an Teams von Crowd-Worker*innen ausgelagert wird, die für die Verschlagwortung und Kommentierung digitaler Objekte zuständig sind, stellt die automatisierte Filterung von Inhalten die wichtigste kuratorische Strategie der meisten Content-Hosting-Plattformen dar, die sich auf Empfehlungssysteme stützen. Darüber hinaus setzen die großen Suchmaschinen routinemäßig die algorithmische Indizierung für das Ranking, die Priorisierung und die Personalisierung von Inhalten ein. Die Gesamtheit dieser Praktiken in Online-Medien-Ökosystemen wurde als ‚Kuration durch Code' (Morris 2015) oder als ‚auf maschinellem Lernen basierende Kuration' (Heuer 2020) bezeichnet, was sich auf die Verwendung von Computerprogrammen, Algorithmen und maschinellem Lernen für die Organisation von Inhalten im Internet bezieht. Das in der Museologie und Bibliothekswissenschaft weit verbreitete Bild des Kuratierens

wird in Zusammenhang mit KI verwendet, um das Zusammenspiel von Wiederherstellung, Pflege, Bewahrung und Ausstellungspraktiken für die langfristige Unterstützung von Sammlungen des Kulturerbes oder natürlicher Ressourcen zu beschreiben. Kuratierungsalgorithmen werden häufig im Internet eingesetzt, um Informationen auszuwählen und zu ordnen, die von journalistischen Newsfeeds über Textdaten bis hin zu nutzergenerierten Bildern, Musikstücken und audiovisuellem Material reichen. Wissenschaftler*innen haben unterschiedliche Standpunkte zur algorithmischen Kuratierung eingenommen. Die Notwendigkeit, Schlagworte zur Veranschaulichung der Dynamik der neuen Medien zu finden, führte zur Übernahme des Begriffs ‚algorithmisches Kuratieren' für die Analyse der Internetkultur – von sozialen Netzwerken bis hin zu Inhaltsaggregatoren und Streaming-Plattformen. Ein theoretischer Rahmen für das Verständnis von Webkuratierung (Davis 2016) unterscheidet beispielsweise zwischen produktiven (d. h. performativen) Praktiken der Produktion von Inhalten und konsumtiven Praktiken (d. h. der Auswahl von Inhalten durch die Nutzer*innen für den Medienkonsum). Diese Bereiche koexistieren innerhalb der Grenzen der Netzwerk-Kuratierung und des kuratorischen Codes (ebd., 6). Netzwerk-Kuratierung ist ein grundlegendes Ergebnis sozialer Interaktion und kollektiver Entscheidungsfindung, das individuelle Entscheidungen beeinflussen kann. Sie kann algorithmisch in Empfehlungsalgorithmen durch kollaborative Filtermodelle implementiert werden, die Inhalte auf der Grundlage persönlicher Vorlieben und Ähnlichkeiten zwischen den Nutzer*innen anordnen. Schließlich werden Formen der maschinengestützten Kuratierung als ‚kuratorischer Code' definiert, d. h. als Verflechtung von Plattformarchitekturen und algorithmischen Operationen, die die Ordnung von Online-Inhalten herstellen (ebenda). Dieser Rahmen stützt sich auf Hogans Herangehensweise an das Ausstellen (2010), der Online-Sites als Ausstellungsräume interpretiert, in denen die Nutzer*innen ihre eigenen Artefakte (Daten) einreichen, die in Depots (Datenbanken) aufbewahrt und durch Algorithmen vermittelt werden, die nicht nur entscheiden, welche Elemente überhaupt sichtbar werden, sondern auch, wie sie geordnet und angezeigt werden sollen. Am Anfang des 21. Jahrhunderts begannen Archivar*innen, Künstler*innen und Institutionen, sich mit diesem Ausstellungsaspekt der Internetkultur vertraut zu machen und ihre Online-Präsenz zu etablieren, indem sie eigene Datenbanken, Websites und Social-Media-Profile im Rahmen einer kollektiven kuratorischen Bewegung erstellten. Gleichzeitig fungieren Kurator*innen nicht nur als Wächter*innen von Archivsammlungen, sondern übernehmen auch die Aufsicht über die Präsentation, die Authentizität und den zugeschriebenen Wert der Aufzeichnungen. Nagler und del Pesco (2011) verweisen auf dieses Phänomen als Teil eines umfassenderen Wandels in der Rolle der Kunstkurator*innen von Archivverwalter*innen und Ausstellungsmacher*innen (Obrist 2014, 25) zu Geschmacksbildner*innen, der mit der Einführung algorithmischer Systeme zur Informationsfilterung einherging. Als Folge dieses Wandels hat der Begriff ‚Kurator*in' eine unscharfe Definition erhalten und wird oft mit mehreren Identitäten in Verbindung gebracht: in Anlehnung an traditionelle akademische Titel als ‚Wissenschaftler*in', Gelehrte*r', ‚Kustos*Kustodin' (Horie 1986) und im Online-Jargon getarnt als ‚Knotenpunkt' (Graham et al. 2010), ‚Software' oder ‚Algorithmus'. Die Künstlerin Anne-Marie Schleiner hat den Begriff ‚Netzkurator*innen' verwendet, um auf die Verschmelzung von Nutzer*innen – „Kulturproduzent*innen (Künstler*innen) und Konsument*innen, Leser*innen und Schriftsteller*innen, Informationsfilter*innen und Sammler*innen" (Schleiner 2003, 1) – zu einer Masse von kuratorischen Akteur*innen im Web hinzuweisen.

Die algorithmische Kuratierung in Plattformumgebungen unterscheidet sich jedoch sowohl im Umfang als auch in der Anwendung wesentlich von archivarischen Kontexten. Der Zweck der Museumsarchivierung ist weit entfernt von der profitorientierten Dynamik, die die Inszenierung personalisierter Erfahrungen und die Online-Medienexposition mittels algorithmischer Technologien regelt. Bevor der Begriff ‚Kuration' in den Bereich der Internetstudien eintrat, wurde er im Zusammenhang mit ‚digitaler Kuration' (Yakel 2007) und dem ‚Lebenszyklus des Datenmanagements' (Higgins 2008) diskutiert. Diese aus der Archiv- und Informationswissenschaft stammenden Studien gehen auf die etymologischen Ursprünge von Kuration (lat.: cūrāre,

dt.: ‚sich kümmern') zurück und befassen sich in erster Linie mit dem Erwerb, der Lagerung, der Pflege und der Bewahrung von Aufzeichnungen sowie mit dem Gesamtprozess der Klassifizierung und Verbreitung von Wissen. Zu einer Zeit, als die digitale Kuratierung im Mittelpunkt der Debatte in Kulturerbeinstitutionen stand, entwickelte sich der Akt des Kuratierens mittels Software aus dem Feld der Medienpraxis. Im Jahr 2008 veröffentlichte Joasia Krysa eine Dissertation über das „Software-Kuratieren" – d. h. Software als kuratorisches Objekt und kuratorische Plattform –, „die einen direkten Bezug zur Softwarekunst herstellt und die festen Unterscheidungen zwischen Kuratieren, Programmieren und künstlerischer Praxis aufhebt" (Krysa 2008, 21). Das Ergebnis ihrer Arbeit ist *kurator,* „ein Vorschlag für eine experimentelle, permutative Softwareanwendung, die in der Lage ist, Ausstellungen zu kuratieren" (Krysa 2014), und zwar durch eine Kombination aus menschlichen Interventionen und Computerinterventionen. Dieses Projekt sticht als einer der Vorläufer zeitgenössischer algorithmischer Kunstexperimente hervor, die mit der Idee der KI als aktive Teilnehmerin am kuratorischen Prozess spielen.

In jüngerer Zeit haben Cairns und Birchall (2013) in ihrem Versuch, die Wurzeln der kuratorischen Arbeit im Kontext der algorithmischen Kultur zu verorten, in *Curating the Digital World* über die Rolle der Museen – und Kurator*innen – als kulturelle Filter für die Fülle der online verfügbaren Inhalte nachgedacht. Da der eigentliche Akt des Kuratierens außerhalb der physischen Räume von Archiven und Ausstellungsgalerien neu definiert werde, „muss ein Museum jetzt genauso viel über Algorithmen wissen wie über Kunstgeschichte" (ebenda). Um diese Dualität weiter zu problematisieren, hat Hendricks (2017) einen umfassenden Überblick über den Begriff des Algorithmus als Kurator im Museum gegeben, der mehrere Themen berührt – nämlich das Zusammenspiel zwischen Computer/Kurator*in, Datenbank/Narrative, digitaler Reproduktion/Wertschätzung. Ihr Essay ist eine abschließende Antwort auf unsere Unsicherheiten: Algorithmische Kuratierung verändert tatsächlich die Art, wie wir Kultur sammeln, speichern, zugänglich machen, konsumieren und interpretieren. Im Jahr 2019 ist die Diskussion über algorithmische Kuratierung in den Bereichen Archiv- und Medienwissenschaften sowie Kunstgeschichte fest etabliert. Akademische Arbeiten über vernetzte Co-Kuratierung (Dekker und Tedone 2019) und menschlich-algorithmische Kuratierung (Tedone 2019) laden uns ein, unseren Blick auf die sozio-technischen Besonderheiten der algorithmischen Filterung zu erweitern und das Konzept des Kuratierens mit oder gegen die Maschine einzuführen. Die kuratorische Funktion wird in der Interaktion mit KI erforscht, indem menschlichen und nicht-menschlichen Akteur*innen, die am co-kreativen Prozess der Auswahl und Organisation von Informationen beteiligt sind, kuratorische Handlungsfähigkeit verliehen wird. Redakteur*innen, Designer*innen, Archivar*innen, Community-Kurator*innen und Algorithmen nehmen alle an der Modellierung des kuratorischen Akts teil.

Wie Dekker und Tedone (2019, 4) darlegen, „entstanden neue Konzepte wie ‚immaterielles Kuratieren' (Krysa 2006), ‚verteiltes Kuratieren' (Krysa 2008; 2013), ‚computergestütztes Kuratieren' (Grubinger 2006) und ‚posthumanes Kuratieren' (Tyżlik-Carver 2016; 2017), um mit der neuen Rolle des/der Kurator*in zurechtzukommen und sie vielleicht von den zeitgenössischen Kunstkurator*innen abzugrenzen." Jenseits spezifischer Definitionen wird davon ausgegangen, dass vernetzte Bilder unweigerlich eine Form der algorithmischen Co-Kuratierung auf der Plattformebene durchlaufen, bei der Nutzer*innen und Algorithmen gleichermaßen zur Sortierung der Inhalte beitragen. Diese interdisziplinäre Debatte tendiert jedoch dazu, die epistemologischen und pragmatischen Grundlagen der kuratorischen Arbeit zu übersehen, bevor Bilder online geteilt werden. Der Prozess des Kuratierens von Sammlungen beruht in seinen Ursprüngen auf der Schaffung taxonomischer Systeme für die korrekte Klassifizierung und Interpretation von Datensätzen. Der Erwerb von Wissen über das kuratierte Objekt beginnt oft mit einem Katalog (z. B. einer Liste von Objekten) und einer taxonomischen Darstellung (z. B. einer Liste von Kategorien), die einem Ordnungs- oder Gruppierungsprozess unterzogen werden. Selbst wenn die Kuratierung zu einer willkürlichen Auswahl von Objekten führt, wird eine eigene Kategorie – ‚Miscellanea' (dt.: Vermischtes) – geschaffen, um dem Mangel an Informationen, Metadaten oder

künstlerischen Zusammenhängen Rechnung zu tragen. Um die Beziehung zwischen der algorithmischen Co-Kuratierung und den taxonomischen Logiken der KI zu erforschen, wird im Folgenden der aktuelle Stand des maschinellen Sehens näher beleuchtet, eine Disziplin, die eine weitere Veränderung im Bereich der KI für die Verwaltung und Bewahrung des kulturellen Erbes darstellt.

Der algorithmische Katalog

Während sich die Debatte über co-kuratorische Praktiken in der Archivwissenschaft, der Museologie und den Medienwissenschaften verzweigt, haben Forscher*innen im Bereich des maschinellen Lernens intensiv an der Weiterentwicklung von KI-Anwendungen im Bereich der Computer Vision gearbeitet. Das maschinelle Sehen ist ein Teilgebiet der KI, das sich mit dem Training von Maschinen zur Erkennung, Interpretation und Ableitung von Informationen aus digitalen Bildern befasst. In den letzten Jahren hat dieser Bereich eine rasante Entwicklung erfahren, was das Verständnis, die Entwicklung und die Anwendung von Deep-Learning-Methoden auf der Grundlage von Convolutional Neural Networks für die Merkmalsextraktion betrifft, wobei vielversprechende Ergebnisse bei der Durchführung einer Vielzahl von Aufgaben, einschließlich Objekterkennung und Bildklassifizierung, erzielt wurden. Computer-Vision-Modelle können bei der optischen Zeichenerkennung zur Transkription hand- oder maschinengeschriebener Textdaten, bei der inhaltsbasierten Bildsuche zur Anreicherung von Metadaten und zur Erleichterung von Abfragen, bei der Bildverarbeitung zur Restaurierung durch Rausch- und Unschärfereduzierung oder zur strukturellen und texturellen Übermalung eingesetzt werden. Diese erworbene Fähigkeit von Computern bringt neue algorithmische Möglichkeiten für die Archivierung mit sich: Computer sind nicht nur in der Lage, Archivmaterial auszuwählen und zu verwalten, sondern auch Textinformationen aus Bildern abzuleiten, Objekte zu labeln oder beschädigte Aufzeichnungen und beschädigte visuelle Daten zu retten. In Bereichen, die sich stark auf die visuelle Kultur stützen, z. B. Kunst und Medien, erweitert die optische Fähigkeit der KI die Reichweite der hybriden kuratorischen Intelligenz auf die gesamte Archivierungskette, einschließlich der Prozesse der Datenarchäologie, Restaurierung und Klassifizierung. Zusätzlich zur Ordnung von Inhalten in hierarchischen Stichworten und organisierten Strömen sind Maschinen nun mit der Fähigkeit ausgestattet, Objekte zu definieren und taxonomische Systeme zu entwickeln.

Der algorithmische Akt des Identifizierens, Beschreibens und Benennens visueller Objekte beginnt mit einem Trainingsdatensatz – einem anfänglichen Katalog, der aus einer Sammlung von Beispielen und exemplarischen Datensätzen besteht, die verwendet werden, um Modellen des maschinellen Lernens beizubringen, auf der Grundlage von visuellen Hinweisen Schlussfolgerungen zu ziehen. In seinem derzeitigen Stadium ist der algorithmische Katalog sehr umfangreich. Die meisten neueren Benchmark-Datensätze – d. h. Datensätze, die sowohl zum Trainieren als auch zum Testen verwendet werden – spiegeln eine Tendenz zur Verallgemeinerung und Anpassungsfähigkeit jenseits spezifischer Aufgaben wider. Deep Learning markiert eine „Verschiebung in Richtung Homogenisierung: Anstatt maßgeschneiderte Feature-Engineering-Pipelines für jede Anwendung zu haben, kann dieselbe tiefe neuronale Netzwerkarchitektur für viele Anwendungen verwendet werden" (Bommasani et al. 2022, 4). In den letzten zehn Jahren wurde dieses Maß an Verallgemeinerung durch den Einsatz von Supervised Learning erreicht, einem Teilbereich des maschinellen Lernens, der auf markierten Datensätzen basiert, um Algorithmen für Vorhersagen zu trainieren. Einer der einflussreichsten Datensätze in der Computer Vision ist ImageNet. ImageNet ist eine groß angelegte Ontologie kommentierter Bilder, die auf der hierarchischen Struktur von WordNet (Fellbaum 1998) aufbaut, einer umfassenden lexikalischen Datenbank, die die semantischen und lexikalischen Beziehungen der englischen Sprache rekonstruiert. Dieser Bilddatensatz wurde als „entscheidende Ressource für die Entwicklung fortschrittlicher, groß angelegter inhaltsbasierter Bildsuch- und Bildverstehensalgorithmen sowie für die Bereitstellung wichtiger Trainings- und Benchmarking-Daten für solche Algorithmen" angekündigt (Deng et al. 2009, 1).

ImageNet zielte darauf ab, für jedes ‚Synset' (d. h. Synonymdatensatz, eine Gruppe von Wörtern, die sich auf ein einziges Konzept beziehen) in WordNet 500–1.000 Bilder bereitzustellen (ebenda), was bei seiner Fertigstellung insgesamt fast 15 Millionen Bilder in Beziehung zu Substantiven und zu zweiundzwanzigtausend Kategorien ergab (Crawford und Paglen 2021). Diese Operation sollte den Prozess der Übertragung der gleichen kognitiven Verhaltensweisen in Maschinen unterstützen, die wir im menschlichen Gehirn finden, wo Lernen mit wiederholter Beobachtung und Erfahrung in der realen Welt verbunden ist. Die automatisierte Interpretation von Bildern auf der Grundlage statistischer Erhebungen von Trainingssätzen ist jedoch mit Herausforderungen und Einschränkungen verbunden. So stammen beispielsweise die meisten Bilder in ImageNet als digitale Datensätze aus dem Internet, die über Web-Scraping gewonnen wurden, während die taxonomischen Beziehungen auf einer englischen Datenbank und auf Kommentaren von Menschen beruhen, was die Objektklassen, die Taxonomie und die Bezeichnungen problematisch machen. In ihrer Archäologie der Datensätze kommen Crawford und Paglen zu dem Schluss, dass die kanonischen „Trainingssätze gelabelter Bilder, die in der zeitgenössischen Computer Vision und KI allgegenwärtig sind, auf einem Fundament unbegründeter und instabiler erkenntnistheoretischer und metaphysischer Annahmen über die Natur von Bildern, Labels, Kategorisierung und Repräsentation aufgebaut sind" (ebenda). Darüber hinaus haben, wie Koch et al. anmerken, „Datensatz-Audits bedenkliche Verzerrungen aufgedeckt, die direkte Auswirkungen für algorithmischen Bias und Gefahren haben […]. Problematische kategoriale Schemata wurden in populären Bilddatensätzen identifiziert, darunter schlecht formulierte Kategorien und die Einbeziehung von abwertenden und beleidigenden Bezeichnungen" (Koch et al. 2021, 1–2).

Abgesehen davon, dass es eine Reihe ethischer Fragen zu den sozialen Auswirkungen der Computer Vision aufwirft, ist das Trainieren von Maschinen zur Durchführung von Aufgaben im Zusammenhang mit der visuellen Verarbeitung ein rechenintensives Unterfangen. Um sowohl die sozialen als auch die rechnerischen Kosten bestehender Benchmark-Datensätze zu begrenzen, erforschen KI-Forscher*innen das selbstüberwachte Lernen, eine Methode, bei der Modelle auf unkommentierten Daten vortrainiert werden. Wissenschaftler*innen der Stanford University bezeichnen diesen Trend als „einen Paradigmenwechsel mit dem Aufkommen von Modellen (z. B. BERT, DALL-E, GPT-3), die auf einer breiten Datenbasis trainiert werden (im Allgemeinen unter Verwendung von Self-Supervised Learning in großem Maßstab) und an eine breite Palette von nachgelagerten Aufgaben angepasst werden können. Wir nennen diese Modelle Foundation Models [...]" (Bommasani et al. 2022, 1). In einem im Jahr 2022 veröffentlichten Blog-Post zu Foundation Models beschreibt IBM diese als „die Zukunft der KI": „flexible, wiederverwendbare KI-Modelle, die auf so gut wie jede Domäne oder Industrieaufgabe angewendet werden können", „trainiert auf großen, unmarkierten Datensätzen und fein abgestimmt für eine Reihe von Anwendungen" (Murphy 2022). Dennoch können diese Modelle nicht dem Risiko entgehen, Fehler in großem Umfang zu reproduzieren, selbst wenn die Datensätze nicht gelabelt sind. Um die Flut automatischer Fehler beim selbstüberwachten Lernen in den Griff zu bekommen, erfordern Foundation Models handwerkliche Fürsorge bei der Aufbereitung der für die Trainingsphase ausgewählten Daten.

Ganz gleich, wie umfangreich der Katalog ist oder wie gründlich die kategorialen Schemata sind, die Betrachtung von Millionen von Bildern kann die fehlenden Daten in digitalen Repositories nicht ausgleichen. Akademische Studien haben wiederholt gezeigt, dass algorithmischer Bias ein direktes Symptom voreingenommener Eingabedatensätze ist – ganz gleich ob gelabelt oder ungelabelt –, die nur eine unvollständige Sicht auf die Welt bieten und die Diskriminierung verstetigen. Pereira und Moreschi (2021, 1) deckten diese Mängel auf, indem sie kommerzielle Bilderkennungsmodelle an Kunstwerken testeten und die „kapitalistische und produktorientierte Lesart der Welt" der KI enthüllten. Jahrelang haben Anwendungen des maschinellen Lernens der Optimierung von Modellarchitekturen Vorrang vor der Datenkuratierung eingeräumt (Jakubik et al. 2022). Aus archivarischer Sicht sind viele beliebte Trainingsdatensätze grundsätzlich unkuratiert. Dies veranlasste Forscher*innen und Praktiker*innen dazu, einen Wechsel von modellzentrierten zu datenzentrierten Ansätzen zu fordern, die „das systematische Design und Engineering von Daten" betonen (ebd., 3). Unter anderem forderten Jo und Gebru (2020) die Einführung von Archivierungsstrategien für soziokulturelle Datensätze, die sich auf ethische und integrative Datensammlungspraktiken, standardisierte Dokumentation und Wartungspläne stützen. Im Kontext der datenzentrierten KI stellen archivierte Datensätze einen entscheidenden Vorteil dar. Sie bieten nicht nur einen praktischen Rahmen für Datenerhebungsstandards, sondern können auch kuratierte Datensätze mit strukturierten Vokabularen und Katalogisierungsreferenzen bereitstellen, um Maschinen zu leiten, eine historische Haltung gegenüber historischen Aufzeichnungen einzunehmen.

Eine historische Betrachtungsweise

Eine historische Betrachtungsweise beim maschinellen Lernen erfordert eindeutige kuratorische Entscheidungen, die der Einbeziehung von Archivsammlungen Vorrang vor Web-Scraping-Daten einräumen. Als das ImageNet-Projekt 2009 erstmals vorgestellt wurde, befanden sich die Digitalisierungsbemühungen in Museen noch in einem frühen Stadium. Selbst als maschinenlesbares Archivmaterial verfügbar wurde, stellten Probleme im Zusammenhang mit der Zugänglichkeit und mangelnder Finanzierung – z. B. Datenqualität, Urheberrechtszuordnung, Bestimmungen zum geistigen Eigentum (Fiorucci et al. 2020) – eine Herausforderung für die Verwendung historischer Aufzeichnungen beim maschinellen Lernen dar. In der Zwischenzeit haben große Technologieunternehmen wie Microsoft und Google ihre eigenen Datensätze mit Tausenden von Bildern erstellt, die Alltagsgegenstände, Gebäude, Menschen, Tiere und Naturszenen zeigen. Die meisten dieser Datensätze werden immer noch von Bildern dominiert, die online gefunden wurden. Das Ergebnis ist, dass das Wissen, das an die KI weitergegeben wird, oft keine historische Tiefe hat. Wie Groys (2016) feststellt, ist „das Internet weniger historisch organisiert als traditionelle Bibliotheken und Museen. Der interessanteste Aspekt des Internets als Archiv sind gerade die Möglichkeiten zur Dekontextualisierung und Rekontextualisierung durch die Operationen des Ausschneidens und Einfügens, die das Internet seinen Nutzer*innen bietet." Die Möglichkeit, Dinge aus ihrem historischen Kontext herauszunehmen, ist zwar verlockend, kann aber zu kontroversen Ergebnissen führen. Geschichte ist der Boden, auf dem wir Wissen organisieren, wiederfinden und abrufen. In Foucaults Worten organisiert die Geschichte selbst das Dokument, „sie organisiert es, zerlegt es, verteilt es, ordnet es, teilt es nach Schichten auf, stellt Serien fest, unterscheidet das, was triftig ist, von dem, was es nicht ist, findet Elemente auf, definiert Einheiten, beschreibt Beziehungen" (Foucault 1969, 14).

Die Tendenz des Internets, Daten in hyper-personalisierten Erfahrungen über mehrere Plattformen zu zerlegen, förderte schädliche Phänomene wie die Verbreitung von Fehlinformationen und Desinformationen oder die Schaffung von Informationssilos. Versuche, Online-Inhalte durch automatisierte algorithmische Kuratierung und nutzergenerierte Warteschlangen zu reaggregieren und zu anthologisieren (Doueihi 2011; Taurino 2020), führten zu uneinheitlichen kuratorischen Praktiken und untergruben die Idee des vernetzten enzyklopädischen Wissens, die das Internet ursprünglich darstellen sollte. Ein Computer-Vision-Modell, das mit einer Fülle von Bildern aus dem Internet gefüttert wird, birgt die Gefahr, dass eine Dynamik entsteht, bei der Kuratierung nur noch ein Synonym für eine mechanische, ungeregelte Klassifizierung von Inhalten ist. Mit den fortschreitenden Programmen zur Digitalisierung der Archive in den GLAM-Institutionen* bieten sich neue Möglichkeiten, die visuelle Syntax der KI zu verfeinern und kuratorische Algorithmen so auszubilden, dass sie Kunst, historische Dokumente und historische Aufzeichnungen erkennen und gleichzeitig fragmentierte Katalogisierungspraktiken vermeiden. Unter Berücksichtigung des maschinellen Lernens als archivarische Wissenschaft (Taurino und Smith 2022) schlagen wir in diesem Kapitel einen Rahmen für das Training von maschinellen Lernmodellen vor, die auf kuratierten Archivdatensätzen und einem taxonomiebasierten Ansatz zur algorithmischen Co-Kuratierung aufbauen. Der Begriff ‚algorithmische Co-Kuratierung' wird hier verwendet, um eine Reihe von co-kuratorischen Aktivitäten zu definieren, die auf der Ebene

* Das englischsprachige Akronym GLAM steht für Galerien, Bibliotheken, Archive sowie Museen und wird als Referenz für Einrichtungen des kulturellen Erbes verwendet, die sich um die Bewahrung und Verbreitung von Wissen bemühen.

des Datensatzdesigns und der Modellschulung ansetzen. In diesem Rahmen werden Sammlungen des kulturellen Erbes in der Computer Vision unter zwei Gesichtspunkten eingesetzt: 1.) Verbesserung des qualitativen Domänenwissens in KI-Modellen (Training des Algorithmus) und 2.) Verbesserung der menschlich-algorithmischen Zusammenarbeit und der computergestützten kuratorischen Arbeit (Training des Archivs).

Auf der Ebene der menschlichen Kognition können wir, nachdem das Gehirn die grundlegende Fähigkeit zur Unterscheidung von Objekten erworben hat, durch ständiges Training die Fähigkeit zur Interpretation visueller Hinweise perfektionieren und verschiedene Ebenen von Fachwissen erwerben. Es handelt sich um einen fortlaufenden Prozess, der alltägliche Lernerfahrungen mit Objekten der realen Welt, vermittelte Formen des Wissens sowie individuelle und kollektive Erinnerungen umfasst. Heutige vortrainierte KI-Modelle können bei einer Vielzahl von visuellen Aufgaben relativ hohe Konfidenzwerte erreichen und ziemlich genaue Rückschlüsse auf das Allgemeinwissen ziehen. Wenn es jedoch um bereichsspezifisches Wissen geht, z. B. die Interpretation von Kunstwerken und historischen Aufzeichnungen, ist die Wahrscheinlichkeit groß, dass sie unterdurchschnittlich abschneiden oder Fehler liefern. Da Foundation Models zunehmend an Bedeutung gewinnen, besteht eine Lösung für die Unzulänglichkeiten der KI darin, Modelle des maschinellen Lernens zu co-kuratieren, indem historische Datensätze entweder im Pre-Trainingsprozess oder in der Phase des Re-Trainings zur Feinabstimmung auf spezielle Aufgaben verwendet werden. Garcia et al. (2020) schlagen beispielsweise vor, den SemArt-Datensatz – eine Sammlung von Bildern von Kunstgemälden, die mit einer Liste von Attributen und künstlerischen Kommentaren gepaart sind (Garcia und Vogiatzis 2018) – zu verwenden, um ein Basismodell (VIKING) zu trainieren, das speziell für die Extraktion von Wissen aus Bildern von Gemälden für das semantische Kunstverständnis und die Beantwortung von Fragen entwickelt wurde. Ihre Studie zeigt, dass das VIKING-Model bestehende Computer-Vision-Modelle bei der Analyse von Kunstwerken übertrifft (Garcia et al. 2020, 14).

Es hat sich gezeigt, dass KI-Systeme, deren Algorithmen spezifischen kulturellen Aufzeichnungen und nuancierteren abstrakten Kategorien ausgesetzt sind, Expert*innenfähigkeiten nachahmen können. Forscher*innen haben diese Art von Wissen als „qualitatives Domänenwissen, das eine intensive Beschäftigung und Interpretation durch Menschen mit anspruchsvollen Fähigkeiten erfordert“ (Deng et al. 2020, 3), beschrieben. Domänenwissen unterstützt Modelle, die statistisch strenger sind und Probleme der Interpretierbarkeit vermeiden, indem sie für den Menschen verständliche Eingaben und Ausgaben beibehalten (ebd., 4). Die Ergänzung der maschinellen Intelligenz durch Fachwissen trägt dazu bei, die herausragende Fähigkeit von Computern, Informationen zu speichern, durch das notwendige menschliche Fachwissen zu ergänzen, das ganze Bereiche wie Kulturwissenschaften oder Kunstgeschichte hervorgebracht hat. Zu diesem Zweck ist eine historische Dokumentation von grundlegender Bedeutung, um die lexikalischen und visuellen Lücken zu füllen, die bei ‚unspezifischen' maschinellen Lernmodellen auftreten. Ein allgemeiner Archivdatensatz als Ganzes reicht jedoch möglicherweise nicht aus. Wir müssen Archivsammlungen in Museen oder Bibliotheken identifizieren, die als Grenzobjekte (Star und Griesemer 1989) fungieren können, d. h. als symbolische – sowohl abstrakte als auch konkrete – Referenzen, die Algorithmen nutzen können, um Formen der Zusammenarbeit mit menschlichen Akteur*innen zu unterstützen. Am Beispiel eines zoologischen Museums beschreiben Star und Griesemer vier Arten von Grenzobjekten, die die interdisziplinäre Zusammenarbeit zwischen Fachleuten unterstützen können: Repositorien, Idealtypen, koinzidente Grenzen und standardisierte Formen (ebd., 411). Insbesondere können Repositorien „die Form einer Reihe modularer Dinge annehmen. Das sind Dinge, die einzeln entfernt werden können, ohne dass die Struktur des Ganzen zusammenbricht oder verändert wird. Eine Bibliothek zum Beispiel oder eine Sammlung von Fallstudien [...] ist ein Repositorium“ (Star 2010, 603).

Wie wir in diesem Kapitel darlegen, stellt die Identifizierung von Grenzobjekten in großen Sammlungen des Kulturerbes einen strategischen Ausgangspunkt für die Sammlung von Trainingsdaten dar, die maschinellen Lernmodellen bei der Beantwortung bereichsbezogener Fragen helfen können. Große, strukturierte Archivbestände

umfassen historische Repertoires in Form von iterativen und repräsentativen Aufzeichnungen, was sie zu einem großartigen Sammelbecken macht, aus dem man relevante Muster für das Erlernen spezifischer Aufgaben ziehen kann. Darüber hinaus ermöglicht die modulare Struktur kuratierter digitaler Archive die Auswahl entweder eines gesamten Repositoriums (z. B. alle Objekte in einer Sammlung zeitgenössischer Kunst) oder von Untergruppen, die von Interesse sind (z. B. ein Ordner mit Kunstwerken eines einzelnen Künstlers). Je nach den Merkmalen des Archivmaterials, das wir hervorheben wollen, können Ordner unterschiedlichen Organisationsprinzipien folgen. In Kunstsammlungen können Ordner beispielsweise nach geografischen Orten, Zeiträumen oder Kunstrichtungen unterteilt werden, die auf zeitlicher, räumlicher oder stilistischer Nähe basieren. Dieselben Prinzipien und zeitlich-räumlichen Grenzen, die auf Fotojournalismus-Archive angewandt werden, können dazu führen, dass ein Sammelsurium von Bildern entsteht, die verschiedene Ereignisse und Themen abdecken. Eine Ordnerstruktur in einem Repositorium kann dabei helfen, die Datensätze in verschiedene Cluster umzuverteilen, je nach dem Umfang der Aufgabe, die dem maschinellen Lernmodell gestellt wird. Wenn wir beispielsweise ein Computer-Vision-Modell trainieren wollen, um Fragen zu beantworten, über welche Art von Ereignis in einer Fotojournalismus-Berichterstattung berichtet wird, müssten wir Ordner kuratieren, die visuelle Aufzeichnungen mit gemeinsamen Themen und Inhalten enthalten, anstatt sie nach Ort, Datum oder Stil zu unterteilen.

Die organisatorische Denkweise von strukturierten Archiven begünstigt den Wissenstransfer zwischen menschlichen und algorithmischen Kurator*innen vom Datensatzdesign bis zur Trainingsphase. Während die Gestaltung des Datensatzes als direkter Akt der Zusammenarbeit zwischen Mensch und Maschine gedacht werden kann, bei dem Archivar*innen oder Forscher*innen Kataloge, Repositorien und Ordnersysteme kuratieren, die als relevante Quelle für visuelle Daten für Trainingssätze dienen können, kann der kooperative Prozess in der eigentlichen Phase des maschinellen Lernens je nach Größe des Datensatzes unterschiedliche Wege einschlagen. Bei kleinen Archiven des Kulturerbes besteht eine Möglichkeit darin, einen taxonomiebasierten Ansatz zu wählen, der nicht nur die Verwendung von ‚Expert*innen'-Datensätzen unterstützt, sondern auch einen Rahmen für die semantische Interpretation bietet. Michaud et al. (2018) diskutieren unter anderem inhaltsbasierte Bildretrieval-Methoden in der Anwendung auf kleine Kulturerbesammlungen. Sie vergleichen die Leistung verschiedener algorithmischer Modelle sowohl auf ‚generischen' Bilddatensätzen (wie ImageNet) als auch auf ‚Expert*innen'-Datensätzen (z. B. ROMANE 1K, Coin Collection Online Catalogue). In ihrem Aufsatz stellen sie ein „unsupervised Bezugssystem vor, das darauf abzielt, automatisch die Informationen verschiedener lokaler Deskriptoren zu kombinieren“ (ebd., 161). Sie schlagen insbesondere eine Reihe von Verfahren zur Merkmalsextraktion vor, die auf dem Modell der Bags-of-Visual-Words basieren, einer Pre-Deep-Learning-Methode, die Bildmerkmale als Wörter (mit Deskriptoren verbundene Schlüsselpunkte) behandelt und Clustering verwendet, um ein hochdimensionales visuelles Vokabular zu erzeugen. Ihre Studie zeigt, dass das Modell der Bags-of-Visual-Words bei der Anwendung auf Sammlungen des kulturellen Erbes genauere Ergebnisse liefert als Modelle der Convolutional Neural Networks (ebd., 169).

Andere Studien deuten darauf hin, dass die Verwendung von domänenspezifischen Vokabularen, Worteinbettungen und taxonomischen Bezeichnungen die Leistung von Bildanalyse, -klassifizierung und -abfrage verbessern könnte. In einer 2019 veröffentlichten Arbeit stellen Garcia et al. zwei Methoden zur Gewinnung kontextbezogener Embeddings vor, eine auf der Grundlage visueller Elemente und eine auf der Grundlage nicht visueller Attribute, die in einem Wissensgraphen definiert sind. Ihre Forschung zeigt, dass „kontextbewusste Embeddings bei vielen Problemen der automatischen Kunstanalyse von Vorteil sind und die Klassifizierungsgenauigkeit von Kunstwerken um bis zu 7,3 % im Vergleich zu Klassifizierungsbaselines verbessern“ (Garcia et al. 2019, 8). Darüber hinaus wird in einer 2021 veröffentlichten Studie vorgeschlagen, „grobe taxonomische Kennzeichnungen einzubeziehen, um Bildklassifizierer in feinkörnigen Domänen zu trainieren“ (Su und Maji 2021, 1), indem Semi-Supervised Learning (d. h. eine Kombination aus gelabelten und nicht-gelabelten Daten)

verwendet wird, um ein Computer-Vision-Modell auf einem Biologiedatensatz zu trainieren. Die Ergebnisse deuten darauf hin, dass die Implementierung von taxonomiesensiblen Benutzer*inneneingaben zu Verbesserungen bei der Feinklassifizierung von Bildern führt (ebd., 10). Obwohl Taxonomien in Kunst und Kultur nicht immer einer starren Hierarchie von Klassen und Ordnungen folgen, können taxonomische Logiken dennoch erfolgreich eingesetzt werden, um menschliches domänenspezifisches Wissen auf algorithmische Systeme zu übertragen. Kontrollierte Vokabulare stellen neben anderen Ressourcen ein System zur Wissensorganisation dar, das auf Indexierungsschemata basiert, die von Expert*innen entwickelt und gepflegt werden, um lexikalische Mehrdeutigkeit zu reduzieren und terminologische Konsistenz zu gewährleisten. In der Bibliothekswissenschaft stellen diese Listen von Begriffen und Wendungen eine Standardreferenz für die Organisation von Inhalten in Katalogen, die Kommentierung von Bildern und die Kennzeichnung von Archivmaterial mit Metadaten dar.

Ähnlich wie andere Klassifizierungssysteme haben kontrollierte Vokabulare auch Nachteile: Sie müssen ständig aktualisiert werden, um eine veraltete Terminologie zu vermeiden; wenn sie für die Informationsbeschaffung verwendet werden, geht die Präzision mit dem Risiko der Nicht-Vollständigkeit einher; die Gestaltung dieser Systeme erfordert ethische Überlegungen zur lexikalischen Auswahl. Nichtsdestotrotz können kontrollierte Vokabulare zusammen mit anderen kategorialen ‚Schemata', die für die systematische Wissensorganisation verwendet werden, beim ‚supervised' und ‚semi-supervised' wortschatzbasierten Lernen wirksam eingesetzt werden, um die Genauigkeit von Computer-Vision-Modellen zu verbessern, die für die Objekterkennung und Bildkategorisierung trainiert werden (Fu und Sigal 2016). Mit ihren eigenen Einschränkungen können kontrollierte Vokabulare und Klassifizierungsschemata bei der Arbeit mit vortrainierten maschinellen Lernmodellen die Implementierung von menschlicher Kuratierung begünstigen, indem sie eine Richtlinie für die Bewertung der Label-Verteilung bieten, indem sie wiederum Label-Sets, die von vortrainierten Modellen erzeugt wurden, gegenüber anderen priorisieren oder ganze Label-Sets ausschließen, die für eine domänenspezifische Anwendung als nicht relevant erachtet werden.

In den vorangegangenen Abschnitten haben wir uns mit der Frage befasst, wie sich archivarische und kuratorische Arbeit und algorithmische Verfahren bei der Anwendung auf Kulturerbearchive überschneiden könnten. Obwohl viele Archive noch dabei sind, analoge Aufzeichnungen zu digitalisieren, umfassen einige historische Sammlungen bereits Millionen von Bildern, was sie zu guten Kandidatinnen für die Erstellung von Trainingsdatensätzen für die KI macht. Darüber hinaus sind in der Kunstgeschichte und in den Museumswissenschaften strenge Praktiken des Kuratierens weit verbreitet und die Datensätze des Kulturerbes werden oft von einer gründlichen Dokumentation begleitet. So hat das Getty Research Institute die Cultural Objects Name Authority (CONA) eingerichtet, ein strukturiertes Vokabular für die Zusammenstellung von Metadaten über bildende Kunst und kulturelles Erbe. Darüber hinaus unterhält das Getty Vocabulary Program die Categories for the Description of Works of Art (CDWA), eine Liste bewährter Verfahren für die Katalogisierung von Kunstwerken, die in Cataloging Cultural Objects (CCO), einem Handbuch für die Beschreibung von Kulturerbeeinträgen, enthalten ist. Ob in Form von Repositorien, Ordnern, Katalogen oder Vokabularen, die Argumentation in Form von Grenzobjekten könnte der Schlüssel für die Definition von Praktiken der algorithmischen Co-Kuratierung sein, die es Forscher*innen und Expert*innen aus verschiedenen Disziplinen ermöglichen, bei der Entwicklung von KI zusammenzuarbeiten, die feldspezifische Schlussfolgerungen ziehen kann.

Während die meisten gängigen Bilderkennungsmodelle durch Web Scraping auf einer zufälligen Auswahl von Objekten trainiert werden, impliziert das Training des Archivs immer ein Prinzip der menschlichen Kuratierung und Organisation des Datensatzes an der Quelle, der dann auf maschinelle Bildverarbeitungsmodelle übertragen wird. In gewisser Weise beginnt der Prozess des Trainings des Archivs mit dem Training des Algorithmus. In diesem Kapitel wurde ein Überblick darüber gegeben, wie ein

gegenseitiges (sowohl menschliches als auch maschinelles) Training in Praktiken der algorithmischen Co-Kuratierung umgesetzt wird. Nachdem wir erste Definitionen der algorithmischen Kuratierung in Bezug auf Internetkultur und Computerkunstpraktiken skizziert hatten, entwickelten wir einen Ansatz zur algorithmischen Co-Kuratierung, der sich auf die Archiv- und Informationswissenschaft stützt. In diesem Rahmen sind Algorithmen aufgerufen, mit Forscher*innen, Archivar*innen und Kurator*innen in der Phase des Modelldesigns und -trainings zusammenzuwirken, ebenso wie menschliche Kurator*innen eingeladen sind, mit Algorithmen während vernetzter co-kuratorischer Praktiken zusammenzuarbeiten. Die Sammlungen des Kulturerbes wurden als bevorzugter Rahmen für diese Zusammenarbeit diskutiert, an der Schnittstelle zwischen Algorithmen (d. h., die Maschinen trainieren, aus gelabelten oder nicht gelabelten Archivalien zu lernen) und menschlichen Kurator*innen (d. h., die Archivar*innen, Kurator*innen, Bibliothekar*innen und Forscher*innen lehren, mit maschinellen Lernmodellen zu arbeiten und maschinell erzeugte Metadaten zu bewerten). Um den Grad des menschlichen Eingreifens in den Trainingsprozess zu bestimmen, haben wir einen taxonomiebasierten Ansatz für die algorithmische Co-Kuratierung entwickelt, der die Besonderheiten jeder Sammlung des Kulturerbes und die Abfragen, die für die Informationsbeschaffung verwendet werden könnten, berücksichtigt.

Wie bereits erwähnt, beruhen die derzeitigen modernen Modelle für die Computer Vision auf mehreren problematischen Annahmen, die auf der Ebene des Datensatzdesigns gemacht werden. Kanonische Trainingsdatensätze wie ImageNet beziehen ihre Bilder aus dem Internet, wobei der Definition von Grenzobjekten, die kollaborative Arbeit und historisch fundiertes Wissen unterstützen können, nur wenig Beachtung geschenkt wird. In diesen groß angelegten Bilddatensätzen wird die menschliche Kuratierung häufig durch eine automatische Datenerfassung ersetzt, und wenn sie vorhanden ist, erfolgt sie in Form einer allgemeinen Crowd-Annotation, wodurch die Bedeutung der spezifischen Fachkenntnisse unterschätzt wird. Schließlich werfen Benchmark-Datensätze, die in großem Umfang von der Community für das Training und die Validierung von Modellen des maschinelles Lernen verwendet werden, Fragen zur institutionellen Vielfalt auf. Wissenschaftler*innen haben festgestellt, dass „sich das gesamte Feld auf Datensätze konzentriert, die von Forscher*innen in einer kleinen Anzahl von Elite-Institutionen eingeführt wurden" (Koch et al. 2021, 1). Dies kann verschiedene Gründe haben, z. B. fehlende finanzielle Mittel und Ressourcen für die Erstellung von diverseren Trainingsdatensätzen, Probleme bei der Zugänglichkeit von Daten aus dem institutionalisierten Kulturerbe aufgrund von Urheberrechts- oder Datenschutzbestimmungen, Herausforderungen bei Digitalisierungsverfahren, die verhindern, dass Archivmaterial maschinenlesbar ist.

Die Verwendung von historischen Archiven als Trainingssätze kann uns dabei helfen, die Fehler von KI-Modellen auf der Grundlage von großen Web-Scraped-Datensätzen zu beheben und gleichzeitig neue Wege für Anwendungen des maschinellen Lernens zu eröffnen. Einerseits wird durch die Beschaffung von Datensätzen aus historischen Archiven der Schwerpunkt der KI auf die Bedeutung der Dokumentation und des taxonomischen Verständnisses verlagert, sodass Geschichte und kulturelles Gedächtnis das Rückgrat der Modelle für maschinelles Lernen bilden. Andererseits fördert dieser Ansatz der algorithmischen Co-Kuratierung ein kooperatives Modell der maschinellen Unterstützung auf der Grundlage von Fachwissen. Als solches diversifiziert er die Trainingsdaten nicht nur in Bezug auf spezialisierte Repositorien, die auf lokalem Wissen beruhen, sondern auch in Bezug auf die institutionellen Wurzeln.

In ähnlicher Weise kann die Archivierungs- und Kurator*innentätigkeit vom Einsatz der KI zur Automatisierung eines Teils des Archivierungs- und Verwaltungsprozesses profitieren. Insbesondere im Hinblick auf die Extraktion von Metadaten (wie mithilfe von OCR zur Transkription handschriftlicher Manuskripte oder Bilderkennung zum Labeln von Objekten) haben sich Modelle des maschinellen Lernens als wichtige Ressource zur Verbesserung der Qualität der Klassifizierung, Indizierung und Abfrage von Inhalten in Archiven erwiesen (Colavizza et al. 2022). In diesem Szenario haben wir den Begriff der algorithmischen Co-Kuration nicht nur als theoretischen Rahmen eingeführt, der zur Beschreibung von Fällen verwendet werden kann, in denen Maschi-

nen Inhalte über das Internet auf mehr oder weniger hybride und interaktive Weise kuratieren. Vielmehr schlugen wir vor, ihn als methodischen Ansatz für die Organisation und Extraktion von Informationen zu verwenden, der sich sowohl auf menschliche als auch auf maschinelle Intelligenz stützt, um das kollektive Wissen und Gedächtnis zu bewahren. Mit dieser aktualisierten Definition zielt unser Aufsatz darauf ab, die wissenschaftliche Aufmerksamkeit vom Konzept des Algorithmus als Kurator*in auf die eigentliche Praktik des Kuratierens des Algorithmus zu verlagern. Das heißt, dass es nicht darum geht, die Rolle der Kurator*innen im Kontext der algorithmischen Kultur neu zu definieren, sondern ein algorithmisches Design zu etablieren, das sich in bereits etablierte kuratorische Traditionen einfügt. Diesem Ansatz folgend, schlagen wir vor, dass Modelle des maschinellen Lernens für Archive so zu entwickeln sind, dass sie erfolgreich in bestehende Arbeitsabläufe integriert werden können. Dadurch kann zu einer KI-gestützten Kuration beitragen werden, die über die Bewältigung von Standardaufgaben hinausgeht und Formen einer spezialisierten algorithmischen Co-Kuration umfasst, die eine feinkörnige Analyse von übergeordneten fachlichen Kategorien und Domänenwissen ermöglicht.

Literaturverzeichnis

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein et al. 2022. „On the Opportunities and Risks of Foundation Models“. arXiv. http://arxiv.org/abs/2108.07258.

Cairns, Susan und Danny Birchall. 2013. „Curating the Digital World: Past Preconceptions, Present Problems, Possible Futures“. In *Museums and the Web*, herausgegeben von Nancy Proctor und Rich Cherry. Silver Spring, MD: Museums and the Web. https://mw2013.museumsandtheweb.com/paper/curating-the-digital-world-past-preconceptions-present-problems-possible-futures.

Colavizza, Giovanni, Tobias Blanke, Charles Jeurgens und Julia Noordegraaf. 2022. „Archives and AI: An Overview of Current Debates and Future Perspectives“. *Journal on Computing and Cultural Heritage* 15 (1): 1–15. https://doi.org/10.1145/3479010.

Crawford, Kate und Trevor Paglen. 2021. „Excavating AI: The Politics of Images in Machine Learning Training Set“. *AI & SOCIETY* 36 (Juni): 1105–16. https://doi.org/10.1007/s00146-021-01162-8.

Davis, Jenny L. 2016. „Curation: A Theoretical Treatment“. *Information, Communication & Society* 20 (5): 770–83. https://doi.org/10.1080/1369118X.2016.1203972.

Dekker, Annet und Gaia Tedone. 2019. „Networked Co-Curation: An Exploration of the Socio-Technical Specificities of Online Curation“. *Arts* 8 (3), 86. https://doi.org/10.3390/arts8030086.

Deng, Changyu, Xunbi Ji, Colton Rainey, Jianyu Zhang und Wei Lu. 2020. „Integrating Machine Learning with Human Knowledge“. *iScience* 23 (11), 101656. https://doi.org/10.1016/j.isci.2020.101656.

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li und Li Fei-Fei. 2009. „ImageNet: A Large-Scale Hierarchical Image Database“. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–55. IEEE. https://doi.org/10.1109/CVPR.2009.5206848.

Doueihi, Milad. 2011. *Digital Cultures*. Amerikanische Edition. Cambridge, MA: Harvard University Press.

Fellbaum, Christiane, Hrsg. 1998. *WordNet: An Electronic Lexical Database*. Eingeleitet von George Miller. Cambridge, MA: The MIT Press.

Fiorucci, Marco, Marina Khoroshiltseva, Massimiliano Pontil, Arianna Traviglia, Alessio Del Bue und Stuart James. 2020. „Machine Learning for Cultural Heritage: A Survey“. *Pattern Recognition Letters* 133 (Mai): 102–8. https://doi.org/10.1016/j.patrec.2020.02.017.

Foucault, Michel. 1969. *Archäologie des Wissens*. Frankfurt am Main: Suhrkamp.

Fu, Yanwei und Leonid Sigal. 2016. „Semi-Supervised Vocabulary-Informed Learning“. arXiv. http://arxiv.org/abs/1604.07093.

Garcia, Noa, Benjamin Renoust und Yuta Nakashima. 2019. „Context-Aware Embeddings for Automatic Art Analysis“. In *Proceedings of the 2019 International Conference on Multimedia Retrieval*, 25–33. https://doi.org/10.1145/3323873.3325028.

Garcia, Noa und George Vogiatzis. 2018. „How to Read Paintings: Semantic Art Understanding with Multi-Modal Retrieval“. arXiv. http://arxiv.org/abs/1810.09617.

Garcia, Noa, Chentao Ye, Zihua Liu, Qingtao Hu, Mayu Otani, Chenhui Chu, Yuta Nakashima und Teruko Mitamura. 2020. „A Dataset and Baselines for Visual Question Answering on Art“. arXiv. http://arxiv.org/abs/2008.12520.

Graham, Beryl, Sarah Cook und Steve Dietz. 2010. *Rethinking Curating: Art after New Media*. Leonardo Books. Cambridge, MA: The MIT Press.

Groys, Boris. 2016. „The Truth of Art“. *e-flux* 71 (März). https://www.e-flux.com/journal/71/60513/the-truth-of-art.

Grubinger, Eva. 2006. „C@C: Computer-Aided Curating (1993–1995)“. In *Curating Immateriality: The Work of the Curator in the Age of Network Systems (DATA Browser 3)*, herausgegeben von Joasia Krysa, 107–15. New York: Autonomedia.

Hendricks, Manique. 2017. „The Algorithm as Curator: In Search of a Non-Narrated Collection Presentation“. *Stedelijk Studies Journal* 1. https://doi.org/10.54533/StedStud.vol005.art11.

Heuer, Hendrik. 2020. Users and Machine Learning-Based Curation Systems. Universität Bremen: Dissertation. https://doi.org/10.26092/ELIB/241.

Higgins, Sarah. 2008. „The DCC Curation Lifecycle Model“. *International Journal of Digital Curation* 3 (1): 134–40. https://doi.org/10.2218/ijdc.v3i1.48.

Hogan, Bernie. 2010. „The Presentation of Self in the Age of Social Media: Distinguishing Performances and Exhibitions Online“. *Bulletin of Science, Technology & Society* 30 (6): 377–86. https://doi.org/10.1177/0270467610385893.

Horie, Charles V. 1986. „Who Is a Curator?“. *International Journal of Museum Management and Curatorship* 5 (3): 267–72. https://doi.org/10.1080/09647778609515029.

Jakubik, Johannes, Michael Vössing, Niklas Kühl, Jannis Walk und Gerhard Satzger. 2022. „Data-Centric Artificial Intelligence“. arXiv. http://arxiv.org/abs/2212.11854.

Jo, Eun Seo und Timnit Gebru. 2020. „Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning“. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 306–16. https://doi.org/10.1145/3351095.3372829.

Koch, Bernard, Emily Denton, Alex Hanna und Jacob G. Foster. 2021. „Reduced, Reused and Recycled: The Life of a Dataset in Machine Learning Research“. arXiv. http://arxiv.org/abs/2112.01716.

Krysa, Joasia, Hrsg. 2006. *Curating Immateriality: The Work of the Curator in the Age of Network Systems*. New York: Autonomedia Press.

———. 2008. *Software Curating: The Politics of Curating in/as (an) Open System(s)*. University of Plymouth: Dissertation. https://doi.org/10.24382/4984.

———. 2013. „Some Questions on Curating as (Public) Interface to the Art Market“. *APRJA* (Archiv), 12. April 2013. https://www.academia.edu/30947090/Some_Questions_on_Curating_as_Public_Interface_to_the_Art_Market.

———. 2014. „Kurator - a Proposal for an Experimental, Permutational Software Application Capable of Curating Exhibitions“. In *Networks*, herausgegeben von Lars Bang Larsen, 118–21. Documents of Contemporary Art. Cambridge, MA: The MIT Press.

Murphy. 2022. „What Are Foundation Models?“. *IBM Research* (Blog), 9. Mai 2022. https://research.ibm.com/blog/what-are-foundation-models.

Michaud, Dorian, Thierry Urruty, Philippe Carré und François Lecellier. 2018. „Adaptive Features Selection for Expert Datasets: A Cultural Heritage Application“. *Signal Processing: Image Communication* 67 (September): 161–70. https://doi.org/10.1016/j.image.2018.06.011.

Morris, Jeremy Wade. 2015. „Curation by Code: Infomediaries and the Data Mining of Taste“. *European Journal of Cultural Studies* 18 (4–5): 446–63. https://doi.org/10.1177/1367549415577387.

Nagler, Christian und Joseph del Pesco. 2011. „Curating in the Time of Algorithms“. *Fillip* 15 (Herbst): 52–61. https://www.academia.edu/15258682/Curating_in_the_Time_of_Algorithms.

Obrist, Hans Ulrich. 2014. *Ways of Curating*. Amer. Erstauflage. New York: Faber and Faber.

Pereira, Gabriel und Bruno Moreschi. 2021. „Artificial Intelligence and Institutional Critique 2.0: Unexpected Ways of Seeing with Computer Vision“. *AI & SOCIETY* 36 (4): 1201–23. https://doi.org/10.1007/s00146-020-01059-y.

Schleiner, Anne-Marie. 2003. „Fluidities and Oppositions among Curators, Filter Feeders and Future Artists“. *intelligent agent* 3, Nr. 1 (März). http://www.intelligentagent.com/archive/v03[1].01.curation.schleiner.pdf.

Star, Susan Leigh. 2010. „This Is Not a Boundary Object: Reflections on the Origin of a Concept“. *Science, Technology, & Human Values* 35 (5): 601–17. https://doi.org/10.1177/0162243910377624.

Star, Susan Leigh und James R. Griesemer. 1989. „Institutional Ecology, ‘Translations’ and Boundary Objects: Amateurs and Professionals in Berkeley’s Museum of Vertebrate Zoology, 1907-39“. *Social Studies of Science* 19 (3): 387–420. https://doi.org/10.1177/030631289019003001.

Su, Jong-Chyi und Subhransu Maji. 2021. „Semi-Supervised Learning with Taxonomic Labels“. arXiv. http://arxiv.org/abs/2111.11595.

Taurino, Giulia. 2020. „Redefining the Anthology: Forms and Affordances in Digital Culture“. https://doi.org/10.6092/UNIBO/AMSDOTTORATO/9365.

Taurino, Giulia und David Smith. 2022. „Machine Learning as an Archival Science: Narratives behind Artificial Intelligence, Cultural Data, and Archival Remediation“. In *NeurIPS AI Cultures Workshop*. https://ai-cultures.github.io/papers/machine_learning_as_an_archiva.pdf.

Tedone, Gaia. 2019. „Human-Algorithmic Curation: Curating with or against the Algorithm“. In *2019 Conference on Computation, Communication, Aesthetics & X*, 125–39. Mailand, Italien: xCoAx. https://2019.xcoax.org/xCoAx2019.pdf.

Tyżlik-Carver, Magda. 2016. Curating in/as Commons. Posthuman Curating and Computational Cultures. Aarhus University: Dissertation.

———. 2017. „| Curator | Curating | the Curatorial | Not-Just-Art Curating: A Genealogy of Posthuman Curating“. *Springerin*, Nr. 1. https://www.springerin.at/en/2017/1/kuratorin-kuratieren-das-kuratorische-nicht-nur-kunst-kuratieren/.

Yakel, Elizabeth. 2007. „Digital Curation“. *OCLC Systems & Services: International Digital Library Perspectives* 23 (4): 335–40. https://doi.org/10.1108/10650750710831466.

In order to incorporate the curators' historical, stylistic, and object contextual knowledge, a process of human-machine interaction is significant. *Training the Archive* investigated whether the process of curating can be broken down into its individual steps to transfer them into statistical procedures. For this reason, the 'curatorial gaze'—understood as a complex gesture of bringing together and selecting artworks—will itself form the basis for the machine learning methods used. The result is a software application, the so-called 'Curator's Machine,' that enables an exploratory search of museum collections, where the recommended artworks are in turn influenced and trained by expert-made groupings, thus placing the objects into (novel) context. In doing so, the Curator's Machine is understood as a generator of ideas that puts the human at its centre and is intended to support processes of rediscovery and revisiting of digital objects in art museum collections. Together with RWTH Aachen University, we developed the software in different prototype phases. To address the needs of curators as potential users, we interviewed eleven of them and thoroughly explored the surveyed process of curating (see working paper: *From Keras Import Curating)*. The prototype phases brought their own findings and were progressed iteratively.

Prototype 1

First stage was the clustering of objects in a museum collection with the use of pre-trained 'off-the-shelf' artificial neural network models. Investigating whether automated visual groupings can be changed by training the algorithm with man-made annotations about hidden patterns of connection between artworks was of particular interest.

Prototype 2

The objective was the development of a recommender system that provides suggestions from the collection depending on a sequence of image selections by an expert. This annotated sequence of artworks that would belong together in an exhibition represents a trajectory through the embedding space that the recommender system is supposed to replicate to continue the 'path' and make meaningful suggestions to the curator by presenting Nearest Neighbor samples. The finding in this prototype phase was the omission of pre-trained artificial neural networks in favor of a self-built auto-encoder due to identified biases towards art-historical image corpora (see working paper: *"Why So Many Windows?")*.

Prototype 3

This iteration was marked by the usage of vision-language models for simultaneous embedding of semantic text and image information—via the neural network CLIP—to be able to draw on extended textual concepts and descriptions for the recommendations. The major benefit in these multimodal models is that the artworks do not have to be labeled with manual annotations and the query is not being processed in a relational sense. Thus, for each possible query, one gets a statistical interpretation by CLIP of which digitized object from the collection best matches the given text prompt. This leads to a constant stream of search results, through which curators are encouraged to discover new content in the collection and could approach an individual exhibition idea by adapting their search prompts.

Prototype 4

The Curator's Machine became an easy-to-use multimodal retrieval system that suggests relevant artworks from the museum collection based on search prompts only. The artworks of interest can be interactively arranged and grouped together. The model learns from the manually set clusters as well as the defined relation patterns on the

canvas and adapts the image search results in real time. The design challenge was to keep the query time short and to develop an appealing and simple interface (see working paper: *Point Clouds. Scatterplots and Tables as User Interfaces of Artificial 'Intelligence')*. The software was tested and given feedback by curators from the project network. The final use case will be the application of the Curator's Machine to the digitized collection of the Ludwig Forum Aachen.

In summary, the key findings of the prototyping process were the use of a multimodal system such as CLIP for exploring museum collections and the adaptation of the search results to the individual user with manual selections by means of so-called 'localized latent updates' to fine-tune the vision-language model (see paper by Ibing et al. in this book). The Curator's Machine is available through an open repository, ready to be used in many museums and other digital archives. To download the framework and start curating yourself, scan the QR code.

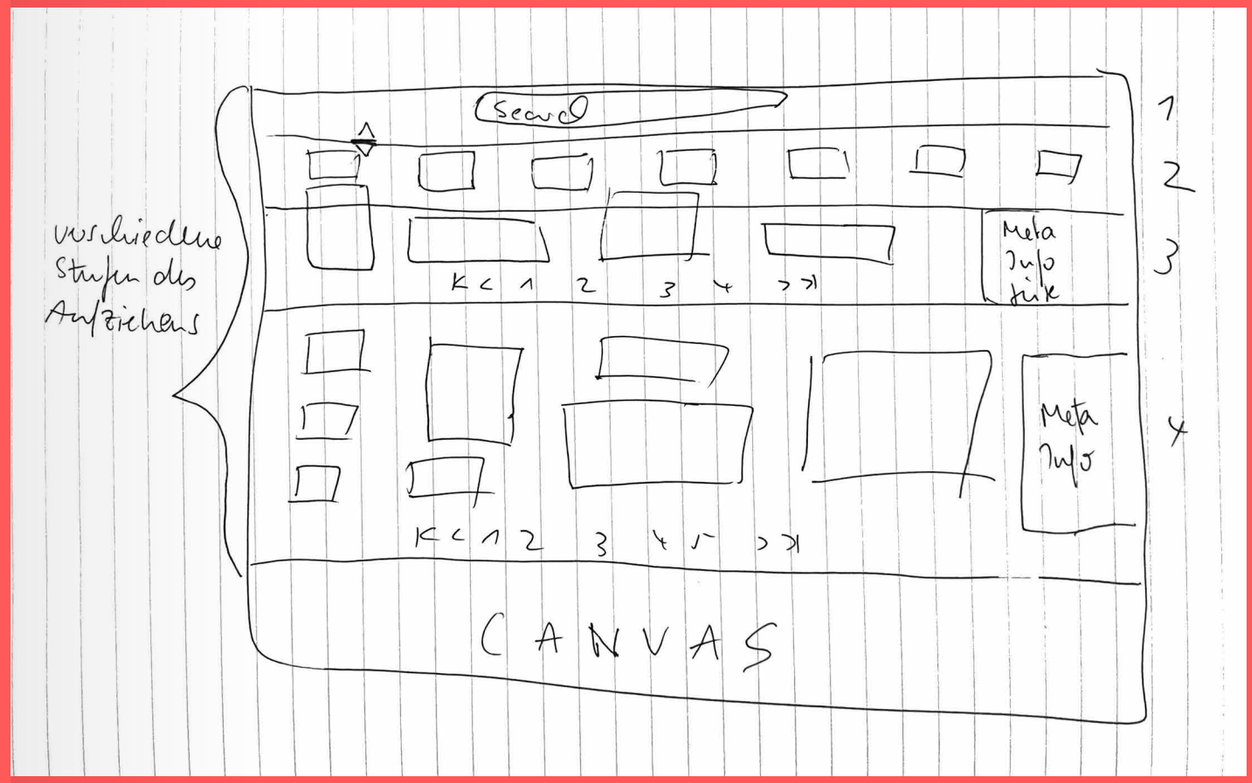
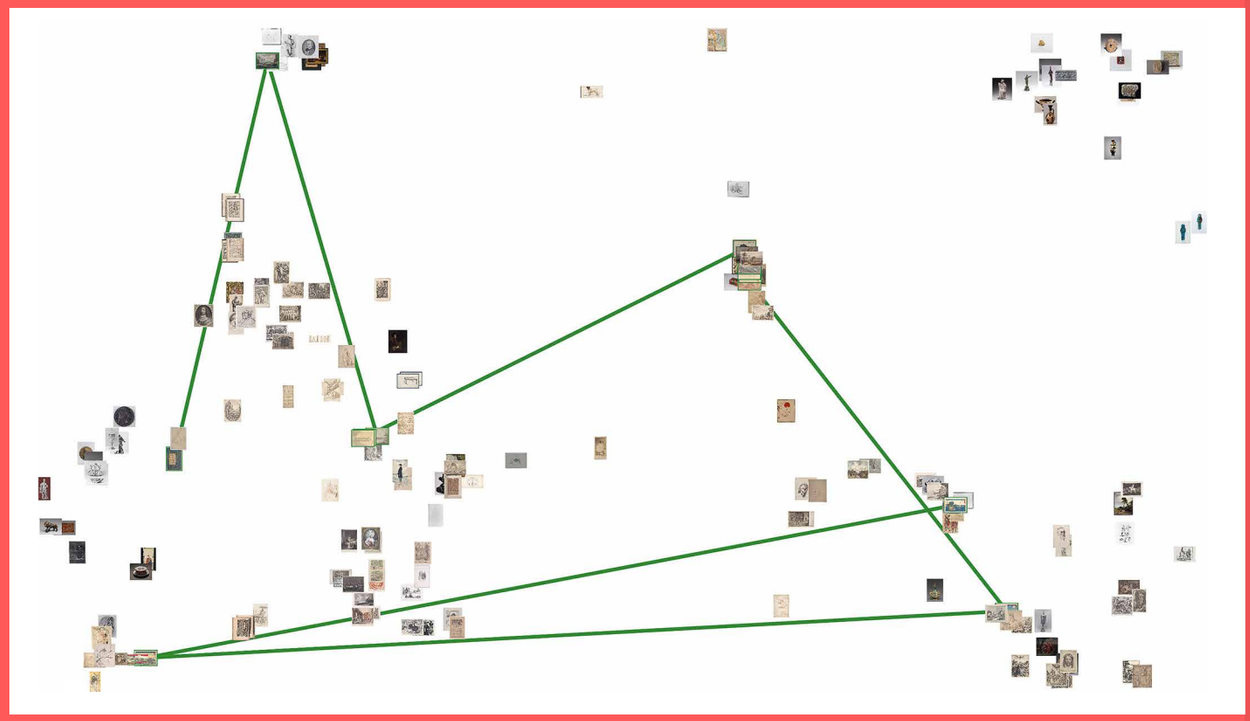
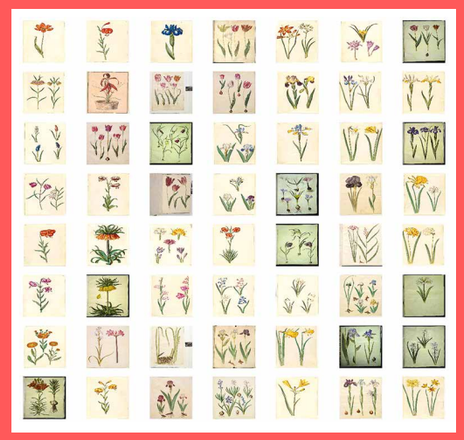


Table of Figures

pp. 97, 100 (bottom): A scatter plot with different busts and a grid plot with drawings of flowers as a cluster visualization examined in the first prototype iteration. Dominik Bönisch, Ludwig Forum Aachen. 2020. All images are due to open-source licensing from the Statens Museum for Kunst (SMK), Copenhagen.
p. 100 (top): Example of a trajectory through the embedding space of recommender system from the second prototype. Visual Computing Institute (VCI), RWTH Aachen University. 2020. All images are open-source data from the Metropolitan Museum of Art, New York.
p. 101: Process of an iteration to obtaining annotations for the first prototype. Dominik Bönisch and Thomas Rost, Ludwig Forum Aachen. 2020. All images are from the SMK.
pp. 99, 104: Early layout scribble of the user interface implemented in the third prototype and sketch of the features and design that were included in the fourth prototype. Dominik Bönisch, Ludwig Forum Aachen. 2022–23.
p. 105: Final interface of the Curator's Machine. VCI, RWTH Aachen University. 2023. All images are from the SMK.
pp. 110, 111: Modes of mapping visual connections between artworks in a collection, e.g., as a Nearest Neighbor relation or as a cluster. Dominik Bönisch and Thomas Rost, Ludwig Forum Aachen. 2020. All images are from the SMK.
pp. 112–13, 114: Views from the exhibition display at the Ludwig Forum Aachen. Mareike Tocha. 2022.
pp. 115, 120: The unfolded program booklet and the designed poster motif for the conference on 'Art & Algorithms.' Off Office. 2022.





enter anchor and positive, separated by comma. (x to quit)

Um das historische, stilistische und objektbezogene Wissen von Kurator*innen in KI-Anwendungen einzubeziehen, ist ein Prozess der Mensch-Maschine-Interaktion von Bedeutung. *Training the Archive* untersuchte, ob sich der Prozess des Kuratierens in seine einzelnen Schritte zerlegen lässt, um sie in statistische Verfahren zu übertragen. Aus diesem Grund wird der ‚kuratorische Blick‘ – verstanden als komplexe Geste der Kontextualisierung und Auswahl von Kunstwerken – selbst die Grundlage für die eingesetzten maschinellen Lernverfahren bilden. Das Ergebnis ist eine Softwareanwendung, die sogenannte ‚Curator’s Machine‘, die eine explorative Suche in Museumssammlungen ermöglicht, wobei die systemseitig empfohlenen Kunstwerke wiederum durch von Expert*innen erstellte Gruppierungen beeinflusst und trainiert werden und so die Objekte in einen (neuen) Kontext stellen. Die Curator’s Machine versteht sich dabei als Ideengeberin, die den Menschen in den Mittelpunkt stellt und Prozesse der Wiederentdeckung und Wiedererfahrung von digitalen Objekten in Kunstmuseumssammlungen unterstützen soll. Gemeinsam mit der RWTH Aachen University haben wir die Software in verschiedenen Prototypenphasen entwickelt. Um die Bedürfnisse von Kurator*innen als potenzielle Nutzer*innen zu adressieren, haben wir elf von ihnen interviewt und den Prozess des Kuratierens erforscht (siehe Working Paper: *From Keras Import Curating)*. Die Prototyp-Phasen brachten jeweils ihre eigenen Erkenntnisse für unsere Forschung und wurden iterativ vorangetrieben.

Prototyp 1

Die erste Phase fokussierte sich auf das Clustering von Objekten in einer Museumssammlung mithilfe von vortrainierten künstlichen neuronalen Netzwerkmodellen. Von besonderem Interesse war dabei die Untersuchung der Frage, ob automatisch erstellte visuelle Gruppierungen durch das Trainieren des Algorithmus mit von Menschen gemachten Annotationen zu verdeckten Beziehungsmustern zwischen Kunstwerken verändert werden können.

Prototyp 2

Das Ziel war die Entwicklung eines Empfehlungssystems, das Vorschläge aus der Sammlung abhängig von einer von Expert*innen ausgewählten, kuratierten Bildsequenz liefert. Diese annotierte Sequenz von Kunstwerken, die in einer Ausstellung zusammengehören würden, stellt eine Trajektorie durch den Einbettungsraum dar, die das Empfehlungssystem nachbilden soll, um den ‚Pfad‘ fortzusetzen und den Kurator*innen sinnvolle Vorschläge zu liefern, indem es ‚Nearest Neighbors‘ präsentiert. Das Ergebnis dieser Prototyp-Phase war der Verzicht auf vortrainierte künstliche neuronale Netze zugunsten eines selbst erstellten Auto-Encoders aufgrund der festgestellten Bias für kunsthistorische Bildkorpora (siehe Working Paper: *„Why So Many Windows?“)*.

Prototyp 3

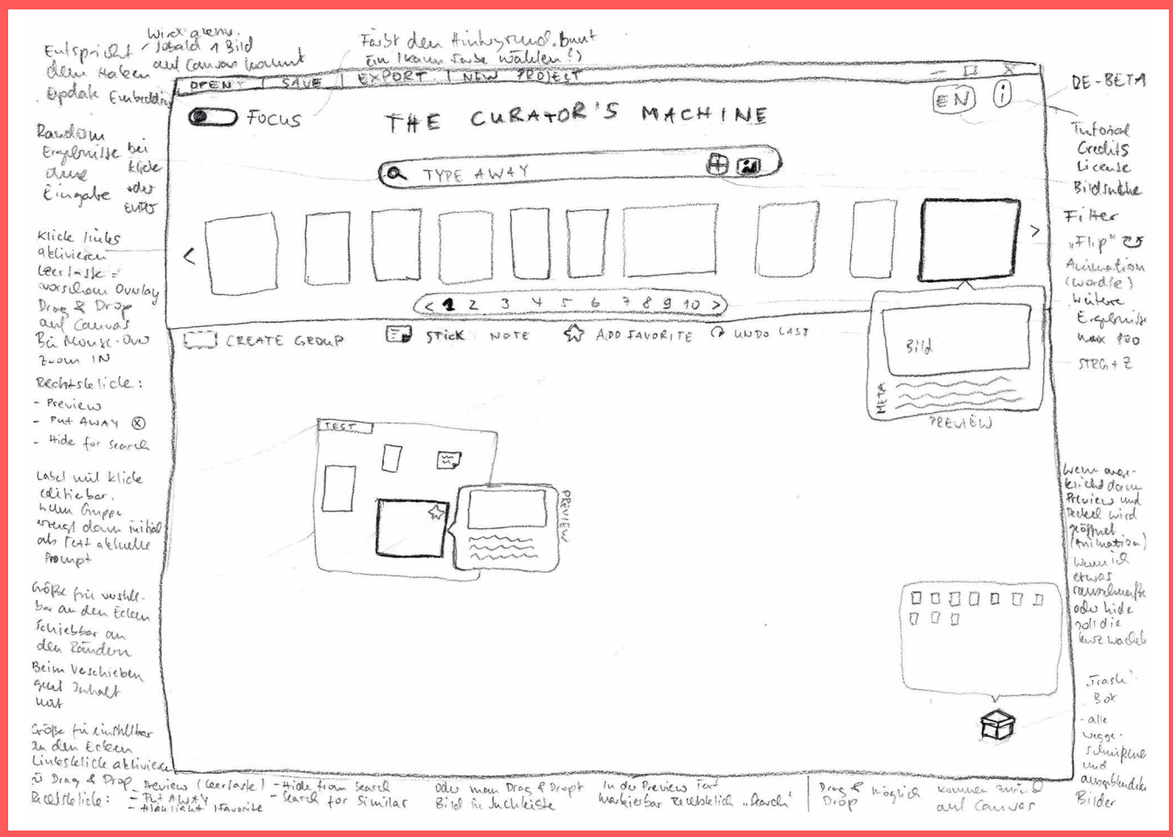
Diese Iteration war gekennzeichnet durch die Verwendung von Vision-Language-Modellen zur gleichzeitigen Einbettung von semantischen Text- und visuellen Bildinformationen – über das neuronale Netzwerk CLIP –, um auf erweiterte textuelle Konzepte und Beschreibungen für die Vorschlagsempfehlungen zurückgreifen zu können. Der große Vorteil dieser multimodalen Modelle besteht darin, dass die Kunstwerke nicht mit manuellen Annotationen versehen werden müssen und die Abfrage nicht in einem relationalen Sinne verarbeitet wird. So erhält man für jede mögliche Suchanfrage eine statistische Interpretation durch CLIP, welches digitalisierte Objekt aus der Sammlung am besten zu dem gegebenen ‚Prompt‘ passt. Dies führt zu einem konstanten Fluss von Suchergebnissen, durch den Kurator*innen ermutigt werden sollen, neue Inhalte in der Sammlung zu entdecken und sich durch Anpassung ihrer Suchanfragen einer individuellen Ausstellungsidee zu nähern.

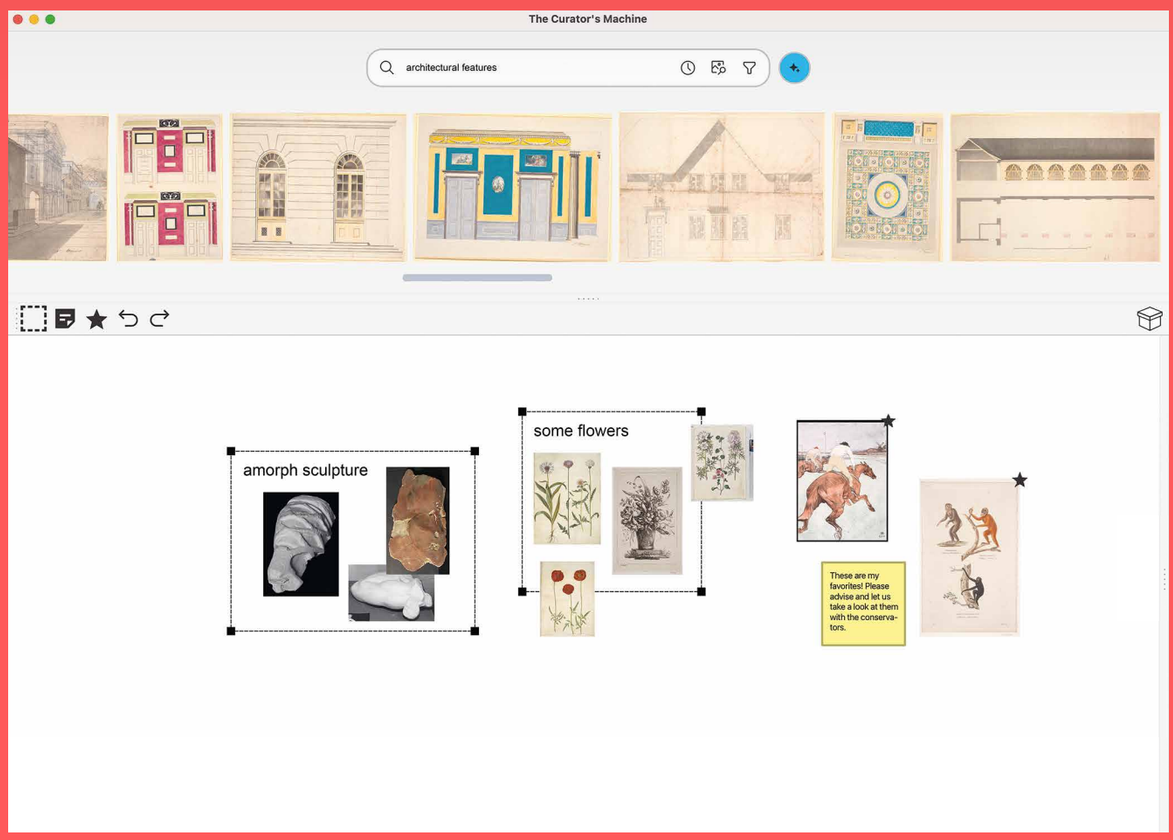
Die Curator’s Machine wurde zu einem einfach zu bedienenden multimodalen Suchabfragesystem, das relevante Kunstwerke aus der Museumssammlung nun auf Basis von Textprompts oder Bilderähnlichkeitssuchen vorschlägt. Die Kunstwerke von Interesse können interaktiv auf einem Arbeitsbereich angeordnet und gruppiert werden. Das Modell lernt aus diesen manuell erstellten Clustern sowie den durch Zuordnung definierten Beziehungsmustern und passt die Bildsuchergebnisse in Echtzeit an. Die Herausforderung an das Design der Nutzer*innenoberfläche bestand darin, die Abfragezeiten kurz zu halten und ein ansprechendes und einfaches Interface zu entwickeln (siehe Working Paper: *Punktwolken. Scatterplots und Tabellen als User-Interfaces Künstlicher ‚Intelligenz‘)*. Die Software wurde ausgiebig getestet und mit Kurator*innen aus unserem Netzwerk rückgekoppelt. Projektabschluss wird die Anwendung der Curator’s Machine auf die digitalisierte Sammlung des Ludwig Forums Aachen bilden.

Zusammenfassend waren die wichtigsten Erkenntnisse aus dem Prototyping-Prozess die Nutzung eines multimodalen Systems wie CLIP/Open-CLIP für die Exploration von Museumssammlungen und die Anpassung der Suchergebnisse an individuelle Nutzer*innen mittels sogenannter ‚lokalisierter latenter Updates‘ zur Feinabstimmung des Vision-Language-Modells (siehe Beitrag von Ibing et al. in diesem Buch). Die Curator’s Machine ist über ein offenes Repository verfügbar und kann in vielen Museen und anderen digitalen Archiven eingesetzt werden. Um das Framework herunterzuladen und selbst mit dem Kuratieren zu beginnen, muss der QR-Code gescannt werden.

Abbildungsverzeichnis

S. 97, 100 (unten): Ein Streudiagramm mit verschiedenen Büsten und ein Rasterdiagramm mit Zeichnungen von Blumen als Clustervisualisierung, die in der ersten Iteration des Prototyps untersucht wurden. Dominik Bönisch, Ludwig Forum Aachen. 2020. Aus lizenzrechtlichen Gründen werden Open-Source-Daten des Statens Museum for Kunst (SMK), Kopenhagen für die Abbildung verwendet.
S. 100 (oben): Beispiel für eine Trajektorie durch den Einbettungsraum des Empfehlungssystems aus dem zweiten Prototyp. Visual Computing Institute (VCI), RWTH Aachen University. 2020. Alle abgebildeten Miniaturansichten sind Open-Source-Daten aus dem Metropolitan Museum of Art, New York.
S. 101: Prozess einer Iteration zur Gewinnung von Annotationen für den ersten Prototyp. Dominik Bönisch und Thomas Rost, Ludwig Forum Aachen. 2020. Alle gezeigten Abbildungen stammen aus offenen Daten vom SMK.
S. 99, 104: Scribble der implementierten Benutzer*innenoberfläche des dritten Prototyps und Skizze der Funktionen sowie vom Layout des vierten Prototyps. Dominik Bönisch, Ludwig Forum Aachen. 2022–23.
S. 105: Finales Interface der Curator’s Machine. VCI, RWTH Aachen University. 2023. Die Abbildungen der gezeigten Kunstwerke stammen aus den offenen Daten des SMK.
S. 110, 111: Arten der Darstellung visueller Zusammengehörigkeit von Kunstwerken aus einer Sammlung, z. B. als Nearest-Neighbor-Beziehung oder als ein Cluster. Dominik Bönisch und Thomas Rost, Ludwig Forum Aachen. 2020. Alle gezeigten Abbildungen stammen aus offenen Daten vom SMK.
S. 112–13, 114: Ansichten aus der Ausstellung am Ludwig Forum Aachen. Mareike Tocha. 2022.
S. 115, 120: Das ausgefaltete Programmheft und das gestaltete Plakatmotiv zur Konferenz ‚Kunst & Algorithmen‘. Off Office. 2022.


THE CURATOR'S MACHINE
OPEN
SAVE
EXPORT
NEW PROJECT
FOCUS
EN
TYPE AWAY
CREATE GROUP
STICK NOTE
ADD FAVORITE
UNDO LAST
PREVIEW
TEST
DE-BETA
Tutorial
Credits
License
Bildsuche
Filter
Trash
Bild
META


The Curator's Machine
architectural features
amorph sculpture
some flowers
These are my favorites! Please advise and let us take a look at them with the conservators.

Published Working Papers

The working papers mark our state of research and discussion within the project group. Seven texts make new knowledge available both to those involved and to a broader public. The results of *Training the Archive* have thus been continuously published throughout the process, introduced into the scholarly discourse, and made accessible to interested art and cultural institutions. The papers are available for download via the respective QR code.

The Curator's Machine: Clustering of Museum Collection Data through Annotation of Hidden Connection Patterns between Artworks—Dominik Bönisch

The discoverability of items within digitized art museums is an ongoing problem for collection managers, who strive to make objects maximally visible to both researchers and the public. Here it is not enough to simply limit the search in databases to narrowly defined keywords. Rather, specific interfaces and visualizations should allow the user to explore an online inventory as well as to 'stroll' through the digital collection. Machine learning can reveal connections and links between artworks that otherwise might not have been fully legible. The working paper presents a first prototype from where the research project *Training the Archive* in the field of digital humanities investigates the machine-aided, explorative (re)discovery of connections within the museum's collection.

"Why So Many Windows?" How the ImageNet Image Database Influences Automated Image Recognition of Historical Images—Francis Hunger

In the field of computer vision, the ImageNet data collection plays a central role as a training dataset. The text discusses the extent to which ImageNet influences the software prototype: The so-called 'Curator's Machine.' The Curator's Machine is designed to facilitate the discovery of relationships and connections between artworks for curators. The text explains how ImageNet, anchored in contemporary image worlds, acts on contemporary and historical artworks by 1.) examining the absence of the classification 'art' in ImageNet, 2.) questioning ImageNet's lack of historicity, and 3.) discussing the relationship between texture and outline in ImageNet-based automated image recognition. This research is important for the genealogical, art historical and coding related usage of the ImageNet dataset in the fields of curating, art history, art studies, and digital humanities.

Curation and Its Statistical Automation by Means of Artificial 'Intelligence'—Francis Hunger

The concept of post-AI curating discussed in this working paper explores curation as a knowledge-creation process, supported by pattern recognition and weighted networks as technical tools of artificial 'intelligence.' The text discusses a number of concepts that build on each other, such as curating, curator, the curatorial, curatorial experimental research, post-human curating, and post-AI curating. It then examines several projects as case studies that approach curation using artificial 'intelligence': *The Next Biennial Should Be Curated by a Machine* (2021) from UBERMORGEN, Leonardo Impett and Joasia Krysa as a meta-artwork about curation and biennials; Tillmann Ohm's project *Artificial Curator* (2020), which resulted in an automatically curated exhibition; and *#Exstrange* (2017) by Rebekah Modrak and Marialaura Ghidini et al., which presents artworks as data objects on the eBay online platform. The text contributes to the field of curatorial studies from a perspective of post-AI-curating.

Can Artificial Intelligence Be Biased? On the Critique of AI's 'Algorithmic Bias' in the Arts—Inke Arns

This working paper is dedicated to artistic positions that critically deal with artificial 'intelligence' and automated pattern recognition through algorithms. Using a series of examples, it shows the social struggles that result from the distortions of bias and how artists react to it. Building on analyses by Harun Farocki and Hito Steyerl, projects by Adam Harvey and Jules LaPlace, Zach Blas and Jemima Wyman, Elisa Giardina Papa, Francis Hunger and Flupke, Erika Scourti, Mimi Onuoha, Nora Al-Badri, and Jan Nikolai Nelles are presented. Various artistic strategies from the field of artistic research are made visible to deal with an actually invisible object: data. The text discusses artistic reactions to societal changes that can be summarized under the notion of artificial 'intelligence.'

Point Clouds. Scatterplots and Tables as User Interfaces of Artificial 'Intelligence'—Francis Hunger

Scatterplots and tabular structures are the main graphical user interfaces for the diagrammatic depiction of large image data collections, which have been processed with visual recognition tools. This study discusses various media visualisations as case examples: *ARTigo* (LMU Munich, 2010–23), *imgs.ai* (UC Santa Barbara & Philipps-Universität Marburg, 2020–), *iArt* (LMU Munich & Leibniz Universität Hannover, 2018–23), *Vikus Viewer* (FH Potsdam, 2014–17), and *X Degrees of Separation* (Mario Klingemann, 2016). It further investigates how these projects employ visualisation algorithms like PCA, t-SNE and UMAP in relation to artificial weighted networks such as VGG19 or CLIP. This comparative study is situated in the fields of software studies, design and user interface critique and related to Big Data applications and the digital humanities.

Unhype Artificial 'Intelligence!' A Proposal to Replace the Deceiving Terminology of AI —Francis Hunger

Artificial intelligence as a field of research and also its criticism is dominated by notions such as 'intelligence,' 'learning' or 'neuronal.' This paper discusses how the use of anthropomorphizing language is fueling AI hype. AI hype involves many promises, such as that 'AI can be creative,' or 'AI can solve world hunger.' This hype is problematic since it covers up the negative consequences of 'AI' use. Instead, the author proposes to use alternative terminology such as: 'automated pattern recognition,' 'machine conditioning,' or 'weighted network.' The working paper delivers pragmatic suggestions, which are positioned in the fields of 'AI' critique and computer sciences and approach the issue with a media-genealogical perspective.

'From Keras Import Curating': An Empirical Study on the Application of Curatorial Practice to Machine Learning Models—Dominik Bönisch & Francis Hunger

The working paper evaluates eleven interviews with professional curators conducted for the project *Training the Archive*. From the interviews, a series of personas were derived that provided insight into the needed functionality of the Curator's Machine and led to corresponding changes in the prototypes. The text then highlights, on the one hand, curatorial approaches that can be realized with the Curator's Machine, including systems of non/order and databases, the process of selection, and thinking in terms of configurations and categories, as well as a reflection on the data basis used. On the other hand, curatorial approaches that are not feasible or only to a limited extent were discussed, for example, issues of representation and data bias, the influence of the institutional setting, and spatial relations in the exhibition conception. The working paper thus conclusively discusses the project *Training the Archive* and is a contribution to the practical application of software in the field of curating.

Veröffentlichte Working Papers

Das Format der Working Papers markiert unseren Arbeits- und Diskussionsstand innerhalb der Forschungsgruppe. Sieben Texte stellen neues Wissen sowohl für die Projektbeteiligten als auch für eine breite Öffentlichkeit zur Verfügung. Die Ergebnisse von *Training the Archive* wurden dadurch kontinuierlich im Prozess veröffentlicht, in den wissenschaftlichen Diskurs eingebracht und für interessierte Kunst- und Kultureinrichtungen zugänglich gemacht. Die Texte stehen über den QR-Code zum Download bereit.

Die Curator's Machine: Clustering von musealen Sammlungsdaten durch Annotieren verdeckter Beziehungsmuster zwischen Kunstwerken – Dominik Bönisch

Die Digitalisierung in Kunstmuseen verspricht einen erweiterten Zugriff auf Sammlungsobjekte sowohl für die Wissenschaft als auch für ein interessiertes Publikum, und das bestenfalls online – ortsunabhängig und jederzeit. Dabei reicht es nicht aus, die Suche in Datenbanken auf eng gedachte Stichworte zu limitieren. Vielmehr sollen über spezielle Interfaces und Visualisierungen eine Exploration von digitalen Beständen, sowie ein ‚Schlendern' durch die Online-Sammlung möglich gemacht werden. Durch maschinelles Lernen können Zusammenhänge und Verbindungen zwischen Kunstwerken offenbart werden, die sich Kurator*innen bislang schwer oder nur unvollständig erschließen konnten. Das Working Paper stellt im Sinne der Digital Humanities einen ersten Prototyp vor, von dem ausgehend das Forschungsprojekt *Training the Archive* ein maschinengestütztes, exploratives (Wieder-)Entdecken von Verknüpfungen innerhalb der eigenen musealen Sammlung untersuchen möchte.

„Why so many windows?" Wie die Bilddatensammlung ImageNet die automatisierte Bilderkennung historischer Bilder beeinflusst – Francis Hunger

Im Feld der Computer Vision hat die Bilddatensammlung ImageNet eine zentrale Rolle als Trainingsdatensatz inne. Der Text erörtert, in welchem Maße ImageNet den Software-Prototypen, die Curator's Machine, beeinflusst. Die Curator's Machine soll Zusammenhänge und Verbindungen zwischen Kunstwerken für Kurator*innen erschließen. Wie das in zeitgenössischen Bilderwelten verankterte ImageNet auf aktuelle und historischer Kunstwerke einwirkt, erläutert der Text, indem er 1.) die Abwesenheit der Klassifikation ‚Kunst' in ImageNet untersucht, 2.) die fehlende Historizität von ImageNet hinterfragt und 3.) das Verhältnis von Textur und Umriss in automatisierter Bilderkennung mit ImageNet diskutiert. Diese Untersuchung ist wichtig für die genealogische, kunsthistorische und programmiertechnische Verwendung der Datensammlung ImageNet in den Feldern des Kuratierens, der Kunstgeschichte, der Kunstwissenschaften und der Digital Humanities.

Kuratieren und dessen statistische Automatisierung mittels Künstlicher ‚Intelligenz' – Francis Hunger

Das in diesem Working Paper diskutierte Konzept des Post-AI-Curating untersucht das Kuratieren als wissensbildendes Verfahren, unterstützt durch Mustererkennung und gewichtete Netze, die als technische Mittel der Künstlichen ‚Intelligenz' zum Einsatz kommen. Dafür diskutiert der Text eine Reihe aufeinander aufbauende Konzepte wie Kuratieren, Kurator*in, das Kuratorische, kuratorische Forschung, Post-Human Curating und Post-AI Curating. Im Anschluss werden eine Reihe von Projekten, die sich dem Kuratieren mit Mitteln Künstlicher ‚Intelligenz' nähern, als Fallbeispiele diskutiert: *The Next Biennial Should Be Curated by a Machine* (2021) von UBERMORGEN, Leonardo Impett und Joasia Krysa als Meta-Kunstwerk über Kuratieren und Biennalen; Tillmann Ohms Projekt *Artificial Curator* (2020), welches in eine automatisiert kuratierte Ausstellung mündete; und *#Exstrange* (2017) von Rebekah Modrak und Marialaura Ghidini et al., welches auf der Plattform Ebay die Kunstwerke als Datenobjekte inszeniert. Der Text ist ein Beitrag zu den Curatorial Studies aus Perspektive des Post-AI-Curating.

Kann Künstliche Intelligenz Vorurteile haben? Zur Kritik des ‚algorithmic bias' von KI in den Künsten – Inke Arns

Das Working Paper widmet sich künstlerischen Positionen, die sich kritisch mit Künstlicher ‚Intelligenz' und automatisierter Mustererkennung durch Algorithmen auseinandersetzen. Anhand einer Reihe von Beispielen zeigt es die gesellschaftliche Problematik auf, die aus den Verzerrungen des Bias resultiert, und wie Künstler*innen darauf reagieren. Ausgehend von Analysen Harun Farockis und Hito Steyerls werden Projekte von Adam Harvey und Jules LaPlace, Zach Blas und Jemima Wyman, Elisa Giardina Papa, Francis Hunger und Flupke, Erika Scourti, Mimi Onuoha, Nora Al-Badri sowie Jan Nikolai Nelles dargestellt. Dabei werden verschiedene künstlerische Strategien aus dem Feld der Artistic Research sichtbar, um mit einem eigentlich unsichtbaren Gegenstand – Daten – umzugehen. Der Text diskutiert künstlerische Reaktionen auf gesellschaftliche Veränderungen, die unter dem Schlagwort KI zusammenzufassen sind.

Punktwolken. Scatterplots und Tabellen als User-Interfaces Künstlicher ‚Intelligenz' – Francis Hunger

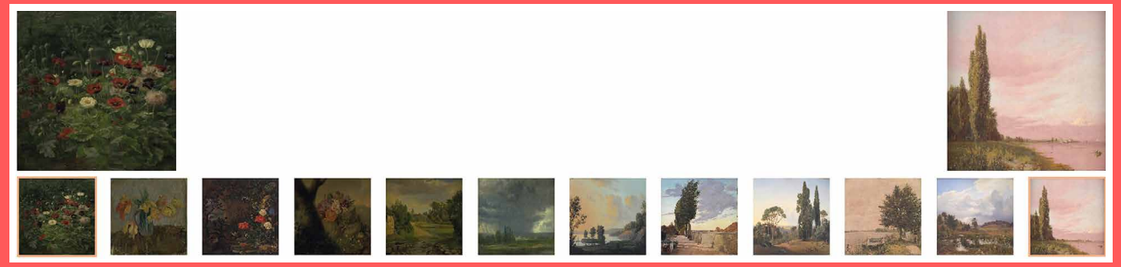
Scatterplots und tabellarische Strukturen sind die wichtigsten grafischen User-Interfaces für die diagrammatische Darstellung großer Bilddatensammlungen, die mit visuellen Erkennungswerkzeugen verarbeitet wurden. In dieser Studie werden verschiedene Medienvisualisierungen als Fallbeispiele diskutiert: *ARTigo* (LMU München, 2010–23), *imgs.ai* (UC Santa Barbara und Philipps-Universität Marburg, 2020–), *iArt* (LMU München und Leibniz Universität Hannover, 2018–23), *Vikus Viewer* (FH Potsdam, 2014–17), und *X Degrees of Separation* (Mario Klingemann, 2016). Darüber hinaus wird untersucht, wie diese Projekte die Visualisierungsalgorithmen PCA, t-SNE und UMAP im Verhältnis zu künstlichen gewichteten Netzen wie VGG19 oder CLIP einsetzen. Diese vergleichende Untersuchung ist ein Beitrag zu den Feldern der Software Studies, der Design- und User-Interface-Kritik in Bezug auf Big-Data-Anwendungen und die Digital Humanities.

Die Künstliche ‚Intelligenz' enthypen! Wie die täuschende Terminologie der KI ersetzt werden kann – Francis Hunger

KI als Forschungsgebiet und auch ihre Kritik werden durch Begriffe wie ‚Intelligenz', ‚Lernen' oder ‚neuronal' beherrscht. In diesem Beitrag wird erörtert, wie dieser anthropomorphisierende Sprachgebrauch den Hype um KI anheizt. KI-Hype macht viele Versprechungen, wie etwa, dass ‚KI kreativ sein kann' oder ‚KI den Welthunger löst'. Dieser Hype ist problematisch, weil er die negativen Auswirkungen von KI verdeckt. Der Autor schlägt vor, stattdessen eine andere Terminologie zu verwenden, z. B.: ‚Automatisierte Mustererkennung', ‚maschinelle Konditionierung' oder ‚gewichtetes Netzwerk'. Das Working Paper liefert pragmatische Vorschläge, die in den Feldern der KI-Kritik und der Informatik angesiedelt sind und aus medientheoretischer Perspektive auf den Genealogien Künstlicher ‚Intelligenz' aufsetzen.

‚From Keras Import Curating': Eine empirische Erhebung zur Übertragung von kuratorischer Praxis auf maschinelle Lernmodelle – Dominik Bönisch & Francis Hunger

Das Working Paper wertet elf Interviews mit professionellen Kurator*innen aus, die für das Projekt *Training the Archive* geführt wurden. Aus den Interviews wurde eine Reihe von Personas abgeleitet, welche Aufschluss über die benötigte Funktionalität der Curator's Machine gaben und zu entsprechenden Änderungen der Prototypen führten. In der Auswertung beleuchtet der Text zum einen kuratorische Ansätze, die mit der Curator's Machine realisiert werden können, darunter Un-/Ordnungssysteme und Datenbanken, den Prozess des Auswählens und des Denkens in An-Ordnungen oder Kategorien sowie eine Reflexion zur verarbeiteten Datengrundlage. Zum anderen wurden kuratorische Ansätze diskutiert, die nicht oder nur eingeschränkt realisierbar sind, z. B. Fragen der Repräsentation und des Datenbias, der Einfluss des institutionellen Rahmens und räumliche Verhältnisse in der Ausstellungskonzeption. Das Working Paper reflektiert damit abschließend das Projekt *Training the Archive* und ist ein Beitrag zur praktischen Anwendung von Software im Feld des Kuratierens.




'geistiger,' 'visueller' und
'organisatorischer' Arbeit beschreiben?
und der MS-CEEB Datensatz
einige der ersten Datensätze,
SENNHEISER

Lab-like Exhibition

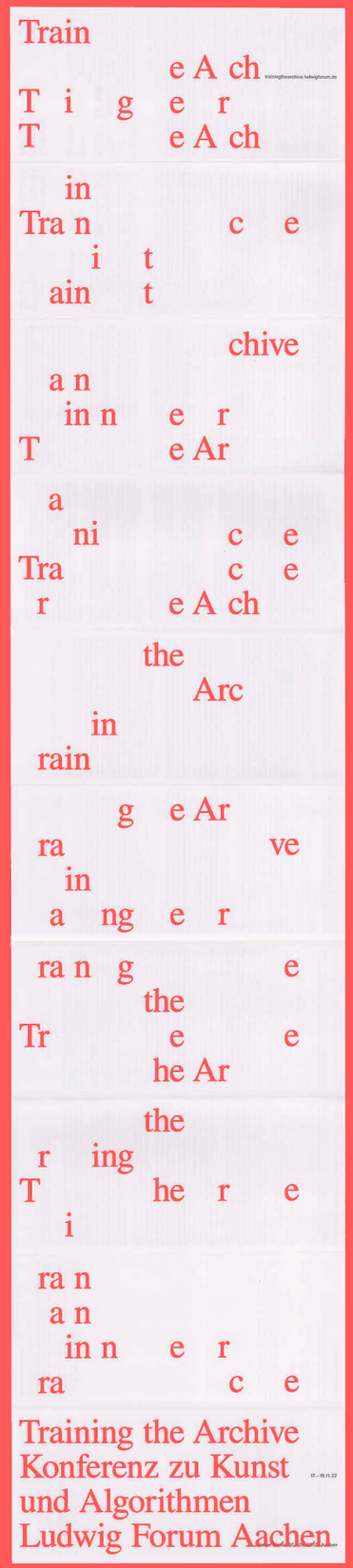
In a lab-like exhibition at the Ludwig Forum Aachen (2022–23), *Training the Archive* presented its findings and provided insights into the discourse of the research team. In addition to a selection of literature for self-study, the published working papers were available to the interested public. The interviews with international artists, curators and theorists—which are included in this book as summarized transcripts—were also shown in the installation. The presentation was understood as a 'work in progress,' and may also display new results in the future or put relevant artistic positions into context.

For more information scan the QR code or visit: trainingthearchive.ludwigforum.de/en.

Labor-Ausstellung

In einer laborhaften Ausstellung am Ludwig Forum Aachen (2022–23) präsentierte *Training the Archive* seine Ergebnisse und gab Einblicke in die fachliche Auseinandersetzung des Forschungsteams. Neben einem Handapparat mit einer Auswahl an Literatur zum Selbststudium standen ebenfalls die veröffentlichten Working Papers dem interessierten Publikum zur Verfügung. In der Installation wurden auch die Interviews mit internationalen Künstler*innen, Kurator*innen und Theoretiker*innen gezeigt, die in diesem Buch als zusammengefasste Transkripte enthalten sind. Die Präsentation verstand sich als ‚Work in Progress' und kann dadurch in Zukunft neue Ergebnisse bündeln, sich verändern oder passende künstlerische Positionen in Zusammenhang bringen.

Für mehr Informationen den QR-Code scannen oder auf: trainingthearchive.ludwigforum.de.



On November 17 and 18, 2022, a conference on the topic of ‘Art & Algorithms’ took place in the scope of the research project *Training the Archive*. At the conference, renowned academics provided an overview of the current state of research on the application of algorithms in the context of art and museums. The first day introduced the history and present state of algorithms and culture in the form of a prologue on the topic ‘Reflect.’ On the second day, three panels discussed the question of how machine learning could change access to museum collections. The first panel addressed ‘Select’ as a gesture in the space of possibilities of current AI models. The following panel, ‘Retrieve,’ focused on multimodal approaches to museum collections and iconography. In the last panel, ‘Combine,’ the topic of curating as a link between man and machine was explored. Mads Pankow hosted the conference, which was free of charge and recorded. To access the contributions in the respective language of the presentation, use the corresponding QR code.

Prologue ‘Reflect’ on 2022–11–17

Digital Culture. Fundamentals and Ambivalences of Algorithmization—Sybille Krämer

We associate the algorithmic with computers. But is this true? In fact, there is an embryonic digitality before electronic networking. Is the transition from computational rules to recommendation algorithms an inversion of empowerment into uncontrollability? The keynote seeks answers to the question of what a ‘cultural technique of digital literacy’ consists of.

Culture between Humans and Machines: Who Curates Whom?—Inke Arns, Sybille Krämer & Yvonne Zindel

How do algorithms influence artistic and curatorial work? What does the data reveal about us and our biases? And who programs whom in the end, the human the algorithm or vice versa? A panel discussion.

Future Actions—Mattis Kuhn

Mattis Kuhn reads excerpts from his books *Selbstgespräche mit einer KI* (2020–21) and *Grasslands for Insects* (2022), the latter in combination with generated images. In the production process, different approaches to datasets and authorship were used, which will be presented comparatively after the lecture performance.

Panel ‘Select’ on 2022–11–18

On the Concept of History (in Foundation Models)—Fabian Offert

This talk discusses the concept of history inherent in so-called ‘foundation models,’ focusing on OpenAI’s CLIP model, a large-scale multimodal model that projects text and image data into a common embedding space. How does the idiosyncratic historical perspective that emerges from such models impact their potential as art historical tools?

Images on Command. Text-Image Relations in DALL·E 2—Roland Meyer

Text-to-image generators like DALL·E 2 promise to produce any possible image on command. The virtual archive of textually annotated and digitally mobilized images of the past serves them as a seemingly inexhaustible data resource—with consequences for our concept of images that are just beginning to become apparent.

Curatorial Algorithms and Collection Spaces—Tillmann Ohm

Cultural heritage such as art collections can be quantified and embedded in shared computational spaces including vector embeddings and networks. Interfaces and algorithms help researchers and curators to explore these possibility spaces. This talk provides an overview of such techniques and introduces examples of algorithmic curation.

Multimodal Models as Cultural Snapshots—Eva Cetinić

Trained on millions of image-text pairs sampled from the Internet at a certain point in time, multimodal models learn not only literal similarities between images and words, but also various cultural references. As synchronic cultural snapshots embedded in a specific technological framework, they become not just tools but also new objects of study.

How to Train the Curator’s Machine—Dominik Bönisch

The Curator’s Machine is a prototype that enables curators to approach museum collections intuitively and free of fixed search terms via a multimodal retrieval system. An explorative process of human-machine interaction also includes the experts’ contextual knowledge in the machine learning.

Pose Estimation for Artworks Retrieval—Vincent Christlein

Human pose detection represents an important factor in scene analysis of artworks. Additionally, ‘pose estimation’ can be used for an interpretable image retrieval. This talk will discuss pose estimation and its role in image retrieval in various fields of application such as classical archaeology and art history.

Panel ‘Combine’ on 2022–11–18

Human-Algorithmic Curation: Between the Visible and Invisible—Gaia Tedone

The presentation discusses human-algorithmic curation as a networked practice that is both reflexive and playful and that can engender new forms of cooperation and value co-creation between humans and technical agents. It speculates on its potential to expose the invisible power structures and technical asymmetries of the networks.

Old Canon, New Bias—Katrin Glinka

In her lecture, Katrin Glinka emphasizes the necessity of designing algorithmic systems in such a way that they support a critically reflective use. She contrasts technology-inherent potentials of algorithmic procedures with regard to accessibility and representation of objects with established structures of museum knowledge representation.

Towards Human-Machine Visual Literacy—Geoff Cox

Referring to *Ways of Seeing* (John Berger, 1972), the presentation argues for an expansion of visual literacy to examine how machine vision further unsettles received humanist notions. When images are made by machines for other machines, and part of vast annotated datasets, how are worldviews reinforced differently, and what kind of literacy applies, if at all?

Konferenz zum Thema ‚Kunst & Algorithmen‘

Am 17. und 18. November 2022 fand im Rahmen des Forschungsprojekts *Training the Archive* eine Konferenz zum Thema ‚Kunst & Algorithmen‘ statt. Auf der Konferenz gaben namhafte Wissenschaftler*innen einen Überblick zum aktuellen Forschungsstand der Anwendung von Algorithmen im Kunst- und Museumskontext. Am ersten Tag wurde in Form eines Prologs zum Thema ‚Reflektieren‘ in die Geschichte und Gegenwart von Algorithmen und Kultur eingeführt. Der zweite Tag verhandelte in drei Panels die Frage, wie das maschinelle Lernen den Zugang zu musealen Sammlungen verändern kann. Das erste Panel vermittelte hierbei das ‚Auswählen‘ als Geste im Möglichkeitsraum aktueller KI-Modelle. Das folgende Panel ‚Erschließen‘ fokussierte auf multimodale Zugänge zu musealen Sammlungen sowie zur Ikonografie. Im letzten Panel ‚Verbinden‘ wurde das Kuratieren als Bindeglied zwischen Mensch und Maschine thematisiert. Die kostenfreie Konferenz wurde von Mads Pankow moderiert und konnte aufgezeichnet werden. Die Beiträge sind über den QR-Code in der jeweiligen Vortragssprache abrufbar.

Prolog ‚Reflektieren‘ am 17–11–2022

Digitale Kultur. Grundlagen und Ambivalenzen der Algorithmisierung – Sybille Krämer
Wir verbinden das Algorithmische mit Computern. Doch trifft das zu? Tatsächlich gibt es eine embryonale Digitalität vor der elektronischen Vernetzung. Ist der Übergang von Rechenregeln zu Empfehlungsalgorithmen ein Umschlag von Ermächtigung in Unkontrollierbarkeit? Die Keynote sucht Antworten auf die Frage, worin eine ‚Kulturtechnik digitaler Literalität‘ besteht.

Kultur zwischen Mensch und Maschine: Wer kuratiert wen? – Inke Arns, Sybille Krämer & Yvonne Zindel
Wie beeinflussen die Algorithmen die künstlerische und kuratorische Arbeit? Was verraten die Daten über uns und unsere Vorurteile? Und wer programmiert am Ende wen, der Mensch den Algorithmus oder umgekehrt? Ein Podiumsgespräch.

Future Actions – Mattis Kuhn
Mattis Kuhn liest Auszüge aus seinen Büchern *Selbstgespräche mit einer KI* (2020–21) und *Grasslands for Insects* (2022). Letzteres in Kombination mit generierten Bildern. Im Produktionsprozess wurden unterschiedliche Umgänge mit Datensets und Autor*innenschaft angewendet, welche im Anschluss an die Lesung vergleichend vorgestellt werden.

Panel ‚Auswählen‘ am 18–11–2022

Über den Begriff von Geschichte (in Foundation Models) – Fabian Offert
Haben Modelle aus dem Bereich des maschinellen Lernens einen Begriff von Geschichte? Der Vortrag beleuchtet diese Frage unter besonderer Berücksichtigung von CLIP, einem von OpenAI publizierten, multimodalen ‚Foundation Model‘, das Text- und Bilddaten in Zusammenhang bringt und für die kunsthistorische Arbeit von besonderer Bedeutung ist.

Bilder auf Kommando. Text-Bild-Verhältnisse bei DALL·E 2 – Roland Meyer
Text-to-Image-Generatoren wie DALL·E 2 versprechen die Produktion beliebiger Bilder auf Kommando. Das virtuelle Archiv der sprachlich erschlossenen und digital mobilisierten Bilder der Vergangenheit dient ihnen dabei als scheinbar unerschöpfliche Datenressource – mit Konsequenzen für unseren Bildbegriff, die erst langsam erahnbar werden.

Kuratorische Algorithmen und Sammlungsräume – Tillmann Ohm
Kulturelles Erbe, z. B. Kunstsammlungen, können quantifiziert und in gemeinsame mathematische Räume wie Vektor-Embeddings und Netzwerke übertragen werden. Interfaces und Algorithmen helfen Forscher*innen und Kurator*innen bei der Erkundung dieser Möglichkeitsräume. Dieser Vortrag gibt einen Überblick über solche Techniken und stellt Beispiele für algorithmische Kuration vor.

Panel ‚Erschließen‘ am 18–11–2022

Multimodale Modelle als kulturelle Momentaufnahmen – Eva Cetinić
Anhand millionenfacher Bild-Text-Paare, die zu einem bestimmten Zeitpunkt im Internet gesammelt wurden, erlernen multimodale Modelle nicht nur die Parallelen zwischen Bildern und Wörtern, sondern auch kulturelle Referenzen. Als synchrone kulturelle Momentaufnahmen, eingebettet in einen spezifischen technologischen Rahmen, werden sie von reinen Werkzeugen zu neuen Studienobjekten.

Wie man die Curator’s Machine trainiert – Dominik Bönisch
Die ‚Curator’s Machine‘ ist ein Prototyp, der es Kurator*innen über ein multimodales Vorschlagssystem ermöglicht, sich musealen Sammlungen intuitiv und frei von festen Suchbegriffen anzunähern. Durch einen explorativen Prozess der Mensch-Maschine-Interaktion wird dabei auch das Kontextwissen der Expert*innen in das maschinelle Lernen einbezogen.

Pose Estimation für die Erschließung von Kunstwerken – Vincent Christlein
Die Erkennung der menschlichen Pose oder Körperhaltung ist ein wichtiger Faktor bei der Analyse von Szenen in Kunstwerken. Darüber hinaus kann die ‚Pose Estimation‘ für eine interpretierbare Bilderkennung verwendet werden. In diesem Vortrag wird die Pose Estimation und ihre Rolle bei der Bilderkennung in verschiedenen Anwendungsbereichen wie der klassischen Archäologie und der Kunstgeschichte diskutiert.

Panel ‚Verbinden‘ am 18–11–2022

Menschlich-algorithmisches Kuratieren: Zwischen Sichtbarem und Unsichtbarem – Gaia Tedone
Der Vortrag erörtert das menschlich-algorithmische Kuratieren als eine vernetzte Praxis, die sowohl reflexiv als auch spielerisch ist und neue Formen der Zusammenarbeit sowie der gemeinsamen Wertschöpfung zwischen Menschen und Software-Agent*innen hervorbringen kann. Dabei wird das Potenzial diskutiert, die unsichtbaren Machtstrukturen und technischen Asymmetrien der Netzwerke aufzudecken.

Alter Kanon, neuer Bias – Katrin Glinka
In ihrem Vortrag betont Katrin Glinka die Notwendigkeit, algorithmische Systeme so zu gestalten, dass sie eine kritisch-reflektierte Nutzung unterstützen. Sie kontrastiert technikinhärente Potenziale algorithmischer Verfahren hinsichtlich Zugänglichkeit und Repräsentation von Objekten mit etablierten Strukturen musealer Wissensrepräsentation.

Hin zu einer visuellen Mensch-Maschine-Literalität – Geoff Cox
Unter Bezugnahme auf *Sehen: Das Bild der Welt in der Bilderwelt* (John Berger, dt. Ed. 1974) plädiert der Vortrag für eine Ausweitung der visuellen Literalität, um zu verdeutlichen, wie das maschinelle Sehen die gängigen humanistischen Vorstellungen ins Wanken bringt. Wenn Bilder von Maschinen für andere Maschinen gemacht werden und Teil riesiger kommentierter Datensätze sind, wie werden dann unterschiedliche Weltanschauungen verstärkt und welche Art von Lesekompetenz kommt zum Tragen – wenn überhaupt?

In Conversation with Matteo Pasquinelli about “A Larger Perspective on Artificial Intelligence”

An overview talk about the ideological form, the logical form, the technical form, and the social form of artificial intelligence. Pasquinelli describes AI as automation of manual, mental, visual, and organizational labor.

Transcript of the interview with Matteo Pasquinelli, conducted by Francis Hunger on 2022–01–21, and edited by Pasquinelli.

Hunger: Welcome Matteo! With my questions I'd like to start with the big picture and then go into details with you. So, to begin with, what's the state of so-called artificial intelligence studies today? What are the main historical paradigms for its interpretation?

Pasquinelli: We can summarize the large spectrum of perspectives about AI under three theses. I call them the 'mechanist,' 'modelist,' and 'workerist' theses. The mechanist thesis is the one championed by Alan Turing, for instance, in his famous paper *Computing Machinery and Intelligence* from 1950. According to this thesis, a machine can perfectly imitate human intelligence. This has been called the mechanist thesis for a few decades (e.g., see Stuart Shanker's book *Wittgenstein's Remarks on the Foundation of AI,* 1998). The mechanist thesis is a central postulate of the tradition of 'symbolic AI,' of the GOFAI ('Good Old Fashioned Artificial Intelligence'), of Turing, as we just said, but also of Warren McCulloch and Walter Pitts, who introduced the idea of the artificial neuron in 1943. However, another completely different approach emerged in the same years: the one of 'connectionism' and artificial neural networks of the statistical kind, that would lay the foundation of machine learning later on. In this tradition of research, engineers and mathematicians were not concerned with encoding human reason as a formalized logic into a machine; instead, their research was about building systems that would be able to imitate the brain's capacity to model the world. I call this approach the modelist thesis of AI. It is based on the idea that human intelligence is not about representing the world through procedural knowledge, but about constructing models of the world. What is a model? A model is not an exact representation of the world, but an approximation, a reduced and incomplete yet still operative form of knowledge. This intuition is found at the core of statistics, connectionism, and then machine learning. Friedrich Hayek's book *The Sensory Order* from 1952 can be considered its philosophical tractate. Another perspective from which we could see the development of AI is the perspective of social praxis, in particular of labor automation and the control of productivity and social relations. I call this perspective the workerist thesis of AI. It is the idea that AI was developed not to fulfill the dream to imitate human intelligence, but rather for the industrial and military agenda to automate labor. This is a thesis that has emerged thanks to the work of decolonial, feminist, and labor studies. You can consider here Kalindi Vora and Neda Atanasoski's book *Surrogate Humanity* (2019), and also my forthcoming *The Eye of the Master: A Social History of Artificial Intelligence* (2023).

Hunger: You proposed to analyze the complex artifact of artificial 'intelligence' in four levels. So what did you mean by separating the 'ideological form,' the 'logical form,' the 'technical form,' and the 'social form' of AI?

Pasquinelli: AI is a complex artifact, and we have to develop a consistent methodology to map such complexity. By distinguishing these levels, I mean that not just AI, but any artifact in the history of science and technology has been developed proceeding from different force fields: economic and social needs, technological drivers, scientific research, and also ideological imaginaries (here you may consider the different mythologies about automata). In studying AI, I think we have to keep all these levels together. I would say that in critical AI studies, there is the tendency (understandably) to focus on one specific factor at a time. For instance, in the history of automation and labor, of course, there is a focus on the dimension of praxis, on the role of the division of labor and economic dynamics. In the tradition of German media theory, to bring another example, there is a tendency to focus only on a sort of hardware a priori (if not military a priori, as in the case of Friedrich Kittler). Other studies based on discourse analysis have only been mapping the circulation of discourse and narratives around AI, and so on. We should take all these different perspectives and research experiences and integrate them, basically to place AI in the larger history of modern technology and science.

Hunger: One of the premises of AI is the promise to automate labor away and to free humans from tedious and boring tasks. Could you describe the history and different layers of automation of manual, mental, visual, and organizational labor?

Pasquinelli: Let's have a look at the different forms of labor automation that drove the making of AI. Scholars such as Lorraine Daston and Simon Schaffer have already explained very well how modern computation and the idea of machine intelligence started with Charles Babbage and his project to automate mental labor, taking the idea of the division of labor from Adam Smith and applying it to the organization and automation of mental labor. Babbage, specifically, aimed at automating the making of logarithmic tables in the industrial age, and this should be remembered as one of the crucial genealogies of computation, extending until Turing and the 20th century project of AI. This lineage interprets mental labor as hand calculation and automates it as a form of symbol manipulation. This intuition was later translated into the project of symbolic AI. That is, symbolic AI proceeds from the very specific project to automate mental labor as hand calculation. On the other hand, the origin of today's deep learning and its artificial neural networks are found in the experiments of Frank Rosenblatt back in the 1950s, sponsored by the US Army and Navy, to automate something different: that is, the labor of vision or supervision, the labor of recognition that thereafter was termed 'pattern recognition.' This is a different kind of labor than mental labor understood as hand calculation. It's not something we can describe as manipulation of symbols or manipulation of numerical signs. It was the capacity to recognize an image. Specifically, in the case of Rosenblatt's experiment for the US Navy, it was the capacity to recognize an enemy vessel or a tank in the photographs of an aerial reconnaissance. Originally designed for visual pattern recognition, one day these systems were then applied to other forms of pattern recognition in non-visual datasets: the recognition of audio patterns, or the recognition of textual patterns, for instance. This happened basically with the development of deep learning, and today it has become a multimodal form of pattern recognition with the so-called 'Large Language Models' (LLMs). There is another important aspect in the role that the division of labor plays in AI. It's not about the automation of labor, but the automation of management. The gig economy platforms used by companies like Uber, Wolt, Foodora, and Deliveroo represent a phenomenon which is not completely new, but typical of the history of automation. They didn't come to replace workers, but to replace managers and therefore to multiply workers. They engendered new figures of the worker such as temporary drivers, riders, online workers (the so-called 'Turkers'), care workers, and so on. The algorithms of these platforms are used to organize the workforce in a very dynamic, flexible, and adaptive way, also using new systems to track and quantify social behaviors and physical movements across the urban space. For this new category of workers, the algorithm is the boss. So, regarding AI, we should discuss the automation of at least four kinds of labors: manual, mental, visual, and the supervision of labor itself, that is, management.

Hunger: I also wanted to ask about the major milestones of artificial weighted networks as experimental parametrized machines, but I'm not even certain if we need to go into that anymore, as we've already been talking about it.

Pasquinelli: Today's AI takes the form of machine learning models (which include artificial neural networks). Technically speaking, today we don't use anymore the term 'AI' but 'models,' as indeed these artifacts are based on a specific technique that is algorithmic modelling. In other terms, they can also be called 'parametrized machines,' as they are computing networks made of a very high number of parameters. The first parametrized machine (by convention) was the artificial neural network—Perceptron—that Frank Rosenblatt experimented with back in 1956. When Frank Rosenblatt tried to defend the idea of the artificial neural network, he considered it an experimental machine. In this way, he intended to defend the experimental method (even making use of Galileo in his writings) against the tradition of symbolic AI that was very allergic to the idea of experiment. Symbolic AI researchers, in fact, thought that they could incarnate human intelligence into a machine straight away without going through an experimental phase. Rosenblatt, on the other hand, had the idea that these machines, these artificial neural networks with different topologies were experiments, meaning that they could not predict what they were capable of, and they were like experimental settings that had to be studied and tested. Somehow, these machines imitated traditional science experiments to study the laws of nature. As in some of the hypotheses of scientific research, they were made of variables, of parameters, to be found. The difference between these machines and rather traditional experiments of science is that Rosenblatt's artificial neural network had, mathematically speaking, a higher number of parameters of the same kind. The Perceptron was basically a machine made of 400 parameters, which corresponded to the 400 pixels of its visual matrix. The optimal value of these 400 parameters or 'weights' had to be found through a training procedure. This number gives you the scale of the experiments that were possible in the 1950s and 1960s in terms of computing power. In the following decades, the parameters of these machines or neural networks would escalate. A second milestone in the history of deep learning can be identified in the convolutional neural network invented by Yann LeCun to automate the recognition of numbers. His artificial neural network was called LeNet, and it featured a few thousand parameters in 1998. A third milestone is the AlexNet deep artificial neural designed by Hinton and his students, which won the ImageNet competition in 2012 and ignited the deep learning revolution. AlexNet counted 60 million parameters. Indeed, 'parametrized machine' is the term through which we should define the epistemic form of machine learning today.

Scan the QR code below to watch the full video interview online.

Im Gespräch mit
Matteo Pasquinelli über „Künstliche Intelligenz in breiterer Perspektive“

Ein Überblicksgespräch über die ideologische Form, die logische Form, die technische Form und die soziale Form der Künstlichen Intelligenz. Pasquinelli beschreibt im Interview KI als Automatisierung von manueller, geistiger, visueller und organisatorischer Arbeit.

Transkript des Interviews mit Matteo Pasquinelli, geführt von Francis Hunger am 21–01–2022 und überarbeitet durch Pasquinelli.

Hunger: Ich möchte zunächst mit meinen Fragen das große Ganze thematisieren und dann mit dir ins Detail gehen. Was ist der heutige Stand der KI-Forschung? Welches sind die wichtigsten historischen Paradigmen für die Interpretation von sogenannter Künstlicher ‚Intelligenz'?

Pasquinelli: Wir können das große Spektrum an Perspektiven auf die Künstliche Intelligenz auf drei Thesen reduzieren. Ich nenne sie die ‚Maschinen'-, die ‚Modell'- und die ‚Arbeiter*innen'-Thesen. Die Maschinenthese ist vor allem von Alan Turing in seinem berühmten Aufsatz *Computing Machinery and Intelligence* (1950) vertreten worden. Sie besagt, dass eine Maschine menschliche Intelligenz perfekt nachahmen kann. Dies wird seit einigen Jahrzehnten als Maschinenthese bezeichnet (siehe u. a. Stuart Shankers Buch *Wittgenstein's Remarks on the Foundation of AI,* 1998). Sie ist zentral für die Tradition der ‚symbolischen KI' (engl.: GOFAI – ‚Good Old Fashioned Artificial Intelligence') von Turing, aber auch von Warren McCulloch und Walter Pitts, welche 1943 auch die Idee künstlicher Neuronen einführten. Aber wir sollten zur Kenntnis nehmen, dass sich in denselben Jahren ein grundlegend anderer Ansatz entwickelte: Es war der Ansatz der ‚Konnektionist*innen' und der künstlichen neuronalen Netzwerke, die auf statistischen Verfahren aufbauten und später zur Grundlage des maschinellen Lernens werden sollten. In diesem Forschungsstrang waren Ingenieur*innen und Mathematiker*innen nicht daran interessiert, die menschliche Vernunft als mathematische Logik in die Maschine zu encodieren, sondern Systeme zu konstruieren, welche die Fähigkeit des Gehirns, die Welt zu modellieren, nachbilden könnten. Ich bezeichne dies als die Modellthese der KI. Sie beruht auf der Idee, dass es bei der menschlichen Intelligenz nicht darum geht, die Welt durch nüchternes Wissen zu repräsentieren, sondern Modelle der Welt zu konstruieren. Was ist ein Modell? Bei einem Modell handelt es sich nicht um eine exakte Repräsentation der Welt, sondern um eine Annäherung, eine reduzierte und unvollständige, jedoch ausreichend operative Form von Wissen. Diese Einsicht ist im Kern der Statistik, des Konnektionismus und schließlich des maschinellen Lernens zu finden. Als philosophische Grundlage kann man Friedrich Hayeks Buch *The Sensory Order* von 1952 einordnen. Eine andere Perspektive, aus der wir die Entwicklung Künstlicher Intelligenz beschreiben könnten, ist die Perspektive der gesellschaftlichen Praxis, spezifisch der Automatisierung von Arbeit und der Kontrolle von Produktivität und sozialen Beziehungen. Ich bezeichne diese Perspektive als die Arbeiter*innenthese. Es handelt sich um die Idee, dass KI nicht entwickelt wurde, um den Traum zu verwirklichen, menschliche Intelligenz nachzubilden, sondern dass man im Grunde eine industrielle und militärische Agenda befolgte, Arbeit zu automatisieren. Diese These ist dank der Dekolonisierungsforschung, der Feministischen Forschung und natürlich auch der Forschungen zu Arbeit entstanden. Hierfür kommen Kalindi Vora und Neda Atanasoskis Buch *Surrogate Humanity* (2019) und auch mein bald erscheinendes *The Eye of the Master: A Social History of Artificial Intelligence* (2023) in Betracht.

Hunger: Du hast kürzlich vorgeschlagen, das komplexe Artefakt der KI auf vier Ebenen zu analysieren. Was meintest du mit der Unterscheidung in die ‚ideologische Form', die ‚logische Form,' die ‚technische Form' und die ‚soziale Form' der Künstlichen ‚Intelligenz'?

Pasquinelli: KI ist ein komplexes Artefakt, und wir müssen eine konsistente Methodologie entwickeln, um diese Komplexität beschreiben zu können. Indem ich diese Ebenen unterscheide, arbeite ich heraus, dass sich nicht nur KI, sondern jedes Artefakt in der Geschichte der Wissenschaft und Technologie aufgrund verschiedener Druckkräfte entwickelt hat: ökonomische und soziale Bedingungen, technologische Triebkräfte, wissenschaftliche Forschung und auch ideologische Vorstellungen (zu denen wir auch verschiedenste Mythologien über Automaten zählen können). Und ich denke, wenn es um eine Auseinandersetzung mit KI geht, müssen wir all diese Ebenen zusammen betrachten. Ich würde sagen, dass es in der kritischen KI-Forschung (verständlicherweise) eine Tendenz gibt, sich auf einen einzelnen spezifischen Faktor zu konzentrieren. Wir können beispielsweise die Sozialgeschichte der Automatisierung und der Arbeit nennen, die sich natürlich auf die Dimension der Praxis, die Arbeitsteilung und die ökonomischen Dynamiken konzentriert. Wir können auch die Tradition der deutschen Medientheorie erwähnen, die sich auf die Apriori von Hardware konzentriert (wenn nicht sogar auf militärische Apriori wie im Fall von Friedrich Kittler). Oder wir können die Diskursanalyse heranziehen, welche allein die Zirkulation von Diskursen und Narrativen um KI oder andere Technologien kartiert hat, und so weiter. Wir sollten all diese unterschiedlichen Perspektiven und wissenschaftlichen Erfahrungen einbeziehen und miteinander integrieren, um KI in der größeren Geschichte der modernen Technologie und Wissenschaft einzuordnen.

Hunger: Eine der Prämissen der KI ist das Versprechen, Arbeit zu automatisieren und den Menschen von langweiligen und ermüdenden Aufgaben zu befreien. Könntest du die Geschichte und die verschiedenen Ebenen der Automatisierung von manueller, geistiger, visueller und organisatorischer Arbeit beschreiben?

Pasquinelli: Wir können die verschiedenen Formen der Automatisierung von Arbeit betrachten, die die Entwicklung der KI vorangetrieben haben. Wissenschaftler*innen wie Lorraine Daston und Simon Schaffer haben bereits sehr gut erklärt, wie die moderne Computertechnik und sogar die Idee der maschinellen Intelligenz mit Charles Babbage und seiner Idee begann, geistige Arbeit zu automatisieren, indem er von Adam Smith die Idee, wie Arbeitsteilung zu organisieren ist, übernahm und auf die Organisation und Automatisierung geistiger Arbeit anwandte. Babbage wollte im Industriezeitalter insbesondere die Erstellung von logarithmischen Tabellen automatisieren, und das sollte als eine der entscheidenden Genealogien des Rechnens in Erinnerung bleiben. Diese Genealogie reicht bis Turing und zur Künstlichen Intelligenz als Projekt des 20. Jahrhunderts. In dieser Abstammungslinie wird Denkarbeit als Rechnen mit der Hand verortet und später deren Automatisierung mittels automatisierter Symbolverarbeitung. Diese Einsicht wurde später in das Projekt der symbolischen Künstlichen Intelligenz übersetzt. Die symbolische KI folgte also aus dem spezifischen Projekt, die Denkarbeit mittels Berechnung per Hand zu automatisieren. Andererseits entstammen das Deep Learning und die künstlichen neuronalen Netze dem Experiment von Frank Rosenblatt in den 1950er Jahren, das von der US-Armee und der US-Marine gefördert wurde, um etwas anderes zu automatisieren: und zwar die Arbeit der Überwachung, die Arbeit der Erkennung oder das, was später als ‚Mustererkennung' bezeichnet wurde. Das ist eine andere Art von Arbeit im Vergleich zur Geistesarbeit im Sinne von manuellen mathematischen Berechnungen. Es ist nicht etwas, das wir als Symbolverarbeitung oder als Verarbeitung von Zahlenfolgen beschreiben können. Es war die Fähigkeit, ein Bild wiederzuerkennen. Genauer: In dem Experiment von Rosenblatt für die US Navy war es die Fähigkeit, ein feindliches Schiff oder einen Panzer auf Fotografien der Luftaufklärung zu erkennen. Später wurden diese ursprünglich für die visuelle Mustererkennung konzipierten Systeme auf andere Formen der Mustererkennung angewandt: Die Erkennung von Audiomustern, die Erkennung von Textmustern zum Beispiel. Dies fand im Grunde mit der Entwicklung von Deep Learning statt und hat sich heute zu einer multi-modalen Form als Text-Bild-Korrelation der Mustererkennung mit den sogenannten ‚Large Language Models' (dt.: Große Sprachmodelle) weiterentwickelt. Es gibt da noch einen anderen wichtigen Aspekt bezüglich der Rolle der Arbeitsteilung für KI. Es geht dabei nicht um die Automatisierung der Arbeit, sondern um die Automatisierung des Managements. Die sogenannten Gig-Economy-Plattformen, die von Unternehmen wie Uber, Wolt, Deliveroo oder Foodora verwendet werden, repräsentieren ein Phänomen, welches nicht völlig neu, sondern für die Geschichte der Automatisierung typisch ist. Sie sind nicht darauf ausgerichtet, die Arbeiter*innen zu ersetzen, sondern die Manager*innen, um somit die Arbeiter*innen zu multiplizieren. Sie ermöglichten neue Figuren von Arbeiter*innen, wie temporäre Fahrer*innen, Kurier*innen, Clickworker*innen von Amazon Mechanical Turk, Care-Arbeiter*innen und so weiter. Die Algorithmen dieser Plattformen werden eingesetzt, um diese Arbeitskräfte dynamisch, flexibel und auf individuelle Weise zu organisieren, indem sie mittels ebenfalls neuer Systeme deren soziales Verhalten und körperliche Bewegungen im gesamten urbanen Raum tracken und so quantifizieren. Für diese neue Kategorie von Arbeiter*innen ist der Algorithmus der ‚Boss'. Wir sollten also in Bezug auf KI mindestens vier Arten von Automatisierung von Arbeit diskutieren: manuelle, mentale, visuelle und die Überwachung der Arbeit selbst, also deren Management.

Hunger: Ich habe hier noch die Frage nach den wichtigsten Meilensteinen der künstlichen, gewichteten Netzwerke als experimentelle parametrisierte Maschinen offen, aber ich bin mir unsicher, ob wir darauf noch eingehen müssen, denn wir haben das bereits angesprochen.

Pasquinelli: Die heutige KI tritt in Form von Machine-Learning-Modellen (inkl. künstlicher neuronaler Netze) auf. Technisch gesprochen benutzen wir inzwischen nicht mehr den Begriff ‚KI', sondern ‚Modelle', da dies tatsächlich Artefakte sind, die auf einer spezifischen Technik von algorithmischer Modellierung basieren. Sie können auch genauso als parametrisierte Maschinen beschrieben werden, da es sich um prozessierende Netzwerke mit einer hohen Anzahl von Parametern handelt. Die erste parametrisierte Maschine war das künstliche neuronale Netzwerk Perzeptron, mit dem Frank Rosenblatt 1956 experimentierte. Zur Verteidigung seiner Idee eines künstlichen neuronalen Netzes – des Perzeptrons – brachte Rosenblatt hervor, dass es sich um eine experimentelle Maschine handelte. Damit wollte er seine experimentelle Methode gegen die Tradition der symbolischen KI verteidigen (wobei er in seinen Schriften sogar Galileo mobilisierte), die sehr allergisch auf die Idee des Experiments reagierte. Die Forscher*innen zu symbolischer KI dachten, sie könnten die menschliche Intelligenz direkt in einer Maschine verkörpern, ohne die experimentelle Phase zu durchlaufen. Rosenblatt hingegen ging davon aus, dass diese Maschinen, diese verschiedenen neuronalen Netze mit unterschiedlichen Topologien Experimente waren, was bedeutet, dass man die genauen Fähigkeiten nicht vorhersagen konnte, und sie sich wie experimentelle Settings verhielten, die untersucht und getestet werden mussten. Auf eine Weise imitierten diese Maschinen die traditionellen wissenschaftlichen Experimente, welche dem Studium der Naturgesetze dienten. Ähnlich wie in einigen Hypothesen der wissenschaftlichen Forschung bestanden sie aus Variablen, aus Parametern, die es zu finden gilt. Der Unterschied zwischen diesen Maschinen und den traditionellen Experimenten der Wissenschaft besteht also darin, dass sie deutlich mehr Parameter der gleichen Art hatten. Und das Perzeptron war im Grunde eine Maschine, die aus 400 Parametern bestand, welche mit 400 Pixeln in der visuellen Matrix korrespondierten. Der optimale Wert für diese 400 Parameter oder ‚Gewichtungen' musste im Rahmen einer Trainingsprozedur gefunden werden. Diese Zahl zeigt die Größenordnung der Experimente, die in den 1950er und 1960er Jahren in Hinsicht auf die Rechenkapazität möglich waren. In den folgenden Dekaden stieg die Anzahl der Parameter in diesen Maschinen bzw. neuronalen Netzwerken stark an. Als zweiten Meilenstein könnte man das Convolutional Neural Network bezeichnen, das Yann LeCun entwickelt hat, um das Erkennen von Zahlen zu automatisieren. Sein künstliches neuronales Netzwerk hieß LeNet und hatte 1998 einige tausend Parameter. Ein dritter Meilenstein ist dann das Deep Artificial Neural Network AlexNet, das von Hinton und seinen Studierenden entwickelt wurde und 2012 den ImageNet-Wettbewerb gewann und somit die Deep-Learning-Revolution anfachte. AlexNet zählte 60 Millionen Parameter. Tatsächlich ist ‚parametrisierte Maschine' der Begriff, den wir als Grundlage nehmen sollten, um die epistemische Form des heutigen maschinellen Lernens zu definieren.

Den QR-Code scannen, um das vollständige Videointerview anzusehen.

In Conversation with
Alexa Steinbrück on “Modes of Representation of AI and How to Teach It”

The interview with Alexa Steinbrück asks how artificial intelligence can be explored creatively, artistically, and critically. It further discusses how the teaching and mediation of AI and its social consequences could look like in an artistic context.

Transcript of the interview with Alexa Steinbrück, conducted by Francis Hunger on 2022–10–25.

Hunger: In recent years, a large number of myths concerning artificial intelligence have been discussed. You developed, together with Daniel Leufer, the project *AImyths.org* (2019), which takes a critical look at such myths. Of all the myths discussed there, which do you think is the most problematic?

Steinbrück: The main myth that perhaps underlies or reinforces all the others is: 'AI has agency' as in 'the ability to act,' a form of proactivity. One could even go in the direction of 'free will' or 'consciousness.' And this myth suggests that AI systems have this kind of autonomy through which they act. And the myth is expressed in various fields. So, if you start with language, you can find headlines such as "AI has generated its own language"—that was a sensation on Facebook—or "An AI has created this work of art." This already contains a great deal of agency and a great deal of autonomy, and behind this lies a belief that these systems actually have these capabilities, that they are more than just tools. And to me it's as if a pocket calculator, which everyone is familiar with, is suddenly granted a consciousness. But in the end there is no consciousness and no agency. What makes this myth so problematic is that believing in these systems hides many other realities. So, a headline that reads "AI wrote this and this," for instance, "AI wrote this text," hides the fact that there is a system behind it that is trained on Big Data. It hides the human labor involved. It hides the power, and the systems of power that produce such technical systems.

Hunger: Better Images of AI (2021) is another exciting project you've been involved in. What's that about? It also essentially ties in with the myths of artificial intelligence.

Steinbrück: When you run an image search for 'AI' these days, because you might need an image for a presentation or to support something visually, you can't help but look at this blue wall of images that are all relatively similar. So, AI is still depicted as humanoid robots; the color blue is very dominant. You have matrix numbers floating around in the air. So, these are all visual topoi that have nothing to do with the reality of AI at all. That's one criticism—we say, "That's not representative." And the other criticism is that these images are actually harmful because this technology has become so important in our lives that you can't just dismiss it as some offshoot of pop culture, as "yeah, that's funny." Instead, these images have an effect and influence how we perceive AI and how we might perceive the regulation of this system. *Better Images of AI* aimed to create a database of 'better' images that were accessible to everyone under Creative Commons licenses. Indeed, it's not easy to get or commission such images. But it's a constantly growing library. In parallel we also conduct awareness-raising campaigns, hold workshops on the topic, and we intervene on social media.

Hunger: And how can I envision the images that you create? You also worked on this topic with students from Burg Giebichenstein University of Art and Design, and you all contributed to it as well. What visual solutions did you find?

Steinbrück: You could look at the human side, for instance. That is the human work behind the system, and then you arrive relatively quickly at the topic of 'clickworkers' or 'gig workers.' In other words, people who do the very important work of labeling the datasets or even creating the datasets that are necessary to train these 'intelligent' systems. And you can also shed light on this topic from two sides. You end up on the dark side relatively quickly when you consider the topic of clickworkers, because you know that they're underpaid, that it is just monotonous work. And a student of mine also interpreted it that way. He used a 3D printer to create little figures working on computers, sat them all at tables, printed them out hundreds of times, and then photographed them using an infinity cove under relatively gloomy lighting. And these pictures naturally tell the narrative of exploitation, of monotonous work, of hidden work. But the same topic has different sides, and we also have other pictures in the database that represent the work of clickworkers, but these are photos of real people from a company called humansintheloop.org that shows strong social commitment, that roughly says: "We know this work is monotonous, so at the same time we teach these people basic digital skills that will enable them to find a better paying job in the future." And these pictures feature cheerful colors. You see people at computers. It has a bit of a stock-photo aesthetic, but they're real people, very diverse, too. In other words, you can highlight and depict every aspect of AI differently.

Hunger: When you were working on the *Better Images of AI* project, were there things that particularly stood out to you during the discussions with the students?

Steinbrück: I was totally grateful to be able to address this topic at an art academy, because you have these fabulous students who all already bring their own artistic, creative experiences, which they can now apply to such a concrete task. At the same time, however, I became aware that the task is incredibly difficult. So, the task is basically: "Make better images. Free yourself from all these myths that you've been socialized with and have grown up with and do it all ideally within two months!" And that's what I understood, that it takes a lot longer. You really need half a year of intensive study in the field, of the various aspects, and perhaps also of the technical realities, especially the discourses.

Hunger: I'd be interested in hearing a bit more about that, because the artistic approach to artificial intelligence is still a very young and experimental field. Have you come up with other approaches for teaching? Perhaps also what works and what doesn't?

Steinbrück: I've found that the basics of machine learning, including the technical ones, are relatively easy to teach. In other words, you can explain supervised learning without going into the details of linear algebra and calculus, and the core functionality can still be made understandable with Runway ML, which is this graphical user interface-based software. We've had good experiences because it is really easy to get started with it. It's simply a drag-and-drop no-coding tool, as easy to use as Photoshop or maybe an app store, because Runway is essentially a 'bucket' of different models. That was quite good for motivating the students and helping them gain a sense of how diverse the field is. If, however, you want to custom program something, then you really have to program, and that's where you need 'eco-

system skills.' So you need to know the command line, for instance, the function of Git, and Python basics, and that's a very steep learning curve. And I don't think it's realistic to teach that to all students or to make them fit to such a degree that they can design their own systems independently. But at the same time there are always two or three people at the art academy who are totally willing to go into depth and even suffer a bit and push themselves through it.

Hunger: You recently presented the project *The Literal Unseen* (2022), which is based on GANs generating these synthetic images that you already mentioned. Please remind us again how GANs work, also in relation to CLIP, and then I'd like to hear more about the project.

Steinbrück: GAN stands for 'generative adversarial network.' And it's actually two networks that are used for training. And that's a super smart architecture that Ian Goodfellow came up with in 2014, because it kind of reproduces the game between the detectives and the forgers who gradually, because they're confronted with each other, get better and better. So, in a GAN there's a generator network and a discriminator network. The generator generates images. The discriminator decides whether the image is real or fake. So, 'real' would mean it comes from the training dataset, and 'fake' would mean it's synthetic. And when both parties are trained, so to speak, with the GAN, you can then disconnect the generator again and use it as an independent network—to generate new images for an art project, for instance. And CLIP originally has nothing to do with GANs. It's simply a different network architecture developed by the company OpenAI, which deals with language, with the relationship between language and images. And you can combine CLIP and GANs and do relatively exciting things this way. CLIP has learned to project images and texts into the same mathematical space and is thus able to measure the distances between the image and the text. And thus, you can say: "Okay, if they are relatively close to each other, then they probably describe the same thing," then this text is probably a description for the image. And that's so powerful that many new architectures can be based on this CLIP technology. Last year, for example, Katherine Crowson conducted an experiment where she combined this GAN with CLIP, which means that suddenly you were able to generate images based on text input, and that's the same principle that DALL-E uses, or Midjourney, or Stable Diffusion. Anyway, at that time I began an artistic project called *The Literal Unseen*. 'Literal,' because I work with text input. 'Unseen,' because it's about the hidden stories behind AI. We've already talked about bad AI images that reproduce the same stereotypes over and over again, the same myths about AI. And I was interested in what other stories you could access, or perhaps how you could represent this technology more realistically with the help of text-to-image algorithms. And *Literal Unseen* is a series of images. Actually, it's a search process. And for me it's less about the artifacts that emerge in the end, the images that you can frame. One topic, for instance, is the collection of data in the public and private spheres. The systems we're talking about right now are of course also trained on scraped data from the Internet, where virtually no attention is paid as to whether this data is subject to copyright or whether personal rights are affected. So, this gesture of 'grabbing everything that's there' has been around for a very long time in the history of AI. Or another image, this is actually based on an original prompt from Kate Crawford and her latest book *Atlas of AI* (2021). She says: "The cloud is […] made of rocks and lithium brine and crude oil." And I queried the prompt literally as it was into such a system and the image was perfect. It shows a landscape with a huge cloud that has the structure of rocks, and thus translates this metaphor that it uses into an existing image and is, in my opinion, also a realistic representation of AI, although the cloud is of course more. But cloud technologies are relatively central to AI products.

Hunger: Stable Diffusion as one of the current diffusion models is based on this dataset LAION-5B. Have you been able to explore this dataset?

Steinbrück: There is an interface [to explore the dataset] from the LAION group itself. You have various parameters and can also filter it. What I found fascinating was the filter called 'aesthetic scoring.' The last Stable Diffusion model was trained in many different steps, and in each step, there was also a different sub-dataset, and the last one it was trained on was actually this aesthetic subset of the larger dataset. And of course, I found that intriguing because I asked myself, what kind of categorization is that? Where do they get this data from? Then I took a look at the filter and realized, the colors are totally saturated, there's a strong depth of focus, everything is a bit more symmetrical. So I wouldn't say more aesthetic, but maybe 'better produced' or 'glossy.' And it builds on another dataset, which is based on a survey of how people rate synthetically generated images—so actually synthetically generated, and not real images. And the question was, I don't remember exactly, but roughly: "How much do you like this picture?" On the basis of this question, they developed this aesthetic score. And I find that totally fascinating, because of course, 'How much do you like this picture?' can mean all kinds of things. It can mean "I don't like the content," or "I don't like the lighting," or "I don't like the resolution," or "I don't like the JPEG compression." So, 'like' leaves so much open and now it's a central part of the models we work with, and yet a lot remains hidden.

Scan the QR code below to watch the full video interview online.

Im Gespräch mit
Alexa Steinbrück zu „Repräsentationsweisen Künstlicher Intelligenz und wie eine künstlerische Lehre aussehen kann“

Das Interview mit Alexa Steinbrück fragt danach, wie Künstliche Intelligenz kreativ, künstlerisch und kritisch erforscht werden kann. Weiter diskutiert Steinbrück, wie Lehre und Vermittlung von KI und deren gesellschaftlichen Folgen im künstlerischen Kontext aussehen können.

Transkript des Interviews mit Alexa Steinbrück, geführt von Francis Hunger am 25–10–2022.

Hunger: In den letzten Jahren ist eine ganze Reihe von Mythen der Künstlichen ‚Intelligenz' diskutiert worden. Du hast zusammen mit Daniel Leufer das Projekt *AImyths.org* (2019) entwickelt, welches solche Mythen kritisch diskutiert. Welcher aus den dort aufgezeigten Mythen ist deiner Meinung nach der problematischste Mythos?

Steinbrück: Der Hauptmythos, der vielleicht allen anderen zugrunde liegt oder diese auch noch bestärkt, ist – und das lässt sich jetzt schwer auf Deutsch übersetzen – ‚AI has agency'. Also ‚Handlungsfähigkeit', eine Form von Proaktivität. Wenn man will, könnte es auch in die Richtung gehen wie: ‚freier Wille' oder ‚Bewusstsein' sogar. Und dieser Mythos suggeriert, dass eben KI-Systeme diese Art von Autonomie haben, aus der sie heraus handeln. Und der Mythos drückt sich in verschiedenen Bereichen aus. Also wenn man mit Sprache anfängt, kann man zum Beispiel sagen, solche Headlines wie „KI hat ihre eigene Sprache generiert", das war mal so ein Eklat um Facebook herum, oder: „Eine KI hat dieses Kunstwerk kreiert". Da steckt schon ganz viel an Handlungsfähigkeit drin und ganz viel Autonomie und dahinter liegt der Glaube daran, dass diese Systeme tatsächlich diese Kapazitäten haben. Dass sie eben mehr sind als bloße Tools, und für mich wirkt es so, als würde man einem Taschenrechner, den ja jeder von uns kennt, auf einmal ein Bewusstsein zusprechen. Aber am Ende steckt da eben kein Bewusstsein drin und keine eigenständige Handlungsfähigkeit. Was diesen Mythos so problematisch macht, ist, dass dieser Glaube an diese Systeme ganz viele andere Realitäten versteckt. Also zum Beispiel versteckt so eine Headline, die sagt: „KI hat das und das geschrieben" – zum Beispiel: „KI hat diesen Text geschrieben" –, dass dahinter ein System steckt, was auf große Datenmengen trainiert ist. Sie versteckt die menschliche Arbeit, versteckt die Macht und die Machtsysteme, die solche technischen Systeme produzieren.

Hunger: Ein anderes spannendes Projekt, an dem du beteiligt warst, ist *Better Images of AI* (2021). Worum geht es da? Das knüpft auch im Prinzip an die Mythen von Künstlicher ‚Intelligenz' an.

Steinbrück: Wenn man heute eine Bildersuche macht nach ‚KI', weil man vielleicht ein Bild braucht für eine Präsentation oder um irgendetwas visuell zu hinterlegen, dann kann man nicht anders, als auf diese blaue Wand zu gucken von Bildern, die sich alle relativ ähneln. Also KI wird immer noch dargestellt als humanoider Roboter, die Farbe Blau dominiert sehr. Man hat Matrixzahlen, die in der Luft rumschwirren, also alles visuelle Topoi, die überhaupt nichts mit der Realität von KI zu tun haben. Das ist die eine Kritik, dass wir sagen: „Das ist nicht repräsentativ". Und die andere Kritik ist, dass diese Bilder eigentlich schädlich sind, weil diese Technologie so wichtig geworden ist in unserem Leben, dass man es eben nicht nur als irgendeinen Auswuchs der Popkultur abtun kann, als: „Ja, das ist lustig". Sondern diese Bilder haben eine Wirkung und beeinflussen, wie wir KI wahrnehmen und wie wir vielleicht die Regulierung von diesem System wahrnehmen. *Better Images of AI* hat sich zum Ziel gesetzt, eine Bilddatenbank zu erschaffen von eben ‚besseren' Bildern, die für alle unter Creative-Commons-Lizenzen zugänglich sind. Es ist auch nicht einfach, solche Bilder zu bekommen oder in Auftrag zu geben. Aber das ist trotzdem eine beständig wachsende Bibliothek, und parallel dazu machen wir Awareness-Raising-Kampagnen. Also wir geben Workshops zu dem Thema, wir mischen uns auf Social Media ein.

Hunger: Und wie kann ich mir konkret dann die Bilder, die ihr schafft, vorstellen? Du hast mit Studierenden der Kunsthochschule Burg Giebichenstein an diesem Thema gearbeitet und ihr habt auch Beiträge dazu geliefert. Welche visuellen Lösungen habt ihr da gefunden?

Steinbrück: Man könnte zum Beispiel auch die menschliche Seite beleuchten, also die menschliche Arbeit hinter dem System und da kommt man relativ schnell auf das Thema ‚Clickworker*innen' oder auch ‚Gig-Worker*innen', wenn man so will. Also Menschen, die sehr wichtige Arbeit machen, die Datensets zu labeln oder überhaupt zu erstellen, die dafür notwendig sind, um diese intelligenten Systeme zu trainieren. Und das Thema selbst kann man auch von zwei Seiten beleuchten. Man kann so relativ schnell zu einer düsteren Seite kommen, wenn man das Thema Clickworker*innen betrachtet, weil man weiß, die sind unterbezahlt, es ist monotone Arbeit. Und in die Richtung hat das ein Studierender von mir auch interpretiert. Der hat mit einem 3D-Drucker kleine Menschen am Computer gebastelt, diese an Tische gesetzt, hundertfach so ausgedruckt, dann in einer Hohlkehle fotografiert und auch relativ düster beleuchtet. Und diese Bilder erzählen natürlich das Narrativ von Ausbeutung, von monotoner Arbeit, von versteckter Arbeit. Dasselbe Thema hat aber auch verschiedene Seiten und wir haben auch andere Bilder in der Datenbank, die die Arbeit von Clickworker*innen darstellen, aber das sind Fotos von echten Menschen, von einer Firma, Humansintheloop.org, die einen sozialen Anspruch hat, die sagt: „Wir wissen, dass diese Arbeit monoton ist, daher lehren wir diesen Menschen gleichzeitig grundlegende digitale Skills, die sie befähigen, vielleicht in ihrem nächsten Job eine bessere Bezahlung zu bekommen." Und diese Bilder haben freudige Farben, man sieht Leute am Computer. Es hat schon so ein bisschen eine Art Stockfoto-Ästhetik, aber es sind existierende Menschen, sehr divers. Also man kann jeden Aspekt von KI unterschiedlich beleuchten und darstellen.

Hunger: Als du an dem Projekt 2021 gearbeitet hast, gab es dabei Sachen, die dir in der Diskussion mit den Studierenden auffielen?

Steinbrück: Ich war total dankbar, dieses Thema so an eine Kunsthochschule zu bringen, weil man natürlich diese fabelhaften Studierenden hat, die alle schon ihre eigene künstlerische, gestalterische Praxis haben, die sie nun auf so eine konkrete Aufgabe anwenden können. Gleichzeitig ist die Aufgabe unheimlich schwierig. Also die Aufgabe heißt ja im Grunde: „Mach bessere Bilder, befreie dich von diesen ganzen Mythen, mit denen man sozialisiert und aufgewachsen ist, und das innerhalb von zwei Monaten!" Und das habe ich verstanden, dass das viel mehr Zeit bedarf. Also eigentlich bräuchte man ein halbes Jahr intensiver Auseinandersetzung mit dem Feld, mit den verschiedenen Aspekten und vielleicht auch mit den technischen Realitäten, vor allen Dingen mit den Diskursen.

Hunger: Das würde mich mehr interessieren, weil eben dieser künstlerische Umgang mit Künstlicher ‚Intelligenz' noch ein sehr junges und experimentelles Feld ist. Hast du noch weitere Vorgehensweisen für die Lehre entwickelt, vielleicht auch, was funktioniert und was nicht?

Steinbrück: Ich habe festgestellt, dass man doch relativ einfach die Basics von Machine Learning, also auch die technischen, gut vermittelt bekommt. Man kann Supervised Learning erklären, ohne in die Tiefen von linearer Algebra und Calculus einzugehen und trotzdem lässt sich die Kernfunktionalität verständlich machen mit Runway ML, also das ist diese GUI-Software. Wir haben damit ganz gute Erfahrungen gemacht, weil es so gut wie keine Einstiegshürde gibt. Es ist ja einfach ein Drag-and-Drop- bzw. No-Coding-Tool, so einfach wie Photoshop zu benutzen oder vielleicht auch einen App-Store, weil Runway ja eigentlich so ein ‚Eimer' von verschiedenen Modellen ist. Das war ganz gut, um die Studierenden so dafür zu begeistern und ein Gefühl dafür zu kriegen, wie vielfältig das Feld auch ist. Wenn man aber irgendwas ganz custom-mäßiges programmieren will, muss man eben programmieren und benötigt ‚Ecosystem Skills'. Also beispielsweise: Kommandozeile, die Funktion von Git oder Python-Grundlagen, und das ist eine sehr steile Lernkurve. Und ich glaube, allen Studierenden das beizubringen, dass sie selbständig ihre eigenen Systeme entwerfen können, ist nicht realistisch. Gleichzeitig gibt es aber immer an der Kunsthochschule so die zwei, drei Personen, die total willig sind, da in die Tiefe zu gehen und auch ein bisschen dafür zu leiden.

Hunger: Zuletzt stelltest du ja das Projekt *The Literal Unseen* (2022) aus, welches auf GANs, also den von dir schon angesprochenen synthetischen Bildern basiert. Bitte ruf uns noch mal in Erinnerung, wie GANs funktionieren, auch im Verhältnis zu CLIP, und dann würde ich gerne mehr über das Projekt erfahren.

Steinbrück: GAN steht für ‚Generative Adversarial Network' und es sind eigentlich zwei Netzwerke, die zum Trainieren verwendet werden. Und das ist eine super smarte Architektur, die sich Ian Goodfellow 2014 ausgedacht hat, weil es so ein bisschen das Spiel zwischen der Polizei und den Geldfälscher*innen reproduziert, die graduell, weil sie miteinander konfrontiert sind, immer besser werden. Also in einem GAN gibt es ein Generator-Netzwerk und ein Diskriminator-Netzwerk. Der Generator generiert Bilder. Der Diskriminator entscheidet, ob das Bild echt ist oder ein Fake. Also ‚echt' würde bedeuten, es kommt aus dem Trainingsdatensatz und ‚Fake' würde bedeuten, es ist synthetisch. Und wenn bei dem GAN beide Parteien trainiert sind, kann man dann den Generator wieder abstöpseln und als eigenständiges Netzwerk benutzen, um beispielsweise neue Bilder für ein künstlerisches Projekt zu generieren. Und CLIP hat erst einmal mit GANs gar nichts zu tun. Also es ist einfach eine andere Netzwerk-Architektur, von der Firma OpenAI entwickelt, die sich mit Sprache, mit der Beziehung von Sprache und Bildern auseinandersetzt und man kann CLIP und GANs kombinieren und dadurch relativ spannende Sachen machen. CLIP hat gelernt, Bilder und Texte in denselben mathematischen Raum zu projizieren, und ist dadurch in der Lage, die Distanzen zwischen dem Bild und dem Text auszumessen. Und dadurch kann man sagen: „Okay, wenn die relativ nah beieinander liegen, dann beschreiben sie wohl dasselbe." Also dann ist wohl dieser Text eine Beschreibung für das Bild. Und das ist so powerful, dass auf dieser CLIP-Technologie jetzt ganz viele neue Architekturen beruhen. Letztes Jahr hat Katherine Crowson ein Experiment gemacht, wo sie dieses GAN mit CLIP kombiniert hat, das heißt, auf einmal war man in der Lage, Bilder zu generieren auf Basis von Text-Input und das ist dasselbe Prinzip, was auch DALL-E macht, Midjourney und Stable Diffusion. Zu dem Zeitpunkt habe ich ein künstlerisches Projekt gestartet, das nennt sich *The Literal Unseen*. ‚Literal', weil ich eben mit Text-Input arbeite. ‚Unseen', weil es um die versteckten Geschichten hinter KI geht. Also wir haben ja vorhin schon über schlechte KI-Bilder gesprochen, die so immer wieder dieselben Stereotypen reproduzieren, dieselben Mythen über KI, und mich hat interessiert, was man denn für andere Geschichten aufrufen oder wie man diese Technologie realistischer mithilfe von Text-zu-Bild-Algorithmen darstellen kann. Und *Literal Unseen* ist eine Bilderserie, eigentlich ist es ein Suchprozess, und mir geht es weniger um die Artefakte, die am Ende rauskommen, die Bilder, die man dann einrahmen kann. Ein Thema ist zum Beispiel die Datensammlung im Öffentlichen sowie im Privaten, da diese Systeme, über die wir hier gerade reden, natürlich auch trainiert sind auf gescrapten Daten aus dem Netz, wo so gut wie keine Aufmerksamkeit dafür existiert, ob jetzt ein Copyright darauf liegt oder Persönlichkeitsrechte betroffen sind. Also diesen Gestus von ‚sich alles greifen, was da ist', den gibt es so ja schon sehr lange in der Geschichte der KI. Oder ein anderes Bild, das beruht tatsächlich auf einem Original-Prompt von Kate Crawford und ihrem Buch *Atlas of AI* (2021). Da sagt sie: „The cloud is […] made of rocks and lithium brine and crude oil". Und den Prompt habe ich eins zu eins in so ein System eingegeben und das Bild war perfekt. Es zeigt eine Landschaft mit einer riesigen Wolke, die die Struktur von Steinen hat, und übersetzt dadurch halt diese Metapher, die sie anwendet, in ein existierendes Bild und ist meiner Meinung nach auch eine realistische Darstellung von KI, obwohl die Cloud natürlich mehr ist. Aber Cloud-Technologien sind für KI-Produkte zentral.

Hunger: Stable Diffusion als eines der aktuellen Diffusion Models, das basiert ja auf diesem Datensatz LAION-5B. Konntest du dich bereits mit dem Datensatz auseinandersetzen?

Steinbrück: Es gibt ein offenes Interface [mit dem man den Datensatz erkunden kann], von der LAION-Gruppe selbst. Was ich da spannend fand, war der eine Filter, der nannte sich ‚ästhetisches Scoring'. Das letzte Stable-Diffusion-Modell wurde ja in verschiedenen Schritten trainiert und in jedem Step gab es auch einen anderen Sub-Datensatz und der letzte, auf den es trainiert wurde, war tatsächlich dieses ästhetische Subset von dem größeren Datensatz, und das fand ich natürlich spannend, weil ich mich gefragt habe, was ist denn das für eine Kategorisierung? Wo nehmen die die Daten her? Ich habe mir dann den Filter mal angeguckt und gemerkt, dass die Farben total satt sind, es gibt eine starke Tiefenschärfe, es ist alles ein wenig symmetrischer. Ich würde nicht sagen ‚ästhetischer', aber vielleicht ‚besser produziert' oder ‚glossy'. Und das beruht wohl auf einem anderen Datensatz, der auf einer Umfrage beruht, wie Menschen synthetisch generierte Bilder bewerten – also tatsächlich synthetisch generiert und nicht echte Bilder. Und da war die Frage einfach auf Englisch: „Wie gut gefällt dir dieses Bild?" Und auf Basis dieser Frage haben sie diesen ästhetischen Score entwickelt. Und das finde ich total spannend,

weil ‚Wie gut gefällt dir dieses Bild?' natürlich alles Mögliche bedeuten kann. Das kann meinen: „Mir gefällt der Inhalt nicht", „Mir gefällt die Belichtung nicht", „Mir gefällt die Auflösung nicht" oder „Mir gefällt die JPEG-Kompression nicht". Das ‚Gefallen' lässt so viel offen und trotzdem ist es jetzt ein zentraler Bestandteil der Modelle, mit denen wir arbeiten; dabei bleibt vieles auch versteckt.

Den QR-Code scannen, um das vollständige Videointerview anzusehen.

In Conversation with Anna Ridler about "From GANs to Stable Diffusion. On Artistic Collaboration with Generative Algorithms"

The interview explores Ridler's relationship to AI art and her artistic engagement with generative visual processes. She talks about the materiality of datasets that she herself creates for artistic purposes and which serve as the basis for her own AI models.

Transcript of the interview with Anna Ridler, conducted by Francis Hunger on 2023–03–20.

Hunger: Maybe we can begin with the provocative question of whether you would consider yourself, now in 2023, an AI artist. I would suggest you wouldn't. And the more interesting question then is, what does AI art do and what do you do differently?

Ridler: That is a slightly provocative question, because I would not necessarily consider myself to be an AI artist, although I very often show in exhibitions about AI or machine learning. But I consider myself to be an artist who is interested in systems and particularly systems of knowledge and how knowledge is formed and kept. And I keep coming back to AI and keep coming back to machine learning as a way of exploring this interest, because for me, machine learning is a very powerful and interesting way of exploring knowledge. The way that it works in an art practice gives you multiple different paths of exploring one idea. Because you've got this training dataset and there are a whole host of issues around that, which we can talk about. Then you've got the algorithm itself, which again you can play with and try different ideas or different concepts, and then you've got the eventual output, which again can do different things or be displayed in a different way. The most interesting thing about these methods and these methodologies is the way that they allow you to convince through concepts. Without sounding too pretentious, I would say that I consider myself to be more a conceptual artist who works with technology and not a 'technology-first' artist.

Hunger: In your 2017 work *Fall of the House of Usher,* you trained a GAN with your own training dataset. Could you describe how this approach emerged and what you also learned during the process?

Ridler: *Fall of the House of Usher* was made using pix2pix, which is a very-small-badge machine learning algorithm where you deliberately need a small dataset, which again opens up the possibilities of what you can do with that. Because normally, when you are working with machine learning you need hundreds, thousands, millions of images or inputs in order for it to work. I've always been interested in the smaller end of that. What's the minimum amount that you need in order for something to make sense, in order for something to happen? But I primarily work with GANs, generative adversarial networks. The way they are most commonly described is: you have a forger network and a detective network, and the forger network is looking at the training set and trying to produce an image that could come from that network. And the detective network is looking at the result and saying whether it is 'real' or 'fake.' And they loop through this cycle over many iterations or epochs until the forger network can produce something that fools the detective network—and that's when you get your image or output. There are these two uncontrollable networks that dance around each other in a way that still isn't that well understood. I've worked with them extensively and they do have their quirks and subtleties. They tend to oversaturate certain colors, so a lot of the output you see can tend to look a little bit garish, they really overfit the color 'red,' for example. But one of the things that I find most interesting, because they are very difficult to control, is how you can try and impose control onto them, using things like labels to shape the way that they behave, to try and get something out that you can then work with. But there is this really lovely quality of chance and control. Even though you know that you can spend a lot of time labeling and working with your data, if you have an image in your head, it will never be that same image.

Hunger: This train of thought also plays into your work *Circadian Bloom* from 2021, which installs GAN-generated imagery of flowers on screens in public spaces, building on Carl von Linné's concept of a flower clock. It seems you are really invested in the order of knowledge.

Ridler: *Circadian Bloom* is thinking about how we measure time, both in a scientific and a human way through clock time and through atomic time, and how it's so ultra-precise. And then the project is thinking about earlier, different ways of how we used to tell time through more cyclical measures, for instance, medieval ways of telling time. They would divide the amount of daylight of the day, which changes depending on your geography and also on the time of the year. And then natural ways of telling time, for instance, when different flowers bloom and close. And for *Circadian Bloom* I was really interested in such a certain group of flowers which have their own chronobiological clock that sits inside them; they bloom and close at fixed points during the day, regardless of whether they are in sunlight or darkness. I created these series of GANs to produce these digital flowers that would bloom and close in synchronicity with their natural counterparts, so that they're constantly opening and closing. They're programmed to start and finish according to the dawn and dusk of where they're installed. It's this very natural way of telling time, but presented in a hyper-accurate, mechanical way. There are these clocks that are made accurate to the atomic millisecond, but visually that accuracy is totally obscured. I designed the flowers to be installed in public spaces and to be there for a long period of time, so that over the seasons people could see them changing and could remember this different way of telling time.

Hunger: *Laws of Ordered Form* from 2020 is a work in which you ceated a downloadable dataset of encyclopedias from the Victorian and Edwardian eras. Why did you think it was necessary to provide such a dataset for download?

Ridler: It's really nice to be open with datasets, because part of the piece was examining the fluidity of image and word. I took these images from historical encyclopedias and scanned them and took them out of their original context and gave them my own labels, but I wanted to give that as a downloadable dataset for people to then explore, and play, and add, and change, and cut, and crop, and do what they wanted with. Particularly, I was thinking about the context of Victorian encyclopedias. People, and particularly women, at that time would use this imagery to create their own worlds, to collage and to make the imagery they wanted. And now, it returns in this nice way, when you allow it to be downloadable and used by whoever finds it, rather than keeping it hidden, and structured, and set.

Hunger: Could you describe the process a little bit more in terms of how you developed this work, where you got the encyclopedias from, and what you did with them to create the dataset?

Ridler: I was very interested in making connections between datasets and how they are basically contemporary encyclopedias. How datasets are trying to describe and record the world around us and then represent it. Rather than being an encyclopedia that is written by a human for a human, the dataset works human to machine. It is slightly different, but there is this thread of trying to capture the world and describe it. And when you start to look through these historic encyclopedias, you can really see in a very stark way some of the problems that happen when you're trying to describe the world: the viewpoints that are given, the casual racism and stereotyping and sexism that just exist there, the erasure, the over-representation. And you can also see the things that are given a lot of prominence and interest, versus the things that are very squeezed in. And presenting this, it allows people to understand in a much sharper way these issues and problems. You can also see the freezing of meaning, as well, because compared to books, datasets are published and once they're published, they're facts. They are very rarely updated, and once they're used to train a model, you just don't know whether that model was up-to-date even if the dataset is updated. And so, you can start to use them as a really nice way to compare and draw out dissimilarities. Encyclopedias, now, nobody uses them. I found them all in charity shops and on eBay, because nobody wants these big books in their houses. And datasets, too, they're working objects, but nobody is really looking after them. It's really hard to find early instances of datasets, because they've been taken down and only small parts of them still exist and so you can start to see connections.

Hunger: I also like the materiality. When I think about it, how you collect, you're taking these books, you're putting them on the scanner, so it really shows how you create data with your own hands, compared to scraping data mechanically from the Internet. So, this work gives a much clearer vision in understanding of how data is being created through manual labor in your case.

Ridler: One of the things that I'm also really interested in as part of my practice is to make the labor visible when the work is displayed. Rather than just presenting something that is literally crisp and clean and final, I like to have the working process out there as well. So, in the case of *Laws of Ordered Form* it's a video work that shows me doing that entire process—looking through the books, scanning the books, working with the imagery, labeling the imagery, working with my computer, moving files, re-labeling files, all of these things. All of this huge amount of work that has to be done before you even get to train anything. And the labor is very often obscured, particularly if you're using ready-made models or ready-made datasets or using a web user interface with machine learning. And it's really important, because there is this hugely human element to working with this stuff. There are choices at every step of the process and these are the choices that can turn something into an artwork versus a 'cool' tech demo.

Hunger: Maybe we can also contrast that with what we currently see as a surge of text-to-image generative networks, such as Stable Diffusion, Midjourney or DALL-E. These are large pre-trained networks, pre-trained on the data scrapings of the Internet. How does image production change in general with these networks and how does it change for you as an artist?

Ridler: I found it very difficult to make something conceptually interesting with that, because for the most part they're just behind APIs [Application Programming Interfaces]. So you can't even get at the code to experiment with that. With Stable Diffusion, you can download the model, you can download the weights and you can start to mess with that a little bit, but there is no sense of how it's being created. You can even just start to work backwards and see how certain object classes might have been trained on different websites by an overproduction of certain styles, but you don't know. And so, for me, because so much of my practice is about active choice and about choosing and being deliberate about all of these different things, I find it really hard to then introduce what I'm interested in into working with these models. These generative models make incredible images, very quickly, but I find it difficult using them in a critical and interesting way, and there is the additional debate around the ownership of data and whether or not the producer's permissions have been given.

Hunger: Well, it gave a lot of people, who might not originally be framed as artists, the motivation to get involved with art and to deal with the art they know. And then, I have the feeling, many even started calling themselves 'AI artists.'

Ridler: It has now led to 'AI art' meaning a very particular thing. If you go onto Twitter or Discord now and talk about AI art, it does mean this text-to-image type of art. And it's a good thing that people want to explore and become creative, but you know, there is always a difference between something being an artistic process and being an artwork. What separates art from a pretty picture or poster is the intention and concept. And the thing I find hardest when working with these tools is how to get that intention and concept in.

Scan the QR code below to watch the full video interview online.

Im Gespräch mit
Anna Ridler zu „Von GANs bis Stable Diffusion. Über die künstlerische Zusammenarbeit mit generativen Algorithmen“

Das Interview erkundet Ridlers Verhältnis zur KI-Kunst und ihre persönliche künstlerische Auseinandersetzung mit generativen visuellen Verfahren. Sie spricht über die Materialität der Datensätze, welche sie für künstlerische Zwecke selbst erstellt und die ihr als Grundlage für eigene KI-Modelle dienen.

Transkript des Interviews mit Anna Ridler, geführt von Francis Hunger am 20–03–2023.

Hunger: Vielleicht können wir mit der etwas provokanten Frage beginnen, ob du dich selber heute, im Jahr 2023, als KI-Künstlerin betrachten würdest, und ich vermute mal, dass dies nicht der Fall ist. Und die interessantere Frage ist dann, was macht KI-Kunst und was machst du anders?

Ridler: Das ist eine leicht provokante Frage, da ich mich nicht unbedingt als KI-Künstlerin bezeichnen würde, obwohl ich sehr oft an Ausstellungen über KI oder maschinelles Lernen teilnehme. Aber ich betrachte mich als eine Künstlerin, die sich für Systeme interessiert, insbesondere für Wissenssysteme und dafür, wie Wissen entsteht und aufbewahrt wird. Und ich komme immer wieder auf KI und maschinelles Lernen zurück, um diesem Interesse nachzugehen. Denn für mich ist das maschinelle Lernen eine sehr machtvolle und interessante Art, Wissen zu erforschen. Denn man hat den Trainingsdatensatz und damit verbunden eine ganze Reihe von Problemen, über die wir reden können. Und dann hat man den Algorithmus selbst, mit dem wiederum verschiedene Ideen oder Konzepte durchgespielt werden können. Schließlich erhält man den letztendlichen Output, der wiederum verschiedene Dinge tun oder auf verschiedene Weise angezeigt werden kann. Für mich ist das Interessanteste an diesen Methoden und Verfahren die Art und Weise, wie man durch Konzepte überzeugen kann. Ohne zu prätentiös zu klingen, würde ich also sagen, dass ich mich zuvorderst als Konzeptkünstlerin sehe, die mit Technologie arbeitet, und nicht als eine Künstlerin, welche die Technologie in den Vordergrund rückt.

Hunger: In der Arbeit *Fall of the House of Usher* (2017) hast du ein GAN mit eigenen Trainingsdaten trainiert. Könntest du beschreiben, wie dein Ansatz entstanden ist und was du während des Prozesses gelernt hast?

Ridler: *Fall of the House of Usher* wurde mit pix2pix produziert, einem Algorithmus für maschinelles Lernen, bei dem man absichtlich einen kleinen Datensatz nimmt, was wiederum die Möglichkeiten erweitert, was man damit machen kann. Normalerweise benötigt man bei der Arbeit mit maschinellem Lernen Hunderte, Tausende, Millionen von Bildern oder Inputs, damit es funktioniert. Ich habe mich schon immer für den kleineren Bereich interessiert. Was ist die Mindestmenge, die man braucht, damit etwas Sinn macht, damit etwas passiert? Aber ich arbeite hauptsächlich mit GANs – ‚Generative Adversarial Networks'. Die gängigste Beschreibung ist, dass es ein Fälschernetzwerk und so ein Detektivnetzwerk gibt, wobei das Fälschernetzwerk den Trainingssatz betrachtet und versucht, ein Bild zu erzeugen, das von diesem Netzwerk stammen könnte. Das Detektivnetz prüft das Ergebnis und sagt, ob es ‚echt' oder ‚gefälscht' ist. Der Zyklus wird in einer Schleife über viele Iterationen oder Epochen durchlaufen, bis das Fälschernetzwerk etwas produziert, das das Detektivnetzwerk erfolgreich täuscht, und dann erhält man das Bild oder den Output. Es gibt diese beiden unkontrollierbaren Netzwerke, die auf eine Art und Weise umeinander herumtanzen, die immer noch nicht wirklich gut verstanden wird. Ich habe also ausgiebig mit diesen gearbeitet, und sie haben ihre Macken und Feinheiten. Sie neigen dazu, bestimmte Farben zu ‚übersteuern', sodass viele der Ergebnisse, die man sieht, ein wenig grell aussehen können, weil sie beispielsweise die Farbe Rot übersteuern. Aber eines der Dinge, die ich am interessantesten finde, ist – weil sie sehr schwer zu kontrollieren sind – wie man nun versuchen kann, ihnen Kontrolle aufzuerlegen, indem man Dinge wie Labels verwendet und versucht, mithilfe von Labels ihr Verhalten zu formen, um etwas zu erreichen, mit dem man dann arbeiten kann. Es gibt diese wirklich schöne Qualität von Zufall und Kontrolle. Auch wenn man weiß, dass man Ewigkeiten damit verbringen kann, seine Daten zu labeln und mit ihnen zu arbeiten, wird man nie wirklich in der Lage sein, ein Bild zu produzieren, das genau so aussehen wird wie das Bild, welches man im Kopf hat.

Hunger: Dieser Gedankengang spielt auch in deiner Arbeit *Circadian Bloom* von 2021 eine Rolle, in der du GAN-generierte Blumenbilder auf Bildschirmen im öffentlichen Raum installiert hast, die auf Carl von Linnés Konzept einer Blumenuhr aufbauen.

Ridler: Bei *Circadian Bloom* geht es darum, wie wir die Zeit messen, und zwar sowohl auf wissenschaftliche als auch auf humane Art mittels der Uhrzeit, der Atomzeit, und wie ultrapräzise das ist. Und dann denke ich in dem Projekt über historische Methoden nach, wie wir die Zeit durch zyklischere Messungen, also mittelalterliche Methoden der Zeitmessung, bestimmt haben. Sie urteilten damals über die Menge des Tageslichts am Tag und dann hängt es von der Geografie und der Jahreszeit ab. Und dann gibt es noch natürliche Methoden der Zeitmessung, zum Beispiel, wann verschiedene Blumen sich öffnen und wieder schließen. Für *Circadian Bloom* habe ich mich für eine bestimmte Gruppe von Blumen interessiert, die ihre eigene chronobiologische Uhr haben, und sie öffnen und schließen sich zu bestimmten Zeitpunkten am Tag, unabhängig davon, ob sie im Sonnenlicht oder im Dunkeln waren. Ich habe also eine Reihe von GANs entwickelt, um diese digitalen Blumen zu produzieren, die sich synchron zu ihren natürlichen Gegenstücken öffnen und schließen und die dann so programmiert sind, dass sie je nach Sonnenaufgang und -untergang entsprechend ihrem Standort beginnen und enden. Es handelt sich sozusagen um eine sehr natürliche Art und Weise, die Zeit zu messen, aber auf eine sehr genaue, mechanische Weise. Es gibt Uhren, die bis auf die atomare Millisekunde genau sind, aber auf der visuellen Ebene der Arbeit ist diese Genauigkeit völlig verdeckt. Ich habe sie also so konzipiert, dass sie im öffentlichen Raum installiert wird und dort über einen langen Zeitraum verbleibt, sodass die Menschen im Laufe der Jahreszeiten sehen können, wie sich die Blüten verändern, und sich so auf diese andere Art des Zeitmessens zurückbesinnen.

Hunger: In der Arbeit *Laws of Ordered Form* (2020) hast du einen herunterladbaren Datensatz von Enzyklopädien aus der viktorianischen und edwardianischen Zeit erstellt. Warum hieltest du es für notwendig, so einen Datensatz als Download anzubieten?

Ridler: Es ist wirklich schön, offen mit Datensätzen umzugehen, denn ein Teil der Arbeit war die Untersuchung der Fluidität von Bild und Wort. Ich habe also diese Bilder aus historischen Enzyklopädien gescannt, sie aus ihrem ursprünglichen Kontext herausgenommen und mit meinen eigenen Labels versehen. Aber ich wollte sie als herunterladbaren Datensatz zur Verfügung stellen, damit andere Menschen sie erforschen und damit spielen, etwas hinzufügen, sie verändern, ausschneiden und beschneiden können. Ich habe dabei vor allem an den Kontext der viktorianischen Enzyklopädien gedacht: Menschen, und hier insbesondere Frauen, nutzten damals diese Bilder, um ihre eigenen Welten zu erschaffen und jene Bilder zu collagieren, die sie haben wollten. Und so kehrt diese Idee auf diese schöne Art und Weise zurück, wenn man zulässt, dass der Datensatz heruntergeladen und von allen, die ihn finden, genutzt werden kann, anstatt ihn versteckt und strukturiert und festgelegt [für sich] zu behalten.

Hunger: Könntest du den Prozess ein wenig genauer beschreiben, wie du diese Arbeit entwickelt hast, woher du die Enzyklopädien bekommen hast und was du mit ihnen gemacht hast, um den Datensatz zu erstellen?

Ridler: Ich war sehr daran interessiert, Verbindungen zwischen Datensätzen herzustellen und zu zeigen, wie sie im Grunde genommen zeitgenössische Enzyklopädien sind. Wie Datensätze die Welt um uns herum beschreiben und aufzeichnen, um sie dann darzustellen. Anstatt also eine Enzyklopädie zu sein, die vom Menschen für den Menschen geschrieben wird, fungieren Datensätze vom Menschen zur Maschine. Also ist also etwas anders, aber es gibt diesen roten Faden, der versucht, die Welt zu erfassen und zu beschreiben. Und wenn man sich diese historischen Enzyklopädien anschaut, sieht man sehr deutlich einige der Probleme, die auftreten können, wie bestimmte Sichtweisen, der beiläufige Rassismus, all die Stereotypen und der Sexismus, die dort einfach vorhanden sind, die Auslöschung, die Überrepräsentation. Und man kann auch sehen, welche Dinge viel Beachtung und Interesse finden und welche nicht. Wenn man dies darstellt, kann man diese Fragen und Probleme viel besser verstehen. Man kann auch sehen, wie Bedeutung eingefroren wird, denn ähnlich zu Büchern, wenn sie veröffentlicht werden, sind Datensätze, sobald sie veröffentlicht sind, Fakten. Sie werden nur sehr selten aktualisiert, und wenn sie einmal zum Trainieren eines Modells verwendet wurden, weiß man nicht, ob das Modell aktuell ist, selbst wenn der Datensatz aktualisiert wurde. Man kann sie also sehr gut zum Vergleichen und zum Herausarbeiten von Unterschieden verwenden. Enzyklopädien, die benutzt heute niemand mehr. Ich habe sie alle in Wohltätigkeitsläden und online bei eBay gefunden, weil niemand diese dicken Bücher im Haus haben will. Und auch die Datensätze sind Arbeitsobjekte, aber niemand kümmert sich wirklich um sie. Es ist schwer, frühe Versionen von Datensätzen zu finden, weil sie offline gestellt wurden und nur noch kleine Teile von ihnen existieren, und somit kann man immer klarer Zusammenhänge erkennen.

Hunger: Mir gefällt auch die Materialität. Du sammelst, nimmst diese Bücher und legst sie auf den Scanner. Das zeigt, wie du Daten im Vergleich zum maschinellen Auslesen aus dem Internet mit deinen eigenen Händen erzeugst. Dieses Projekt vermittelt also eine klarere Vorstellung davon, wie Daten durch manuelle Arbeit erzeugt werden.

Ridler: Eines der Dinge, an denen ich als Teil meiner Praxis wirklich interessiert bin, ist es, den Arbeitsprozess sichtbar zu machen, wenn das Projekt ausgestellt wird. Anstatt also nur etwas zu präsentieren, das im wahrsten Sinne des Wortes klar, bereinigt und fertig ist, möchte ich auch den Prozess sichtbar machen. Im Fall von *Laws of Ordered Form* handelt es sich um ein Video, das mich bei der Arbeit zeigt, beim Durchsehen der Bücher, beim Einscannen der Bücher, bei der Arbeit mit den Bildern, beim Labeln der Bilder, bei der Arbeit mit meinem Computer, beim Verschieben von Dateien, beim Umlabeln von Dateien, bei all diesen Dingen. Alles das ist eine riesige Menge an Arbeit, die geleistet werden muss, bevor man überhaupt mit dem Training beginnen kann. Und diese Arbeit wird sehr oft vernachlässigt, besonders wenn man mit vorgefertigten Modellen oder Datensätzen arbeitet oder ein Web-Interface für maschinelles Lernen verwendet. Und das ist wirklich sehr wichtig, denn die Arbeit mit diesem Material hat auch ein enormes menschliches Element. Bei jedem Schritt des Prozesses gibt es Entscheidungen, und diese Entscheidungen können dazu führen, dass etwas als Kunst funktioniert und sich von einer ‚coolen' Tech-Demo unterscheidet.

Hunger: Vielleicht können wir das auch damit in Kontrast setzen, was wir derzeit als eine Welle von generativen Text-Bild-Netzen wie Stable Diffusion, Midjourney oder DALL-E sehen. Dabei handelt es sich um große vortrainierte Netze, die mit Daten aus dem Internet gespeist wurden. Wie verändert sich deiner Meinung nach die Bildproduktion im Allgemeinen mit diesen Netzwerken und wie verändert sie sich für dich als Künstlerin?

Ridler: Für mich war es schwierig, damit etwas konzeptionell Interessantes zu machen, weil sie größtenteils nur durch APIs [Application Programming Interfaces] zugänglich sind. Man kommt also nicht an den Code heran, um damit zu experimentieren. Bei Stable Diffusion kann man das Modell und die Gewichte herunterladen und damit herumspielen, aber man hat kein Gefühl dafür, wie es erstellt wurde. Man kann es sogar rückwärts analysieren, um nachzuvollziehen, wie etwa bestimmte Objektklassen anhand verschiedener Websites durch eine Überproduktion bestimmter Stile trainiert wurden, aber man weiß es letztendlich nicht genau. Da es in meiner Praxis so sehr um die aktive Auswahl geht und darum, all diese verschiedenen Dinge bewusst auszuwählen, finde ich es schwierig, das, was mich interessiert, in die Arbeit mit diesen Modellen einzubringen. Diese generativen Modelle erstellen unglaubliche Bilder sehr schnell, aber ich finde es schwierig, sie auf eine kritische und interessante Weise zu verwenden. Daneben gibt es die Debatte über das Urheberrecht an Daten und darüber, ob die Ersteller*innen die Erlaubnis dazu gegeben haben.

Hunger: Es hat allerdings vielen Menschen, die bisher nicht als Künstler*innen angesehen werden, die Motivation gegeben, sich mit Kunst zu beschäftigen und anzufangen, sich selbst ‚KI-Künstler*innen' zu nennen.

Ridler: Das hat jetzt dazu geführt, dass ‚KI-Kunst' etwas ganz Bestimmtes bedeutet. Wenn man jetzt auf Twitter oder Discord über KI-Kunst spricht, ist damit diese Art von Text-Bild-Kunst gemeint. Es ist gut, dass die Leute etwas erforschen und kreativ werden wollen, aber es gibt noch immer einen Unterschied zwischen einem künstlerischen Prozess und einem Kunstwerk. Was sich von einem hübschen Bild oder Poster unterscheidet, ist die Intention und das Konzept. Und am schwierigsten bei der Arbeit mit diesen Tools finde ich die Frage, wie man Intention und Konzept einbringen kann.

Den QR-Code scannen, um das vollständige Videointerview anzusehen.

Localized Latent Updates for Fine-Tuning Vision-Language Models

Moritz Ibing, Isaak Lim, Leif Kobbelt

1 Introduction

When it comes to vision tasks, such as classification, object detection or segmentation, much of the success of deep learning is due to ever larger models trained on an increasingly large quantity of data.

One popular approach to make use of immense sources of uncurated data in the form of images with textual descriptions is vision-language models (Jia et al. 2021; Radford et al. 2021). Here, both image and description are individually mapped into a joint embedding space. This embedding is optimized so that matching pairs are close, and the distance between all other pairs large. A model trained in this fashion can be used for zero-shot classification, as the language model can deal with every conceivable class by embedding a textual description (e.g., "a picture of [CLASS]").

The performance of these models on other tasks and datasets, however, such as judging the similarity of artworks, can be suboptimal. In these cases, one common technique is to fine-tune the pre-trained model for the task at hand. Updating the complete model is, however, quite expensive. Two solutions are proposed in the literature to tackle this problem. One is prompt-learning (Zhou et al. 2022a, 2022b; Zhu et al. 2022), where the context around the class ("a picture of" in the last example) is optimized for a specific dataset, rather than hand-crafted. The other is to use adapters (Gao et al. 2021; Zhang et al. 2021): light-weight models (usually small multilayer perceptrons) that modify the embedding produced by either the visual or language model (or both), thus updating the predictions without the need to update the original networks' parameters.

Both these approaches still, however, have the problem that though the performance is improved for the specific domain and task for which the fine-tuning was done, this comes at the cost of a decrease in performance on other tasks/domains compared to the original model (Goodfellow et al. 2013).

The goal of this work is to reap the benefits of fine-tuning on a specific task without losing the generalization ability of the original model. Work in this direction has already been done in the form of CoCoOp (Zhou et al. 2022a), where prompt-learning is employed, but the context is not fine-tuned on a specific dataset, but rather a suitable context is predicted from the image to be classified. Another approach using prompt-learning is ProGrad (Zhu et al. 2022), where the context update is restricted in order not to lose information from the pre-training stage. Although both methods decrease the performance loss in the zero-shot setting, they still do not reach the abilities of the original model.

In contrast, we choose a simple method based on adapters. The idea is to only update the embedding where we actually have training data, and leave it unchanged everywhere else, thus retaining the original predictions of the model where we cannot improve on them. Furthermore, even where we have data, we want to change the embedding as little as possible to allow sensible interpolation between fine-tuned and original embedding. This approach is extremely lightweight, as we need to tune only a small amount of parameters, and back-propagating through the original model is not necessary. Nonetheless, we show an improvement in performance compared to the previous state of the art.

2 Method

Before introducing our approach in more detail, we will give a short overview on how vision-language models work in general using the example of CLIP, which is used in all our experiments.

2.1 Vision-Language Models

Vision-language models consist of two networks: an image encoder f_I and a text encoder f_T. Their exact implementation is of no interest to us in this context. All we need to know is that these models embed an image or a text respectively to a (normalized) feature vector of the same dimension. The cosine distance between an embedded image

and text should then correspond to their similarity, i.e., how well the text describes the image. During training, we are given a batch of n images x and their textual descriptions y. We make the simplifying assumption that each text is a perfect description of the corresponding image, and all other texts are completely unrelated. Thus, we want to minimize the cosine distance between embeddings of matching image/text pairs $f_I(x_i), f_T(y_i)$ and maximize the distance between all other pairs within the batch $f_I(x_i), f_T(y_j)$ with $i \neq j$. Another view would be to regard the cosine distance as the likelihood that a given text y describes the corresponding image x, or vice versa. We can compute the normalized probabilities, where τ is a learned temperature parameter, as:

$$p(y|x) = \frac{\exp(f_I(x)^T f_T(y)/\tau)}{\sum_{i=1}^{n} \exp(f_I(x)^T f_T(y_i)/\tau)}$$

As we assume the embeddings to be normalized, the dot product is equivalent to the cosine similarity. The probability $p(x|y)$ differs only in the normalization.

In this view, it now makes sense to maximize the probability for the correct pairs, for which we can use cross-entropy loss. As we want to maximize the probabilities in both directions, the loss is given as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}(\log(p(x_i|y_i)) + \log(p(y_i|x_i)))/2$$

Zero-Shot Classification: This approach leads to a semantically meaningful embedding of both images and text that can be used for downstream tasks. Alternatively, we can use it directly for zero-shot classification. To do so, we embed a text for each class (e.g., a picture of [CLASS]) to a feature vector $f_T(y_k)$. Then, to classify a given image, we embed it and compute the distance between its feature vector $f_I(x)$ and all class embeddings. The class probability for a class k is then, similarly as before, given as $p(y_k|x)$.

2.2 Local Linear Updates

One major benefit of vision-language models is their generalization capability. As they are usually trained on immense datasets, they tend to show great zero-shot capabilities even on unseen domains. However, they are not necessarily well-tuned for small specific datasets that were not well represented in the original training data.

Although we could fine-tune the networks on this specific set, this would come at the cost of the model's ability to generalize, and would be quite expensive. Instead, we add additional functions on the output of the model $g_I \circ f_I$ and $g_T \circ f_T$ respectively and optimize only those. This is much cheaper, as g has many fewer parameters (we will omit the subscript for both f and g if both the image and textual models are meant). Furthermore, as g is applied onto the output of f, we do not even have to back-propagate through the big models. We could even precompute f on the given dataset to save further computation cost. In our case, we use distinct networks for g_I and g_T, but it is possible to use the same function as well ($g_I = g_T$), as they are applied on the same domain.

The training works slightly differently than before. Since we now do not have unique image/text pairs, but instead images with their classes, we leave out one half of the loss, leading to the classical cross-entropy loss:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log(p(y_i|x_i))$$

A similar idea was presented in CLIP Adapter (Gao et al. 2021), where only g_l is used. This approach significantly reduces the cost of fine-tuning, but does not solve the problem of losing capabilities on the previous training dataset while overfitting on the new one.

Local Interpolation: To lessen the overfitting, WISE-FT (Wortsman et al. 2022) interpolates the weights between the original model and a fine-tuned version. The results of f and g are interpolated with a similar aim in CLIP Adapter (Gao et al. 2021):

$$\alpha(g \circ f) + (1-\alpha)f$$

Here, α is a global parameter. It would be more sensible to localize this interpolation to the area of the feature space, where we obtained new data. Only there do we actually have information on how to sensibly update the embeddings. As g is a global function but we only supervise it at our training samples, it is unlikely that it represents a sensible modification away from these samples. Thus, in our case α is not a global parameter, but a function

$$\alpha(x, D) = \beta \cdot \max_{d \in D}(\exp(-\gamma(1 - x^T d)))$$

where D is the set of datapoints we fine-tuned on and β is a global parameter similar to how α was defined previously. γ concentrates the focus of the interpolation mask and thus influences in what range our updated embedding should be applied. Were we to let it go towards zero, we would have a global α parameter, similar to CLIP Adapter. On the other hand, were we to let it go towards infinity, we would update only exactly the images and classes on which we fine-tuned. Whenever an image is close to one already seen during training, we thus use our updated features; otherwise, we utilize the general knowledge of the pre-trained model. Note that we do this separately for the text and image encoder, so there are separate sets D_{text} and D_{image}.

Clustering: This approach requires us to save the feature vectors of all datapoints seen during fine-tuning. As long as the dataset used is indeed very small, this is not a problem. If this is not the case, however, we cluster the feature vectors to find sensible representatives. For this we use agglomerative clustering, where we start by regarding each datapoint as an individual cluster, and then iteratively merge pairs based on the maximum distance between their members until we have reached the desired number of clusters. Each cluster needs a position, which is computed as the (normalized) mean of all its members.

We choose this approach, as it does not make any assumptions about cluster shapes, nor does it require a sensible initialization. Furthermore, we expect the number of clusters to be in an order of magnitude similar to the number of datapoints, so we do not need many merge operations.

Identity Regularization: So far, we have restricted the region where we change the feature space, but not the magnitude of the update, which can become arbitrarily large. It is desirable, however, that the update be as small as possible, while minimizing

the training loss. As we assume the original pre-trained features to already be useful, we want to stay as close to them as possible to retain generality. Furthermore, the interpolation between f and $g \circ f$ should result in sensible embeddings, which is more likely to be the case if they are close to each other. In other words, g should stay as close to the identity as possible.

This is easy to enforce if we simply choose g as an affine function $g = Wx + b$. In this case, our regularization takes the form

$$\lambda(||W - I||_2 + ||b||_2)$$

where I is the identity matrix and λ a weighting parameter. Of course, we could choose a more complex function and regularize g to stay close to the identity at a set of sample points. However, a dense sampling of the embedding space is infeasible, and we are interested in retaining this property wherever the interpolation weight α is non-zero. Furthermore, we initialize g as the identity function, which is trivial for affine functions, but not for non-linear MLPs.

3 Evaluation

For our evaluation, we follow CoCoOp (Zhou et al. 2022a), where three problem settings are investigated:

1. Generalization to new classes within a given dataset.
2. Generalization to new datasets after fine-tuning.
3. Generalization to domain-shift.

Before presenting the conclusions, we will introduce the datasets used, and explain the training procedure.

Datasets: Similar to CoCoOp, we follow CoOp (Zhou et al. 2022b) in the choice of datasets used in evaluation. To be precise, we use 11 datasets that cover a wide range of tasks: ImageNet (Fei-Fei et al. 2009) and Caltech101 (Fei-Fei, Fergus, and Perona 2004) for generic object classification; OxfordPets (Parkhi et al. 2012), StanfordCars (Krause et al. 2013), Flowers102 (Nilsback and Zisserman 2008), Food101 (Bossard, Guillaumin, and Gool 2014), and FGVCAircraft for more specific object classification (Maji et al. 2013); SUN397 (Xiao et al. 2010), DTD (Cimpoi et al. 2014), EuroSAT (Helber et al. 2019), and UCF101 (Soomro, Zamir, and Shah 2012) for a diverse set of tasks. Furthermore, to evaluate domain generalization, we regard ImageNet as source and four different versions under different types of domain shift as target. The four datasets are: ImageNetV2 (Recht et al. 2019), ImageNet-Sketch (Wang et al. 2019), ImageNet-A (Hendrycks, Zhao, et al. 2021) and ImageNet-R (Hendrycks, Basart, et al. 2021).

The set of images for few-shot training is randomly sampled for each dataset, while using the original test set for testing. We average the results over three runs for approaches that need training.

Training: Our implementation is based on the published code of CoOp. We use the same learning rate and number of epochs as they do. Following CoCoOp we use ViT-B/16 as the vision backbone of CLIP. As ProGrad has been evaluated on a different backbone, we retrained it for a fair comparison. Note that both CoOp and CoCoOp have a context length of 4 initialized as the prompt: 'a photo of a,' whereas for CLIP Adapter and us the context is class-dependent; ProGrad has a context length of 16 with a class-dependent initialization. If not otherwise stated, we choose the parameters of our approach as $\beta = 0.5$, $\gamma = 20$, $\lambda = 1e3$ and the number of clusters as 512.

3.1 Base to New Generalization

On each dataset, the classes are split equally into a set of 'base' classes on which the adapter is trained and unseen 'new' classes, where we only evaluate. Thus, no matter how many shots are given for the training, on the new classes we will always do zero shot inference. We show results for different numbers of shots in Figure 2* and report exact values in Table 1. Here we also give the harmonic mean between the evaluation on base and new classes to compare respective trade-offs of the approaches more easily.

As can be seen, on the 16-shot evaluation our approach outperforms all other methods on 8 out of 11 datasets, when regarding the harmonic mean. Here we have on average an improvement of almost 3 percentage points over CoCoOp (the next best method). Furthermore, (on average) our localized adapter beats all other methods regarding new classes independent of the number of shots, and is first or second on the base classes.

Our method is the only one that reaches the performance of CLIP when it comes to zero-shot performance on unseen classes. All other methods show a drop in performance here that usually increases with the number of shots, hinting at overfitting.

3.2 Cross-Dataset Generalization

In this experiment (Table 2) the models are fine-tuned on ImageNet and then evaluated on the other datasets; thus, an improvement on ImageNet (compared to CLIP) is expected. Interestingly, both CoCoOp and our approach show an improvement (on average) on the other datasets as well. Apparently, the training samples of ImageNet are numerous and diverse enough to avoid overfitting, and the data distribution of ImageNet is closer to the other datasets regarded than CLIP's original training set. Although our method does not reach the results of CoCoOp on this evaluation, we come very close.

3.3 Domain Generalization

In this last comparison (Table 3), the models are again fine-tuned on ImageNet and then evaluated on different versions with a clear domain shift. Here we can see a slight drop of performance between our method and prompt-based approaches. This might be due to the fact that prompt-based approaches only fine-tune the input of the text encoder. As the class names and thus the text encodings are not affected by domain shift, their performance generalizes better. On the other hand, we directly update the text and image embedding (and CLIP Adapter only updates the image embedding), which might be problematic, as changes caused by the domain-shift have a more direct effect here.

3.4 Training Speed

A comparison of the training speed between our approach and prompt-based methods depends on both the batch size and the number of classes. As we can precompute the class embeddings, the training time of our method is almost independent of their number. Prompt-based approaches need instead to compute the class embeddings in every iteration. On the other hand, the number of class embeddings is independent of the batch size, whereas our adapter needs to be applied to every training sample. In Figure 3,* we show the timing for a single forward and backward pass depending on the batch size. As can be seen, our method and CLIP Adapter are consistently the fastest and the difference in their timings is negligible (it is barely possible to differentiate their lines).

4 Conclusion

As the requirements in size, data and compute for state-of-the-art AI models increases, it becomes more and more important to be able to use available pre-trained networks for complex downstream tasks. In order to do this, we need to be able to fine-tune these models in an efficient manner, preferably without losing the generalization capability that makes them so useful in the first place. We have introduced an extremely simple approach for this task, introducing small linear updates to the embedding space, localized to the datapoints, where we fine-tune. Our model is fast to train and needs a minimal amount of extra parameters, but still reaches state-of-the-art results on both fine-tuned and unseen classes.

This work was originally published at the *Efficient Deep Learning for Computer Vision CVPR Workshop 2023*. For more information and detailed, colored figures, please scan the QR code to the original paper.

* Further explanations and detailed figures can be found in the full text publication at: https://www.graphics.rwth-aachen.de/publication/03349.

	Base	New	H
CLIP	69.34	74.22	71.70
CoOp	82.69	63.22	71.66
CoCoOp	80.47	71.69	75.83
ProGrad	82.79	68.55	74.46
CLIP Ada.	82.62	70.97	76.02
LLU	**83.48**	**74.47**	**78.46**

(a) Average over 11 datasets

	Base	New	H
CLIP	72.43	68.14	70.22
CoOp	76.47	67.88	71.92
CoCoOp	75.98	**70.43**	**73.10**
ProGrad	**77.03**	68.8	72.68
CLIP Ada.	76.53	66.67	71.26
LLU	76.77	69.00	72.68

(b) ImageNet

	Base	New	H
CLIP	96.84	**94.00**	95.40
CoOp	98.00	89.81	93.73
CoCoOp	97.96	93.81	95.84
ProGrad	**98.50**	91.90	95.09
CLIP Ada.	98.20	93.20	95.63
LLU	98.17	93.93	**96.00**

(c) Caltech101

	Base	New	H
CLIP	91.17	97.26	94.12
CoOp	93.67	95.29	94.47
CoCoOp	**95.20**	**97.69**	**96.43**
ProGrad	94.40	95.10	94.75
CLIP Ada.	94.40	94.10	94.25
LLU	94.47	97.00	95.72

(d) OxfordPets

	Base	New	H
CLIP	63.37	74.89	68.65
CoOp	78.12	60.40	68.13
CoCoOp	70.49	73.59	72.01
ProGrad	79.00	67.93	73.05
CLIP Ada.	77.13	69.23	72.97
LLU	**79.27**	**75.50**	**77.34**

(e) StanfordCars

	Base	New	H
CLIP	72.08	**77.80**	74.83
CoOp	97.60	59.67	74.06
CoCoOp	94.87	71.75	81.71
ProGrad	96.27	71.07	81.77
CLIP Ada.	97.70	70.83	82.13
LLU	**97.83**	76.03	**85.57**

(f) Flowers102

	Base	New	H
CLIP	90.10	91.22	90.66
CoOp	88.33	82.26	85.19
CoCoOp	**90.70**	91.29	**90.99**
ProGrad	90.17	89.53	89.85
CLIP Ada.	90.40	90.40	90.40
LLU	90.20	**91.33**	90.76

(g) Food101

	Base	New	H
CLIP	27.19	**36.29**	31.09
CoOp	40.44	22.30	28.75
CoCoOp	33.41	23.71	27.74
ProGrad	42.63	26.97	33.04
CLIP Ada.	39.57	32.27	35.55
LLU	**43.87**	34.67	38.72

(h) FGVCAircraft

	Base	New	H
CLIP	69.36	75.35	72.23
CoOp	80.60	65.89	72.51
CoCoOp	79.74	**76.86**	78.27
ProGrad	80.70	71.03	75.56
CLIP Ada.	**81.67**	73.93	77.61
LLU	81.27	76.67	**78.90**

(i) SUN397

	Base	New	H
CLIP	53.24	59.90	56.37
CoOp	79.44	41.18	54.24
CoCoOp	77.01	56.00	64.85
ProGrad	76.70	46.67	58.03
CLIP Ada.	80.47	52.23	63.35
LLU	**80.56**	**60.63**	**69.19**

(j) DTD

	Base	New	H
CLIP	56.48	64.05	60.03
CoOp	**92.19**	54.74	68.69
CoCoOp	87.49	60.04	71.21
ProGrad	91.37	56.53	69.85
CLIP Ada.	86.93	64.20	73.86
LLU	90.33	**66.30**	**76.37**

(k) EuroSAT

	Base	New	H
CLIP	70.53	77.50	73.85
CoOp	84.69	56.05	67.46
CoCoOp	82.33	73.45	77.64
ProGrad	83.90	68.50	75.42
CLIP Ada.	85.80	73.63	79.25
LLU	**85.83**	**78.13**	**81.80**

(l) UCF101

1

	Source	Target										
	ImageNet	Caltech101	OxfordPets	StanfordCars	Flowers102	Food101	FGCVAircraft	SUN397	DTD	EuroSAT	UCF101	Average
CLIP	66.7	92.8	89.2	65.3	68.2	85.8	**24.9**	62.6	44.5	47.7	66.7	64.77
CoOp	71.51	93.70	89.14	64.51	68.71	85.30	18.47	64.15	41.92	46.39	66.55	63.88
CoCoOp	71.02	**94.43**	**90.14**	65.32	**71.88**	86.06	22.94	**67.36**	**45.73**	45.37	**68.21**	**65.74**
ProGrad	72.00	92.67	89.73	64.00	68.37	85.27	20.30	64.60	43.07	44.53	65.20	63.78
CLIP Adapter	71.77	92.17	86.47	60.50	67.63	82.53	22.90	62.77	42.23	47.67	63.37	62.82
LLU	**72.13**	92.00	89.10	**65.37**	71.23	**86.10**	24.87	64.93	44.63	**47.77**	67.10	65.31

2

	Source	Target			
	ImageNet	ImageNetV2	ImageNet-Sketch	ImageNet-A	ImageNet-R
CLIP	66.73	60.83	46.15	47.77	73.96
CoOp	71.51	64.20	47.99	49.71	75.21
CoCoOp	71.02	64.07	**48.75**	**50.63**	**76.18**
ProGrad	72.00	**64.70**	48.37	49.73	75.57
CLIP Adapter	71.77	63.97	46.27	47.80	72.10
LLU	**72.13**	64.53	47.17	48.87	74.30

3

1 Comparison in the intra-class generalization setting. All methods except for CLIP (CoOp, CoCoOp, ProGrad, CLIP Adapter and our method LLU (Localized Linear Updates)) are trained on the base classes with 16 shots. H denotes the harmonic mean. The best results are marked in bold.

2 Comparison for cross dataset generalization capability. All approaches are trained on ImageNet (16 shots) and then evaluated on all 11 datasets. The best results are marked in bold.

3 Comparison for domain generalization capability. All approaches are trained on the standard version of ImageNet (16 shots) and then evaluated on 4 different types of domain shift. The best results are marked in bold.

Bibliography

Bossard, Lukas, Matthieu Guillaumin, and Luc Gool. 2014. "Food-101–Mining Discriminative Components with Random Forests." In *European Conference on Computer Vision*, 446–61. Heidelberg: Springer.

Cimpoi, Mircea, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. 2014. "Describing Textures in the Wild." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3606–13.

Fei-Fei, Li, Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, and Kai Li. 2009. "Imagenet: A Large-Scale Hierarchical Image Database." In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–55. IEEE. doi:10.1109/CVPR.2009.5206848.

Fei-Fei, Li, Rob Fergus, and Pietro Perona. 2004. "Learning Generative Visual Models from Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object Categories." In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 178–178. IEEE. doi:10.1109/CVPR.2004.383.

Gao, Peng, Shijie Geng, Renrui Zhang, Teli Ma, Rongyao Fang, Yongfeng Zhang, Hongsheng Li, and Yu Qiao. 2021. "Clip-Adapter: Better Vision-Language Models with Feature Adapters." arXiv. https://arxiv.org/abs/2110.04544.

Goodfellow, Ian J., Mehdi Mirza, Xiao, Aaron Courville, and Yoshua Bengio. 2013. "An Empirical Investigation of Catastrophic Forgetting in Gradient-Based Neural Networks." arXiv. https://arxiv.org/abs/1312.6211.

Helber, Patrick, Benjamin Bischke, Andreas Dengel, and Damian Borth. 2019. "Eurosat: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12 (7): 2217–26. doi:10.1109/JSTARS.2019.2918242.

Hendrycks, Dan, Steven Basart, Norman Mu, Saurav Kadavath, Frank Wang, Evan Dorundo, Rahul Desai, Tyler Zhu, Samyak Parajuli, and Mike Guo. 2021. "The Many Faces of Robustness: A Critical Analysis of Out-of-Distribution Generalization." In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 8340–49.

Hendrycks, Dan, Kevin Zhao, Steven Basart, Jacob Steinhardt, and Dawn Song. 2021. "Natural Adversarial Examples." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15262–71.

Jia, Chao, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc Le, Yun-Hsuan Sung, Zhen Li, and Tom Duerig. 2021. "Scaling up Visual and Vision-Language Representation Learning with Noisy Text Supervision." In *Proceedings of Machine Learning Research*, 4904–16.

Krause, Jonathan, Michael Stark, Jia Deng, and Li Fei-Fei. 2013. "3D Object Representations for Fine-Grained Categorization." In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 554–61.

Maji, Subhransu, Esa Rahtu, Juho Kannala, Matthew Blaschko, and Andrea Vedaldi. 2013. "Fine-Grained Visual Classification of Aircraft." arXiv. https://arxiv.org/abs/1306.5151.

Nilsback, Maria-Elena, and Andrew Zisserman. 2008. "Automated Flower Classification over a Large Number of Classes." In *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, 722–29. IEEE. doi:10.1109/ICVGIP.2008.47.

Parkhi, Omkar M., Andrea Vedaldi, Andrew Zisserman, and C.V. Jawahar. 2012. "Cats and Dogs." In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 3498–3505. IEEE. doi:10.1109/CVPR.2012.6248092.

Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, and Jack Clark. 2021. "Learning Transferable Visual Models from Natural Language Supervision." In *Proceedings of Machine Learning Research*, 8748–63.

Recht, Benjamin, Rebecca Roelofs, Ludwig Schmidt, and Vaishaal Shankar. 2019. "Do Imagenet Classifiers Generalize to Imagenet?" In *Proceedings of Machine Learning Research*, 5389–5400.

Soomro, Khurram, Amir Roshan Zamir, and Mubarak Shah. 2012. "UCF101: A Dataset of 101 Human Actions Classes from Videos in the Wild." arXiv. https://arxiv.org/abs/1212.0402.

Wang, Haohan, Songwei Ge, Zachary Lipton, and Eric P. Xing. 2019. "Learning Robust Global Representations by Penalizing Local Predictive Power." *Advances in Neural Information Processing Systems* 32: 10506–18.

Wortsman, Mitchell, Gabriel Ilharco, Jong Wook Kim, Mike Li, Simon Kornblith, Rebecca Roelofs, Raphael Gontijo Lopes, Hannaneh Hajishirzi, Ali Farhadi, and Hongseok Namkoong. 2022. "Robust Fine-Tuning of Zero-Shot Models." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7959–71.

Xiao, Jianxiong, James Hays, Krista A. Ehinger, Aude Oliva, and Antonio Torralba. 2010. "Sun Database: Large-Scale Scene Recognition from Abbey to Zoo." In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3485–92. IEEE. doi:10.1109/CVPR.2010.5539970.

Zhang, Renrui, Rongyao Fang, Peng Gao, Wei Zhang, Kunchang Li, Jifeng Dai, Yu Qiao, and Hongsheng Li. 2021. "Tip-adapter: Training-free CLIP-Adapter for Better Vision-Language Modeling." arXiv. https://arxiv.org/abs/2111.03930.

Zhou, Kaiyang, Jingkang Yang, Chen Change Loy, and Ziwei Liu. 2022a. "Conditional Prompt Learning for Vision-Language Models." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16816–25.

———. 2022b. "Learning to Prompt for Vision-Language Models." *International Journal of Computer Vision* 130: 2337–48.

Zhu, Beier, Yulei Niu, Yucheng Han, Yue Wu, and Hamwang Zhang. 2022. "Prompt-Aligned Gradient for Prompt Tuning." arXiv. https://arxiv.org/abs/2205.14865.

Lokalisierte latente Updates für die Feinabstimmung von Vision-Language-Modellen

Moritz Ibing, Isaak Lim, Leif Kobbelt

Ein großer Teil des Erfolgs von Deep Learning bei Bildverarbeitungsaufgaben wie Klassifizierung, Objekterkennung oder Segmentierung ist auf immer größere Modelle zurückzuführen, die auf immer größeren Datenmengen trainiert werden.

Ein beliebter Ansatz, um die immensen Quellen unkuratierter Daten in Form von Bildern mit textuellen Beschreibungen zu nutzen, sind Vision-Language-Modelle (Jia et al. 2021; Radford et al. 2021). Hier werden sowohl Bild als auch Beschreibung individuell in einem gemeinsamen Einbettungsraum abgebildet. Diese Einbettung wird so optimiert, dass übereinstimmende Paare nahe beieinanderliegen und der Abstand zwischen allen anderen Paaren groß ist. Ein auf diese Weise trainiertes Modell kann für die Zero-Shot-Klassifikation verwendet werden, da das Sprachmodell durch die Einbettung einer textuellen Beschreibung (z. B. ‚ein Bild von [KLASSE]') jede denkbare Klasse behandeln kann.

Die Leistung dieser Modelle bei anderen Aufgaben und Datensätzen, z. B. der Beurteilung der Ähnlichkeit von Kunstwerken, kann suboptimal sein. In diesen Fällen besteht eine gängige Technik darin, das vortrainierte Modell für die jeweilige Aufgabe fein abzustimmen. Das Training des gesamten Modells ist jedoch recht kostspielig. In der Literatur gibt es zwei Lösungsvorschläge für dieses Problem. Einer davon ist das Prompt-Learning (Zhou et al. 2022a; Zhou et al. 2022b; Zhu et al. 2022), bei dem der Kontext um die Klasse (z. B. ‚ein Bild von') für einen bestimmten Datensatz optimiert und nicht von Hand erstellt wird. Die andere Option ist die Verwendung von Adaptern (Gao et al. 2021; Zhang et al. 2021), leichtgewichtigen Modellen (in der Regel kleine Multilayer Perceptrons), die die vom visuellen oder sprachlichen Modell (oder von beiden) erzeugte Einbettung ändern und so die Vorhersagen aktualisieren, ohne dass die ursprünglichen Netzwerkparameter aktualisiert werden müssen. Beide Ansätze haben jedoch das Problem, dass, obwohl die Leistung für die spezifische Domäne und Aufgabe, für die die Feinabstimmung vorgenommen wurde, verbessert wird, dies um den Preis eines Leistungsabfalls bei anderen Aufgaben/Domänen im Vergleich zum ursprünglichen Modell geschieht (Goodfellow et al. 2013).

Ziel dieser Arbeit ist es, die Vorteile der Feinabstimmung für eine bestimmte Aufgabe zu nutzen, ohne die Generalisierungsfähigkeit des ursprünglichen Modells zu verlieren. Arbeiten in dieser Richtung wurden bereits in Form von CoCoOp (Zhou et al. 2022a) durchgeführt, bei dem Prompt-Learning eingesetzt, aber der Kontext nicht auf einen bestimmten Datensatz abgestimmt wird, sondern ein geeigneter Kontext aus dem zu klassifizierenden Bild vorhergesagt wird. Ein weiterer Ansatz, der Prompt-Learning einsetzt, ist ProGrad (Zhu et al. 2022), bei dem die Aktualisierung des Kontexts eingeschränkt wird, um keine Informationen aus der Pre-Trainingsphase zu verlieren. Obwohl beide Methoden den Leistungsverlust in der Zero-Shot-Einstellung verringern, erreichen sie immer noch nicht die Fähigkeiten des ursprünglichen Modells.

Im Gegensatz dazu wählen wir eine einfache Methode, die auf Adaptern basiert. Die Idee besteht darin, die Einbettung nur dort zu aktualisieren, wo wir tatsächlich Trainingsdaten haben, und sie überall sonst unverändert zu lassen, sodass die ursprünglichen Vorhersagen des Modells erhalten bleiben, wo wir diese nicht verbessern können. Darüber hinaus wollen wir auch dort, wo wir Daten haben, die Einbettung so wenig wie möglich ändern, um eine sinnvolle Interpolation zwischen der feinabgestimmten und der ursprünglichen Einbettung zu ermöglichen. Dieser Ansatz ist sehr einfach, da wir nur eine kleine Anzahl von Parametern abstimmen müssen. Dennoch zeigen wir eine Verbesserung der Leistung im Vergleich zum bisherigen Stand der Technik.

2 Methode

Bevor wir unseren Ansatz im Detail vorstellen, geben wir einen kurzen Überblick über die Funktionsweise von Vision-Language-Modellen im Allgemeinen am Beispiel von CLIP, das in allen unseren Experimenten verwendet wird.

Vision-Language-Modelle bestehen aus zwei Netzwerken: einem Bild-Encoder f_I und einem Text-Encoder f_T. Ihre genaue Implementierung ist für uns in diesem Zusammenhang nicht von Interesse. Alles, was wir wissen müssen, ist, dass diese Modelle ein Bild bzw. einen Text in einen (normalisierten) Merkmalsvektor der gleichen Dimension einbetten. Der Kosinusabstand zwischen einem eingebetteten Bild und einem Text sollte dann ihrer Ähnlichkeit entsprechen, d. h. wie gut der Text das Bild beschreibt.

Während des Trainings erhalten wir eine Reihe von n Bildern x und deren Textbeschreibungen y. Wir gehen vereinfachend davon aus, dass jeder Text eine perfekte Beschreibung des zugehörigen Bildes ist und alle anderen Texte in keinem Zusammenhang stehen. Wir wollen also den Kosinusabstand zwischen den Beschreibungen der übereinstimmenden Bild/Text-Paare $f_I(x_i), f_T(y_i)$ minimieren und den Abstand zwischen allen anderen Paaren innerhalb einer Gruppe $f_I(x_i), f_T(y_j)$ mit $i \neq j$ maximieren.

Eine andere Sichtweise wäre, den Kosinusabstand als die Wahrscheinlichkeit zu betrachten, dass ein gegebener Text y das entsprechende Bild x beschreibt, oder umgekehrt. Wir können die normalisierten Wahrscheinlichkeiten wie folgt berechnen, wobei τ ein erlernter Temperaturparameter ist:

$$p(y|x) = \frac{\exp(f_I(x)^T f_T(y)/\tau)}{\sum_{i=1}^{n} \exp(f_I(x)^T f_T(y_i)/\tau)}$$

Da wir davon ausgehen, dass die Einbettungen normalisiert sind, ist das Skalarprodukt gleich der Kosinusähnlichkeit. Die Wahrscheinlichkeit $p(x|y)$ unterscheidet sich nur in der Normalisierung. Aus dieser Sicht ist es nun sinnvoll, die Wahrscheinlichkeit für die richtigen Paare zu maximieren, wofür wir die Cross-Entropie verwenden können. Da wir die Wahrscheinlichkeiten in beiden Richtungen maximieren wollen, ergibt sich die zu optimierende Funktion wie folgt:

$$L = -\frac{1}{N}\sum_{i=1}^{N}(\log(p(x_i|y_i)) + \log(p(y_i|x_i)))/2$$

Zero-Shot Klassifikation: Dieser Ansatz führt zu einer semantisch sinnvollen Einbettung – sowohl von Bildern als auch von Text –, die für nachgelagerte Aufgaben verwendet werden kann. Alternativ können wir sie direkt für die Zero-Shot-Klassifikation verwenden. Dazu betten wir einen Text für jede Klasse (z. B. ein Bild von [KLASSE]) in einen Merkmalsvektor $f_T(y_k)$) ein. Um dann ein bestimmtes Bild zu klassifizieren, betten wir es ein und berechnen den Abstand zwischen seinem Merkmalsvektor $f_I(x)$ und allen Klasseneinbettungen. Die Klassenwahrscheinlichkeit für eine Klasse k ist dann ähnlich wie zuvor als $p(y_k|x)$ gegeben.

2.2 Lokale lineare Updates

Ein großer Vorteil von Vision-Language-Modellen ist ihre Generalisierungsfähigkeit. Da sie in der Regel auf riesigen Datensätzen trainiert werden, neigen sie dazu, auch auf unbekannten Gebieten großartige Zero-Shot-Fähigkeiten zu zeigen. Allerdings sind sie nicht unbedingt gut auf kleine spezifische Datensätze abgestimmt, die in den ursprünglichen Trainingsdaten nicht gut repräsentiert waren. Obwohl wir die Netze auf diesen speziellen Datensatz feinabstimmen könnten, würde dies auf Kosten der Generalisie-

rungsfähigkeit des Modells gehen und ziemlich teuer sein. Stattdessen fügen wir zusätzliche Funktionen am Ausgang des Modells $g_I \circ f_I$ bzw. $g_T \circ f_T$ hinzu und optimieren nur diese. Das ist viel günstiger, da g viel weniger Parameter hat. Wir lassen den Index für f und g weg, wenn sowohl das Bild- als auch das Textmodell gemeint sind. Da g nur auf die Ausgabe von f angewandt wird, müssen wir außerdem nicht Gradienten für das große Modell berechnen. Wir können sogar f auf dem gegebenen Datensatz vorberechnen, um weitere Rechenkosten zu sparen. In unserem Fall verwenden wir unterschiedliche Netze für g_I und g_T, aber es ist auch möglich, dasselbe Netz zu verwenden ($g_I = g_T$), da sie auf dieselbe Domäne angewendet werden.

Das Training funktioniert etwas anders als zuvor. Da wir jetzt keine eindeutigen Bild-/Text-Paare haben, sondern Bilder mit ihren Klassen, lassen wir die eine Hälfte der Kosten-Funktion weg, was zur klassischen Cross-Entropie-Kosten-Funktion führt:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log(p(y_i|x_i))$$

Eine ähnliche Idee wurde in CLIP Adapter (Gao et al. 2021) vorgestellt (hier wurde nur g_I verwendet). Dieser Ansatz reduziert die Kosten für die Feinabstimmung erheblich, löst aber nicht das Problem, dass die Fähigkeiten auf dem ursprünglichen Trainingsdatensatz nachlassen, während das Modell an den neuen Datensatz zu gut angepasst wird.

Lokale Interpolation: Um die Überanpassung zu verringern, interpoliert WISE-FT (Wortsman et al. 2022) die Gewichte zwischen dem ursprünglichen Modell und einer feinabgestimmten Version. Mit einem ähnlichen Ziel werden in CLIP Adapter (Gao et al. 2021) die Ergebnisse von f und g interpoliert:

$$\alpha(g \circ f) + (1-\alpha)f$$

Dabei ist α ein globaler Parameter. Es wäre sinnvoller, diese Interpolation auf den Bereich des Merkmalsraums zu beschränken, in dem wir neue Daten erhalten haben. Nur dort haben wir tatsächlich Informationen darüber, wie wir die Einbettungen sinnvoll aktualisieren können. Da g eine globale Funktion ist, wir sie aber nur an wenigen Datenpunkten trainieren können, ist es unwahrscheinlich, dass sie eine sinnvolle Änderung außerhalb dieser Datenpunkte darstellt. In unserem Fall ist α also kein globaler Parameter, aber die folgende Funktion, wobei D die Menge der Datenpunkte ist, für die wir eine Feinabstimmung vorgenommen haben, und β ein globaler Parameter, ähnlich wie α zuvor definiert wurde:

$$\alpha(x, D) = \beta \cdot \max_{d \in D}(\exp(-\gamma(1 - x^T d)))$$

γ konzentriert den Fokus der Interpolationsmaske und beeinflusst somit, in welchem Bereich unsere aktualisierte Einbettung angewendet werden sollte. Wenn ein Bild nahe an einem bereits beim Training gesehenen Bild liegt, verwenden wir also unsere aktualisierten Merkmale, andernfalls nutzen wir das allgemeine Wissen des vortrainierten

Modells. Es ist zu beachten, dass wir dies getrennt für den Text- und den Bild-Encoder tun, sodass es separate Datensätze D_{text} und D_{image} gibt.

Clustering: Dieser Ansatz erfordert die Speicherung der Merkmalsvektoren aller Datenpunkte, die während der Feinabstimmung gesehen wurden. Dies stellt kein Problem dar, solange der verwendete Datensatz tatsächlich sehr klein ist. Ist dies jedoch nicht der Fall, clustern wir die Merkmalsvektoren, um sinnvolle Repräsentanten zu finden. Dazu verwenden wir das agglomerative Clustering, bei dem wir zunächst jeden Datenpunkt als einen einzelnen Cluster betrachten und dann iterativ Paare auf der Grundlage des maximalen Abstands zwischen ihren Mitgliedern zusammenführen, bis wir die gewünschte Anzahl von Clustern erreicht haben. Jeder Cluster benötigt eine Position, die als (normalisierter) Mittelwert aller seiner Mitglieder berechnet wird.

Identitätsregularisierung: Bisher haben wir den Bereich, in dem wir den Merkmalsraum verändern, eingeschränkt, nicht aber die Größe der Aktualisierung, die beliebig groß werden kann. Es ist jedoch wünschenswert, dass die Aktualisierung so klein wie möglich ist, während die Trainings-Fehler-Funktion minimiert wird. Da wir davon ausgehen, dass die ursprünglich trainierten Merkmale bereits nützlich sind, wollen wir so nahe wie möglich an ihnen bleiben, um die Generalisierungsperformanz zu wahren. Außerdem sollte die Interpolation zwischen f und $g \circ f$ zu sinnvollen Einbettungen führen, was eher der Fall ist, wenn sie nahe beieinander liegen. Mit anderen Worten: g sollte so nahe wie möglich an der Identität bleiben.

Dies ist leicht zu erreichen, wenn wir g als affine Funktion $g = \mathrm{W}x + b$ wählen. In diesem Fall nimmt unsere Regularisierung die folgende Form an, wobei I die Identitätsmatrix ist und λ ein Gewichtungsparameter ist:

$$\lambda(||W - I||_2 + ||b||_2)$$

Für unsere Auswertung folgen wir CoCoOp (Zhou et al. 2022a), wo drei Problemstellungen untersucht werden:

1. Generalisierung auf neue Klassen innerhalb eines gegebenen Datensatzes
2. Verallgemeinerung auf neue Datensätze nach Feinabstimmung
3. Verallgemeinerung bei Domänenverschiebung

Bevor wir die Schlussfolgerungen präsentieren, werden wir die verwendeten Datensätze vorstellen und das Trainingsverfahren erklären.

Datensätze: Ähnlich wie bei CoCoOp folgen wir CoOp (Zhou et al. 2022b) bei der Auswahl der Datensätze, die für die Evaluation verwendet werden. Um genau zu sein, verwenden wir 11 Datensätze, die ein breites Spektrum an Aufgaben abdecken: ImageNet (Fei-Fei et al. 2009) und Caltech101 (Fei-Fei, Fergus und Perona 2004) für die allgemeine Objektklassifikation, OxfordPets (Parkhi et al. 2012), StanfordCars (Krause et al. 2013), Flowers102 (Nilsback und Zisserman 2008), Food101 (Bossard, Guillaumin und Gool 2014) und FGVCAircraft für die spezifischere Objektklassifikation (Maji et al. 2013), SUN397 (Xiao et al. 2010), DTD (Cimpoi et al. 2014), EuroSAT (Helber et al. 2019) und UCF101 (Soomro, Zamir und Shah 2012) für eine Vielzahl von Aufgaben. Darüber hinaus betrachten wir zur Bewertung der Domänengeneralisierung ImageNet als Quelle und vier verschiedene Versionen mit unterschiedlichen Arten von Domänenverschiebungen als Ziel. Die vier Datasets sind: ImageNetV2 (Recht et al. 2019), ImageNet-Sketch (Wang et al. 2019), ImageNet-A (Hendrycks, Zhao et al. 2021) und ImageNet-R (Hendrycks, Basart et al. 2021). Die Bilder für das Few-Shot-

Training werden für jeden Datensatz zufällig ausgewählt, während für das Testen der ursprüngliche Testsatz verwendet wird. Bei Ansätzen, die ein Training erfordern, werden die Ergebnisse über drei Durchläufe gemittelt.

Training: Unsere Implementierung basiert auf dem veröffentlichten Code von CoOp. Wir verwenden die gleiche Lernrate und Anzahl von Epochen. In Anlehnung an CoCoOp verwenden wir ViT-B/16 als Vision-Backbone von CLIP. Da ProGrad auf einem anderen Backbone evaluiert wurde, haben wir es für einen fairen Vergleich neu trainiert. Zu beachten ist, dass sowohl CoOp als auch CoCoOp eine Kontextlänge von 4 haben, die mit dem Prompt ‚a photo of a' initialisiert werden, während bei CLIP Adapter und unserer Methode der Kontext klassenabhängig ist und ProGrad eine Kontextlänge von 16 mit einer klassenabhängigen Initialisierung hat. Wenn nicht anders angegeben, wählen wir die Parameter unseres Ansatzes als $\beta = 0{,}5$, $\gamma = 20$, $\lambda = 1e3$ und die Anzahl der Cluster als 512.

3.1 Generalisierung auf neuen Klassen

Bei jedem Datensatz werden die Klassen zu gleichen Teilen in einen Satz von ‚Basisklassen', auf denen der Adapter trainiert wird, und in ‚neue' Klassen, die wir nur auswerten, aufgeteilt. Unabhängig davon, wie viele Shots für das Training abgegeben werden, werden wir bei den neuen Klassen immer eine Zero-Shot-Inferenz durchführen. Wir zeigen die Ergebnisse für verschiedene Anzahlen von Shots in Abbildung 2* und geben die genauen Werte in Tabelle 1 an.

Wie zu sehen ist, übertrifft unser Ansatz bei der Auswertung mit 16 Shots alle anderen Methoden in 8 von 11 Datensätzen, wenn man das harmonische Mittel betrachtet. Hier haben wir im Durchschnitt eine Verbesserung von fast 3 Prozentpunkten gegenüber CoCoOp (der nächstbesten Methode). Darüber hinaus schlägt unser lokalisierter Adapter (im Durchschnitt) alle anderen Methoden bei den neuen Klassen unabhängig von der Anzahl der Shots und liegt bei den Basisklassen an erster oder zweiter Stelle.

Unsere Methode ist die einzige, die die Leistung von CLIP erreicht, wenn es um die Zero-Shot-Leistung bei ungesehenen Klassen geht, während alle anderen Methoden hier einen Leistungsabfall zeigen, der in der Regel mit der Anzahl der Shots zunimmt, was auf eine Überanpassung hindeutet.

3.2 Generalisierung auf neue Datensätze

In diesem Experiment (Tabelle 2) werden die Modelle auf ImageNet feinabgestimmt und dann auf den anderen Datensätzen bewertet, sodass eine Verbesserung auf ImageNet (im Vergleich zu CLIP) erwartet wird. Interessanterweise zeigen sowohl CoCoOp als auch unser Ansatz eine Verbesserung (im Durchschnitt) auf den anderen Datensätzen. Offensichtlich sind die Trainingsbeispiele von ImageNet zahlreich und vielfältig genug, um eine Überanpassung zu vermeiden, und die Datenverteilung von ImageNet ist näher an den anderen betrachteten Datensätzen als der ursprüngliche Trainingssatz von CLIP. Obwohl unsere Methode bei dieser Auswertung nicht die Ergebnisse von CoCoOp erreicht, kommen wir ihr sehr nahe.

3.3 Generalisierung bei Domänenverschiebung

In diesem letzten Vergleich (Tabelle 3) werden die Modelle erneut auf ImageNet feinabgestimmt und dann auf verschiedenen Versionen mit einer deutlichen Verschiebung der Domäne bewertet. Hier ist ein leichter Leistungsabfall zwischen unserer Methode und prompt-basierten Ansätzen zu erkennen. Dies könnte auf die Tatsache zurückzuführen sein, dass prompt-basierte Ansätze nur die Eingabe des Text-Encoders feinabstimmen. Da die Klassennamen und damit die Textcodierungen nicht von der Domänenverschiebung betroffen sind, ist ihre Leistung besser verallgemeinerbar. Andererseits aktualisieren wir direkt die Text- und Bildeinbettung (und CLIP Adapter aktualisiert nur die Bildeinbettung), was problematisch sein könnte, da sich hier die durch die Domänenverschiebung verursachten Änderungen direkter auswirken.

* Für die detaillierten Abbildungen siehe Volltextveröffentlichung unter: https://www.graphics.rwth-aachen.de/publication/03349.

3.4 Trainingsgeschwindigkeit

Ein Vergleich der Trainingsgeschwindigkeit zwischen unserem Ansatz und prompt-basierten Methoden hängt sowohl von der Batchgröße als auch von der Anzahl der Klassen ab. Da wir die Klasseneinbettungen vorberechnen können, ist die Trainingszeit unserer Methode fast unabhängig von ihrer Anzahl. Prompt-basierte Ansätze hingegen müssen die Klasseneinbettungen in jeder Iteration berechnen. Andererseits ist die Anzahl der Klasseneinbettungen unabhängig von der Batchgröße, wohingegen unser Adapter auf jede Trainingsstichprobe angewendet werden muss. In Abbildung 3* zeigen wir die Zeiten für einen einzelnen Vorwärts- und Rückwärtsdurchlauf durch die Netzwerke in Abhängigkeit von der Batchgröße. Wie zu sehen ist, sind unsere Methode und CLIP Adapter durchweg am schnellsten und der Unterschied zwischen ihren Zeiten ist vernachlässigbar (es ist kaum möglich, ihre Linien zu unterscheiden).

4 Fazit

Da die Anforderungen an die Größe, die Daten und die Rechenleistung für moderne KI-Modelle steigen, wird es immer wichtiger, verfügbare vortrainierte Netzwerke für komplexe nachgelagerte Aufgaben verwenden zu können. Zu diesem Zweck müssen wir in der Lage sein, diese Modelle auf effiziente Weise fein abzustimmen, vorzugsweise ohne die Generalisierungsfähigkeit zu verlieren, die sie überhaupt erst so nützlich macht. Wir haben einen extrem einfachen Ansatz für diese Aufgabe entwickelt, indem wir kleine lineare Aktualisierungen des Einbettungsraums einführen, die auf die Datenpunkte beschränkt sind, an denen wir die Feinabstimmung vornehmen. Unser Modell ist schnell zu trainieren und benötigt nur eine minimale Menge an zusätzlichen Parametern, erreicht aber dennoch die besten Ergebnisse sowohl bei feinabgestimmten als auch bei ungesehenen Klassen.

Diese Arbeit wurde ursprünglich auf dem *Efficient Deep Learning for Computer Vision CVPR Workshop 2023* veröffentlicht.
Für weitere Informationen und detaillierte, farbige Abbildungen den QR-Code zum Paper scannen.

	Base	New	H
CLIP	69.34	74.22	71.70
CoOp	82.69	63.22	71.66
CoCoOp	80.47	71.69	75.83
ProGrad	82.79	68.55	74.46
CLIP Ada.	82.62	70.97	76.02
LLU	**83.48**	**74.47**	**78.46**

(a) Average over 11 datasets

	Base	New	H
CLIP	72.43	68.14	70.22
CoOp	76.47	67.88	71.92
CoCoOp	75.98	**70.43**	**73.10**
ProGrad	**77.03**	68.8	72.68
CLIP Ada.	76.53	66.67	71.26
LLU	76.77	69.00	72.68

(b) ImageNet

	Base	New	H
CLIP	96.84	**94.00**	95.40
CoOp	98.00	89.81	93.73
CoCoOp	97.96	93.81	95.84
ProGrad	**98.50**	91.90	95.09
CLIP Ada.	98.20	93.20	95.63
LLU	98.17	93.93	**96.00**

(c) Caltech101

	Base	New	H
CLIP	91.17	97.26	94.12
CoOp	93.67	95.29	94.47
CoCoOp	**95.20**	**97.69**	**96.43**
ProGrad	94.40	95.10	94.75
CLIP Ada.	94.40	94.10	94.25
LLU	94.47	97.00	95.72

(d) OxfordPets

	Base	New	H
CLIP	63.37	74.89	68.65
CoOp	78.12	60.40	68.13
CoCoOp	70.49	73.59	72.01
ProGrad	79.00	67.93	73.05
CLIP Ada.	77.13	69.23	72.97
LLU	**79.27**	**75.50**	**77.34**

(e) StanfordCars

	Base	New	H
CLIP	72.08	**77.80**	74.83
CoOp	97.60	59.67	74.06
CoCoOp	94.87	71.75	81.71
ProGrad	96.27	71.07	81.77
CLIP Ada.	97.70	70.83	82.13
LLU	**97.83**	76.03	**85.57**

(f) Flowers102

	Base	New	H
CLIP	90.10	91.22	90.66
CoOp	88.33	82.26	85.19
CoCoOp	**90.70**	91.29	**90.99**
ProGrad	90.17	89.53	89.85
CLIP Ada.	90.40	90.40	90.40
LLU	90.20	**91.33**	90.76

(g) Food101

	Base	New	H
CLIP	27.19	**36.29**	31.09
CoOp	40.44	22.30	28.75
CoCoOp	33.41	23.71	27.74
ProGrad	42.63	26.97	33.04
CLIP Ada.	39.57	32.27	35.55
LLU	**43.87**	34.67	38.72

(h) FGVCAircraft

	Base	New	H
CLIP	69.36	75.35	72.23
CoOp	80.60	65.89	72.51
CoCoOp	79.74	**76.86**	78.27
ProGrad	80.70	71.03	75.56
CLIP Ada.	**81.67**	73.93	77.61
LLU	81.27	76.67	**78.90**

(i) SUN397

	Base	New	H
CLIP	53.24	59.90	56.37
CoOp	79.44	41.18	54.24
CoCoOp	77.01	56.00	64.85
ProGrad	76.70	46.67	58.03
CLIP Ada.	80.47	52.23	63.35
LLU	**80.56**	**60.63**	**69.19**

(j) DTD

	Base	New	H
CLIP	56.48	64.05	60.03
CoOp	**92.19**	54.74	68.69
CoCoOp	87.49	60.04	71.21
ProGrad	91.37	56.53	69.85
CLIP Ada.	86.93	64.20	73.86
LLU	90.33	**66.30**	**76.37**

(k) EuroSAT

	Base	New	H
CLIP	70.53	77.50	73.85
CoOp	84.69	56.05	67.46
CoCoOp	82.33	73.45	77.64
ProGrad	83.90	68.50	75.42
CLIP Ada.	85.80	73.63	79.25
LLU	**85.83**	**78.13**	**81.80**

(l) UCF101

1

	Source	Target										
	ImageNet	Caltech101	OxfordPets	StanfordCars	Flowers102	Food101	FGCVAircraft	SUN397	DTD	EuroSAT	UCF101	Average
CLIP	66.7	92.8	89.2	65.3	68.2	85.8	**24.9**	62.6	44.5	47.7	66.7	64.77
CoOp	71.51	93.70	89.14	64.51	68.71	85.30	18.47	64.15	41.92	46.39	66.55	63.88
CoCoOp	71.02	**94.43**	**90.14**	65.32	**71.88**	86.06	22.94	**67.36**	**45.73**	45.37	**68.21**	**65.74**
ProGrad	72.00	92.67	89.73	64.00	68.37	85.27	20.30	64.60	43.07	44.53	65.20	63.78
CLIP Adapter	71.77	92.17	86.47	60.50	67.63	82.53	22.90	62.77	42.23	47.67	63.37	62.82
LLU	**72.13**	92.00	89.10	**65.37**	71.23	**86.10**	24.87	64.93	44.63	**47.77**	67.10	65.31

2

	Source	Target			
	ImageNet	ImageNetV2	ImageNet-Sketch	ImageNet-A	ImageNet-R
CLIP	66.73	60.83	46.15	47.77	73.96
CoOp	71.51	64.20	47.99	49.71	75.21
CoCoOp	71.02	64.07	**48.75**	**50.63**	**76.18**
ProGrad	72.00	**64.70**	48.37	49.73	75.57
CLIP Adapter	71.77	63.97	46.27	47.80	72.10
LLU	**72.13**	64.53	47.17	48.87	74.30

3

1 Vergleich bei der Generalisierung. Alle Methoden mit Ausnahme von CLIP (CoOp, CoCoOp, ProGrad, CLIP Adapter und unsere Methode LLU (Localized Linear Updates)) werden auf den Basisklassen mit 16 Shots trainiert. H steht für das harmonische Mittel. Die jeweils besten Ergebnisse sind fett hervorgehoben.

2 Vergleich der datensatzübergreifenden Generalisierungsfähigkeit. Alle Ansätze werden auf ImageNet (16 Shots) trainiert und dann auf allen 11 Datensätzen bewertet. Die jeweils besten Ergebnisse sind fett hervorgehoben.

3 Vergleich der Domänengeneralisierungsfähigkeit. Alle Ansätze werden auf der Standardversion von ImageNet (16 Shots) trainiert und dann auf 4 verschiedene Arten von Domänenverschiebungen bewertet. Die jeweils besten Ergebnisse sind fett hervorgehoben.

Literaturverzeichnis

Bossard, Lukas, Matthieu Guillaumin und Luc Gool. 2014. „Food-101—Mining Discriminative Components with Random Forests“. In *European Conference on Computer Vision*, 446–61. Heidelberg: Springer.

Cimpoi, Mircea, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed und Andrea Vedaldi. 2014. „Describing Textures in the Wild“. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3606–13.

Fei-Fei, Li, Jia Deng, Wei Dong, Richard Socher, Li-Jia Li und Kai Li. 2009. „Imagenet: A Large-Scale Hierarchical Image Database“. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–55. IEEE. doi:10.1109/CVPR.2009.5206848.

Fei-Fei, Li, Rob Fergus und Pietro Perona. 2004. „Learning Generative Visual Models from Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object Categories“. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 178. IEEE. doi:10.1109/CVPR.2004.383.

Gao, Peng, Shijie Geng, Renrui Zhang, Teli Ma, Rongyao Fang, Yongfeng Zhang, Hongsheng Li und Yu Qiao. 2021. „Clip-Adapter: Better Vision-Language Models with Feature Adapters“. arXiv. https://arxiv.org/abs/2110.04544.

Goodfellow, Ian J., Mehdi Mirza, Xiao, Aaron Courville und Yoshua Bengio. 2013. „An Empirical Investigation of Catastrophic Forgetting in Gradient-Based Neural Networks“. arXiv. https://arxiv.org/abs/1312.6211.

Helber, Patrick, Benjamin Bischke, Andreas Dengel und Damian Borth. 2019. „Eurosat: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification“. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12, Nr. 7: 2217–26. doi:10.1109/JSTARS.2019.2918242.

Hendrycks, Dan, Steven Basart, Norman Mu, Saurav Kada- vath, Frank Wang, Evan Dorundo, Rahul Desai, Tyler Zhu, Samyak Parajuli und Mike Guo. 2021. „The Many Faces of Robustness: A Critical Analysis of Out-of-Distribution Generalization“. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 8340–49.

Hendrycks, Dan, Kevin Zhao, Steven Basart, Jacob Steinhardt und Dawn Song. 2021. „Natural Adversarial Examples“. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15262–71.

Jia, Chao, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc Le, Yun-Hsuan Sung, Zhen Li und Tom Duerig. 2021. „Scaling up Visual and Vision-Language Representation Learning with Noisy Text Supervision“. In *Proceedings of Machine Learning Research*, 4904–16.

Krause, Jonathan, Michael Stark, Jia Deng und Li Fei-Fei. 2013. „3D Object Representations for Fine-Grained Categorization“. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 554–61.

Maji, Subhransu, Esa Rahtu, Juho Kannala, Matthew Blaschko und Andrea Vedaldi. 2013. „Fine-Grained Visual Classification of Aircraft“. arXiv. https://arxiv.org/abs/1306.5151.

Nilsback, Maria-Elena und Andrew Zisserman. 2008. „Automated Flower Classification over a Large Number of Classes“. In *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, 722–29. IEEE. doi:10.1109/ICVGIP.2008.47.

Parkhi, Omkar M., Andrea Vedaldi, Andrew Zisserman und C.V. Jawahar. 2012. „Cats and Dogs“. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 3498–3505. IEEE. doi:10.1109/CVPR.2012.6248092.

Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin und Jack Clark. 2021. „Learning Transferable Visual Models from Natural Language Supervision“. In *Proceedings of Machine Learning Research*, 8748–63.

Recht, Benjamin, Rebecca Roelofs, Ludwig Schmidt und Vaishaal Shankar. 2019. „Do Imagenet Classifiers Generalize to Imagenet?“. In *Proceedings of Machine Learning Research*, 5389–400.

Soomro, Khurram, Amir Roshan Zamir und Mubarak Shah. 2012. „UCF101: A Dataset of 101 Human Actions Classes from Videos in the Wild“. arXiv. https://arxiv.org/abs/1212.0402.

Wang, Haohan, Songwei Ge, Zachary Lipton und Eric P. Xing. 2019. „Learning Robust Global Representations by Penalizing Local Predictive Power“. *Advances in Neural Information Processing Systems* 32: 10506–18.

Wortsman, Mitchell, Gabriel Ilharco, Jong Wook Kim, Mike Li, Simon Kornblith, Rebecca Roelofs, Raphael Gontijo Lopes, Hannaneh Hajishirzi, Ali Farhadi und Hongseok Namkoong. 2022. „Robust Fine-Tuning of Zero-Shot Models“. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7959–71.

Xiao, Jianxiong, James Hays, Krista A. Ehinger, Aude Oliva und Antonio Torralba. 2010. „Sun Database: Large-Scale Scene Recognition from Abbey to Zoo“. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3485–3492. IEEE. doi:10.1109/CVPR.2010.5539970.

Zhang, Renrui, Rongyao Fang, Peng Gao, Wei Zhang, Kunchang Li, Jifeng Dai, Yu Qiao und Hongsheng Li. 2021. „Tip-Adapter: Training-Free Clip-Adapter for Better Vision-Language Modeling“. arXiv. https://arxiv.org/abs/2111.03930.

Zhou, Kaiyang, Jingkang Yang, Chen Change Loy und Ziwei Liu. 2022a. „Conditional Prompt Learning for Vision-Language Models“. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16816–25.

———. 2022b. „Learning to Prompt for Vision-Language Models“. *International Journal of Computer Vision* 130: 2337–48.

Zhu, Beier, Yulei Niu, Yucheng Han, Yue Wu und Hamwang Zhang. 2022. „Prompt-Aligned Gradient for Prompt Tuning“. arXiv. https://arxiv.org/abs/2205.14865.

In Conversation with
Nick Couldry & Ulises Mejias on “Data Colonialism”

In the interview, Nick Couldry and Ulises Mejias discuss data as an abstraction of life, and describe how data is extracted and colonially exploited. They claim: “Whether we’re talking about value or labor or subjectivity or social relations, it’s all becoming digital information that can be amassed, can be processed, and used to control not just workers in the factories. But also to control and to capitalize on people who are not working.”

Transcript of the interview with Nick Couldry & Ulises Mejias, conducted by Francis Hunger on 2021–09–21.

Hunger: To approach the topic of your 2019 book — *The Costs of Connection. How Data is Colonizing Human Life and Appropriating it for Capitalism* — I'd like to ask you for a few definitions. So, the most obvious is: What is data? Do you have a specific understanding of data that refers to other scholars or differs from them?

Couldry: Basically, data is the material that is produced by abstracting the world into categories and other forms and formats. In theory, any aspects of the world whatsoever can be converted into data, by even being a related category, and that category being activated in a database. So we are interested in all sorts of data, not just social media data, personal data, but data about nature. Any form of data is part of it, colonized in the world for data extraction.

Mejias: What does extraction mean in particular? Very precisely: our social lives. So, data is a means to abstract information from our social lives and to quantify it and use it to generate a profit. Let me just give you an example: The social graph, as you know, is what Facebook basically built its business model on, is just a network visualization or representation of all our social connections. It's basically my family, my friends, and my coworkers — all that data represented as a series of connections. It's data that can be used very specifically to target advertisements, to predict behavior, the way corporations are now using it. Basically, as an instrument to extract knowledge from this domain of our life, and to use it to generate profits for themselves.

Hunger: In your specific perspective, what is colonialism, and how have you arrived at this understanding?

Mejias: I think when we look at the definition of colonialism, basically we think of it as economic dominance at a distance. And of course, it's not just an economic phenomenon. It also of course involves politics and culture. For us, what was interesting was to look at colonialism as a global system. A global system of extractivism, based on these power differentials between the colonizers and the colonized. For us then, data colonialism is this emerging social order for the continuous extraction of data from our lives, for the exclusive purpose of generating wealth.

Couldry: There is a German word for this, a rather nice one, 'Landnahme,' the land grab. Without the land grab, there was no colonialism. It was the grabbing of the land, the discovery of the so-called New World in the late 15th century which was not known to exist. Suddenly it did exist, and it was there for the taking by Spain and Portugal. That made the idea of modern colonialism. It's that initial land grab that we're interested in. Because it's that initial stage of the unfolding of a colonial power, of appropriation, of grabbing assets that we think is going on, and is the beginning of a new phase of data colonialism, which in that sense is precisely parallel to the beginnings of the historical period of colonialism. We are looking at the parallels between the land grab that all of us have lived through in our lives when our data suddenly becomes available to corporations, and that initial period of grabbing land, animals, and bodies in the late 15th century. That's the core of the comparison.

Hunger: So, maybe we can have a third definition, because in your book you bring these three elements together: data, colonialism, and capitalism. And there are many notions of capitalism. I wonder how you're using 'capitalism' and 'capital.'

Couldry: The difference between this new phase of colonialism, deemed data colonialism, and the original phase is that capitalism now exists. In fact, historically, it came to exist through the land grab of centuries of colonialism. But of course, it can't be wished away now, it continues. So, the new colonialism takes place in that context of two or three centuries of historically developing capitalism. We broadly take on board Marx's and David Harvey's approach to capitalism. So, capital is basically value, economic value that's in motion. That is able to be circulated and translated into other forms of value which can expand through various forms of investments and production, and so on and so forth. There is a debate about whether data literally is capital. The purpose [of capitalism] is the social system that exists around the maximizing of the value of capital through its circulation, through its investment, through intensification of production, and I think in a broad sense it's clear that data is at the heart of contemporary capitalism.

Mejias: There is a concept that we develop in the book, namely the concept of 'data relations.' Because we did want to be very clear about the fact that if we limit ourselves to a framework of capitalism to try to understand what is happening with data, that framework might not be enough. And so I think historically we try to understand what is happening with information and with digital technologies by expanding this concept of labor relations. Which means that we're not looking only at what happens when we work. A lot of what we're describing happens actually when we're not working. When we are relaxing, when we are sending pictures to our friends or posting status updates on Facebook or on Twitter. We're not working. We are socializing. We are outside of work. And yet, data is being collected at those moments as well and used for the generation of wealth in equitable conditions. So, that's why we also use this concept of data relations to hint at all these other processes that are happening within capitalism, but that go beyond this traditional framework of just looking at things from the perspective of labor relations.

Hunger: One could argue that whenever data is being produced it points toward labor having been expended. But it seems to me that you are not completely agreeing with that, because you are maintaining that there is something like leisure time or non-productive time, even when data is being created.

Couldry: In thinking about the core of Marx's concept, [Michael] Postone comes up with a fundamental insight which is: the core is abstraction. It's about abstracting from the messy flow, whatever I do in my office or my workshop and extract from that a unit that I can produce and is sold for a fixed price on a market. We realized this is exactly what's going on with data. It is an act of abstracting from a flow. And so—if that's true—why wouldn't it be possible, at a later stage in history, 150 years after Marx was writing, for us to now have new types of extraction relations which are data relations? Which we enter into every time when we accept that our daily life, when we are finally not working after an extended working week, our absolutely not working is nonetheless extractable for value. It's not a labor relation. It is a data relation. It is a non-labor extractive form of relating to capitalism through abstraction.

Hunger: The question for me is now, does the data subjectivity relationship simply replace the value labor abstraction, or how do the two interact?

Couldry: This is again the double nature of the argument: We're not saying: "Everything we knew about capitalism is wrong," or equally: "Everything we knew about historical colonialism is irrelevant." We're saying they all remain extremely relevant. But at this moment in history, we have a new layer of capitalism, a new type of invention like a new type of relation that could feed into the machine of capital growth. As it does this, we have a new type of colonialism. So we both have the abstraction of value from our non-labor activities which, since the rest of our life by the definition of what labor means, there is nothing in everyday life which isn't abstracted now for value. But at the same time, we have an intensification of abstraction of data from our labor activities. So, if you're a worker in the Amazon warehouse, every single move you make will be tracked and categorized by a robot. So we're seeing both a massive intensification of the gathering of data within labor relations and a gathering of data from what [formerly] were not labor relations. But the two together make up the land grab that we call data colonialism as part of this huge intensification of capitalism at this point.

Mejias: We are basically speaking about a reconfiguration of history. A new way of trying to understand history that allows us to see that overlap, that way in which colonialism continues to inform capitalism. Yes, capitalism is no longer just about producing things, it is no longer just industrial capitalism, it's no longer just about services, informational capitalism. So, to your question about whether we are seeing data subjectivity is just replacing the value of labor, that kind of abstraction, I think that in some way it's becoming the same thing at a larger scale. Whether we're talking about value or labor or subjectivity or social relations, it's all becoming digital information that can be amassed, can be processed, and used to control not just workers in the factories, although it is used to do that as well, as Nick mentioned in the example of the Amazon warehouse. But also, to control and to capitalize on people who are not working. So, it's now outside of that domain of work.

Hunger: There is a lot of hype about notions of machine learning and artificial intelligence right now, and with *Training the Archive* we're participating in that hype and also working on pattern recognition within a machine learning project. We're developing a software prototype on the one hand, but also doing some kind of in-house reflection on it. What is the role of so-called artificial 'intelligence' or machine 'learning,' from your perspective as a colonial data practice?

Mejias: I was thinking actually of the title of Lisa Gitelman's book: "*Raw Data" is an Oxymoron* (2013). When we think about data, we think of something that exists in a pure form and then can just be analyzed for this or another purpose. But the collection and the analysis of data always implies an intention and always implies a politics. So, if we have an artificial intelligence algorithm to decide whether people are going to get coverage by insurance, those algorithms can be biased. They are trained with the data from the real world and the real world is prejudiced. So you think about those tools to make decisions about who gets insurance, who gets a loan, who gets selected for a security check, all of those decisions, of those predictions, of those ways to nudge certain kinds of behavior can in fact discriminate. And if we think about it, this is only going to get worse. I think we're now saying that in a few years, by 2025, we are going to be generating about 175 zettabytes of data a year. We're producing more and more data, but we can't analyze all of it. Humans can't look at all of it. So, that's where artificial intelligence comes in, as a way to help us make sense of all of these huge amounts of data, but in a way that's not unbiased, that's not disinterested, that always already implies an intention and a politics.

Hunger: You developed your book as a sort of a critique, so maybe it's a bit difficult for you to answer this question, but it would be interesting to hear your take on how do you think these relations can be decolonized?

Couldry: We want to see better ways of thinking about machine learning and artificial intelligence, so that human values are consulted, so that these processes are oriented to human values, not just the values of a few elite humans, but all the communities whose data is already being extracted. That would be a starting point to rethinking how data relations should operate, not just in accordance with a sort of responsible code developed in an elite institution which may never have been democratically consulted on outside that institution. But in genuine processes, in social thinking and reflection that enables data to be gathered and then used in a way that those people themselves would like to happen and continues to be accountable to those social purposes. I know that's abstract, but I think that's a starting point at least for thinking about what decolonizing might mean.

Scan the QR code below to watch the full video interview online.

Im Gespräch mit
Nick Couldry & Ulises Mejias über „Datenkolonialismus“

Gemeinsam diskutieren Nick Couldry und Ulises Mejias im Interview über Daten als Abstraktion des Lebens und beschreiben, wie Daten extrahiert und kolonial ausgebeutet werden. Sie erklären dazu: „Egal, ob wir nun über Wert, Arbeit, Subjektivität oder soziale Beziehungen sprechen, alles wird zu digitalen Informationen, die gesammelt, verarbeitet und genutzt werden können, um nicht allein die Arbeiter*innen in den Fabriken zu kontrollieren. Sondern auch, um Menschen, die nicht arbeiten, zu kontrollieren und aus ihnen Kapital zu schlagen.“

Transkript des Interviews mit Nick Couldry & Ulises Mejias, geführt von Francis Hunger am 21–09–2021

Hunger: Um uns dem Thema eures gemeinsamen Buches – *The Costs of Connection. How Data is Colonizing Human Life and Appropriating it for Capitalism* (2019) – zu nähern, würde ich euch gern um ein paar Definitionen bitten. Die offensichtlichste ist die erste: Was sind Daten? Habt ihr ein bestimmtes Verständnis von Daten, das sich auf andere Forscher*innen bezieht oder sich von ihnen abgrenzt?

Couldry: Im Grunde genommen sind Daten das Material, welches durch das Abstrahieren der Welt in Kategorien und andere Formen und Formate entsteht. Theoretisch kann also jeder beliebige Aspekt der Welt in Daten umgewandelt werden, indem man ihn in eine entsprechende Kategorie einordnet und diese Kategorie in einer Datenbank aktiviert. Wir sind also an allen Arten von Daten interessiert, nicht nur an Daten aus den sozialen Medien, persönlichen Daten, sondern auch an Daten über die Natur. Jede Form von Daten ist Teil einer Kolonisierung [der Welt], um Daten zu extrahieren.

Mejias: Was bedeutet Extraktion genau? Wir meinen damit unser soziales Leben. Daten sind also ein Mittel, um Informationen aus unserem sozialen Leben zu abstrahieren, sie zu quantifizieren und daraus einen Profit zu schlagen. Lass mich ein Beispiel geben: Der Social Graph, auf dem Facebook sein Geschäftsmodell aufgebaut hat, ist im Grunde eine Netzwerkvisualisierung aller meiner sozialen Verbindungen. Es handelt sich im Grunde um meine Familie, meine Freund*innen, meine Kolleg*innen – all diese Daten werden als eine Reihe von Verbindungen dargestellt. Es sind Daten, die ganz spezifisch für gezielte Werbung und Verhaltensvorhersagen genutzt werden können, so wie es die Unternehmen jetzt tun. Im Grunde genommen sind sie ein Instrument, um Wissen aus diesem Bereich unseres Lebens zu extrahieren und es zu nutzen, um Gewinne zu erzielen.

Hunger: Was ist in eurer Lesart Kolonialismus und wie kommt ihr zu diesem Verständnis?

Mejias: Wenn wir uns die Definition von Kolonialismus ansehen, dann denken wir im Grunde an wirtschaftliche Dominanz aus der Ferne. Und natürlich ist es nicht nur ein ökonomisches Phänomen. Es geht natürlich auch um Politik und Kultur. Für uns war es interessant, den Kolonialismus als ein globales System zu betrachten. Ein globales System der Extraktion, das auf diesen Machtunterschieden zwischen den Kolonisator*innen und den Kolonisierten beruht. Für uns ist der Datenkolonialismus diese entstehende soziale Ordnung für die fortwährende Extraktion von Daten aus unserem Leben, mit dem einzigen Zweck, Reichtum zu schaffen.

Couldry: Es gibt dafür ein deutsches Wort, ein ziemlich schönes, nämlich ‚Landnahme'. Ohne Landnahme gab es keinen Kolonialismus. Es war die Landnahme, die Entdeckung der sogenannten Neuen Welt im späten 15. Jahrhundert, von der man so nicht wusste, dass sie existiert. Plötzlich gab es sie und Spanien und Portugal konnten sie erobern. Es ist diese ursprüngliche Landnahme, für die wir uns interessieren. Denn es ist diese Anfangsphase der Entfaltung einer kolonialen Macht, der Aneignung, des Plünderns von Vermögenswerten, die wir für den Beginn einer neuen Phase des Datenkolonialismus halten, der in diesem Sinne genau parallel zu den Anfängen der historischen Periode des Kolonialismus verläuft. Wir beobachten verstärkt die Parallelen zwischen der Landnahme, die wir alle in unserem Leben erlebt haben, als unsere Daten plötzlich für Konzerne verfügbar wurden, und dieser ersten Periode des Land-, Tier- und Körperraubes im späten 15. Jahrhundert. Das ist der Kern des Vergleichs.

Hunger: Vielleicht können wir noch eine dritte Definition bekommen, denn in eurem erwähnten Buch bringt ihr diese drei Elemente zusammen: Daten, Kolonialismus und Kapitalismus. Und es gibt viele Vorstellungen von Kapitalismus. Ich frage mich, wie ihr ‚Kapitalismus' und ‚Kapital' verwendet.

Couldry: Der Unterschied zwischen dieser neuen Phase des Kolonialismus, des Datenkolonialismus, und der ursprünglichen Phase ist, dass der Kapitalismus jetzt existiert. Tatsächlich ist er durch den Landraub des jahrhundertelangen Kolonialismus entstanden. Aber natürlich kann er jetzt nicht weggewünscht werden, er geht weiter. Der neue Kolonialismus findet also in diesem Kontext von zwei oder drei Jahrhunderten des sich historisch entwickelnden Kapitalismus statt. Wir übernehmen im Großen und Ganzen den Ansatz von Karl Marx und David Harvey zum Kapitalismus. Kapital ist also grundsätzlich ein Wert, ökonomischer Wert, der in Bewegung ist. Er kann zirkulieren und in andere Formen von Wert umgewandelt werden, der sich durch verschiedene Formen von Investitionen und Produktion und so weiter ausdehnen kann. Es gibt eine Debatte darüber, ob Daten buchstäblich Kapital sind. Das zielt auf ein gesellschaftliches System ab, welches auf die Maximierung des Kapitals durch dessen Zirkulation, durch dessen Investition, und durch die Intensivierung der Produktion aus ist. Insgesamt ist klar, dass Daten das Herzstück des heutigen Kapitalismus sind.

Mejias: Es gibt ein Konzept, das wir in dem Buch entwickeln, das Konzept der Datenbeziehungen. Wir wollten uns nämlich darüber im Klaren sein, dass es nicht ausreicht, wenn wir uns nur auf einen kapitalistischen Rahmen beschränken, um zu verstehen, was mit Daten geschieht. Daher versuchen wir aus historischer Perspektive zu verstehen, was mit Informationen und digitalen Technologien geschieht, indem wir das Konzept der Arbeitsverhältnisse erweitern. Das bedeutet, dass wir nicht nur betrachten, was passiert, wenn wir arbeiten. Vieles von dem, was wir hier beschreiben, geschieht auch, wenn wir nicht arbeiten. Wenn wir Freizeit haben, wenn wir Fotos an unsere Freund*innen schicken oder die Status-Updates auf Facebook oder Twitter posten. Wir arbeiten nicht. Wir unterhalten uns. Wir befinden uns klar außerhalb der Arbeit. Und doch werden auch in diesen Momenten Daten gesammelt und für die Schaffung von Wohlstand unter kapitalisierten Bedingungen genutzt. Aus diesem Grund verwenden wir dieses Konzept der Datenverhältnisse auch, um auf all die anderen Prozesse hinzuweisen, die im Kapitalismus ablaufen, die aber über diesen traditionellen Rahmen hinausgehen, in dem man die Dinge nur aus der Perspektive der Arbeitsbeziehungen betrachtet.

Hunger: Man könnte argumentieren, dass immer dann, wenn Daten produziert werden, es ein Hinweis darauf ist, dass Arbeit geleistet wurde. Aber ich habe den Eindruck, dass ihr damit nicht ganz einverstanden seid, denn ihr schlagt eine Sichtweise vor, dass es sich um Freizeit oder so etwas wie unproduktive Zeit handelt, auch wenn Daten produziert werden.

Couldry: Wenn [Michael] Postone über den Kern des Marx'schen Konzepts nachdenkt, kommt er zu einer grundlegenden Einsicht, nämlich dass der Kern die Abstraktion ist. Es ist die Abstraktion von dem chaotischen Prozess, was auch immer ich in meinem Büro oder meiner Werkstatt tue, um daraus etwas zu extrahieren, das ich produzieren kann und das zu einem festen Preis auf dem Markt verkauft wird. Als wir das lasen, wurde uns klar, dass genau das mit den Daten passiert. Es ist ein Akt der Abstraktion eines Prozesses. Wenn das stimmt, warum sollte es dann nicht möglich sein, dass wir zu einem späteren Zeitpunkt in der Geschichte – 150 Jahre nachdem Marx geschrieben hat – jetzt neue Arten von Extraktionsbeziehungen haben, die Datenverhältnisse sind. Auf die wir uns einlassen, wenn wir jedes Mal akzeptieren, dass unser tägliches Leben, wenn wir nach einer überlangen Arbeitswoche endlich nicht mehr arbeiten, dass unser absolutes Nicht-Arbeiten dennoch als Wert extrahierbar ist. Es handelt sich nicht um ein Arbeitsverhältnis. Es ist ein Datenverhältnis. Es ist eine arbeitsfreie, extrahierende Form, ein Verhältnis zum Kapitalismus durch Abstraktion.

Hunger: Die Frage wäre dann, ob die Beziehung zwischen Daten und Subjektivität einfach die Abstraktion von Wert und Arbeit ersetzt oder wie beide zusammenwirken?

Couldry: Nun, das ist wieder die doppelte Natur des Arguments. Wir sagen nicht: „Alles, was wir über den Kapitalismus wussten, ist falsch", oder auch: „Alles, was wir über den historischen Kolonialismus wussten, ist nun irrelevant". Denn wir finden, dass dies alles äußerst relevant bleibt. Aber momentan entsteht eine neue Schichtung des Kapitalismus, eine neue Art von Erfindung, wie auch eine neue Art von Verhältnis, welches in die Maschinerie des Kapitalwachstums einfließen könnte. Ebenso haben wir eine neue Form von Kolonialismus. Wir haben also die Abstraktion von Wert aus unseren Nicht-Arbeitsaktivitäten, da der Rest unseres Lebens durch die Definition von Arbeit bestimmt ist und es zurzeit nichts im täglichen Leben gibt, das nicht zu Wert gerinnt. Aber gleichzeitig haben wir eine zunehmende Abstraktion von Daten aus unseren Arbeitstätigkeiten. Wenn du also ein*e Arbeiter*in in einem Amazon-Lagerhaus bist, wird jede einzelne Bewegung, die du machst, von einem Roboter kategorisiert, der dich wahrscheinlich beobachtet und überprüft. Wir erleben also beides, eine massive Intensivierung der Datenerfassung innerhalb solcher Arbeitsbeziehungen und eine Datenerfassung in Bereichen, die [früher] nicht zu den Arbeitsbeziehungen gehörten. Aber beide zusammen bilden die Landnahme, die wir als Datenkolonialismus bezeichnen, als Teil dieser enormen Intensivierung des Kapitalismus an diesem spezifischen Punkt.

Mejias: Wir sprechen im Grunde von einer Neukonfiguration der Geschichte. Eine neue Art und Weise, die Geschichte zu verstehen, die es uns erlaubt, die Überschneidungen zu sehen, die Art und Weise, in der der Kolonialismus weiterhin den Kapitalismus beeinflusst. Im Kapitalismus geht es nicht mehr nur darum, Dinge zu produzieren, es geht nicht mehr nur um Industriekapitalismus, es geht nicht mehr nur um Dienstleistungen, um Informationskapitalismus. Zu der Frage, ob wir sehen, dass die Datensubjektivität den Wert der Arbeit, diese Art von Abstraktion, einfach ersetzt, denke ich, dass es in gewisser Weise in einem größeren Maßstab ineinander fällt. Egal, ob wir über Wert, Arbeit, Subjektivität oder soziale Beziehungen sprechen, alles wird zu digitalen Informationen, die gesammelt, verarbeitet und genutzt werden können, um nicht allein die Arbeiter*innen in den Fabriken zu kontrollieren, obwohl sie dafür natürlich auch genutzt werden, wie Nick am Beispiel von Amazon erwähnte. Sondern auch, um Menschen, die nicht arbeiten, zu kontrollieren und aus ihnen Kapital zu schlagen. Das liegt also jetzt außerhalb des Bereichs der Arbeit.

Hunger: Im Moment gibt es einen Hype um den Begriff des maschinellen Lernens und der KI. Mit *Training the Archive* nehmen wir an diesem Trend teil und arbeiten auch mittels Mustererkennung und Computer Vision. Einerseits entwickeln wir einen Prototyp, andererseits machen wir auch eine Art interne Reflexion. Was ist die Rolle der sogenannten ‚KI' oder des maschinellen ‚Lernens' aus eurer Sicht als koloniale Datenpraxis?

Mejias: Ich dachte gerade an den Titel von Lisa Gitelmans Buch: *„Raw Data" is an Oxymoron* (2013). Wenn wir an Daten denken, denken wir an etwas, das in einer reinen Form existiert und dann einfach für diesen oder jenen Zweck analysiert werden kann. Aber die Sammlung und Analyse von Daten impliziert immer eine Absicht und eine Politik. Wenn wir also einen Algorithmus der Künstlichen Intelligenz einsetzen, um zu entscheiden, ob Menschen Versicherungsschutz erhalten, können diese Algorithmen voreingenommen sein. Sie werden mit Daten aus der realen Welt trainiert, und die reale Welt ist voreingenommen. Man denkt also, dass diese Werkzeuge, die Entscheidungen darüber treffen, wer eine Versicherung erhält, wer einen Kredit bekommt, wer für eine Sicherheitsprüfung ausgewählt wird, all diese Entscheidungen, also diese Vorhersagen und diese Möglichkeiten, bestimmte Verhaltensweisen anzustoßen, tatsächlich diskriminierend sein können. Ich glaube, wir sagen jetzt schon, dass wir in ein paar Jahren – im Jahr 2025 – etwa 175 Zettabytes an Daten pro Jahr erzeugen werden. Wir produzieren also mehr und mehr Daten, aber können sie nicht alle auswerten. Menschen können sie nicht alle durchsehen. Hier kommt die KI ins Spiel, um uns zu helfen, solche riesigen Datenmengen sinnvoll zu nutzen. Allerdings auf eine Weise, die nicht unvoreingenommen ist, die nicht uneigennützig ist, die immer schon auch eine Absicht und eine Politik impliziert.

Hunger: Euer Buch hat vor allem eine Kritik entwickelt. Deshalb ist es vielleicht ein bisschen schwierig, diese Frage zu beantworten, aber es wäre trotzdem interessant zu hören, was ihr denkt, auf welche Weise diese Beziehungen wieder dekolonisiert werden können?

Couldry: Wir wollen bessere Denkansätze für das maschinelle Lernen und die Künstliche Intelligenz, sodass menschliche Werte berücksichtigt werden, sodass sich diese Prozesse an menschlichen Werten orientieren, nicht nur an den Werten einer kleinen Elite, sondern der ganzen Gemeinschaften, deren Daten extrahiert werden. Das wäre ein Ansatzpunkt, um zu überdenken, wie Datenverhältnisse funktionieren sollten. Nicht nur als eine Art verantwortungsvoller Code, der von einer elitären Institution entwickelt wurde, der jenseits dieser Institution vielleicht nie demokratisch beraten worden ist. Sondern in echten Prozessen, in einem ge-

sellschaftlichen Denken und durch Reflexion, die es erlauben, Daten zu sammeln und dann in einer Weise zu nutzen, welche die Betroffenen selbst gerne hätten und die fortwährend für diese gesellschaftlichen Zwecke Verantwortung zeigen. Ich weiß, das ist abstrakt, aber ich denke, das ist ein Ansatzpunkt, um darüber nachzudenken, was Dekolonisierung bedeuten könnte.

Den QR-Code scannen, um das vollständige Videointerview anzusehen.

In Conversation with Elisa Giardina Papa on “The Myth of Universality, Transparency, and Truth for What Regards Emotion in Artificial Intelligence”

The interview centers on the artistic work of Elisa Giardina Papa, who investigates emotions and data as a productive force of artificial intelligence. Giardina Papa develops a critique of AI by focusing on the precarious labor conditions that occur in the production of algorithms from a feminist perspective. She records how emotional labor and care work are organized through global platforms and interact with clickworker services.

Transcript of the interview with Elisa Giardina Papa, conducted by Francis Hunger on 2021–10–14, and edited by Giardina Papa.

Hunger: Your recent works *Cleaning Emotional Data* (2021), *Labor of Sleep* (2017), and *Technologies of Care* (2016) deal with the materiality and performativity of data in general and of artificial 'intelligence' in particular. Before we go into the details of these works individually, could you describe your motivation, how and why did you get interested in the social repercussions of data and AI?

Giardina Papa: I'm interested in the material aspect of data and artificial intelligence, therefore, in what lies underneath and hides under the fancy interfaces of AI capitalism. Particularly, I'm interested in pushing the critique of data and AI towards a radical engagement with the material conditions that are supporting the development and thus the functioning of these systems. Among these material conditions my work addresses the question of labor. So I'm posing the question of what are the new forms of labor that are made invisible, but at the same time are sustaining AI and data systems. What my research is telling me is that once you open the black box of these algorithmic enchanted machines, what you find inside is the intensification of the old entanglements of capitalism and colonialism. You find a laboring multitude whose work is once more exploited or poorly paid, along the lines of the historical technique of differential dispossession and exploitation. And I guess this is what *Technologies of Care* and *Cleaning Emotional Data* are about.

Hunger: Let's talk a bit about the artistic strategy. Both data and artificial 'intelligence' are invisible objects—not even objects—rather assemblages. How did you arrive at your artistic decisions to visualize data and AI and the surrounding practices?

Giardina Papa: When I was working on *Technologies of Care,* I did some research on an app that is called The Invisible Boyfriend. This app is a human-powered chatbot that employs anonymous freelancers and writers as surrogate for a conversational algorithm. The Invisible Boyfriend app started as a chatbot-based service, but once it became clear that the clients were not falling for this algorithmically enchanted lover, The Invisible Boyfriend's founders decided to switch to use invisible human workers instead. So basically, this American startup partnered with an outsourcing company to offshore the romantic conversations to a globally dispersed micro-task workforce. In other words, the company determined that the computer architecture of deep neural networks necessary to perform a convincing emotional conversation was actually economically disadvantageous when compared to outsourcing the job to precarious workers. Now, when clients of The Invisible Boyfriend are connecting to the app they aren't connecting to a chatbot, but to a globally dispersed workforce of around 600 writers. This is the human-machine assemblage that I'm thinking about when I address artificial intelligence systems.

Hunger: Could you address a bit more how your artistic choices reflect the question of the invisibility of data and of data practices?

Giardina Papa: *Technologies of Care* is a video installation that explores the ways in which affective and care labor are being outsourced through Internet platforms and apps. And the video is based on conversations with precarious caregivers who work online. That is, digital workers who through a variety of platforms, websites, and apps provide clients with customized goods, experiences, erotic stimulation, companionship, and emotional support. The workers in the videos are all anonymous, because they asked me to be so, and that's why in the final rendering I used an almost synthetic voice for the conversation, and I also decided to portrait the workers as abstract 3D human shapes. The texture of the three-dimensional renderings in the videos echoes small details that I would notice during the conversation in their domestic environment, which is also their working environment. So, the 3D textures are coming from the surface of a pillow in the living room, or a curtain in the bedroom, or a tablecloth in the kitchen. In the structure of the video installation, I also use similar textures to create temporary work cubicles. These structures, made of precarious material, also provide an enclosing space that allows the visitor to experience the videos in a protected and intimate environment. *Technologies of Care* actually started after I spoke with a friend from Sicily—that is where I'm originally from. This friend had recently lost her job and had started to work online as a gig worker. She was doing short translations from English to Italian. And then through the same gig economy platform she was also providing private video chat services. For example, she would talk twice a week with a client from the UK in the Sicilian language or dialect to aid the sleep of the client. It was interesting for me that she was using the same platform to do data entry, translation, and private online video chats. She explained to me how the gig economy operates in relation to care labor, and then she helped me to get in contact with other workers online.

Hunger: Let's look more deeply into the work *Technologies of Care* from 2016 now. Can you first describe the installation, its materiality, the individual parts, and how they interconnect?

Giardina Papa: *Technologies of Care* traces pre-existing and current inequalities in care work, for example: the feminization of caregiving paired with this lack of recognition as wage work, its social devaluation through its proximity to intimacy, and the international division of care labor between the Global North and Global South. It traces how these pre-existing inequalities have been aggravated, dissimulated, and rendered even more invisible within digital economies. The social and economic asymmetries that have shaped care labor in the past still shape digital care labor in the present. And now these social and economic asymmetries are even more amplified by the process of anonymization, fragmentation, and abstraction that the gig economy imposes upon its workers.

Hunger: Talking about clickworker platforms—do the workers have any software or forums which allow them to get in touch with each other, or should we imagine them as isolated individuals?

Giardina Papa: I think isolation is a big part of this kind of labor. Because what the gig economy does in terms of reshaping labor is really imposing a process of anonymization, fragmentation, and abstraction of the worker. But then of course workers are always smarter than any system that is designed to exploit them. For example, I had a conversation with a worker from Greece who created fetish videos on demand via a gig economy platform. She told me how she would connect with other woman who were offering similar services to exchange information about the viability of some of the clients: which clients paid poorly or tried to exploit them, which clients complied with ethical rules that they decided for their work, and so on. They organized themselves autonomously, in a way that disregarded how the system of the gig economy platform was designed.

Hunger: *Cleaning Emotional Data* (2021) is a complex, media-rich installation with sculptural elements, embroidered texts and several videos.

Giardina Papa: The installation is composed of three sculptural elements made of light reflectors held by a light stand. The monitors in which the videos are playing, are protected by a privacy filter. The filters obscure the screen with a gold-pink gradient and make it visible only if you stand right in front of it at the same height. Therefore, the filters are gesturing to the topic of the video: the invisibility of the labor of the data cleaner. And the audience is also asked to do a little bit of labor with their body to be able to watch the videos. On the other side of the installation there are these textiles that are hanging from the same structure, but they are embroidered and printed. The embroidered and the printed parts are abstract lines of facial micro expressions used by AI to detect emotions and are juxtaposed with untranslatable emotional vernacular coming from the Sicilian dialect, my mother tongue. I'm using words such as "ricriju," "raggia," or "babbaria"—which are Sicilian words—to express emotions that are not easily or fully translatable into English, not even into Italian. I wanted to embroider these words in the textile to problematize the supposedly universal understanding of emotions that is too often uncritically embedded in AI systems. So, I wanted to use something that was utterly local and extremely untranslatable to think about what should remain untranslatable, queer, and incomputable even within AI models.

Hunger: What happens on those monitors?

Giardina Papa: While I was doing the preparatory research, I basically ended up working as a data cleaner for three months. So, what you see in the monitors is the work I performed during that period: the cleaning of data that are later used to train AI systems to recognize human emotions. I guess the question really is: when are data considered data? It seems that no matter how big Big Data are, if they are unstructured, the algorithms they use to accumulate, discern, and predict cannot but fail. And that's one of the reasons why in recent years, new forms of precarious labor emerged around AI economies. The data cleaners are contracted by so-called human-in-the-loop companies to process data that are later used to train machine vision systems. These workers label, categorize, annotate, and validate massive amounts of data. If machine learning has a data problem due to the lack of labeled data, affective computing has an even bigger data problem, because nobody is quite sure how to label emotional data in the first place. Too many of these systems are based on the myth of universality, transparency, and truth, according to which emotions are universally expressed and not historically or contextually determined, and that they can be fully revealed, made transparent, reduced, and measured through an ideal scale which can then be the ground truth to make comparisons and judgements. No opacity is allowed within this system. And this I believe is problematic, specifically when we are talking about emotions. This is what *Cleaning Emotional Data* is about.

Scan the QR code below to watch the full video interview online.

Im Gespräch mit
Elisa Giardina Papa zu „Emotionen und der Mythos von Universalität, Transparenz und Wahrheit durch Künstliche Intelligenz“

Im Zentrum des Gespräches steht die künstlerische Arbeit Elisa Giardina Papas, welche Emotionen und Daten als Produktionsverhältnis im Zuge Künstlicher Intelligenz erkundet. Giardina Papa entwirft eine KI-Kritik, indem sie aus feministischer Perspektive die prekären Arbeitsverhältnisse in den Blick nimmt, die mit der Produktion algorithmischer Modelle einhergehen. Sie zeigt auf, wie emotionale Arbeit und Care-Arbeit durch globale Vermittlungsplattformen organisiert werden und mit den Dienstleistungen von Klickarbeiter*innen zusammenfließen.

Transkript des Interviews mit Elisa Giardina Papa, geführt von Francis Hunger am 14–10–2021 und überarbeitet von Giardina Papa.

Hunger: Deine Arbeiten *Cleaning Emotional Data* (2021), *Labor of Sleep* (2017) und *Technologies of Care* (2016) beschäftigen sich mit der Materialität und Performativität von Daten im Allgemeinen und von Künstlicher ‚Intelligenz' im Besonderen. Könntest du, bevor wir im Einzelnen auf diese Arbeiten eingehen, deine Motivation beschreiben, wie und warum du dich für die sozialen Auswirkungen von Daten und KI interessierst?

Giardina Papa: Ich interessiere mich für den materiellen Aspekt von Daten und Künstlicher Intelligenz, also für das, was sich unter den glänzenden Oberflächen des KI-Kapitalismus verbirgt. Insbesondere bin ich daran interessiert, die Kritik an Daten und KI in Richtung einer radikalen Auseinandersetzung mit den materiellen Bedingungen voranzutreiben, die die Entwicklung und damit das Funktionieren dieser Systeme ermöglichen. Bezüglich dieser materiellen Bedingungen beschäftigt sich mein Werk mit der Frage der Arbeit. Ich stelle also die Frage nach den neuen Formen der Arbeit, die unsichtbar gemacht werden, aber gleichzeitig die Systeme der Künstlichen Intelligenz und die Datensysteme ermöglichen. Meine Forschung zeigt mir, dass man, sobald man die Blackbox dieser algorithmischen, verwunschenen Maschinen öffnet, darin die Intensivierung der alten Verstrickungen von Kapitalismus und Kolonialismus findet. Man findet also eine arbeitende Multitude, deren Arbeit erneut mittels der historischen Methode der differenzierten Enteignung und Ausbeutung ausgenutzt oder schlecht bezahlt wird. Und ich denke, darum geht es vor allem bei *Technologies of Care* und bei *Cleaning Emotional Data*.

Hunger: Lass uns ein wenig über die künstlerische Strategie sprechen. Sowohl Daten als auch Künstliche ‚Intelligenz' sind unsichtbare Objekte, nicht einmal Objekte, eher Assemblagen. Wie bist du zu deinen künstlerischen Entscheidungen gekommen, wie du Daten und KI und die sie umgebenden Praktiken sichtbar machst?

Giardina Papa: Als ich an *Technologies of Care* arbeitete, recherchierte ich über eine App namens The Invisible Boyfriend. Diese App ist ein von Menschen betriebener Chatbot, der anonyme Freiberufler*innen und Autor*innen als Ersatz für einen Konversationsalgorithmus einsetzt. Die Invisible-Boyfriend-App startete als chatbotbasierter Dienst. Als jedoch klar wurde, dass die Kund*innen nicht auf diesen algorithmisch verzauberten Liebhaber, den Invisible Boyfriend, hereinfielen, beschlossen die Gründer*innen, stattdessen im Hintergrund auf unsichtbare menschliche Mitarbeiter*innen umzusteigen. Das amerikanische Startup hat sich also mit einem Outsourcing-Unternehmen zusammengetan, um die pseudoromantischen Chats an weltweit verstreute Mitarbeiter*innen mittels Mikroaufgaben auszulagern. Mit anderen Worten, das Startup hat festgestellt, dass die Computerarchitektur von den tiefen neuronalen Netzen, die notwendig sind, um überzeugende emotionale Gespräche zu führen, tatsächlich wirtschaftliche Nachteile hat, verglichen mit der Auslagerung der Arbeit an prekäre Arbeiter*innen. Wenn sich also die Kund*innen von The Invisible Boyfriend mit der App verbinden, interagieren sie nicht mit einem Bot, sondern mit einer weltweit verstreuten Belegschaft von etwa 600 Autor*innen. Es handelt sich also um eine Mensch-Maschine-Assemblage, über die ich nachdenke, wenn ich mich mit Systemen der Künstlichen Intelligenz befasse.

Hunger: Könntest du darauf eingehen, wie deine künstlerischen Entscheidungen die Frage der Unsichtbarkeit von Daten und Datenpraktiken reflektieren?

Giardina Papa: *Technologies of Care* ist eine Videoinstallation, die die Art und Weise erforscht, in der die affektive und Care-Arbeit durch Internetplattformen und Apps ausgelagert wird. Das Video basiert auf Gesprächen mit prekären Care-Arbeit*innen, die online arbeiten. Das sind digitale Arbeitskräfte, die über eine Vielzahl von Plattformen, Websites und Apps Kunden mit maßgeschneiderten Waren oder Erfahrungen, erotischer Stimulation, Begleitung und emotionaler Unterstützung versorgen. Die Arbeiter*innen in den Videos sind alle anonym, weil sie mich darum gebeten haben, und deshalb habe ich in meiner Darstellung eine fast synthetische Stimme für die Konversation verwendet und beschlossen, die Arbeiter*innen als abstrakte menschliche 3D-Formen zu porträtieren. Die Textur dieser dreidimensionalen Darstellung in den Videos spiegelt kleine Details wider, die mir während der Gespräche in ihrer häuslichen Umgebung, die auch ihre Arbeitsumgebung ist, aufgefallen sind. Die 3D-Texturen stammen also von der Oberfläche eines Kissens im Wohnzimmer, eines Vorhangs im Schlafzimmer oder eines Tischtuchs in der Küche. In der Struktur der Videoinstallation nutze ich ähnliche Texturen für temporäre Arbeitsplätze. Diese Strukturen, die aus prekärem Material bestehen, schaffen einen umschließenden Raum, der es den Besucher*innen ermöglicht, das Kunstwerk so in einer geschützten und intimen Umgebung zu erleben. *Technologies of Care* begann, nachdem ich mit einer Freundin aus Sizilien gesprochen hatte, woher ich ursprünglich stamme. Diese Freundin hatte kürzlich ihren Job verloren und begann dann, online als Gig-Workerin zu arbeiten. Sie machte kurze Übersetzungen vom Englischen ins Italienische. Und dann bot sie über dieselbe Gig-Economy-Plattform auch private Videochatdienste an. Sie sprach zum Beispiel zweimal pro Woche mit einem Kunden aus dem Vereinigten Königreich in sizilianischer Sprache oder im sizilianischen Dialekt, um dem Kunden beim Einschlafen zu helfen. Interessant war für mich, dass sie die gleiche Plattform für die Dateneingabe und die Übersetzung und auch für private Online-Videochats verwendete. Sie hat mir also erklärt, wie die Gig-Economy auch in Bezug auf die Care-Arbeit funktioniert, und dann half sie mir, mit anderen Arbeiter*innen in Kontakt zu treten.

Hunger: Lass uns noch etwas tiefer in die Arbeit *Technologies of Care* von 2016 einsteigen. Kannst du zunächst die Installation beschreiben, ihre Materialität, die einzelnen Teile und wie sie miteinander verbunden sind?

Giardina Papa: *Technologies of Care* zeichnet die bereits bestehende Ungleichheiten in der Pflegearbeit nach, zum Beispiel die Feminisierung der Pflege, gepaart mit der fehlenden Anerkennung als Lohnarbeiter*innen, und auch ihre soziale Abwertung durch ihre Nähe zur Intimität, und die internationale Arbeitsteilung von Pflegearbeit zwischen dem Globalen Norden und dem Globalen Süden. Die Arbeit zeigt also, wie diese bereits bestehenden Ungleichheiten sowohl verschärft, verdeckt, als auch durch die digitale Wirtschaft noch unsichtbarer gemacht wurden. Die sozialen und wirtschaftlichen Asymmetrien, die die Pflegearbeit in der Vergangenheit geprägt haben, prägen meiner Meinung nach auch die digitale Pflegearbeit in der Gegenwart. Und jetzt werden diese sozialen und wirtschaftlichen Asymmetrien durch den Prozess der Anonymisierung, Fragmentierung und Abstraktion, den die Gig-Economy ihren Arbeiter*innen auferlegt, sogar noch verstärkt.

Hunger: Wenn wir über die Clickwork-Plattformen sprechen, gibt es da Software oder Foren, die es den Arbeiter*innen ermöglichen, in Kontakt zu treten, oder müssen wir sie uns als isolierte Individuen vorstellen?

Giardina Papa: Ich denke, dass Isolation ein wichtiger Bestandteil dieser Art von Arbeit ist. Denn was die Gig-Economy in Bezug auf die Neuformierung der Arbeit bewirkt, ist ein Prozess der Anonymisierung, Fragmentierung und Abstraktion der Arbeitenden. Aber natürlich sind die Arbeitnehmer immer schlauer als jedes System, das sie ausnutzen soll. Ich hatte zum Beispiel ein Gespräch mit einer Arbeiterin aus Griechenland, die Fetisch-Videos ‚on Demand' auf einer Plattform der Gig-Economy produzierte. Sie erzählte mir z. B., wie sie mit anderen Frauen, die einen ähnlichen Service anboten, in Kontakt trat, und sie tauschten untereinander Informationen über die Zahlungsfähigkeit einiger Kund*innen aus, welche Kund*innen nicht genug zahlten, wer sich ausbeuterisch verhielt, welche Kund*innen sich an die ethischen Regeln hielten, die sie für ihre Arbeit beschlossen hatten. Aber das passierte nur, weil sie sich autonom organisierten, auf eine Weise, die gewissermaßen die Art, wie das System der Gig-Economy-Plattform gestaltet war, ignorierte.

Hunger: *Cleaning Emotional Data* von 2021 ist eine komplexe, medienreiche Installation mit skulpturalen Elementen, gestickten Texten und mehreren Videos.

Giardina Papa: Die Installation besteht aus drei skulpturalen Elementen, die aus Lichtreflektoren bestehen und von einem Lichtständer gehalten werden. Die Monitore, auf denen die Videos abgespielt werden, sind durch einen Sichtschutzfilter geschützt. Im Grunde genommen verdunkeln diese Filter den Bildschirm mit einem gold-rosafarbenen Farbverlauf und machen ihn nur sichtbar, wenn man direkt davor auf gleicher Höhe steht. Die Filter sind also auf das Thema des Videos abgestimmt, nämlich die Unsichtbarkeit der Arbeit von Datenreiniger*innen (engl.: ‚Data Cleaner'). Und so wird das Publikum auch aufgefordert, ein wenig mit ihren Körpern zu arbeiten, um die Videos sehen zu können. Auf der anderen Seite des Gerüsts hängen diese Textilien an der gleichen Struktur, aber sie sind bestickt und bedruckt. Im Druck und in den Stickereien werden die abstrakten Linien eines Gesichtsausdrucks, der zur Erkennung von Emotionen erfasst wird, der unübersetzbaren emotionalen Umgangssprache aus dem sizilianischen Dialekt, meiner Muttersprache, gegenübergestellt. Ich verwende also Wörter wie „ricriju", „raggia" oder „babbaria", also sizilianische Wörter, die Emotionen ausdrücken, die sich nicht leicht oder gar nicht ins Englische, ja nicht einmal ins Italienische übersetzen lassen. Ich wollte diese Wörter in das Textil einarbeiten, um diese vermeintlich universelle Erkennbarkeit von Gefühlen zu problematisieren, die allzu oft unkritisch in künstliche intellektuelle Systeme eingebettet ist. Ich wollte also etwas verwenden, das lokal und extrem unübersetzbar ist, um darüber nachzudenken, was unübersetzbar, queer und unberechenbar bleiben sollte, selbst bei Systemen Künstlicher Intelligenz.

Hunger: Was passiert auf diesen Monitoren?

Giardina Papa: Während ich diese vorbereitende Forschung betrieb, arbeitete ich für drei Monate als Datenreinigerin. Was du auf den Monitoren siehst, ist im Grunde die Arbeit, die ich während dieser drei Monate geleistet habe, nämlich das Bereinigen von jenen Daten, die später zum Trainieren von KI-Systemen zur Erkennung menschlicher Emotionen verwendet werden. Ich denke, die Frage ist wirklich, wann Daten als Daten gelten. Denn es scheint, dass, egal wie groß Big Data ist, wenn sie unstrukturiert sind, die Algorithmen, die verwendet werden, um zu akkumulieren, zu unterscheiden, und vorherzusagen, nur scheitern können. Und das ist einer der Gründe, warum ich glaube, dass in den letzten Jahren neue Formen von prekärer Arbeit rund um die Ökonomie der Künstlichen Intelligenz entstanden sind. Die Data Cleaner, die Datenreiniger*innen, werden von sogenannten Human-in-the-Loop-Unternehmen beauftragt, Daten zu verarbeiten, die später zum Trainieren von maschinellen Bildverarbeitungssystemen verwendet werden. Im Grunde genommen beschriften, kategorisieren, kommentieren und validieren diese Arbeiter*innen riesige Datenmengen. Wenn das maschinelle Lernen bereits ein Datenproblem hat, weil es an sinnvoll gelabelten Daten mangelt, dann hat die Informatik ein noch größeres Problem, weil niemand so recht weiß, wie man emotionale Daten überhaupt labeln soll. Zu viele dieser Systeme basieren auf dem Mythos der Universalität, der Transparenz und der Wahrheit, demzufolge Emotionen universell sind, universell ausgedrückt werden und nicht historisch oder kontextuell bedingt sind, und darauf, dass Emotionen vollständig offengelegt, transparent gemacht und auf eine ideale Skala reduziert werden können, welche so zur Grundlage für Vergleiche und Entscheidungen wird. In dieses System ist keine Opazität eingebaut. Und das halte ich für problematisch, gerade wenn wir von Emotionen sprechen. Das ist die Kernbotschaft von *Cleaning Emotional Data*.

Den QR-Code scannen, um das vollständige Videointerview anzusehen.

In Conversation with
Mar Hicks on “The Politics of Artificial Intelligence and Algorithmic Bias”

This interview with Mar Hicks puts current technological developments in the field of artifical intelligence in perspective with the history of computing. It discusses the gendered power structures behind computing — not just as technology, but also as cultural technique.

Transcript of the interview with Mar Hicks, conducted by Francis Hunger on 2022–05–20.

Hunger: In your recent text *Sexism Is a Feature, Not a Bug* (2021) you observe and I'm quoting: "Electronic computing technology has long been an abstraction of power into machine form." Could you dissect these concepts a bit for us: What is power, what is the computer as a machine, and how does the abstraction of power emerge?

Hicks: Computers or technologies, they are tools that people create in order to do something more efficiently, or to do something that could not be done before. And computers in particular, by nature of being tools, but also because of their history of funding from military interests and government interests, they have been designed as tools that help certain people and groups wield power. Now sometimes that power is in the form of information. It's not the hard power of, say, guns and boats and armies. When we look at a new technology, especially something that is very expensive and cutting-edge, we see that the people who are funding that technology and going to have access to using that technology to execute their plans and their desires, they're going to be people and institutions that tend to have money, power, and privilege already. Today, if you look at how, for instance, corporations design many communications technologies and deploy them, there is a lot of rhetoric about how these might be democratizing technologies or potentially could be used for this or that social good, but we often see that it's rhetoric and an unfulfilled potential. To clarify it a bit: power isn't just political power, of course, power is expressed any time you have somebody in a position to exert their will over other people or over systems and machines. Let me discuss an example where a machine was creating difference in real-world outcomes. The first two Colossus computers materially changed the course of World War Two because the speed at which they were able to decode information was so much faster than the electromechanical machines that had been built immediately before. And so Bletchley Park, where Colossus was developed in 1943, was instrumental in concentrating and wielding power through these computing devices, so that the Allies could make different and more advantageous warfaring decisions. You know the second Colossus, in particular, which went into operation a week or so before the D-Day landings, materially changed when and how those landings went. There is more computing power in our phones today, for instance, than in one Colossus computer, but those were very powerful machines, technologically and politically, at the time.

Hunger: In another part of your text, you observe early computing in Britain as a feminized and devaluated field. Could you also tap into this historically, why do 'feminized' and 'devaluated' go hand in hand?

Hicks: In the very early history of computing going from the electromechanical to the early electronic age, what we see is that this very technical work is seen—on both sides of the Atlantic—when it comes to things like early programming and even construction, manufacturing or troubleshooting of hardware, as just something that people who are almost functionaries can do or should be able to learn how to do. And it's seen that way despite the fact that in many cases the people doing the work had or were required to have advanced math training, and certain sets of skills that were not just normal or usual things that anybody would have. Simultaneously, it's straddling these two spheres of being seen as not important, but not unskilled. And women, I should say, primarily white women and primarily middle-class white women, are sort of seen as the perfect labor force for work like this, because they're good enough, but not too good. Whenever there is work that does not yet have an established track record and therefore career progression, you will see men who are privileged enough to avoid it and go into more stable work with a clear career progression, and women for whom this kind of work is made available. They are recruited into these jobs, because the idea has been for a very long time and unfortunately, I think still is to a certain extent, that women don't need a career. They, supposedly, don't actually need money to support themselves in the same way that a middle-class white man might, and so they can do work and it's sort of a nice extra and then they leave and go onto the real work of their lives—which is raising a family. And this means there is built-in labor turnover. The social contract determines what work is seen as skilled or not, and that will change depending on who goes into the work, so that the feminization of a field can drag down wages. And that's one of the reasons that the history of computing is so interesting to me, because usually the introduction of more machinery into a field results in more labor feminization and depression of wages. In the history of computing in the middle of the 20th century, we actually see the reverse: we see professionalization and masculinization of the field occurring.

Hunger: If you take this thought further, today's invisible devaluated workforce as clickworkers who create training data for artificial intelligence projects, what parallels do you see?

Hicks: That's a great question, because inherent in that question is both the historical comparison and this unspoken idea on the part of powerful interest management classes, who deploy more and more computing tools, that there should be, in the long run, a diminution of need for human workers. That if you just design the technology right, if you design it well enough, if you train the AI models well enough, then you can basically get rid of the need for people to do an awful lot of jobs that are out there. And that analogy that you used to clickworkers and I might also extend it to gig workers as well, it's running in these very well-worn path of introduction of a technology that has the potential to, let's say, make something better. That technology being deployed in ways that does not necessarily deliver on the original promises, but does something else. Maybe it holds down costs in a certain sector. Maybe it causes costs to rise, but it breaks the power of labor unions in that sector, or maybe it's just, in fact, a way to manipulate stock markets for a short period of time so a lot of people can make a lot of money before a tech bubble, for instance, bursts or crashes. In the case of the gig work economy, one example is ride-sharing apps. Enormous sums of money were poured into ride-sharing apps by venture capitalists on the assumption that the long-term goal would be to essentially automate delivery driving, make it so that taxis drove themselves. And we see now that it's probably not going to happen. But in the meantime, the sort of labor arbitrage that those apps engendered has broken the power of taxi unions, it's made it impossible for many people who used those apps early on to make a somewhat livable wage to continue to make a livable wage, and it's made so that things like congestion and, worse, pollution in city centers have just gone up and up and up, since the idea of a gig app for ride-sharing is that there will be constant availability. So, those cars are constantly circling the same city blocks using gas, outputting CO2, and so on.

Hunger: The prominent case of Timnit Gebru comes to my mind who tried to install an ethical AI position within Google and was ultimately dismissed in 2020. How did you, on the background of what we were talking about just now, perceive this case?

Hicks: That is such a fascinating case and a really distressing flashpoint in the debates around artificial intelligence and ethics in the field of machine learning, because as you point out, Dr. Gebru was hired specifically to be an AI ethicist. She has a robust training in computer science and is a very technical person and was also going to bring this element that is seen as additional, but really should be integral to these fields, to Google's AI projects. And she was effectively pushed out of the company, fired for doing the job that she had been hired for. She raised some concerns about the way that large language models might have unintended consequences and about the scale of harmful effects in the ways that they worked. So, not just how they might be deployed to create harm, but in fact, just the model itself as it's scaled up, it tended to create echo chambers of certain kinds and to do other things that were not necessarily going to lead to positive outcomes. And in that 2021 research paper *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* she and her co-authors were not being alarmist and they presented, I think, very measured and sound findings. However, the paper was effectively suppressed. Google did not even want criticism of the relatively light level in that paper to be out in the world. That was really what caused this massive schism, not just between Dr. Gebru and Google, but within Google and within the AI machine learning community as a whole. And that schism was also caused by the question: Can we criticize this gravy train? If we're getting paid by all of the money that is being poured into artificial intelligence and machine learning, does that mean that critique is off the table? And so, I think that this incident is in some ways a perfect encapsulation of how far you can go working within the logic of a particular technological system before that logic starts to break down and eat the system itself.

Hunger: Your current research is about transgender identity and computing, which you address in the paper *Hacking the Cis-tem* (2019); you describe how trans persons clashed with computer systems that would not allow them to ascribe a transgender status. And I would like you to dive a bit into this question — and maybe this is also a provocation — whether the major problem here is the computer, or maybe the bureaucracy which just happens to use computers?

Hicks: I'll say right upfront the problem isn't the computer itself. The computer, of course, could have been programmed very easily, in fact more easily to put transgender persons into the system in that particular instance in the early 1960s, to put trans people in as their gender, as their lived identity, and it was explicitly programmed in a way so as to not do that. That was a political decision. That was a choice made by high-level civil servants in the British government for reasons that were inherently and explicitly political, and about not being willing to recognize certain people as full members of the state who deserved full rights and deserved recognition, deserved their civil rights and deserved to have the financial safety net that the British welfare state provided — I'll point out after people had paid into it, these people were paying into it their entire working lives. As much as we can use a computer to do different things, it tends to do certain things better than others. It tends to have affordances to, for instance, categorize people in certain ways. If you look at more recent computing systems, for instance, they try to categorize people and collect a lot of personal data on people to categorize them, usually so that they can make money by serving target advertisements. That does become very problematic, because when a system is designed to run on more and more sort of stylized forms of information and to chop people down into smaller and smaller categories, that those people themselves, in fact, might not agree with those forms of labels that are being applied to them. So I think you're absolutely right, it's not the technology or it's not the computer in this situation, but neither is the technology immaterial or sort of endlessly plastic.

Scan the QR code below to watch the full video interview online.

Im Gespräch mit Mar Hicks über „Künstliche Intelligenz und algorithmischer Bias als politisches Feld“

Dieses Interview mit Mar Hicks setzt die aktuellen technologischen Entwicklungen im Bereich der Künstlichen Intelligenz mit der Geschichte der Informatik in Beziehung. Es erörtert die geschlechtsspezifischen Machtstrukturen, die hinter dem Computing – nicht nur als Technologie, sondern auch als Kulturtechnik – stehen.

Transkript des Interviews mit Mar Hicks, geführt von Francis Hunger am 20–05–2022.

Hunger: In deinem jüngsten Text *Sexism Is a Feature, Not a Bug* (2021) stellst du fest, ich zitiere: „Elektronische Rechentechnik ist seit langem eine Abstraktion von Macht in Maschinenform.“ Könntest du diese Konzepte ein wenig für uns aufschlüsseln: Also was ist Macht, was ist der Computer als Maschine und wie entsteht die Abstraktion von Macht?

Hicks: Computer oder Technologien sind von Menschen geschaffene Werkzeuge, um etwas effektiver oder effizienter zu erledigen oder etwas zu tun, was vorher nicht möglich war. Und die Computer sind von Natur aus Werkzeuge, und weil sie in der Vergangenheit von militärischen Interessen und Regierungsinteressen finanziert wurden, wurden sie als Werkzeuge entwickelt, die bestimmten Leuten und Gruppen helfen, Macht auszuüben. Manchmal liegt diese Macht in Form von Informationen vor, nicht in Form von Waffen, Schiffen und Armeen. Wenn wir eine neue Technologie betrachten, besonders eine, die sehr teuer und wegbereitend ist, sehen wir, dass die Leute, die diese Technologie finanzieren und Zugang zu dieser Technologie haben werden, um ihre Pläne und Wünsche auszuführen, oft diejenigen Leute und Institutionen sind, die bereits über Geld, Macht und Privilegien verfügen. Auch heute, wenn man sich anschaut, wie Unternehmen Kommunikationstechnologien entwickeln und einsetzen, dann gibt es viele Diskussionen darüber, wie diese Technologien demokratiefördernd sein könnten oder potenziell für dieses oder jenes soziale Gut genutzt werden könnten, aber wir sehen oft, dass das nur Rhetorik und ein unerfülltes Potenzial ist. Um es ein bisschen einzuordnen: Macht ist natürlich nicht allein politische Macht, Macht kommt immer dann zum Ausdruck, wenn jemand in der Lage ist, seinen Willen über andere Menschen oder über Systeme und Maschinen auszuüben. Ich möchte auf ein sehr frühes historisches Beispiel zurückkommen. Wie du weißt, war das erste digitale elektronische Programm oder der erste Computer nicht allein eine technische Neuheit. Es handelte sich um eine Situation, in der diese Maschine in der realen Welt etwas bewirkte. Die ersten beiden Colossus-Computer veränderten den Verlauf des Zweiten Weltkriegs wesentlich, weil sie Informationen so viel schneller entschlüsseln konnten als die elektromechanischen Maschinen, die unmittelbar zuvor gebaut worden waren. Und so war Bletchley Park – wo Colossus 1943 entwickelt wurde – maßgeblich daran beteiligt, Macht durch diese Computer auszuüben, sodass die Alliierten andere und bessere Entscheidungen im Krieg treffen konnten. Und du weißt, dass insbesondere der zweite Colossus, der etwa eine Woche vor der Landung am D-Day in Betrieb genommen wurde, den Verlauf dieser Landungen wesentlich veränderte. Unsere heutigen Smartphones haben zum Beispiel mehr Rechenleistung als ein Colossus-Computer, doch zu jener Zeit waren das technologisch und politisch sehr mächtige Maschinen.

Hunger: In einem anderen Teil deines Textes betrachtest du die frühe Informatik in Großbritannien als feminisiertes und abgewertetes Feld. Könntest du das historisch beleuchten, warum gehen Feminisierung und Abwertung Hand in Hand?

Hicks: In der frühen Geschichte der Informatik, vom elektromechanischen bis zum frühen elektronischen Zeitalter, ist zu beobachten, dass diese sehr technische Arbeit auf beiden Seiten des Atlantiks stattfand. Wenn es um Dinge ging wie das frühe Programmieren und die Konstruktion, Herstellung oder Fehlerbehebung von Hardware, das wurde wirklich als etwas angesehen, das nur Leute, die fast schon Funktionär*innen sind, tun können oder lernen sollten. Die Menschen, die diese Arbeit verrichteten, hatten in vielen Fällen eine fortgeschrittene mathematische Ausbildung und bestimmte Fähigkeiten, die nicht normal oder üblich waren. Gleichzeitig wurden diese Bereiche als nicht wichtig angesehen, für die man aber auch nicht unqualifiziert sein sollte. Und Frauen, vor allem weiße Frauen aus der Mittelschicht, wurden als die perfekten Arbeitskräfte für diese Art von Arbeit angesehen, weil sie gut genug waren, aber eben auch nicht zu gut. Und wann immer es eine Arbeit gibt, die noch nicht voll etabliert ist und somit kein Karriereentwicklungspotenzial hat, gibt es Männer, die privilegiert genug sind, diese Arbeit zu meiden und in eine stabilere Arbeit mit einem klaren Karriereentwicklungspotenzial zu gehen. Es gibt eine Tendenz, dass diese noch nicht etablierte Arbeit für Frauen zugänglicher gemacht wird. Sie werden für diese Jobs rekrutiert, denn die Vorstellung war lange Zeit, und ich glaube heute leider immer noch bis zu einem gewissen Grad, dass Frauen keine Karriere brauchen. Sie brauchen, angeblich, nicht so viel Geld wie ein weißer Mann aus der Mittelschicht, um ihren Lebensunterhalt zu bestreiten. Und somit können sie arbeiten, was eine Art nettes Zubrot bringt, und dann gehen sie und widmen sich der eigentlichen Arbeit ihres Lebens, nämlich der Familie und Erziehung. Und das bedeutet, dass es da eine eingebaute Fluktuation gibt. Die gesellschaftliche Konvention bestimmt, welche Arbeit als qualifiziert oder nicht qualifiziert angesehen wird. Dies variiert, je nachdem, wer die Arbeit annimmt, sodass die Feminisierung eines Bereichs die Löhne drücken kann. Das ist einer der Gründe, warum die Geschichte der Informatik für mich so interessant ist, denn normalerweise führt die Einführung von mehr Maschinen in einem Bereich zu einer stärkeren Feminisierung der Arbeit und einem Rückgang der Löhne. In der Geschichte der Informatik ab der Mitte des zwanzigsten Jahrhunderts ist das Gegenteil der Fall, wir sehen eine Professionalisierung und Maskulinisierung des Bereichs.

Hunger: Wenn du diesen Gedanken weiterführst – die unsichtbar entwerteten Arbeitskräfte von heute, die Klickarbeiter*innen, welche die Trainingsdaten für Projekte der Künstlichen ‚Intelligenz‘ erstellen –, welche Parallelen siehst du?

Hicks: Das ist eine großartige Frage, denn in dieser Frage steckt sowohl ein historischer Vergleich als auch die unausgesprochene Vorstellung mächtiger Interessen der Managementklasse, die immer mehr Computerwerkzeuge einsetzt, um den Bedarf an menschlichen Arbeitskräften langfristig zu senken. Also im Sinne von: Wenn man die Technologie nur richtig gut entwickeln würde, wenn man die KI-Modelle gut genug trainiert, dann könne man im Grunde genommen den Bedarf an Menschen für eine Reihe von Jobs senken. Und die Analogie, die du zu den Klickarbeiter*innen verwendet hast und die ich auch auf die Gig-Worker*innen ausdehnen könnte, verläuft auf diesem üblichen Pfad, der durch die Einführung einer Technologie gekennzeichnet ist, die das Potenzial besitzt, eine bestimmte Sache zu

verbessern. Diese Technologie wird auf eine Weise eingesetzt, die nicht unbedingt die ursprünglichen Versprechen erfüllt, aber etwas anderes bewirkt. Vielleicht hält sie die Kosten in einem bestimmten Sektor niedrig. Vielleicht treibt sie die Kosten in die Höhe, bricht aber die Macht der Gewerkschaften in diesem Sektor, oder sie ist einfach nur eine Möglichkeit, die Aktienmärkte für eine kurze Zeit zu manipulieren, sodass viele Leute viel Geld verdienen können, bevor beispielsweise eine Technologieblase platzt oder zusammenbricht. In der Gig-Work-Economy ist es oft so wie beim Beispiel der Mitfahrgelegenheiten-Apps. Risikokapitalgeber*innen haben enorme Summen in Mitfahrgelegenheiten-Apps gesteckt, weil sie davon ausgingen, dass das langfristige Ziel darin bestehen würde, Taxi- und Lieferfahrten zu automatisieren und dafür zu sorgen, dass die Taxis autonom fahren. Und wir sehen jetzt, dass das wahrscheinlich nicht passieren wird. Aber in der Zwischenzeit hat diese Art von Arbeitsarbitrage, die diese Apps hervorgebracht haben, die Macht der Taxigewerkschaften gebrochen. Sie hat es für viele Fahrer*innen, die diese Apps schon früh genutzt haben, um einen einigermaßen existenzsichernden Lohn zu verdienen, unmöglich gemacht, weiterhin ihren existenzsichernden Lohn zu verdienen. Und sie hat dazu geführt, dass Dinge wie Staus und die schlimmste Umweltverschmutzung in den Stadtzentren einfach immer weiter zugenommen haben, da die Idee einer Gig-App für Mitfahrgelegenheiten darin besteht, dass es eine ständige Verfügbarkeit gibt. Diese Autos fahren also ständig um dieselben Häuserblocks, verbrauchen Benzin und stoßen dabei CO2 aus und so weiter.

Hunger: Mir kommt der prominente Fall von Timnit Gebru in den Sinn, die versuchte, eine KI-Ethik-Position bei Google einzuführen und schließlich 2020 entlassen wurde. Wie hast du diesen Fall wahrgenommen – vor dem Hintergrund dessen, worüber wir gerade sprechen?

Hicks: Das ist ein faszinierender Fall und ein wirklich beunruhigendes Spannungsgebiet in den Debatten um Künstliche Intelligenz und Ethik im Bereich des maschinellen Lernens. Denn wie du angemerkt hast, wurde Dr. Gebru speziell als KI-Ethikerin eingestellt. Sie hat eine solide Ausbildung in Informatik und ist eine sehr technikaffine Person. Sie sollte ein Ethik-Element, das als zusätzlich angesehen wird, aber in Wirklichkeit unabdingbar in diesem Bereich sein sollte, in Googles KI-Projekte einbringen. Und sie wurde tatsächlich aus dem Unternehmen gedrängt, gefeuert, weil sie den Job machte, für den sie eingestellt worden war. Sie äußerte Bedenken über die Weise, wie große Machine-Learning-Modelle, große Sprachmodelle, unbeabsichtigte Folgen haben könnten. Und auch über das Ausmaß schädlicher Auswirkungen aufgrund ihrer Funktionsweise. Es ging also nicht nur darum, wie sie eingesetzt werden könnten, um Schaden anzurichten, sondern auch darum, dass das Modell selbst, wenn es hochskaliert wird, dazu neigt, Echokammern bestimmter Art zu schaffen und andere Dinge zu tun, die nicht unbedingt zu positiven Ergebnissen führen würden. Und in diesem wissenschaftlichen Aufsatz *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* (2021) waren sie und ihre Mitautorinnen nicht alarmistisch und sie präsentierten sehr maßvolle und fundierte Ergebnisse. Aber ihre Arbeit wurde effektiv unterdrückt. Google wollte nicht einmal, dass die relativ gemäßigte Kritik der Studie an die Öffentlichkeit gelangt. Dies führte zu einer massiven Spaltung, nicht nur zwischen Dr. Gebru und Google, sondern auch innerhalb von Google und der KI-Community für maschinelles Lernen. Und diese Spaltung wurde auch durch die Frage verursacht: Dürfen wir diese Gelddruckmaschine kritisieren? Oder bedeutet das, dass Kritik vom Tisch ist, wenn wir von all dem Geld, das in Künstliche Intelligenz und maschinelles Lernen fließt, bezahlt werden? Und so denke ich, dass dieser Vorfall in gewisser Weise ein perfektes Beispiel dafür ist, wie weit man innerhalb der Logik eines bestimmten technologischen Systems gehen kann, bevor diese Logik zusammenbricht und sich das System selbst zerstört.

Hunger: Deine aktuelle Forschung dreht sich um Trans-Identität und Computer, die du in dem Aufsatz *Hacking the Cis-tem* (2019) behandelt hast. Du beschreibst, wie Transgender-Personen mit Computersystemen kollidieren, die es nicht erlauben, sich einem Trans-Status zuzuordnen. Ich würde dich bitten, ein wenig auf die Frage einzugehen und – vielleicht ist das auch eine Provokation –, ob das Hauptproblem hier der Computer ist oder vielleicht die Bürokratie, die schlichtweg mit Computern arbeitet?

Hicks: Ich sage gleich im Voraus, dass das Problem nicht der Computer selbst ist. Der Computer hätte natürlich sehr einfach programmiert werden können, und zwar so, dass er in diesem speziellen Fall in den frühen 1960er Jahren Trans-Personen in das System aufgenommen hätte, und zwar als ihr Geschlecht, als ihre gelebte Identität. Er wurde aber ausdrücklich so programmiert, dass er das nicht tat. Das war eine politische Entscheidung. Das war eine Entscheidung, die von hochrangigen Beamt*innen in der britischen Regierung aus Gründen getroffen wurde, welche natürlich ausdrücklich politisch waren und bei denen es darum ging, bestimmte Menschen als vollwertige Mitglieder des Staates anzuerkennen oder nicht anzuerkennen. Menschen, die volle Rechte und Anerkennung verdienten, die ihre Bürger*innenrechte verdienten und die es verdienten, das finanzielle Sicherheitsnetz zu haben, das der britische Wohlfahrtsstaat zur Verfügung stellte – ich möchte darauf hinweisen, dass diese [davon ausgeschlossenen] Menschen ihr ganzes Arbeitsleben lang eingezahlt hatten. So sehr uns ein Computer bei verschiedenen Dingen nützt, so sehr kann er auch bei anderen schaden. Er tendiert zu bestimmten Affordanzen, zum Beispiel dabei, Menschen auf bestimmte Weise zu kategorisieren. Wenn man sich neuere Computersysteme ansieht, die versuchen, Menschen zu kategorisieren und eine Menge persönlicher Daten über sie zu sammeln, um sie zu kategorisieren, in der Regel, damit man damit Geld verdienen kann, indem gezielte Werbung geschaltet wird, dann wird es sehr problematisch. Denn wenn ein System so konzipiert ist, dass es mit immer mehr vorgefertigten Arten von Informationen arbeitet und die Menschen in immer kleinteiligere Schubladen sortiert, dann sind diese Menschen vielleicht selbst nicht mit den Kategorien einverstanden, die auf sie projiziert werden. Ich denke also, du hast völlig Recht, es liegt nicht an der Technologie oder am Computer in dieser Situation, doch ist die Technologie auch nicht völlig immateriell oder unendlich formbar.

Den QR-Code scannen, um das vollständige Videointerview anzusehen.

Whoever Controls Language Models Controls Politics

Hannes Bajohr

The world of artificial intelligence thinks both big and simple at the same time—and has done so from the very beginning. When the workshop that launched the concept and the field of 'artificial intelligence' was held at Dartmouth College in the summer of 1956, the self-imposed task was to figure out "how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves" (McCarthy et al. 2006). The expected duration of the project: two months.

Almost seventy years later—on March 22, 2023—an open letter was published on the website of the longtermist Future of Life Institute, with, at the time of writing, more than thirty-three thousand signatures ('Pause Giant AI Experiments: An Open Letter' 2023). It included the likes of Elon Musk and many renowned AI researchers, and called for a moratorium on the development of large language models (LLMs) for at least six months. Systems such as ChatGPT, the authors claim, have now become too powerful and too dangerous, and "profound risks to society and humanity" are posed by "human-competitive AI" (ibidem). Until there is agreement on how to regulate this complex, all AI labs should refrain from further research.

If Dartmouth spectacularly underestimated how difficult the automation of intelligence would prove to be, the open letter is equally bombastic in drawing the wrong conclusions from the power of current language technology. First, even today, Dartmouth's goal remains unmet: for all their successes, ChatGPT and other LLMs do not operate at "human level" by any meaningfully robust measure (Mitchell and Krakauer 2023; Floridi 2023). Such fantasies fuel AI hype,* a tendency that has afflicted the industry since its early days (Aggarwal 2018), and which ultimately serves the companies that sell it (Luitse and Denkena 2021). What better proof of a developer's power than their ability to distribute a product that could potentially destroy the world? Insiders soon speculated that the goal was in fact to subvert the industry's long-held rule of open research and to continue working in secret for the suggested six months. And indeed, Musk announced the launch of his own AI company, called X.AI, on April 14; his signature very soon no longer seemed to count for much (Jin and Seetharaman 2023).

Second, however, and more importantly, the letter also speaks to a disastrous understanding of the interplay between technology and politics, both in terms of its dangers and the means to address them. While the fear that AI-generated text could flood information channels with falsehoods and "propaganda-as-a-service" (Bagdasaryan and Shmatikov 2022) is entirely valid, the letter is otherwise driven by apocalyptic fantasies about the total replacement of humans by machines and the "loss of control of our civilization" ('Pause Giant AI Experiments: An Open Letter' 2023).

This "x-risk" (the "x" standing for "existential," Bostrom 2002) is the primary concern of 'longtermists'—a libertarian, transhumanist, and ultra-utilitarian school of thought that gives possible future humans an incomparably greater moral weight than actual present ones (Torres 2021). Its proponents, to whom Musk, too, feels a close affinity, think in millennia and in terms of narrow utility maximization. For this reason, the threat of a hyper-intelligent machine—"we are all going to die" stated one particularly alarmist op-ed (Yudkowsky 2023)—worries them much more than, for example, the immediate damage caused by climate change, social injustice, or poverty—problems that, to them, are either non-issues or will be resolved through the very AI they perceive as existential threat (Klein 2023).

A Democratic Disaster

The risk LLMs like ChatGPT pose, however, is not so much the technical catastrophe of malicious computers. Much more concretely, language models threaten democratic disaster—through the privatization of language technologies as the future site of political public spheres, and the purely technocratic approach to solving its problems. This is where politics and civil society need to step up, and where democratic political theory needs to shift its focus.

Technological progress over the last few years has shown that the more data an AI system is fed, the more powerful it becomes—but also the more expensive it is to develop. While it is difficult to predict future trends, it is not impossible that competition to build ever larger models could result in only a handful of companies remaining in the race (Vincent 2023a), such as OpenAI/Microsoft, Google's Deepmind, or Anthropic. And while there are open-source efforts to 'democratize' language models, they have yet to prove successful compared to Big Tech (after all, most people will not train their own LLMs on their home computers, but will rely on a large company's packaged and serviced product). And at least for now, smaller non-commercial ventures and universities play a negligible role in achieving current scale and performance records.**

Two issues appear to be the most worrisome. What we may be facing, and this is the first issue, is an oligopoly that concentrates language technologies in the hands of a few private companies. These powerful players do not exert dominance over any old product. Rather, it is the future of political opinion-forming and deliberation that will be decided in LLMs, which poses a direct challenge to democratic political theory.

Why this is so can be shown by looking at what until now was seen as the biggest political problem with AI systems, their biases (Bender et al. 2021). LLMs model their output on the texts they have been trained on, which largely comprise the writing found across the Internet and other sources—including the prejudices, racism, and sexism that constitute much of this content. Because "'raw data' is an oxymoron" (Gitelman 2013)—that is, data is always situated within a particular context, made for specific purposes, and shaped by the tools and systems used to generate, capture, and represent it—and because reality itself harbors a "world bias" (Pasquinelli 2019, 9)—that is, inequities in society are simply reflected, repeated, and reinforced in data meant to represent a 'neutral' stance—language models are inherently ideological, even in their ostensibly 'innocent' state of conception. This is so even if there is no conscious or malicious intent on the side of their creators.

But there is more: attempts at debiasing to achieve a 'neutral' outcome—trying to 'de-ideologize' LLMs, as it were—are always in vain, regardless of which end of the process is addressed. One can either censor the output, as is done (to some degree) with ChatGPT and its implementation in Bing (running the risk of rendering it unusable; Apprich forthcoming). Or, as is also practiced, one can sift through the input—the dataset—to remove undesirable components before training commences (Miller et al. 2022). Both filtering the results and curating the dataset amount to correcting the model based on a norm, a view of a better world, which is an eminently political choice. De-ideologizing AI thus necessarily involves formulating a social vision—and is thus again ideological.

ChatGPT happens to represent more progressive values, and conservative media have been quick to get excited about 'woke AI.'*** It is not unlikely that this 'progressiveness' is just due to PR considerations: sexist insults, extremist political positions, or racist output simply have a negative impact on tech companies' bottom lines. But even if one buys into the idea that true conviction stands behind OpenAI's choices, neither the uncurated nor the curated versions of LLMs are 'value-free.' To repeat: AI is always ideological (Bajohr 2021; Weatherby 2023).****

For this reason, it should concern us that decisions about the social vision that language models articulate are in the hands of a few companies not subject to democratic control and accountable to no one but their shareholders. They thus become, to misappropriate a term from philosopher Elizabeth Anderson, "private government"

* Especially dubious but popular is Bubeck et al. (2023): AI hype at its finest that is methodologically and rhetorically flawed, but embraced by some media and researchers.

** At the time of writing, there are at least some rumblings that the big players are concerned about open-source models. It is still hard to say whether these concerns are justified. While it is correct that models like LLaMA have produced some encouraging results, there are two facts that should make one somewhat skeptical about a golden open-source future: The first is that LLaMA was leaked from Meta, and is thus only secondarily a product of free development; instead, it is "stand[ing] on the shoulders of giants" who are ultimately able to provide the necessary groundwork (Heaven 2023). The second is that "foundation models" (Bommasani et al. 2021) are increasingly part of a licensing economy in which the responsibility and 'servicing' for the underlying model will be a selling point, favoring large providers over open-source ones. At the U.S. Senate hearing on AI regulation on May 16, 2023, OpenAI CEO Sam Altman agreed that "there will be a relatively small number of providers that can make models," suggesting, however, that this would be positive for effective regulation: "The fewer of us that you really have to keep an eye on … there's benefits there" (Zakrzewski et al. 2023).

*** This concept was also invoked by Musk as a reason for starting his own AI firm (Perrigio 2023a).

**** "Ideology," here, is not to be understood as value judgment. It simply refers, as political philosopher Judith N. Shklar put it, "to political preferences, some very simple and direct, others more comprehensive. … In no case is there any effort to use the word 'ideology' as one of simple opprobrium. On the contrary, it may well be doubted whether political theory … can be written without some sort of ideological impetus. Nor is there any reason to feel that the expression of personal preferences is an undesirable flaw. It must seem so only to those who equate objectivity with remoteness from their own experiences and especially from those they share with their contemporaries. However, if one thinks of ideology as merely a matter of emotional reactions, both negative and positive, to direct social experiences and to the views of others, it is clear that ideology is as inevitable as it is necessary in giving any thinking person a sense of direction" (Shklar 1986, 4; see also Bajohr 2020).

(Anderson 2017).[*****] At first blush, this may not sound so new. "Artifacts have politics," as Landon Winner put it (1980), and so do digital ones: simply through the way it makes information accessible, Google has already had an outsized influence on what appears as reality to users (Noble 2018). However, with the emergent private government that is capitalized machine learning, an even deeper capture has taken hold. For the product of AI companies is the main resource that makes for a vital democracy: language. It is language through which we negotiate political alternatives at the only level where this is possible—the political public sphere. With LLMs, instead of debating what kind of world we want to live in, that decision is already made even before a single word has been uttered, because the language at one's disposal has itself already been subjected to a preliminary political decision. The more language produced by such models permeates the finest capillaries of everyday life in the future, the more dire such an outlook must seem.

Machines of Epistemic Injustice

It provides no comfort that such LLMs can of course also be steered toward the right, as computer scientist David Rozado recently showed by creating *RightWingGPT* (Rozado 2023). In fact, this points to a second worry from the vantage point of democratic theory. A future in which a conservative language AI coexists with a progressive one would not lead to some kind of balance or trajectory toward ever more nuanced positions. Nor would many factions represented by LLMs constitute a wholesome "variety of sects dispersed over the entire face" of the Internet, to cite James Madison's republican panacea against the dominance of one group over another (Madison 2003, 45). For it would no longer be sects or factions talking to each other, but modeled speech itself. Immediately, political LLMs would eliminate the discussion among social groups whose conflicts ideally contribute to the formation of the opinions of an informed public. Instead of exchange, there would be only the reinforcement of already existing opinions; unlike the much-vaunted echo chambers of social media, it would not even be people who set the parameters of that discussion, but a complex system of natural language processing and profit-driven private corporations.

The detrimental effects on democratic politics in particular can be illustrated by an argument posited by political philosopher Judith Shklar. She held that the most important duty of liberal democracies is to structure their public institutions and public forums so that the voices of the marginalized can be heard. This is not simply a moral imperative, but a democratic one. Since hegemonic notions of justice, equity, fairness, and so on are positive concepts, and, as such, limited to cases these concepts explicitly 'allow for'—that is, cases the majority can imagine as relevant—it is essential to listen to negative insights, that is, cases that slip through the cracks of the official conceptual matrix. Sometimes something is not even thought of as an injustice until it is pointed out as such by those affected by it.

The "sense of injustice" (Shklar 1990, 83) that the marginalized articulate—the immediate feeling of injury preceding all explicitly formulated concepts of justice—is not only a cry for its concrete alleviation; it also expands the democratic project by broadening the notion of what is understood as a possible injustice. Hearing the marginalized means taking them seriously as an epistemic resource and at the same time including them as citizens and thus representing more people in the polity. Miranda Fricker (2007) has termed the phenomenon whereby citizens cannot shift the frame of what constitutes justice in a given democracy because they remain unheard and thus unincluded "epistemic injustice." The hegemonic view that LLMs formulate and encode does not allow for this epistemic correction that comes from listening to the sense of injustice. Thus, LLMs are by their very structure machines of epistemic injustice. This turns into a serious problem if, in the long run, LLMs themselves become a surrogate, a 'synthetic' public sphere. As AI systems generate more and more of the texts that populate our discourse—which seems highly likely—the proportion of discourse produced by humans may steadily decrease. Because language models are difficult to

***** I say "misappropriate" because Anderson was referring to companies' power to regulate the lives of their employees when they, for instance, set standards for their speech even in their private time (Anderson 2017, 39). With LLMs, one need not work for a company to be affected. Nevertheless, the limits on and redirection of speech have a similarly quasi-governmental quality.

change once trained—and because they infer norms about the future from facts of the past (O'Neil 2016; Eubanks 2017), there is a danger of what Bender et al. have called a "value lock" (2021). This means that opinions, values, norms, and tendencies that are otherwise open to modification through discourse or ideological combat, including minorities' sense of injustice, become fixed in place due to the system's inability to adapt. No amount of discussion or hegemonic struggle can alter these baked-in values; the result is a technologically-induced political stagnation that includes an increasingly narrow epistemic horizon. Thus, the content of contributions to such a synthetic public sphere is not only predetermined by technological systems and capital interest, but also bound to remain the same, as whatever engagement they encounter runs parallel to and is unaffected by non-LLM public discourse.

For these considerations, it does not matter whether one follows what political philosophy calls the deliberative or the agonistic approach to democratic theory. From the standpoint of deliberative democratic theory—which understands politics as the process of collective reasoning among citizens—the "requirement of free deliberation" depends on the "discursive quality of the contributions" to the public sphere and the possible inclusion of all citizens (Habermas 2022, 150), both of which would be at stake in a synthetic public sphere generated by ideologically predetermined and value-locked LLMs. From the standpoint of an agonistic political theory—which thinks of politics as a domain of struggle and contestation that might not result in any clear-cut consensus—such a synthetic public sphere would likewise eliminate the "legitimate political channels for dissenting voices" that translate the "struggle between opposing hegemonic projects" into the channels of democratic agonistics (Mouffe 2005, 21).

Again, this need not be the result of malicious intent. A good example of how the engineering spirit of 'fixing things with technology' may in fact squash democratic debate, both deliberative and agonistic, is the study by Argyle et al. (2023). The engineers attempted to "improve the quality of divisive conversations" by interposing an AI system between the exchanges of two debating parties that restated their positions in more "neutral" language. As the authors believe in seemingly Habermasian fashion, "improving the quality of political discourse"—that is, an enforced civil tone—"will have broader benefits related to social cohesion and democracy" (Argyle et al., 3). Not only is this solution an example of techno-paternalism, however, since it does not respect the deliberate communicative choices of the participants in the debate but simply 'fixes' them. It is also, projected onto a larger scale, for instance, as a feature in messengers or discussion forums, very much not neutral, but rather again the result of a prior decision about what neutrality and civility—i.e., the limits and conditions of discourse—entail.[******] And this, again, is a deeply political choice.

Last Resort: Communization

Whoever controls language models controls politics. The regulation of AI—which Big Tech ostensibly calls for, but only as voluntary self-supervision by the industry as a type of "regulatory capture" (Vincent 2023b)[*******]—cannot settle for mere ethical guidelines (Stark 2023). As the firing of Timnit Gebru and Margaret Mitchell from Google in late 2020 shows, "ethics departments" are, at best, a fig leaf of accountability that companies can discard at will (Simonite 2021).

To be sure, it is absolutely necessary to create legal regulations (Noble 2018), beginning with prohibiting using AI for deceptive purposes and banning LLM training without the data source owner's consent (AI Now Institute 2023). Moreover, antitrust law could allow breaking up large companies (Srnicek 2017; Zuboff 2019). It might,

****** This is not a hypothetical worry, see Jakesch et al. 2023.

******* Former Google CEO Eric Schmidt admitted as much in no uncertain terms in an interview with NBC leading up to the Senate AI hearings (in which OpenAI CEO Sam Altman suggested a pro-business regulatory framework to senators). Schmidt not only advocated what amounts to self-regulation through a business-to-law pipeline—that is, a legal privilege—but also claimed for his industry the sole competence to grasp the intricacies of what is to be regulated—that is, an epistemic privilege. Conveniently, the rule of experts coincides with the rule of Big Tech.
"Eric Schmidt: When this technology becomes more broadly available, which it will, very quickly, the problems will get much worse. I would much rather have the companies define reasonable boundaries.
Reporter: It shouldn't be a regulatory framework, it maybe shouldn't even be a sort of a democratic vote, it should be the expertise within the industry to help to sort that out?
Schmidt: The industry will first do that, because there is no way a non-industry person can understand what is possible. It's just too new, too hard, there's not the expertise. There is no one in the government who can get it right, but the industry can roughly get it right, and then the government can put a regulatory structure around it" (NBC News 2023, sc. 9:03).

for instance, be desirable to keep dataset collecting and the training process apart in two separate legal entities. The EU's Digital Markets and Digital Services Acts are better positioned than current US efforts in this regard. The oft-heard argument that any regulation would hamper 'innovation' is misplaced, too: local regulatory efforts within strong economic blocks exert global effects, and the EU's aggressive competition laws have often given it an advantage over US efforts and have led them to be emulated abroad (Bradford et al. 2019). However, as a report in *Time* magazine has shown, OpenAI lobbyists have already succeeded in influencing the EU AI Act legislation in their favor (Perrigio 2023b).

For this reason, it is necessary to think big here as well. If these more traditional regulatory measures are ineffective and AI systems become the site of articulating social visions, a dominant factor in the make-up of the public sphere, or even a political infrastructure themselves, there is much to be said for actually subjecting LLMs to public control as well. If this is taken to its logical conclusion, the last resort would be communization.

If one accepts the idea that the infrastructures of technologically mediated communication are infrastructures just like any other—water, electricity, roads—and that their construction depends to a considerable extent on direct or indirect public funding (Zuboff 2019) as well as on "data as labor" from unpaid users (Posner and Weyl 2018, 205; Whittaker 2021), then the socialization of certain technologies appears less a shocking overreach of the state than the forceful realization that public goods and services should also be in the hands of a self-governing public (Taylor 2014).

Seen in this way, the open letter appears in a different light: not merely as the technological catastrophism of a group of fearmongering 'longtermists,' but as an attempt to distract from the political consequences of this technology. For those consequences are far more concrete than a robot takeover, even in the stark, heuristically overstated way I have presented them here. Regulating these technologies poses a much more dangerous threat to the companies and individuals profiting most from the hype around AI. But from the standpoint of democratic theory, this is precisely what is needed now.

This essay first appeared as "Das Ende der menschlichen Politik" in *Neue Zürcher Zeitung*, April 25, 2023. The present version has been substantially expanded and updated.

Bibliography

Aggarwal, Alok. 2018. "The Birth of AI and the First AI Hype Cycle." *KDnuggets* (blog), February 13, 2018. https://www.kdnuggets.com/the-birth-of-ai-and-the-first-ai-hype-cycle.html.

AI Now Institute. 2023. "Antitrust and Competition: It's Time for Structural Reforms to Big Tech." *AI Now Institute* (blog), April 11, 2023. https://ainowinstitute.org/publication/antitrust-and-competition.

Anderson, Elizabeth. 2017. *Private Government: How Employers Rule Our Lives (and Why We Don't Talk About It)*. University Center for Human Values Series. Princeton, Oxford: Princeton University Press.

Apprich, Clemens. forthcoming. "Ja-Sager." In *ChatGPT und andere "Quatschmaschinen": Gespräche mit Künstlicher Intelligenz*, edited by Anna Tuschling, Andreas Sudmann, and Bernhard J. Dotzler. Bielefeld: Transcript.

Argyle, Lisa P., Ethan Busby, Joshua Gubler, Chris Bail, Thomas Howe, Christopher Rytting, and David Wingate. 2023. "AI Chat Assistants Can Improve Conversations about Divisive Topics." arXiv. https://arxiv.org/abs/2302.07268.

Bagdasaryan, Eugene, and Vitaly Shmatikov. 2022. "Spinning Language Models: Risks of Propaganda-As-A-Service and Countermeasures." In *2022 IEEE Symposium on Security and Privacy (SP)*, 769–86. https://doi.org/10.1109/SP46214.2022.9833572.

Bajohr, Hannes. 2020. "On Liberal Disharmony: Judith N. Shklar and the 'Ideology of Agreement.'" *Journal of the History of Ideas Blog*, May 6, 2020. https://jhiblog.org/2020/05/06/on-liberal-disharmony-judith-n-shklar-and-the-ideology-of-agreement/.

———. 2021. "Wer sind wir? Warum künstliche Intelligenz immer ideologisch ist." *Republik*, April 6, 2021. https://www.republik.ch/2021/04/06/warum-kuenstliche-intelligenz-immer-ideologisch-ist.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. Association for Computing Machinery.

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, et al. 2021. "On the Opportunities and Risks of Foundation Models." arXiv. http://arxiv.org/abs/2108.07258.

Bostrom, Nick. 2002. "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards." *Journal of Evolution and Technology* 9. http://jetpress.org/volume9/risks.html.

Bradford, Anu, Adam Chilton, Katerina Linos, and Alexander Weaver. 2019. "The Global Dominance of European Competition Law Over American Antitrust Law." *Journal of Empirical Legal Studies* 16 (4): 731–66. https://doi.org/10.1111/jels.12239.

Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, et al. 2023. "Sparks of Artificial General Intelligence: Early Experiments with GPT-4." arXiv. https://doi.org/10.48550/arXiv.2303.12712.

Eubanks, Virginia. 2017. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. First Edition. New York, NY: St. Martin's Press.

Floridi, Luciano. 2023. "AI as Agency Without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models." *Philosophy & Technology* 36 (1): 15. https://doi.org/10.1007/s13347-023-00621-y.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Gitelman, Lisa, ed. 2013. *"Raw Data" Is an Oxymoron*. Cambridge, MA: MIT Press. https://doi.org/10.1080/1369118x.2014.920042.

Habermas, Jürgen. 2022. "Reflections and Hypotheses on a Further Structural Transformation of the Political Public Sphere." *Theory, Culture & Society* 39 (4): 145–71. https://doi.org/10.1177/02632764221112341.

Heaven, Will Douglas. 2023. "The Open-Source AI Boom Is Built on Big Tech's Handouts. How Long Will It Last?" *MIT Technology Review*, May 12, 2023. https://www.technologyreview.com/2023/05/12/1072950/open-source-ai-google-openai-eleuther-meta/.

Jakesch, Maurice, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. 2023. "Co-Writing with Opinionated Language Models Affects Users' Views." In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–15. Hamburg, Germany: ACM. doi:10.1145/3544548.3581196.

Jin, Berber, and Deepa Seetharaman. 2023. "Elon Musk Creates New Artificial Intelligence Company X.AI." *The Wall Street Journal*, April 14, 2023. https://www.wsj.com/articles/elon-musks-new-artificial-intelligence-business-x-ai-incorporates-in-nevada-962c7c2f.

Klein, Naomi. 2023. "AI Machines Aren't 'Hallucinating'. But Their Makers Are." *The Guardian*, May 8, 2023. https://www.theguardian.com/commentisfree/2023/may/08/ai-machines-hallucinating-naomi-klein.

Luitse, Dieuwertje, and Wiebke Denkena. 2021. "The Great Transformer: Examining the Role of Large Language Models in the Political Economy of AI." *Big Data & Society* 8 (2): 205395172110477. https://doi.org/10.1177/20539517211047734.

Madison, James. 2003. "The Federalist No. 10." In *The Federalist: With Letters of "Brutus,"* edited by Terence Ball. Cambridge: Cambridge University Press. 40–46.

McCarthy, John, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon. 2006. "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence." *AI Magazine* 27 (4): 12–14.

Miller, Erin, Roudabeh Kishi, Clionadh Raleigh, and Caitriona Dowd. 2022. "An Agenda for Addressing Bias in Conflict Data." *Scientific Data* 9 (1): 593. https://doi.org/10.1038/s41597-022-01705-8.

Mitchell, Melanie, and David C. Krakauer. 2023. "The Debate Over Understanding in AI's Large Language Models." *Proceedings of the National Academy of Sciences* 120 (13). https://doi.org/10.1073/pnas.2215907120.

Mouffe, Chantal. 2005. *On the Political*. London: Routledge.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University.

NBC News, dir. 2023. "'Life and Death Decisions Are Being Made' by Artificial Intelligence." YouTube, May 15, 2023. Video, 10:42. https://www.youtube.com/watch?v=VgJnqJ9WxAw.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Pasquinelli, Matteo. 2019. "How a Machine Learns and Fails—A Grammar of Error for Artificial Intelligence." *Spheres*, no. 5. http://spheres-journal.org/how-a-machine-learns-and-fails-a-grammar-of-error-for-artificial-intelligence/.

"Pause Giant AI Experiments: An Open Letter." 2023. *Future of Life Institute* (blog), March 22, 2023. https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

Perrigio, Billy. 2023a. "Elon Musk Is Bringing the Culture Wars to AI." *Time*, March 3, 2023. https://time.com/6260185/elon-musk-ai-culture-wars/.

———. 2023b. "Exclusive: OpenAI Lobbied the E.U. to Water Down AI Regulation." *Time*, June 20, 2023. https://time.com/6288245/openai-eu-lobbying-ai-act/.

Posner, Eric A. und E. Glen Weyl. 2018. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton: Princeton University Press.

Rozado, David. 2023. "RightWingGPT—An AI Manifesting the Opposite Political Biases of ChatGPT." Substack newsletter. *Rozado's Visual Analytics* (blog), February 16, 2023. https://davidrozado.substack.com/p/rightwinggpt.

Shklar, Judith N. 1986. *Legalism: Law, Morals, and Political Trials*. Cambridge: Harvard University Press.

———. 1990. *The Faces of Injustice*. New Haven: Yale University Press.

Simonite, Tom. 2021. "What Really Happened When Google Ousted Timnit Gebru." *Wired*, June 8, 2021. https://www.wired.com/story/google-timnit-gebru-ai-what-really-happened/.

Srnicek, Nick. 2017. *Platform Capitalism*. London: Polity.

Stark, Luke. 2023. "Breaking Up (with) AI Ethics." *American Literature* 95, no. 2: 365–79. doi:10.1215/00029831-10575148.

Taylor, Astra. 2014. *The People's Platform: Taking Back Power and Culture in the Digital Age*. Toronto: Random House Canada.

Torres, Èmile P. 2021. "Against Longtermism." *Aeon*, October 19, 2021. https://aeon.co/essays/why-longtermism-is-the-worlds-most-dangerous-secular-credo.

Vincent, James. 2023a. "AI Is Entering an Era of Corporate Control." *The Verge*, April 3, 2023. https://www.theverge.com/23667752/ai-progress-2023-report-stanford-corporate-control.

———. 2023b. "The Senate's Hearing on AI Regulation Was Dangerously Friendly." *The Verge*, May 19, 2023. https://www.theverge.com/2023/5/19/23728174/ai-regulation-senate-hearings-regulatory-capture-laws.

Whittaker, Meredith. 2021. "The Steep Cost of Capture." *Interactions* 28 (6): 50–55. https://doi.org/10.1145/3488666.

Winner, Langdon. 1980. "Do Artifacts Have Politics?" *Daedalus* 109, no. 1: 121–36.

Weatherby, Leif. 2023. "ChatGPT Is an Ideology Machine." *Jacobin*, April 17, 2023. https://jacobin.com/2023/04/chatgpt-ai-language-models-ideology-media-production.

Yudkowsky, Eliezer. 2023. "The Open Letter on AI Doesn't Go Far Enough: We Need to Shut It All Down." *Time*, March 29, 2023. https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/.

Zakrzewski, Cat, Nitasha Tiku, Cristiano Lima, and Will Oremus. 2023. "OpenAI CEO Tells Senate That He Fears AI's Potential to Manipulate Views." *Washington Post*, May 16, 2023. https://www.washingtonpost.com/technology/2023/05/16/ai-congressional-hearing-chatgpt-sam-altman/.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Public Affairs.

Wer die Sprachmodelle beherrscht, beherrscht auch die Politik

Hannes Bajohr

Die Welt Künstlicher Intelligenz denkt groß und zugleich simpel, und zwar von Anfang an. Als im Sommer 1956 am Dartmouth College jener Workshop stattfand, der den Begriff und das Feld der ‚artificial intelligence' ins Leben rief, lautete die selbstgesetzte Aufgabe, herauszufinden, „wie man Maschinen dazu bringen kann, Sprache zu verwenden, Abstraktionen und Begriffe zu formen, Probleme zu lösen, die bisher Menschen vorbehalten sind, und sich selbst zu verbessern" (McCarthy et al. 2006). Angesetzte Arbeitszeit: zwei Monate.

Fast siebzig Jahre später – am 22. März 2023 – erschien auf der Website des Future of Life Institute ein offener Brief, der zum Zeitpunkt der Niederschrift dieses Artikels um die dreiunddreißigtausend Unterschriften zählte („Pause Giant AI Experiments: An Open Letter" 2023). Darunter befinden sich Persönlichkeiten wie Elon Musk und viele renommierte KI-Forscher*innen, die ein Moratorium für die Entwicklung großer Sprachmodelle (‚large language models', LLMs) für mindestens sechs Monate fordern. Systeme wie ChatGPT seien inzwischen zu mächtig und zu gefährlich geworden und es bestünden „fundamentale Risiken für die Gesellschaft und die Menschheit", die von „auf Menschenniveau agierender KI" (ebenda) ausgingen. Bis man sich nicht darauf geeinigt habe, wie das zu regulieren sei, sollten alle KI-Labore auf weitere Forschung verzichten.

Unterschätzte man in Dartmouth noch spektakulär, als wie schwierig sich die Automatisierung von Intelligenz herausstellen würde, zieht der offene Brief mit ähnlich viel Bombast die falschen Konsequenzen aus den Möglichkeiten heutiger Sprachtechnologie. Denn erstens ist auch heute das Ziel von Dartmouth nicht erreicht – aller Erfolge zum Trotz agieren ChatGPT und Co nicht „auf Menschenniveau" (Mitchell und Krakauer 2023; Floridi 2023). Solche Fantasien sind Teil des Hypes um KI,* eine Tendenz, die die Branche seit ihren Anfängen plagt (Aggarwal 2018) und die am Ende vor allem den Firmen dient, die sie verkaufen (Luitse und Denkena 2021). Was beweist die Macht der Entwickler*innen mehr als die Fähigkeit, ein Produkt zu vertreiben, das potenziell die Welt zerstören könnte? Insider*innen spekulierten ohnehin bald, hier gehe es darum, die in der Branche lange geltenden Regeln offener Forschung zu unterlaufen und in den sechs Monaten im Geheimen weiterzuarbeiten. Und tatsächlich kündigte Musk am 14. April 2023 die Gründung seines eigenen KI-Unternehmens mit dem Namen X.AI an; seine Unterschrift schien schon bald nicht mehr viel zu gelten (Jin und Seetharaman 2023).

Zweitens – und viel wichtiger – spricht aus dem Brief auch ein katastrophales Verständnis vom Zusammenwirken von Technik und Politik – sowohl was ihre Gefahren als auch deren Bekämpfung betrifft. Während die Befürchtung, KI-generierter Text könnte Informationskanäle mit Unwahrheiten und „propaganda-as-a-service" (Bagdasaryan und Shmatikov 2022) überschwemmen, durchaus berechtigt ist, ist der Brief ansonsten von apokalyptischen Fantasien über die völlige Ersetzung des Menschen durch Maschinen und den „Kontrollverlust über unsere Zivilisation" getragen („Pause Giant AI Experiments: An Open Letter" 2023).

Das „x-risk" – wobei das „x" für „existential" steht (Bostrom 2002) – ist die primäre Sorge sogenannter ‚Longtermists' – einer libertären, transhumanistischen und ultrautilitaristischen Denkschule, die möglichen zukünftigen Menschen ein ungleich größeres moralisches Gewicht einräumt als realen gegenwärtigen (Torres 2021). Ihre Vertreter*innen, denen auch Musk nahesteht, denken in Jahrtausenden und in Begriffen einer reinen Nutzenmaximierung. Aus diesem Grund beunruhigt sie die Bedrohung durch hyperintelligente Maschinen – „Wir werden alle sterben", heißt es in einem besonders alarmistischen Meinungsartikel (Yudkowsky 2023) – viel mehr als beispielsweise die unmittelbaren Schäden des Klimawandels, der sozialen Ungerechtigkeit oder der globalen Armut; Probleme, die für sie entweder überhaupt keine sind oder durch eben jene KI gelöst werden, die sie als existenzielle Bedrohung wahrnehmen (Klein 2023).

Ein demokratisches Desaster

Was aber mit großen Sprachmodellen wie ChatGPT tatsächlich auf uns zukommt, ist keine technische Katastrophe bösartiger Computer. Viel konkreter drohen Sprachmodelle ein demokratisches Desaster zu werden – durch die Privatisierung von Sprachtechnologien als zukünftigem Ort politischer Öffentlichkeit und den rein technokratischen Ansatz zur Lösung dieser Probleme. Genau an dieser Stelle kommen Politik und Zivilgesellschaft ins Spiel, und auch die Demokratietheorie ist hier gefragt.

Die technische Entwicklung der letzten Jahre hat gezeigt, dass ein KI-System umso leistungsfähiger wird, mit je mehr Daten man es füttert. Entsprechend teurer ist es dann auch in der Entwicklung. Es ist zwar schwierig, künftige Trends vorherzusagen, aber es ist nicht ausgeschlossen, dass der Wettbewerb um immer größere Modelle dazu führen könnte, dass nur noch eine Handvoll Unternehmen im Rennen bleibt (Vincent 2023a), z. B. OpenAI/Microsoft, Googles Deepmind oder Anthropic. Und obwohl es Open-Source-Bestrebungen gibt, Sprachmodelle zu ‚demokratisieren', müssen sie sich im Vergleich zu Big Tech erst noch als erfolgreich erweisen (schließlich werden die meisten Menschen nicht ihre eigenen LLMs auf ihren Heimcomputern trainieren, sondern sich auf das fertig verpackte und gewartete Produkt eines großen Unternehmens verlassen). Und zumindest im Moment spielen kleinere nichtkommerzielle Unternehmen und Universitäten beim Erreichen immer neuerer Größenrekorde so gut wie keine Rolle mehr.**

Zwei Probleme stehen hier im Vordergrund. Was uns möglicherweise bevorsteht, und das ist das erste Problem, ist ein Oligopol, das die Sprachtechnologien in den Händen einiger weniger Privatunternehmen konzentriert. Diese mächtigen Akteur*innen üben nicht die Vorherrschaft über ein beliebiges Produkt aus. Vielmehr wird die Zukunft der politischen Meinungsbildung und Deliberation in LLMs entschieden, was eine direkte Herausforderung für die demokratische politische Theorie darstellt.

Warum das so ist, lässt sich anhand dessen zeigen, was bisher als größtes politisches Problem von KI-Systemen galt: ihre „Biases" (Bender et al. 2021). Sprachmodelle modellieren ihren Output entlang der Texte, auf die sie trainiert wurden und die größtenteils aus Dokumenten bestehen, die im Internet und anderen Quellen zu finden sind – einschließlich der Vorurteile, Rassismen und Sexismen, die einen Großteil dieser Inhalte ausmachen. Aber da gilt: „‚Raw Data' Is an Oxymoron" (Gitelman 2013) – Daten stehen immer in einem bestimmten Kontext und sind zu bestimmten Zwecken gedacht, und sie werden von den Werkzeugen und Systemen geformt, die zu ihrer Erzeugung, Erfassung und Darstellung verwendet werden –, und da die Realität selbst einen „world bias" (Pasquinelli 2019, 9) aufweist – Ungleichheiten in der Gesellschaft werden in Daten, die eine ‚neutrale' Haltung darstellen sollen, schlicht wiederholt und verstärkt –, sind Sprachmodelle von Natur aus ideologisch, selbst in ihrem scheinbar ‚unschuldigen' Ausgangszustand. Dies gilt selbst dann, wenn ihre Schöpfer*innen keine bewusste oder böswillige Absicht verfolgen.

Alle Versuche, ein ‚neutrales' Ergebnis zu erzielen – LLMs sozusagen zu ‚entideologisieren' – sind immer vergeblich, unabhängig davon, an welchem Ende des Prozesses angesetzt wird. So kann man entweder den Output zensieren, wie es (bis zu einem gewissen Grad) mit ChatGPT und seiner Implementierung in Bing geschieht (mit dem Risiko, das Programm unbrauchbar zu machen, vgl. Apprich i. E.). Oder man kann, wie ebenfalls praktiziert, den Input – den Datensatz – durchforsten, um unerwünschte Bestandteile zu entfernen, bevor das Training beginnt (Miller et al. 2022). Sowohl das Filtern der Ergebnisse als auch das Kuratieren des Datensatzes laufen darauf hinaus, das Modell auf der Grundlage einer Norm, einer Vorstellung von einer besseren Welt, zu korrigieren. Das aber ist eine eminent politische Entscheidung. Die Entideologisierung der KI beinhaltet notwendigerweise die Formulierung einer gesellschaftlichen Vision – und ist damit ebenfalls ideologisch. Weil GPT etwa tendenziell eher progressive Werte vertritt, haben konservative Medien sich schnell über ‚woke KI'

* Besonders fragwürdig – aber in Longtermist-Kreisen beliebt – ist Bubeck et al. (2023); ein KI-Hype vom Feinsten, der bei ChatGPT Bewusstsein ausgemacht haben will und von einigen Medien und Forscher*innen begrüßt wird, aber methodisch fehlerhaft und rhetorisch irreführend ist.

** Zum Zeitpunkt der Fertigstellung dieses Essays gibt es zumindest einige Gerüchte, dass sich die großen Akteur*innen über Open-Source-Modelle Sorgen machen. Es ist schwer zu sagen, ob diese Bedenken gerechtfertigt sind. Es ist zwar richtig, dass Modelle wie LLaMA einige ermutigende Ergebnisse hervorgebracht haben, aber es gibt zwei Tatsachen, die skeptisch gegenüber einer goldenen Open-Source-Zukunft machen sollten: Die erste ist, dass LLaMA von der Firma Meta geleakt wurde und somit nur in zweiter Linie ein Produkt der freien Entwicklung ist, sondern vielmehr „auf den Schultern von Riesen steht", die allein in der Lage sind, die notwendigen technischen Grundlagen zu schaffen (Heaven 2023). Zweitens sind „Foundation Models" (Bommasani et al. 2021) zunehmend Teil einer Lizenzwirtschaft, in der die Verantwortung und die Wartung für das zugrundeliegende Modell zum Verkaufsargument werden, das große Anbieter*innen gegenüber Open-Source-Anbieter*innen bevorzugt. In der Anhörung des US-Senats zur KI-Regulierung am 16. Mai 2023 stimmte Sam Altman, CEO von OpenAI, zu, dass „es eine relativ kleine Anzahl von Anbieter*innen geben wird, die Modelle herstellen können", was sich jedoch positiv auf eine wirksame Regulierung auswirken würde: „Es gibt weniger von uns, die man wirklich im Auge behalten muss … das hat Vorteile" (Zakrzewski et al. 2023).

ereifert.*** In Wirklichkeit aber stehen eher PR-Erwägungen hinter dieser Kuratierung: Sexistische Beleidigungen, politische Extrempositionen oder rassistische Outputs wirken sich schlicht negativ auf die Gewinnmarge der Techunternehmen aus. Aber selbst, wenn man davon ausgeht, dass hinter OpenAIs Entscheidungen echte Überzeugung steht, sind weder die unkuratierten noch die kuratierten Versionen von LLMs ‚wertfrei'. Ich wiederhole: KI ist immer ideologisch (Bajohr 2021; Weatherby 2023).****

Aus diesem Grund sollte es uns beunruhigen, dass Entscheidungen über den gesellschaftlichen Entwurf, den Sprachmodelle artikulieren, in den Händen einiger weniger Unternehmen liegt, die keiner demokratischen Kontrolle unterstehen und niemandem gegenüber rechenschaftspflichtig sind außer ihren Anteilseigner*innen. Sie werden damit, um einen Begriff der Philosophin Elizabeth Anderson zweckzuentfremden, zu „privater Regierung" (Anderson 2017).***** Auf den ersten Blick mag dies nicht ganz so neu klingen. „Artifacts have politics", wie Langdon Winner es formulierte (1980), und das gilt auch für digitale Artefakte: Allein durch die Art, wie es Informationen zugänglich macht, hat Google bereits einen übergroßen Einfluss darauf, was Nutzer*innen als Realität erscheint (Noble 2018). Doch mit der neu entstehenden privaten Regierung, die maschinelles Lernen kapitalisiert, hat eine noch tiefere Vereinnahmung eingesetzt. Denn das Produkt der KI-Unternehmen ist exakt jene Ressource, die eine lebendige Demokratie ausmacht: Sprache. Es ist die Sprache, durch die wir politische Alternativen auf der einzigen Ebene aushandeln, auf der das möglich ist, auf der Ebene der politischen Öffentlichkeit. Statt zu debattieren, in welcher Welt wir leben wollen, ist mit LLMs ein Teil dieser Entscheidung schon getroffen, bevor ein einziges Wort gesprochen ist, weil die Sprache selbst schon einer politischen Vorentscheidung unterlag. Je mehr die Sprache, die von solchen Modellen produziert wird, zukünftig die feinsten Verästelungen unseres Alltagslebens durchdringt, desto düsterer muss eine solche Perspektive erscheinen.

Maschinen epistemischer Ungerechtigkeit

Da hilft es wenig, dass man ein Sprachmodell logischerweise auch nach rechts dirigieren kann, wie der Informatiker David Rozado gezeigt hat, als er kürzlich *RightWingGPT* schuf (Rozado 2023). Dies verweist auf eine zweite demokratietheoretische Sorge. Eine Zukunft, in der es neben einer progressiven auch eine konservative Sprach-KI gäbe, würde nicht zu einer Art Gleichgewicht oder einer Entwicklung hin zu immer nuancierteren Positionen führen. Ebenso wenig würde eine Vielzahl per LLMs vertretene Fraktionen eine gesunde „Vielfalt der Sekten" darstellen, um James Madisons republikanisches Allheilmittel gegen die Dominanz einer Gruppe über eine andere zu zitieren (Madison 1993, 100). Es wären nicht länger Sekten oder Fraktionen, die miteinander sprechen, sondern stochastisch modellierte Rede selbst. Politische LLMs würden so die Diskussion zwischen gesellschaftlichen Gruppen beenden, deren Konflikte idealerweise zur Meinungsbildung einer informierten Öffentlichkeit beitragen. Statt Austausch gäbe es nur die Verstärkung bereits bestehender Meinungen; und anders als bei den vielbeschworenen Echokammern sozialer Medien, wären es hier nicht einmal mehr Menschen, die die Parameter dieser Diskussion festlegten. An ihre Stelle träte ein komplexes System aus maschineller Sprachverarbeitung und profitorientierten Privatunternehmen.

Die nachteiligen Auswirkungen speziell auf demokratische Politik lassen sich anhand eines Arguments der politischen Philosophin Judith Shklar veranschaulichen. Sie hielt es für die wichtigste Aufgabe liberaler Demokratien, ihre öffentlichen Institutionen und Foren so zu gestalten, dass die Stimmen der Marginalisierten gehört

*** Diese Wendung wurde auch von Musk als Grund für die Gründung seines eigenen KI-Unternehmens angeführt (Perrigio 2023a).

**** ‚Ideologie' ist hier nicht als Werturteil zu verstehen. Er bezieht sich einfach, wie die politische Philosophin Judith N. Shklar es ausdrückt, „auf politische Präferenzen, einige sehr einfach und direkt, andere umfassender. … Auf keinen Fall wird versucht, das Wort ‚Ideologie' als ein einfaches Schimpfwort zu verwenden. Im Gegenteil, es kann durchaus bezweifelt werden, dass politische Theorie … ohne eine Art ideologischen Impetus geschrieben werden kann. Es gibt auch keinen Grund zu der Annahme, dass der Ausdruck persönlicher Präferenzen ein unerwünschter Makel ist. Dies kann nur denjenigen so erscheinen, die Objektivität mit der Entfernung von ihren eigenen Erfahrungen und insbesondere von denen, die sie mit ihren Zeitgenoss*innen teilen, gleichsetzen. Betrachtet man Ideologie jedoch lediglich als eine Angelegenheit emotionaler Reaktionen, sowohl negativer als auch positiver Art, auf unmittelbare soziale Erfahrungen und auf die Ansichten anderer, so ist es klar, dass Ideologie ebenso unvermeidlich wie notwendig ist, um jedem denkenden Menschen eine Orientierung zu geben" (Shklar 1986, 4; siehe auch Bajohr 2018).

***** Ich sage „zweckentfremden", weil Anderson auf die Macht der Unternehmen verwies, das Leben ihrer Mitarbeiter*innen zu regulieren, wenn sie beispielsweise Sprachnormen auch für ihre private Zeit festlegten (Anderson 2017, 39). Bei LLMs muss man nicht für ein Unternehmen arbeiten, um betroffen zu sein. Nichtsdestotrotz haben die Begrenzung und Umlenkung von Sprache eine ähnlich regierungshafte Qualität.

werden können. Dies sei nicht nur ein moralisches Gebot, sondern ein demokratisches: Da es sich bei den hegemonialen Vorstellungen von Gerechtigkeit, Gleichheit, Fairness usw. um positive Konzepte handele, die sich auf Fälle beschränken, die von diesen Konzepten ausdrücklich ‚berücksichtigt' werden – und so auf Fälle, die sich die Mehrheit als relevant vorstellen kann –, ist es unerlässlich, auch negativen Erkenntnissen Gehör zu schenken, also Fällen, die durch das Raster der offiziellen Begriffsmatrix fallen. Manchmal wird etwas gar nicht als Ungerechtigkeit wahrgenommen, bis es von den Betroffenen als solche benannt wird.

Der „Sinn für Ungerechtigkeit" (Shklar 2021, 135), den die Marginalisierten artikulieren – das unmittelbare Gefühl der Verletzung, das allen explizit formulierten Konzepten von Gerechtigkeit vorausgeht –, ist nicht nur ein Ruf nach konkreter Linderung; er bringt auch das demokratische Projekt voran, indem er den Begriff dessen ausweitet, was als mögliche Ungerechtigkeit erkannt werden kann. Die Marginalisierten anzuhören, bedeutet also einerseits, sie als epistemische Ressource ernst zu nehmen, und andererseits, sie effektiv als Bürger*innen anzuerkennen und somit mehr Menschen im Gemeinwesen zu vertreten. In Anschluss an Shklar hat Miranda Fricker (2023) „epistemische Ungerechtigkeit" als das Phänomen bezeichnet, bei dem Bürger*innen den Rahmen dessen, was Gerechtigkeit in einer gegebenen Demokratie ausmacht, nicht verändern können, weil sie ungehört bleiben und nicht berücksichtigt werden. Die hegemoniale Sichtweise, die LLMs formulieren und codieren, lässt diese epistemische Korrektur, die sich aus dem Hören auf den Sinn für Ungerechtigkeit ergibt, nicht zu. LLMs sind damit bereits ihrer Struktur nach Maschinen epistemischer Ungerechtigkeit. Dies wird zu einem ernsten Problem, wenn LLMs langfristig selbst zu einem Surrogat der Öffentlichkeit, also einer ‚synthetischen' Öffentlichkeit werden. Werden immer mehr Texte durch KI-Systeme generiert – und davon ist auszugehen –, sinkt der Anteil an menschenproduzierten Diskursbeiträgen stetig. Da Sprachmodelle, einmal trainiert, schwer zu verändern sind und zudem aus Fakten der Vergangenheit Normen für die Zukunft ableiten (O'Neil 2016; Eubanks 2017), besteht die Gefahr eines „value lock", wie Bender et al. es nennen (2021). Meinungen, Werte, Normen und Tendenzen, die ansonsten durch Diskurs oder ideologische Auseinandersetzung verändert werden können – einschließlich des artikulierten Unrechtsempfindens von Minderheiten –, werden aufgrund der Unfähigkeit des Systems, sich anzupassen, festgeschrieben. Keine noch so lange Diskussion und kein noch so energischer Hegemoniekampf können diese eingebrannten Werte verändern; das Ergebnis ist eine technologisch bedingte politische Stagnation, die einen immer engeren Erkenntnishorizont einschließt. Der Inhalt der Beiträge zu einer solchen synthetischen Öffentlichkeit ist also nicht nur durch technologische Systeme und Kapitalinteressen vorbestimmt, sondern er bleibt auch zwangsläufig derselbe, da er parallel zu und unbeeinflusst von nicht-LLM-generiertem öffentlichem Diskurs verläuft.

Für diese Überlegungen spielt es keine Rolle, ob man einem deliberativen oder einem agonistischen Ansatz der Demokratietheorie folgt. Aus der Sicht deliberativer Theorie – die Politik als Prozess kollektiver, rationaler Argumentation unter Bürger*innen versteht – hängt das „Erfordernis der freien Deliberation" von der „diskursiven Qualität der Beiträge" in der Öffentlichkeit und der möglichen Einbeziehung aller Betroffenen ab (Habermas 2022, 22, 26); in einer synthetischen Öffentlichkeit, die durch ideologisch grundierte LLMs im ‚value lock' erzeugt wird, stünde beides auf dem Spiel. Vom Standpunkt einer agonistischen politischen Theorie aus betrachtet – die Politik als einen Bereich des Kampfes und der Anfechtungen ansieht, der nicht unbedingt zu einem eindeutigen Konsens führt –, würde eine solche synthetische Öffentlichkeit „für widerstreitende Stimmen legitime […] Artikulationsmöglichkeiten" abschaffen, die normalerweise den „Kampf zwischen unvereinbaren hegemonialen Projekten" in die Kanäle der demokratischen Agonistik übertragen sollen (Mouffe 2007, 30–31).

Auch dies muss nicht unbedingt das Ergebnis böswilliger Absicht sein. Ein gutes Beispiel dafür, wie der technische Geist des ‚fixing things with technology' die demokratische Debatte, egal, ob deliberativ oder agonistisch, abwürgen kann, bieten Argyle et al. (2023). Die Ingenieur*innen versuchten, „die Qualität von kontroversen Gesprächen zu verbessern", indem sie ein KI-System zwischen den Dialog zweier

debattierender Parteien schalteten, das ihre Positionen in einer ‚neutraleren' Sprache umformulierte. Die Autor*innen hegen die explizite, scheinbar Habermas'sche Hoffnung, dass „die Verbesserung der Qualität des politischen Diskurses" – d. h. ein erzwungener ziviler Ton – „einen größeren Nutzen für den sozialen Zusammenhalt und die Demokratie haben" werde (ebd., 3). Diese Lösung ist jedoch nicht nur ein Beispiel von Techno-Paternalismus, da sie die bewussten kommunikativen Entscheidungen der Debattenteilnehmer*innen nicht respektiert, sondern schlicht ‚repariert'; sie ist auch, auf einen größeren Maßstab projiziert – z. B. in Messengern oder Diskussionsforen integriert – ganz und gar nicht neutral, sondern wiederum das Ergebnis einer vorherigen Entscheidung darüber, was Neutralität und Höflichkeit sind, wie also die Grenzen und Bedingungen des Diskurses aussehen sollen.****** Und auch dies ist eine zutiefst politische Entscheidung.

Letztes Mittel Vergesellschaftung

Wer die Sprachmodelle beherrscht, beherrscht auch die Politik. Die Regulierung von KI – die Big Tech vorgeblich fordert, aber nur als freiwillige Selbstkontrolle der Industrie, als „regulatory capture" (Vincent 2023b) akzeptieren will******* – kann sich nicht mit bloßen ethischen Richtlinien begnügen (Stark 2023). Wie die Entlassung von Timnit Gebru und Margaret Mitchell bei Google Ende 2020 gezeigt hat, sind ‚Ethikabteilungen' bestenfalls ein Feigenblatt der Verantwortlichkeit, das sich Unternehmen nach Belieben anheften und das sie stets wieder ablegen können (Simonite 2021).

Es ist unbestritten notwendig, gesetzliche Regelungen zu schaffen (Noble 2018), angefangen mit dem Verbot der Nutzung von KI für betrügerische Zwecke und von LLM-Training ohne Zustimmung der Dateneigentümer*innen (AI Now Institute 2023). Außerdem könnte das Kartellrecht die Zerschlagung großer Unternehmen ermöglichen (Srnicek 2017; Zuboff 2019) – beispielsweise mag es wünschenswert sein, das Sammeln von Datensätzen und den Trainingsprozess in zwei getrennten rechtlichen Einheiten zu halten. Der EU Digital Markets and Services Act ist in dieser Hinsicht besser dazu in der Lage als derzeitige Bemühungen der USA. Auch das oft gehörte Argument, dass jegliche Regulierung ‚Innovation' behindern würde, ist unangebracht: Lokale Regulierungsbemühungen innerhalb starker Wirtschaftsblöcke haben globale Auswirkungen, und die aggressiven Wettbewerbsgesetze der EU haben ihr oft einen Vorteil gegenüber den Bemühungen der USA verschafft und dazu geführt, dass sie im Ausland nachgeahmt wurden (Bradford et al. 2019). Wie ein Bericht des *Time Magazine* allerdings gezeigt hat, ist es OpenAI-Lobbyist*innen bereits gelungen, die Gesetzgebung des EU AI Acts in ihrem Sinne zu beeinflussen (Perrigio 2023b).

Aus diesem Grund es ist notwendig, auch hier größer zu denken. Wenn diese traditionelleren Regulierungsmaßnahmen unwirksam sind und KI-Systeme zum Ort von Gesellschaftsentwürfen, zu dominanten Faktoren der Öffentlichkeit oder gar zu einer politischen Infrastruktur selbst werden, spricht viel dafür, sie tatsächlich auch öffentlicher Kontrolle zu unterstellen. Sollten bloße Leitplanken sich als zu schwach dafür erweisen, weil etwa der Einsatzbereich von Allzweck-KIs gar nicht mehr klar abzustecken ist, bleibt als letztes Mittel die Vergesellschaftung.

Akzeptiert man die Vorstellung, dass die Infrastrukturen technologisch vermittelter Kommunikation Infrastrukturen wie jede andere sind – Wasser, Strom, Straßen – und dass ihr Aufbau in erheblichem Maße von direkter oder indirekter öffentlicher Finanzierung (Zuboff 2019) sowie von „data as labor" von unbezahlten Nutzer*innen abhängt (Posner und Weyl 2018, 205; Whittaker 2021), dann erscheint die Vergesellschaftung bestimmter Technologien weniger als schockierende Übergriffigkeit des Staates, sondern vielmehr als die zwingende Folge der Einsicht, dass öffentliche Güter und Dienstleistungen in den Händen einer selbstverwalteten Öffentlichkeit liegen sollten (Taylor 2014).

So erscheint auch der offene Brief in einem anderen Licht: nicht bloß als technologischer Katastrophismus einer Riege von Longtermists, sondern als Versuch, von den politischen Konsequenzen dieser Technik abzulenken. Denn die sind sehr viel konkreter als das ‚x-risk' und die Diktatur der Maschinen, selbst in der krassen und heuristisch überspitzten Weise, in der ich sie hier dargestellt habe. Die Überwachung dieser Technologien stellt eine viel gefährlichere Bedrohung für die Unternehmen und Einzelpersonen dar, die am meisten von dem Hype um die KI profitieren. Aus demokratietheoretischer Sicht aber ist sie unabdingbar.

Dieser Aufsatz erschien zuerst unter dem Titel „Das Ende der menschlichen Politik" in der *Neuen Zürcher Zeitung* vom 25. April 2023. Die vorliegende Fassung wurde wesentlich erweitert und aktualisiert.

****** Das ist keine hypothetische Sorge, siehe Jakesch et al. 2023.

******* Der ehemalige Google-CEO Eric Schmidt hat dies in einem Interview mit NBC im Vorfeld der KI-Anhörungen des US-Senats (in dem der CEO von OpenAI, Sam Altman, den Senator*innen einen wirtschaftsfreundlichen Regulierungsrahmen vorschlug) unmissverständlich zugegeben. Schmidt plädierte nicht nur für eine Selbstregulierung durch eine Business-to-Law-Pipeline – also ein rechtliches Privileg –, sondern beanspruchte für seine Branche auch die alleinige Kompetenz, die Feinheiten dessen, was reguliert werden soll, zu erfassen – also ein epistemisches Privileg. Praktischerweise deckt sich die Herrschaft der Expert*innen mit der Herrschaft von Big Tech:
„Eric Schmidt: Wenn diese Technologie breiter verfügbar wird, was sehr schnell der Fall sein wird, werden die Probleme noch viel größer werden. Mir wäre es viel lieber, wenn die Unternehmen vernünftige Grenzen festlegen würden.
Reporter: Es soll kein regulatorischer Rahmen sein, es soll vielleicht nicht einmal eine Art demokratische Abstimmung sein, es soll das Fachwissen innerhalb der Branche sein, das hilft, das zu klären?
Schmidt: Das wird zuerst die Industrie tun, denn es gibt keine Möglichkeit für Nicht-Insider*innen zu verstehen, was möglich ist. Es ist einfach zu neu, zu schwierig, es gibt nicht das nötige Fachwissen. Es gibt niemanden in der Regierung, der/die es richtig machen kann, aber die Industrie kann es in etwa richtig machen, und dann kann die Regierung eine Regulierungsstruktur um das herum aufbauen" (NBC News 2023, 9:03).

Literaturverzeichnis

Aggarwal, Alok. 2018. „The Birth of AI and the First AI Hype Cycle“. *KDnuggets* (Blog), 13. Februar 2018. https://www.kdnuggets.com/the-birth-of-ai-and-the-first-ai-hype-cycle.html.

AI Now Institute. 2023. „Antitrust and Competition: It’s Time for Structural Reforms to Big Tech“. *AI Now Institute* (Blog), 11. April 2023. https://ainowinstitute.org/publication/antitrust-and-competition.

Anderson, Elizabeth. 2017. *Private Government: How Employers Rule Our Lives (and Why We Don’t Talk About It)*. University Center for Human Values Series. Princeton, Oxford: Princeton University Press.

Apprich, Clemens. Im Erscheinen. „Ja-Sager“. In *ChatGPT und andere „Quatschmaschinen“: Gespräche mit Künstlicher Intelligenz,* herausgegeben von Anna Tuschling, Andreas Sudmann und Bernhard J. Dotzler. Bielefeld: Transcript.

Argyle, Lisa P., Ethan Busby, Joshua Gubler, Chris Bail, Thomas Howe, Christopher Rytting und David Wingate. 2023. „AI Chat Assistants Can Improve Conversations about Divisive Topics“. arXiv. https://arxiv.org/abs/2302.07268.

Bagdasaryan, Eugene und Vitaly Shmatikov. 2022. „Spinning Language Models: Risks of Propaganda-As-A-Service and Countermeasures“. In *2022 IEEE Symposium on Security and Privacy (SP)*, 769–86. https://doi.org/10.1109/SP46214.2022.9833572.

Bajohr, Hannes. 2018. „Harmonie und Widerspruch: Mit Judith N. Shklar gegen die ‚Ideologie der Einigkeit‘“. In *Distanzierung und Engagement: Wie politisch sind die Geisteswissenschaften?*, herausgegeben von Hendrikje Schauer und Marcel Lepper (Stuttgart/Weimar: Works & Nights 2018), 75–85.

———. 2021. „Wer sind wir? Warum künstliche Intelligenz immer ideologisch ist“. *Republik*, 06. April 2021. https://www.republik.ch/2021/04/06/warum-kuenstliche-intelligenz-immer-ideologisch-ist.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major und Shmargaret Shmitchell. 2021. „On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?“. In *FAccT ’21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. Association for Computing Machinery.

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein et al. 2021. „On the Opportunities and Risks of Foundation Models“. arXiv. http://arxiv.org/abs/2108.07258.

Bostrom, Nick. 2002. „Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards“. *Journal of Evolution and Technology* 9. http://jetpress.org/volume9/risks.html.

Bradford, Anu, Adam Chilton, Katerina Linos und Alexander Weaver. 2019. „The Global Dominance of European Competition Law Over American Antitrust Law“. *Journal of Empirical Legal Studies* 16 (4), 731–66. https://doi.org/10.1111/jels.12239.

Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee et al. 2023. „Sparks of Artificial General Intelligence: Early Experiments with GPT-4“. arXiv. https://doi.org/10.48550/arXiv.2303.12712.

Eubanks, Virginia. 2017. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York, NY: St. Martin’s Press.

Floridi, Luciano. 2023. „AI as Agency Without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models“. *Philosophy & Technology* 36 (1), 15. https://doi.org/10.1007/s13347-023-00621-y.

Fricker, Miranda. 2023. *Epistemische Ungerechtigkeit: Macht und die Ethik des Wissens*. Übersetzt von Antje Korsmeier. München: Beck.

Gitelman, Lisa, Hrsg. 2013. *„Raw Data“ Is an Oxymoron*. Cambridge, MA: MIT Press. https://doi.org/10.1080/1369118x.2014.920042.

Habermas, Jürgen. 2022. „Reflections and Hypotheses on a Further Structural Transformation of the Political Public Sphere“. *Theory, Culture & Society* 39 (4), 145–71. https://doi.org/10.1177/02632764221112341.

Heaven, Will Douglas. 2023. „The Open-Source AI Boom Is Built on Big Tech’s Handouts. How Long Will It Last?“. *MIT Technology Review*, 12. Mai 2023. https://www.technologyreview.com/2023/05/12/1072950/open-source-ai-google-openai-eleuther-meta/.

Jakesch, Maurice, Advait Bhat, Daniel Buschek, Lior Zalmanson und Mor Naaman. 2023. „Co-Writing with Opinionated Language Models Affects Users’ Views“. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–15. Hamburg: ACM. doi:10.1145/3544548.3581196.

Jin, Berber und Deepa Seetharaman. 2023. „Elon Musk Creates New Artificial Intelligence Company X.AI“. *The Wall Street Journal*, 14. April 2023. https://www.wsj.com/articles/elon-musks-new-artificial-intelligence-business-x-ai-incorporates-in-nevada-962c7c2f.

Klein, Naomi. 2023. „AI Machines Aren’t ‘Hallucinating’. But Their Makers Are“. *The Guardian*, 8. Mai 2023. https://www.theguardian.com/commentisfree/2023/may/08/ai-machines-hallucinating-naomi-klein.

Luitse, Dieuwertje und Wiebke Denkena. 2021. „The Great Transformer: Examining the Role of Large Language Models in the Political Economy of AI“. *Big Data & Society* 8 (2): 205395172110477. https://doi.org/10.1177/20539517211047734.

Madison, James. 1993. „The Federalist Nr. 10“. In *Die Federalist Papers*, übersetzt, eingeleitet und mit Anmerkungen versehen von Barbara Zehnpfennig. Darmstadt: Wissenschaftliche Buchgesellschaft, 93–100.

McCarthy, John, Marvin L. Minsky, Nathaniel Rochester und Claude E. Shannon. 2006. „A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence“. *AI Magazine* 27 (4), 12–14.

Miller, Erin, Roudabeh Kishi, Clionadh Raleigh und Caitriona Dowd. 2022. „An Agenda for Addressing Bias in Conflict Data“. *Scientific Data* 9 (1), 593. https://doi.org/10.1038/s41597-022-01705-8.

Mitchell, Melanie und David C. Krakauer. 2023. „The Debate Over Understanding in AI’s Large Language Models“. In *Proceedings of the National Academy of Sciences* 120 (13). https://doi.org/10.1073/pnas.2215907120.

Mouffe, Chantal. 2007. *Über das Politische: Wider die kosmopolitische Illusion*. Frankfurt am Main: Suhrkamp.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University.

NBC News, Reg. 2023. „‘Life and Death Decisions Are Being Made’ by Artificial Intelligence“. Youtube, 15. Mai 2023. Video, 10:42. https://www.youtube.com/watch?v=VgJnqJ9WxAw.

O’Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Pasquinelli, Matteo. 2019. „How a Machine Learns and Fails – A Grammar of Error for Artificial Intelligence“. *Spheres,* Nr. 5. http://spheres-journal.org/how-a-machine-learns-and-fails-a-grammar-of-error-for-artificial-intelligence/.

„Pause Giant AI Experiments: An Open Letter“. 2023. *Future of Life Institute* (Blog), 22. März 2023. https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

Perrigio, Billy. 2023a. „Elon Musk Is Bringing the Culture Wars to AI“. *Time*, 3. März 2023. https://time.com/6260185/elon-musk-ai-culture-wars/.

———. 2023b. „Exclusive: OpenAI Lobbied the E.U. to Water Down AI Regulation“. *Time*, 20. Juni 2023. https://time.com/6288245/openai-eu-lobbying-ai-act/.

Posner, Eric A. und E. Glen Weyl. 2018. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton: Princeton University Press.

Rozado, David. 2023. „RightWingGPT – An AI Manifesting the Opposite Political Biases of ChatGPT“. Substack newsletter. *Rozado’s Visual Analytics* (Blog), 16. Februar 2023. https://davidrozado.substack.com/p/rightwinggpt.

Shklar, Judith N. 1986. *Legalism: Law, Morals, and Political Trials*. Cambridge: Harvard University Press.

———. 2021. *Über Ungerechtigkeit. Erkundungen zu einem moralischen Gefühl*. Berlin: Matthes und Seitz.

Simonite, Tom. 2021. „What Really Happened When Google Ousted Timnit Gebru“. *Wired*, 8. Juni 2021. https://www.wired.com/story/google-timnit-gebru-ai-what-really-happened/.

Srnicek, Nick. 2017. *Platform Capitalism*. London: Polity.

Stark, Luke. 2023. „Breaking Up (with) AI Ethics“. *American Literature* 95, Nr. 2, 365–79. doi:10.1215/00029831-10575148.

Taylor, Astra. 2014. *The People’s Platform: Taking Back Power and Culture in the Digital Age*. Toronto: Random House Canada.

Torres, Émile P. 2021. „Against Longtermism“. *Aeon*, 19. Oktober 2021. https://aeon.co/essays/why-longtermism-is-the-worlds-most-dangerous-secular-credo.

Vincent, James. 2023a. „AI Is Entering an Era of Corporate Control“. *The Verge*, 3. April 2023. https://www.theverge.com/23667752/ai-progress-2023-report-stanford-corporate-control.

———. 2023b. „The Senate’s Hearing on AI Regulation Was Dangerously Friendly“. *The Verge*, 19. Mai 2023. https://www.theverge.com/2023/5/19/23728174/ai-regulation-senate-hearings-regulatory-capture-laws.

Whittaker, Meredith. 2021. „The Steep Cost of Capture“. *Interactions* 28 (6), 50–55. https://doi.org/10.1145/3488666.

Winner, Langdon. 1980. „Do Artifacts Have Politics?“. *Daedalus* 109, Nr. 1, 121–36.

Weatherby, Leif. 2023. „ChatGPT Is an Ideology Machine“. *Jacobin*, 17. April 2023. https://jacobin.com/2023/04/chatgpt-ai-language-models-ideology-media-production.

Yudkowsky, Eliezer. 2023. „The Open Letter on AI Doesn’t Go Far Enough: We Need to Shut It All Down“. *Time*, 29. März 2023. https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/.

Zakrzewski, Cat, Nitasha Tiku, Cristiano Lima und Will Oremus. 2023. „OpenAI CEO Tells Senate That He Fears AI’s Potential to Manipulate Views“. *Washington Post*, 16. Mai 2023. https://www.washingtonpost.com/technology/2023/05/16/ai-congressional-hearing-chatgpt-sam-altman/.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Public Affairs.

Further Reading

We would like to provide a list of books and media that we believe can help those interested delve deeper into the complex topic of AI and its critical reflection. A non-complete list of books that inspired us, projects that caught our eye, podcasts we listened to and documentaries we watched. We refer to the respective original titles.

Books

Arns, Inke, Francis Hunger, and Marie Lechner. 2022. *House of Mirrors: Artificial Intelligence as Phantasma*. Dortmund: Kettler.

Atanasoski, Neda, and Kalindi Vora. 2019. *Surrogate Humanity: Race, Robots, and the Politics of Technological Futures*. Durham: Duke University Press.

Bajohr, Hannes. 2018. *Halbzeug*. Berlin: Suhrkamp.

———. 2022. *Schreibenlassen. Texte zur Literatur im Digitalen*. Berlin: August Verlag.

Berger, John. 1972. *Ways of Seeing*. Harmondsworth, London: Penguin Books.

Blackwell, Alan F., Emma Cocker, Geoff Cox, Alex McLean, and Thor Magnusson. 2022. *Live Coding: A User's Manual*. Cambridge, MA: The MIT Press.

Brayne, Sarah. 2021. *Predict and Surveil: Data, Discretion, and the Future of Policing*. New York, NY: Oxford University Press.

Couldry, Nick, and Ulises A. Mejias. 2019. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford, CA: Stanford University Press.

Crawford, Kate. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press.

Dekker, Annet, ed. 2021. *Curating Digital Art: From Presenting and Collecting Digital Art to Networked Co-curation*. Amsterdam: Valiz.

Dewdney, Andrew, and Katrina Sluis, eds. 2023. *The Networked Image in Post-digital Culture*. London: Routledge.

Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor*. New York, NY: St. Martin's Press.

Gitelman, Lisa, ed. 2013. *"Raw Data" Is an Oxymoron*. Cambridge, MA: The MIT Press.

Gröner, Stefan, and Stephanie Heinecke. 2019. *Kollege KI. Künstliche Intelligenz verstehen und sinnvoll im Unternehmen einsetzen*. München: Redline.

Hicks, Mar. 2018. *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing*. Cambridge, MA: The MIT Press.

Hu, Tung-Hui. 2022. *Digital Lethargy: Dispatches from the Age of Disconnection*. Cambridge, MA: The MIT Press.

Krämer, Sybille. 2008. *Medium, Bote, Übertragung. Kleine Metaphysik der Medialität*. Frankfurt/Main: Suhrkamp.

Landwehr, Dominik, ed. 2018. *Machines and Robots*. Basel: Christoph Merian Verlag.

Manovich, Lev. 2020. *Cultural Analytics*. Cambridge, MA: The MIT Press.

Meyer, Roland. 2021. *Gesichtserkennung*. Berlin: Verlag Klaus Wagenbach.

Mullaney, Thomas S., Benjamin Peters, Mar Hicks, and Kavita Philip, eds. 2021. *Your Computer Is on Fire*. Cambridge, MA: The MIT Press.

Noble, Safiya U. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York, NY: NYU Press.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Harmondsworth, London: Penguin Books.

———. 2022. *The Shame Machine*. New York, NY: Crown.

Ohm, Tillmann. 2018. *The Artist's Machine*. https://thesiscommons.org/tj6yf/.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Pasquinelli, Matteo. 2023. *The Eye of the Master: A Social History of Artificial Intelligence*. London: Verso.

Pereira, Gabriel. 2021. *Struggling with the Algorithmic Seeing: Hegemonic Computer Vision and Antagonistic Practices*. Aarhus University: Dissertation.

Stalder, Felix. 2016. *Kultur der Digitalität*. Berlin: Suhrkamp.

Thylstrup, Nanna Bonde, Daniela Agostinho, Annie Ring, Catherine D'Ignazio, and Kristin Veel, eds. 2020. *Uncertain Archives: Critical Keywords for Big Data*. Cambridge, MA: The MIT Press.

Tyżlik-Carver, Magda. 2016. *Curating in/as Common/s. Posthuman Curating and Computational Culture*. Aarhus University: Dissertation.

Wasielewski, Amanda. 2023. *Computational Formalism: Art History and Machine Learning*. Cambridge, MA: The MIT Press.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.

Online

AI: A Museum Planning Toolkit (2019–22), The Museums + AI Network. https://themuseumsai.network/toolkit/.

Anatomy of an AI System (2018), Vladan Joler and Kate Crawford. https://anatomyof.ai/.

Creative AI Lab (2019–), Serpentine Galleries and King's College London. https://creative-ai.org/.

Elements of AI (2018–), MinnaLearn and The University of Helsinki. https://www.elementsofai.com/.

Excavating.ai (2021), Kate Crawford and Trevor Paglen. https://excavating.ai/.

Exposing.ai (2021–), Adam Harvey and Jules LaPlace. https://exposing.ai/.

KI für Alle (2022–), Heinrich-Heine-Universität Düsseldorf. https://www.heicad.hhu.de/lehre/ki-fuer-alle.

openHPI (2012–), Hasso-Plattner-Institut. https://open.hpi.de/channels/ai-service-center.

Public Interest AI (2022–), Alexander von Humboldt Institute for Internet and Society. https://publicinterest.ai/.

The Nooscope Manifested (2020), Vladan Joler and Matteo Pasquinelli. http://nooscope.ai.

Software

AI.LAB (2023–), Fraunhofer Institut IAIS Sankt Augustin. https://www.ai-lab.nrw/.

Collection Space Navigator (2020–24), Tillmann Ohm et al. https://collection-space-navigator.github.io/.

The Curator's Machine (2020–23), Ludwig Forum Aachen, HMKV Hartware MedienKunstVerein, Dortmund, and RWTH Aachen University. https://github.com/VCI-RWTH/TrainingTheArchive.

Digital Curator (2019–22), Lukas Pilka. https://digitalcurator.art/.

iArt (2018–23), TIB, Paderborn University, and Ludwig-Maximilians-Universität München. https://www.iart.vision/.

ImageGraph (2020–), Leonardo Impett. https://www.imagegraph.cc/.

imgs.ai (2020–), Fabian Offert and Peter Bell. https://imgs.ai/.

openArtBrowser (2023–), Hochschule Darmstadt University of Applied Sciences. https://openartbrowser.org/.

Media

Hogan, Mél. 2023. "The Data Fix." Podcast of conversations with scholars, thinkers and feelers. Audio. https://www.thedatafix.net/episodes.

Kantayya, Shalini. 2020. "Coded Bias." 7th Empire Media, New York, NY. Documentary, 90:00. https://www.codedbias.com/.

Leonardi, Massimo. 2020. "Naked AI: What happens when Artificial Intelligence and human creativity meet?" Cisco Italia and Logotel, Milan, IT. Documentary, 57:25. https://naked-ai.com/.

Pankow, Mads. 2021. "Mensch, Maschine!" Podcast on the topic of artificial intelligence, people and society. Audio. http://www.menschmaschine.org.

Schweinberger, Julia. 2019. "Wenn die Zukunft zu klug für uns wird." BR Fernsehen, Munich, DE. Documentary, 43:02. https://www.br.de/fernsehen/ard-alpha/programmkalender/sendung-2832774.html.

Witz, Bernhard. 2022. "Digitale Sammlungen." Kunstmuseum Basel. Forum. https://www.vernetzt.museum/forum/.

Weiterführende Medien

Wir möchten an dieser Stelle eine Liste an Büchern und Medien zur Verfügung stellen, von der wir glauben, dass sie helfen kann, in das komplexe Thema KI und dessen kritische Betrachtung tiefer einzusteigen. Eine unvollständige Liste an Büchern, die uns begeisterten, Projekten, die uns aufgefallen sind, Podcasts, die wir gehört oder Dokumentationen, die wir gesehen haben. Wir zitieren die Originaltitel.

Bücher

Arns, Inke, Francis Hunger und Marie Lechner. 2022. *House of Mirrors: Künstliche Intelligenz als Phantasma*. Dortmund: Kettler.

Atanasoski, Neda und Kalindi Vora. 2019. *Surrogate Humanity: Race, Robots, and the Politics of Technological Futures*. Durham: Duke University Press.

Bajohr, Hannes. 2018. *Halbzeug*. Berlin: Suhrkamp.

———. 2022. *Schreibenlassen. Texte zur Literatur im Digitalen*. Berlin: August Verlag.

Berger, John. 1972. *Ways of Seeing*. Harmondsworth, London: Penguin Books.

Blackwell, Alan F., Emma Cocker, Geoff Cox, Alex McLean und Thor Magnusson. 2022. *Live Coding: A User's Manual*. Cambridge, MA: The MIT Press.

Brayne, Sarah. 2021. *Predict and Surveil: Data, Discretion, and the Future of Policing*. New York, NY: Oxford University Press.

Couldry, Nick und Ulises A. Mejias. 2019. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford, CA: Stanford University Press.

Crawford, Kate. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press.

Dekker, Annet, Hrsg. 2021. *Curating Digital Art: From Presenting and Collecting Digital Art to Networked Co-curation*. Amsterdam: Valiz.

Dewdney, Andrew und Katrina Sluis, Hrsg. 2023. *The Networked Image in Post-digital Culture*. London: Routledge.

Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor*. New York, NY: St. Martin's Press.

Gitelman, Lisa, Hrsg. 2013. *"Raw Data" Is an Oxymoron*. Cambridge, MA: The MIT Press.

Gröner, Stefan und Stephanie Heinecke. 2019. *Kollege KI. Künstliche Intelligenz verstehen und sinnvoll im Unternehmen einsetzen*. München: Redline.

Hicks, Mar. 2018. *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing*. Cambridge, MA: The MIT Press.

Hu, Tung-Hui. 2022. *Digital Lethargy: Dispatches from the Age of Disconnection*. Cambridge, MA: The MIT Press.

Krämer, Sybille. 2008. *Medium, Bote, Übertragung. Kleine Metaphysik der Medialität*. Frankfurt/Main: Suhrkamp.

Landwehr, Dominik, Hrsg. 2018. *Machines and Robots*. Basel: Christoph Merian Verlag.

Manovich, Lev. 2020. *Cultural Analytics*. Cambridge, MA: The MIT Press.

Meyer, Roland. 2021. *Gesichtserkennung*. Berlin: Verlag Klaus Wagenbach.

Mullaney, Thomas S., Benjamin Peters, Mar Hicks und Kavita Philip, Hrsg. 2021. *Your Computer Is on Fire*. Cambridge, MA: The MIT Press.

Noble, Safiya U. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York, NY: NYU Press.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Harmondsworth, London: Penguin Books.

———. 2022. *The Shame Machine*. New York, NY: Crown.

Ohm, Tillmann. 2018. *The Artist's Machine*. https://thesiscommons.org/tj6yf/.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Pasquinelli, Matteo. 2023. *The Eye of the Master: A Social History of Artificial Intelligence*. London: Verso.

Pereira, Gabriel. 2021. *Struggling with the Algorithmic Seeing: Hegemonic Computer Vision and Antagonistic Practices*. Aarhus University: Dissertation.

Stalder, Felix. 2016. *Kultur der Digitalität*. Berlin: Suhrkamp.

Thylstrup, Nanna Bonde, Daniela Agostinho, Annie Ring, Catherine D'Ignazio und Kristin Veel, Hrsg. 2020. *Uncertain Archives: Critical Keywords for Big Data*. Cambridge, MA: The MIT Press.

Tyżlik-Carver, Magda. 2016. *Curating in/as Common/s. Posthuman Curating and Computational Culture*. Aarhus University: Dissertation.

Wasielewski, Amanda. 2023. *Computational Formalism: Art History and Machine Learning*. Cambridge, MA: The MIT Press.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.

Online

AI: A Museum Planning Toolkit (2019–22), The Museums + AI Network. https://themuseumsai.network/toolkit/.

Anatomy of an AI system (2018), Vladan Joler und Kate Crawford. https://anatomyof.ai/.

Creative AI Lab (2019–), Serpentine Galleries und King's College London. https://creative-ai.org/.

Elements of AI (2018–), MinnaLearn und The University of Helsinki. https://www.elementsofai.com/.

Excavating.ai (2021), Kate Crawford und Trevor Paglen. https://excavating.ai/.

Exposing.ai (2021–), Adam Harvey und Jules LaPlace. https://exposing.ai/.

KI für Alle (2022–), Heinrich-Heine-Universität Düsseldorf. https://www.heicad.hhu.de/lehre/ki-fuer-alle.

openHPI (2012–), Hasso-Plattner-Institut. https://open.hpi.de/channels/ai-service-center.

Public Interest AI (2022–), Alexander von Humboldt Institute for Internet and Society. https://publicinterest.ai/.

The Nooscope Manifested (2020), Vladan Joler und Matteo Pasquinelli. http://nooscope.ai.

Software

AI.LAB (2023–), Fraunhofer Institut IAIS Sankt Augustin. https://www.ai-lab.nrw/.

Collection Space Navigator (2020–24), Tillmann Ohm et al. https://collection-space-navigator.github.io/.

The Curator's Machine (2020–23), Ludwig Forum Aachen, HMKV Hartware MedienKunstVerein, Dortmund und RWTH Aachen University. https://github.com/VCI-RWTH/TrainingTheArchive.

Digital Curator (2019–22), Lukas Pilka. https://digitalcurator.art/.

iArt (2018–23), TIB, Paderborn University und Ludwig-Maximilians-Universität München. https://www.iart.vision/.

ImageGraph (2020–), Leonardo Impett. https://www.imagegraph.cc/.

imgs.ai (2020–), Fabian Offert und Peter Bell. https://imgs.ai/.

openArtBrowser (2023–), Hochschule Darmstadt University of Applied Sciences. https://openartbrowser.org/.

Medien

Hogan, Mél. 2023. „The Data Fix". Podcast zu Gesprächen mit Akademiker*innen, Denker*innen und Gefühlsmenschen. Audio. https://www.thedatafix.net/episodes.

Kantayya, Shalini. 2020. „Coded Bias". 7th Empire Media, New York, NY. Dokumentation, 90:00. https://www.codedbias.com/.

Leonardi, Massimo. 2020. „Naked AI: What happens when Artificial Intelligence and human creativity meet?" Cisco Italia und Logotel, Mailand, IT. Dokumentation, 57:25. https://naked-ai.com/.

Pankow, Mads. 2021. „Mensch, Maschine!" Podcast zum Thema Künstliche Intelligenz, Mensch und Gesellschaft. Audio. http://www.menschmaschine.org.

Schweinberger, Julia. 2019. „Wenn die Zukunft zu klug für uns wird". BR Fernsehen, München, DE. Dokumentation, 43:02. https://www.br.de/fernsehen/ard-alpha/programmkalender/sendung-2832774.html.

Witz, Bernhard. 2022. „Digitale Sammlungen". Kunstmuseum Basel. Forum. https://www.vernetzt.museum/forum/.

Contributors

Inke Arns
Inke Arns has been the director of HMKV Hartware MedienKunstVerein, Dortmund, since 2005. She is also a freelance curator and author specialising in media art and theory, net cultures and Eastern Europe. Since 2021, she has been a visiting professor for curatorial practice at the Kunstakademie Münster. More at: inkearns.de and hmkv.de.

Hannes Bajohr
Hannes Bajohr holds a postdoc position at the Seminar for Media Studies at the University of Basel. His work focuses on theories of the digital, political theory, and 20th century history of ideas. Recent publications include *Schreibenlassen. Texte zur Literatur im Digitalen* (Berlin 2022) and *(Berlin, Miami)* (Berlin 2023). Next year will see the publication of *Postartifizielle Texte: Schreiben nach KI*.

Eva Birkenstock
Eva Birkenstock is trained as an art historian and cultural anthropologist. Since October 2021, she has been director of the Ludwig Forum Aachen.

Dominik Bönisch
Dominik Bönisch studied Cultural Studies at the University of Hildesheim and Moholy-Nagy University Budapest. He is currently questioning the connections between artificial intelligence and curating as the scientific project manager of the research project *Training the Archive* at the Ludwig Forum Aachen. As a curatorial assistant, Bönisch also investigated virtual reality in the exhibition *Thrill of Deception. From Ancient Art to Virtual Reality*. His research interests focus on the study of new technologies applied to the arts, as well as to museum collections and exhibition formats. Since 2019, Bönisch has been lecturing on these topics at the HSD Hochschule Düsseldorf.

Eva Cetinić
Eva Cetinić is a postdoctoral fellow at the Center for Digital Visual Studies, hosted by the University of Zurich. Her research interests focus on exploring deep learning techniques for computational image understanding and multimodal reasoning in the context of visual art.

Vincent Christlein
Vincent Christlein studied Computer Science and received his PhD (Dr.-Ing.) from the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). Since 2018, he has been working as a research associate at the Pattern Recognition Lab (FAU) and was promoted to Academic Councilor in 2020, heading the Computer Vision group.

Nick Couldry
Nick Couldry is professor of media communications and social theory at the London School of Economics and Political Science and since 2017 has been a faculty associate at Harvard's Berkman Klein Center for Internet & Society. His research interests include media and communications, culture and power and social theory and the consequences for everyday reality of concentrating symbolic power in particular institutions. Together with Ulises Mejias, Couldry wrote the book *The Costs of Connection. How Data is Colonizing Human Life and Appropriating it for Capitalism*, published in 2019 by Stanford University Press.

Geoff Cox
Geoff Cox is Professor of Art and Computational Culture at London South Bank University, co-Director of Centre for the Study of the Networked Image, co-Director of MA Curating Art and Public Programmes (in collaboration with Whitechapel Gallery) and Adjunct at Aarhus University.

Tim Elsner
Tim Elsner is a machine learning researcher at the Visual Computing Institute at RWTH Aachen University, working on generative models and neural radiance fields. His research focuses on using these models for representing and editing 3D geometry.

Elisa Giardina Papa
Elisa Giardina Papa is an Italian artist whose work investigates gender, sexuality, care, and labor in relation to neoliberal capitalism and the borders of the Global South. Her work has been exhibited and screened at the 59th Venice Biennale, MoMA, Whitney Museum, Gropius Bau, Seoul Mediacity Biennale 2018, BFI London, Flaherty NYC, M+ Hong Kong, ICA Milano, and many more. Giardina Papa is an Assistant Professor of Modern Culture and Media at Brown University.

Katrin Glinka
Katrin Glinka is a cultural scientist who researches digital technologies in cultural contexts, data visualisation and human-computer interaction. She heads the HCC Data Lab in the Human-Centered Computing research group at Freie Universität Berlin and is working on her PhD at Humboldt University of Berlin.

Adam Harvey
Adam Harvey is a researcher and artist based in Berlin, focused on computer vision, privacy and surveillance. He's a graduate of the interactive Telecommunications Program at New York University and he's currently a digital fellow at Weizenbaum Institute, a research fellow at Karlsruhe Hochschule für Gestaltung and a future fellow with Eyebeam Rapid Response for a Better Digital Future. Harvey is also founder of the *VFRAME* computer vision project and cofounder of *MegaPixels*.

Mar Hicks
Mar Hicks is an author, historian, and professor at Illinois Institute of Technology in Chicago doing research on the history of computing and labor, as well as Queer Science, and Technology Studies. Hick's award-winning first book *Programmed Inequality* was published by MIT Press in 2017.

Mél Hogan
Mél Hogan is the Director of the Environmental Media Lab and an Associate Professor in Communication, Media and Film at the University of Calgary (since 2016). Hogan is also host and editor of *The Data Fix* podcast. Since 2012, Hogan's research has been on data storage/archives/repositories, the political-social implications and environmental impacts of server farms and data centers, culminating in the *Data Center Industrial Complex* (2021) and *Critical Studies of the Cloud* (2022). As an extension of research that looks at alternatives to water- and energy-intensive data storage, Hogan is PI for a SSHRC IG project (2021–26) about DNA-based data storage following an IDG project (2018–21) about genomics in the cloud. More at: melhogan.com.

Francis Hunger
Francis Hunger's practice combines artistic research and media theory with the capabilities of narration through installations, radio plays and performances, and Internet-based art. Currently he is a researcher for the project *Training the Archive* at Hartware MedienKunstVerein, Dortmund, critically examining the use of AI, statistics and pattern recognition for art and curating. In 2022 he co-curated with Inke Arns and Marie Lechner the exhibition *House of Mirrors: Artificial Intelligence as Phantasm* at HMKV, Dortmund. His PhD at Bauhaus University Weimar developed a media archeological genealogy of database technology and practices. In 2022–23 Hunger was guest professor at the Intermedia program of the Hungarian Academy for Visual Arts, Budapest. Hunger's artistic work is exhibited internationally. Numerous festival participations, talks, lectures, publications, screenings and academic lectures. He occasionally curates exhibitions, teaches at universities regularly, and publishes daily on twitter: irmielin.org.

Moritz Ibing
Moritz Ibing is a machine learning engineer at Gixel, where he is working on the development of 3D avatars. Previously he worked at the Visual Computing Institute at RWTH Aachen University, where his research was focused on machine learning approaches for the representation and generation of 3D data.

Maya Indira Ganesh
Maya Indira Ganesh is Course Co-leader of the Master of AI Ethics and Society program at Leverhulme Centre for the Future of Intelligence within the University of Cambridge and a media and digital cultures theorist, researcher, and writer. Her dissertation examined the re-shaping of what we mean by the 'ethical' and the shifting role of the human in the emergence of the driverless car.

Leif Kobbelt
Leif Kobbelt is distinguished professor of Computer Science at RWTH Aachen University and head of the Institute for Computer Graphics and Multimedia. After his PhD in 1994 from Karlsruhe Institute of Technology he worked at University of Wisconsin in Madison, University of Erlangen-Nürnberg, and the Max Planck Institute of Computer Science before he moved to RWTH in 2001. His major research interests include 3D reconstruction, efficient geometry processing and shape analysis. He has received a number of academic awards including an ERC Advanced Grant in 2013 and the Gottfried Wilhelm Leibniz Prize in 2014. He has been named a Fellow of the Eurographics Association (2008) as well as a Distinguished Professor (2013) and a Fellow (2019) of RWTH Aachen University. In 2015 he became a member of the Academia Europaea and in 2016 a member of the North Rhine Westphalian Academy of Sciences.

Sybille Krämer
Sybille Krämer (Emerita) was Professor of Philosophy at FU Berlin. In addition to visiting professorships in Tokyo, Yale, Santa Barbara, Santiago de Chile, Vienna and Zurich, she is a senior professor at Leuphana University Lüneburg. She was a member of the German Science and Humanities Council, the European Research Council and the Senate of the DFG.

Mattis Kuhn
Mattis Kuhn is an artist, researcher, and curator. He works on the reciprocal design of humans, machines, and the environment, primarily by means of (executable) text. He is an artistic associate for Creative Coding at the Bauhaus-Universität Weimar and part of the research group ground zero at KHM Cologne.

Isaak Lim
Isaak Lim is a researcher at the Visual Computing Institute with the Graphics, Geometry & Multimedia Group, where he works with Prof. Leif Kobbelt. He received his PhD in Computer Science from the RWTH Aachen University in 2021. His research area is the domain of shape analysis, where he is interested in applying machine learning techniques to geometry processing tasks.

Ulises Mejias
Ulises Mejias is a professor of Commuication Studies and director of the Institute of Global Engagement at the State University of New York at Oswego. His research interests include critical data studies, philosophy and sociology of technology, and the political economy of digital media. Together with Nick Couldry, Mejias wrote the book *The Costs of Connection. How Data is Colonizing Human Life and Appropriating it for Capitalism*, published in 2019 by Stanford University Press.

Roland Meyer
Roland Meyer is a media and visual culture scholar. He is currently researching virtual image archives as a member of the CRC 1567 *Virtual Life Worlds* at Ruhr University Bochum. In 2021, his book *Gesichtserkennung* (Face Recognition) was published in the series Digitale Bildkulturen (Digital Image Cultures) by Wagenbach Verlag.

Off Office
Johannes von Gross, Markus Lingemann, and their team at the Munich-based design studio Off Office create visual identities, websites, books, and exhibition design for arts and culture. Among others, for the Haus der Kunst and the Pinakothek der Moderne in Munich, the Städel Museum in Frankfurt, and the Salzburg Easter Festival. The design for *Training the Archive* at the Ludwig Forum Aachen was awarded in the competition *100 Best Posters* in 2023 and exhibited internationally.

Fabian Offert
Fabian Offert is Assistant Professor for the History and Theory of the Digital Humanities at the University of California, Santa Barbara. His research and teaching focuses on the visual digital humanities, with a special interest in the epistemology and aesthetics of computer vision and machine learning.

Tillmann Ohm
Tillmann Ohm is a Creative Technologist with focus on curatorial systems. He develops experimental tools and methods for the analysis and interaction with cultural collections through his venture *ARCU. technology*. Ohm holds a diploma in Fine Art from the Bauhaus-Universität Weimar and worked as a Cultural Data Analytics Research Fellow at Tallinn University.

Mads Pankow
Mads Pankow is a freelance author and moderator on topics related to artificial intelligence. He is the founder of *Die Epilog*, a magazine for contemporary culture, and from 2014 to 2019 he organized the annual *Digital Bauhaus*, a conference at the nexus of technology, design, and society in Weimar.

Matteo Pasquinelli
Matteo Pasquinelli is a professor in media philosophy at the University of Arts and Design Karlsruhe, where he is coordinating the research group on Artificial Intelligence and Media Philosophy KIM. His research focuses on the intersection of cognitive sciences, the digital economy, and machine intelligence.

Gabriel Pereira
Gabriel Pereira has just completed his PhD at Aarhus University in Denmark with a thesis titled *Struggling with the Algorithmic Seeing: Hegemonic Computer Vision and Antagonistic Practices*. His research investigates data and algorithm infrastructures, especially how computer vision algorithms mediate our relationship with the world. Projects by Pereira have been exhibited in venues such as the 33rd São Paulo Art Biennial, the Van Abbemuseum, the IDFA DocLab, and Itaú Cultural.

Anna Ridler
Anna Ridler is an artist and researcher who is particularly interested in ideas around measurement and quantification and how this relates to the natural world. Ridler holds an MA in information experience design from the Royal College of Art, along with fellowships at the Creative Computing Institute at University of the Arts London.

Marian Schneider
Marian Schneider is a Computer Science Bachelor student at RWTH Aachen University. Since 2022, he has been working at the Visual Computing Institute as a student assistant on the project *Training the Archive*. He is interested in machine learning in the context of computer graphics.

Alexa Steinbrück
Alexa Steinbrück studied Fine Arts in Dresden and in the south of France and then earned a degree in Artificial Intelligence at the University of Amsterdam. From 2020 to 2022, she was researcher at XLAB, the Artificial Intelligence and Robotics Lab at Burg Giebichenstein University of Art and Design in Halle, and she has recently started working at the AI and Design Lab of the University of Design in Schwäbisch-Gmünd.

Giulia Taurino
Giulia Taurino is a Postdoctoral Associate at Northeastern University. Her research focuses on forms of content organization on online platforms and digital archives, cultural implications of algorithmic technologies, and applications of artificial intelligence in the arts, culture and heritage sector. From 2019 to 2022, she developed a series of computational art projects, exploring the intersection between AI and curatorial practices in museums. Past and present affiliations include NULab for Texts, Maps, and Networks, MIT Data + Feminism Lab, metaLAB (at) Harvard, Brown University's Virtual Humanities Lab. Taurino received her PhD in Media Studies and Visual Arts from the University of Bologna and the University of Montreal.

Gaia Tedone
Gaia Tedone is a curator and researcher with an expansive interest in the technologies and apparatuses of image formation. In 2019 she completed her PhD at the Centre for the Study of the Networked Image. She writes and teaches on the topics of digital culture and post-critical museology and curates independently.

Magda Tyżlik-Carver
Magda Tyżlik-Carver researches relational networks of relationships between humans and the non-human. She relates these notions to concepts of the posthuman as curatorial entities. That is, to algorithms, bots, software, and computing infrastructure. Tyżlik-Carver currently teaches at Aarhus University in the Department for Digital Design and Information Studies and published her dissertation in 2016 under the title *Curating in/as Common/s. Posthuman Curating and Computational Culture*.

Yvonne Zindel
Yvonne Zindel works as a curator and mediator especially on possibilities of digitality and sustainability. She deals with NFTs, blockchain and degrowth, as well as the possibilities for decolonial, anti-racist, feminist, and inclusive curating and mediating.

Mitwirkende

Inke Arns
Inke Arns ist seit 2005 Direktorin des HMKV Hartware MedienKunstVerein, Dortmund. Daneben ist sie freie Kuratorin und Autorin mit den Schwerpunkten Medienkunst und -theorie, Netzkulturen und Osteuropa. Seit 2021 ist sie Gastprofessorin für Kuratorische Praxis an der Kunstakademie Münster. Mehr auf: inkearns.de und hmkv.de.

Hannes Bajohr
Hannes Bajohr ist Postdoc am Seminar für Medienwissenschaft der Universität Basel. Er arbeitet zu Theorien der digitalen, politischen Theorie sowie der Ideengeschichte des 20. Jahrhunderts. Zuletzt erschienen: *Schreibenlassen. Texte zur Literatur im Digitalen* (Berlin 2022) und *(Berlin, Miami)* (Berlin 2023). Im kommenden Jahr erscheint *Postartifizielle Texte. Schreiben nach KI*.

Eva Birkenstock
Eva Birkenstock ist ausgebildete Kunsthistorikerin und Kulturanthropologin und leitet seit Oktober 2021 das Ludwig Forum Aachen.

Dominik Bönisch
Dominik Bönisch studierte Kulturwissenschaften an der Universität Hildesheim und der Moholy-Nagy University Budapest. Derzeit hinterfragt er als wissenschaftlicher Projektleiter des Forschungsprojekts *Training the Archive* am Ludwig Forum Aachen die Zusammenhänge von Künstlicher Intelligenz und dem Kuratieren. Als kuratorische Assistenz setzte sich Bönisch auch mit der Virtual Reality in der Ausstellung *Lust der Täuschung. Von antiker Kunst bis zur Virtual Reality* auseinander. Sein Forschungsinteresse liegt in den Auswirkungen neuer Technologie auf die Kunst sowie auf den musealen Sammlungs- und Ausstellungsbetrieb. Seit 2019 hält Bönisch Lehrveranstaltungen zum Thema an der HSD Hochschule Düsseldorf.

Eva Cetinić
Eva Cetinić ist Postdoktorandin am Center for Digital Visual Studies der Universität Zürich. Ihre Forschungsinteressen konzentrieren sich auf die Erforschung von Deep-Learning-Techniken für ein computergestütztes Bildverständnis und für eine multimodale Argumentation im Kontext visueller Kunst.

Vincent Christlein
Vincent Christlein studierte Informatik und promovierte zum Dr.-Ing. an der Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). Seit 2018 arbeitet er als wissenschaftlicher Mitarbeiter am Lehrstuhl für Mustererkennung (FAU), wo er 2020 zum Akademischen Rat befördert wurde und die Gruppe Computer Vision leitet.

Nick Couldry
Nick Couldry ist Professor für Medienkommunikation und Gesellschaftstheorie an der London School of Economics and Political Science und seit 2017 Fakultätsmitglied am Berkman Klein Center for Internet & Society an der Harvard University. Seine Forschungsinteressen umfassen Medien und Kommunikation, Kultur und Macht sowie Gesellschaftstheorie und die Folgen der Konzentration symbolischer Macht in bestimmten Institutionen des Alltags. Gemeinsam mit Ulises Mejias hat Couldry das Buch *The Costs of Connection. How Data is Colonizing Human Life and Appropriating it for Capitalism* verfasst, das 2019 bei Stanford University Press erschienen ist.

Geoff Cox
Geoff Cox ist Professor für Kunst und Computerkultur an der London South Bank University, Co-Direktor des Centre for the Study of the Networked Image, Co-Direktor des MA Curating Art and Public Programmes (in Zusammenarbeit mit der Whitechapel Gallery) und Lehrbeauftragter an der Universität Aarhus.

Tim Elsner
Tim Elsner ist Forscher für maschinelles Lernen am Visual Computing Institute der RWTH Aachen University und arbeitet an generativen Modellen und neuronalen Strahlungsfeldern. Seine Forschung konzentriert sich auf die Verwendung dieser Modelle zur Darstellung und Bearbeitung von 3D-Geometrie.

Elisa Giardina Papa
Elisa Giardina Papa ist eine italienische Künstlerin, die in ihrem Schaffen Geschlecht, Sexualität, Care und Arbeit in Bezug auf den neoliberalen Kapitalismus und die Grenzen des Globalen Südens untersucht. Ihre Arbeiten wurden unter anderem bei der 59. Biennale Venedig, im MoMA, Whitney Museum, Gropius Bau, Seoul Mediacity Biennale 2018, BFI London, Flaherty NYC, M+ Hong Kong und ICA Milano ausgestellt. Giardina Papa ist Assistant Professor für Modern Culture and Media an der Brown University.

Katrin Glinka
Katrin Glinka ist Kulturwissenschaftlerin und forscht zu digitalen Technologien im kulturellen Kontext, Datenvisualisierung und Human-Computer Interaction. Sie leitet das HCC Data Lab in der Forschungsgruppe Human-Centered Computing an der Freien Universität Berlin und promoviert an der Humboldt-Universität zu Berlin.

Adam Harvey
Adam Harvey ist ein in Berlin lebender Forscher und Künstler, der sich mit Computer Vision, Privatsphäre und Überwachung beschäftigt. Er ist Absolvent des interaktiven Telekommunikationsprogramms der New York University und ist derzeit Digital Fellow am Weizenbaum Institut, Forschungsstipendiat an der Hochschule für Gestaltung Karlsruhe und Future Fellow bei Eyebeam Rapid Response for a Better Digital Future. Er ist Gründer des *VFRAME* Computer Vision Projekts und Mitbegründer von *MegaPixels*.

Mar Hicks
Mar Hicks ist Autorin, Historikerin und Professorin am Illinois Institute of Technology in Chicago. Sie forscht zu den Themen Geschichte des Computers und der Arbeit sowie zu Queer Science und Technology Studies. Ihr preisgekröntes erstes Buch *Programmed Inequality* wurde 2017 bei MIT Press veröffentlicht.

Mél Hogan
Mél Hogan ist Direktorin des Environmental Media Lab und außerordentliche Professorin für Kommunikation, Medien und Film an der University of Calgary (seit 2016). Hogan ist außerdem Gastgeberin und Herausgeberin des Podcasts *The Data Fix*. Seit 2012 befasst sich Hogans Forschung mit Datenspeicherung/Archiven/Repositorien, den politisch-sozialen Implikationen und Umweltauswirkungen von Serverfarmen und Rechenzentren, die in *Data Center Industrial Complex* (2021) und *Critical Studies of the Cloud* (2022) gipfelten. Als Erweiterung der Forschung, die sich mit Alternativen zur wasser- und energieintensiven Datenspeicherung befasst, ist Hogan Leiterin des SSHRC IG-Projekts (2021–26) über DNA-basierte Datenspeicherung im Anschluss an ein IDG-Projekt (2018–21) über Genomik in der Cloud. Mehr unter: melhogan.com.

Francis Hunger
Francis Hungers Praxis verbindet künstlerische Forschung und Medientheorie mit den Möglichkeiten der Narration durch Installationen, Hörspiele, Performances und internetbasierte Kunst. Derzeit ist er wissenschaftlicher Mitarbeiter des Projekts *Training the Archive* im Hartware MedienKunstVerein, Dortmund, und setzt sich kritisch mit der Nutzung von KI, Statistik und Mustererkennung für Kunst und das Kuratieren auseinander. Im Jahr 2022 kuratierte er zusammen mit Inke Arns und Marie Lechner die Ausstellung *House of Mirrors: Künstliche Intelligenz als Phantasma* im HMKV, Dortmund. In seiner Promotion an der Bauhaus-Universität Weimar entwickelte er eine medienarchäologische Genealogie von Datenbanktechnologien und -praktiken. In 2022/23 war Hunger Gastprofessor am Intermedia-Programm der Ungarischen Akademie für Bildende Künste, Budapest. Hungers künstlerische Arbeit wird international ausgestellt. Zahlreiche Festivalteilnahmen, Vorträge, Vorlesungen, Publikationen, Screenings und akademische Vorlesungen. Er kuratiert gelegentlich Ausstellungen, unterrichtet regelmäßig an Hochschulen und publiziert täglich auf Twitter: irmielin.org.

Moritz Ibing
Moritz Ibing ist ein Machine Learning Ingenieur bei Gixel, wo er an der Entwicklung von 3D-Avataren arbeitet. Zuvor war er am Visual Computing Institute der RWTH Aachen University tätig, wo sich seine Forschung auf maschinelles Lernen für die Darstellung und Generierung von 3D-Daten fokussiert hat.

Maya Indira Ganesh
Maya Indira Ganesh ist Co-Leiterin des Master of AI Ethics and Society Prgramms am Leverhulme Centre for the Future of Intelligence an der Universität Cambridge. Sie ist Theoretikerin, Forscherin und Autorin im Bereich Medien und digitale Kulturen. In ihrer Dissertation untersuchte sie die Neugestaltung dessen, was wir unter ‚ethisch' verstehen, und die sich verändernde Rolle des Menschen im Zusammenhang mit dem Aufkommen autonomer Fahrzeuge.

Leif Kobbelt
Leif Kobbelt ist Inhaber des Lehrstuhls für Computergrafik und Multimedia an der RWTH Aachen. Nach seiner Promotion im Jahr 1994 am Karlsruher Institut für Technologie arbeitete er an der University of Wisconsin in Madison, der Universität Erlangen-Nürnberg und dem Max-Planck-Institut für Informatik, bevor er 2001 an die RWTH wechselte. Zu seinen Forschungsschwerpunkten gehören 3D-Rekonstruktion, effiziente Geometrieverarbeitung und Shape-Analysis. Er erhielt eine Reihe von akademischen Auszeichnungen, darunter einen ERC Advanced Grant im Jahr 2013 und den Gottfried Wilhelm Leibniz-Preis im Jahr 2014. Er wurde zum Fellow der Eurographics Association (2008) sowie zum Distinguished Professor (2013) und Fellow (2019) der RWTH Aachen University ernannt. Seit 2015 ist er Mitglied der Academia Europaea und seit 2016 Mitglied der Nordrhein-Westfälischen Akademie der Wissenschaften.

Sybille Krämer
Sybille Krämer (Emerita) war Professorin für Philosophie an der FU Berlin. Neben Gastprofessuren in Tokyo, Yale, Santa Barbara, Santiago de Chile, Wien und Zürich ist sie Seniorprofessorin an der Leuphana Universität Lüneburg. Sie war Mitglied des Wissenschaftsrates, des European Research Council und des Senats der DFG.

Mattis Kuhn
Mattis Kuhn ist Künstler, Forscher und Kurator. Er arbeitet zur wechselwirkenden Gestaltung von Menschen, Maschinen und Umwelt, primär mittels (ausführbarem) Text. Er ist künstlerischer Mitarbeiter für Creative Coding an der Bauhaus-Universität Weimar und Teil der Forschungsgruppe ground zero an der KHM Köln.

Isaak Lim
Isaak Lim ist wissenschaftlicher Mitarbeiter am Visual Computing Institute in der Graphics, Geometry & Multimedia Gruppe, wo er mit Prof. Leif Kobbelt zusammenarbeitet. Er erhielt seinen Doktortitel der Informatik von der RWTH Aachen im Jahr 2021. Sein Forschungsgebiet ist der Bereich Shape Analysis, wo er sich für die Anwendung von Techniken des maschinellen Lernens auf Geometrieverarbeitungsaufgaben interessiert.

Ulises Mejias
Ulises Mejias ist Professor für Kommunikationswissenschaften und Direktor des Institute of Global Engagement an der State University of New York in Oswego. Zu seinen Forschungsinteressen gehören neben einer kritischen Datenforschung, Philosophie und Soziologie der Technologie sowie die politische Ökonomie digitaler Medien. Gemeinsam mit Nick Couldry hat Mejias das Buch *The Costs of Connection. How Data is Colonizing Human Life and Appropriating it for Capitalism* verfasst, das 2019 bei Stanford University Press erschienen ist.

Roland Meyer
Roland Meyer ist Bild- und Medienwissenschaftler. Derzeit forscht er als wissenschaftlicher Mitarbeiter im SFB 1567 *Virtuelle Lebenswelten* der Ruhr-Universität Bochum zu virtuellen Bildarchiven. 2021 erschien sein Buch *Gesichtserkennung* in der Reihe Digitale Bildkulturen des Wagenbach-Verlags.

Off Office
Johannes von Gross, Markus Lingemann und ihr Team im Münchner Designstudio Off Office entwerfen Erscheinungsbilder, Websites, Bücher und Ausstellungsdesign für Kunst und Kultur. So etwa für das Haus der Kunst und die Pinakothek der Moderne in München, das Städel Museum in Frankfurt und die Osterfestspiele Salzburg. Das Design für *Training the Archive* am Ludwig Forum Aachen wurde 2023 beim Wettbewerb *100 Beste Plakate* ausgezeichnet und international ausgestellt.

Fabian Offert
Fabian Offert ist Assistant Professor für Geschichte und Theorie der Digital Humanities an der University of California, Santa Barbara. Seine Forschungs- und Lehrtätigkeit konzentriert sich auf die visuellen digitalen Geisteswissenschaften mit einem besonderen Interesse an der Epistemologie und Ästhetik des maschinellen Sehens und Lernens.

Tillmann Ohm
Tillmann Ohm ist ein Creative Technologist mit Schwerpunkt auf kuratorische Systeme. Er entwickelt experimentelle Tools und Methoden für die Analyse und Interaktion mit digitalen Sammlungen im Rahmen seines Unternehmens *ARCU.technology*. Ohm hat ein Diplom der Freien Kunst von der Bauhaus-Universität Weimar und arbeitete als Cultural Data Analytics Research Fellow an der Universität Tallinn.

Mads Pankow
Mads Pankow ist selbstständiger Autor und Moderator zu Themen der Künstlichen Intelligenz. Er ist Gründer der Zeitschrift für Gegenwartskultur *Die Epilog* und von 2014 bis 2019 organisierte er jährlich das *Digital Bauhaus*, eine Konferenz an der Schnittstelle zwischen Technologie, Design und Gesellschaft in Weimar.

Matteo Pasquinelli
Matteo Pasquinelli ist Professor für Medienphilosophie an der Hochschule für Gestaltung Karlsruhe, wo er die Forschungsgruppe für Künstliche Intelligenz und Medienphilosophie KIM koordiniert. Seine Forschung konzentriert sich auf die Überschneidung von Kognitionswissenschaften, digitaler Ökonomie und maschineller Intelligenz.

Gabriel Pereira
Gabriel Pereira hat unlängst seine Promotion an der Universität Aarhus in Dänemark beendet mit einer Arbeit zum Thema *Mit dem algorithmischen Sehen hadern: Hegemoniale Computer-Vision und antagonistische Praktiken*. In seiner Forschung untersucht er Daten- und Algorithmusinfrastrukturen, insbesondere wie Computer-Vision-Algorithmen unsere Beziehung zur Welt vermitteln. Projekte von Pereira wurden unter anderem auf der 33. Kunstbiennale von São Paulo, im Van Abbemuseum, im IDFA DocLab und im Itaú Cultural ausgestellt.

Anna Ridler
Anna Ridler ist eine Künstlerin und Forscherin, die sich für Ideen rund um Daten und Quantifizierung sowie deren Beziehungen zur natürlichen Welt interessiert. Ridler erhielt einen MA-Abschluss in Information Experience Design vom Royal College of Art, und war Fellow am Creative Computing Institute der University of Arts London.

Marian Schneider
Marian Schneider ist ein Informatik-Bachelorstudent an der RWTH Aachen University. Seit 2022 wirkt er am Visual Computing Institute als studentische Hilfskraft am Projekt *Training the Archive* mit. Er interessiert sich für maschinelles Lernen im Zusammenhang mit Computergrafik.

Alexa Steinbrück
Alexa Steinbrück hat in Dresden und in Südfrankreich Bildende Kunst studiert und anschließend an der Universität Amsterdam einen Abschluss in Künstlicher Intelligenz absolviert. Von 2020 bis 2022 leitete sie den Bereich KI am XLAB, dem Lab für Künstliche Intelligenz und Robotik der Kunsthochschule Burg Giebichenstein in Halle, und seit kurzem arbeitet sie im AI and Design Lab der Hochschule für Gestaltung in Schwäbisch-Gmünd.

Giulia Taurino
Giulia Taurino ist Postdoctoral Associate an der Northeastern University. Ihre Forschungsschwerpunkte sind Formen der Inhaltsorganisation auf Online-Plattformen und in digitalen Archiven, die kulturellen Implikationen algorithmischer Technologien und Anwendung Künstlicher Intelligenz im Bereich der Kunst und Kultur sowie des Kulturerbes. Von 2019 bis 2022 entwickelte sie eine Reihe von Computerkunstprojekten, die die Schnittstelle zwischen KI und kuratorischen Praktiken in Museen erforschen. Sie war und ist u. a. am NULab for Texts, Maps, and Networks, am MIT Data + Feminism Lab, am metaLAB (at) Harvard und am Virtual Humanities Lab der Brown University tätig. Taurino hat an der Universität von Bologna und der Universität von Montreal in Medienwissenschaften und visueller Kunst promoviert.

Gaia Tedone
Gaia Tedone ist Kuratorin und Forscherin mit einem weitreichenden Interesse an den Technologien und Apparaten der Bildgestaltung. Im Jahr 2019 schloss sie ihre Promotion am Centre for the Study of the Networked Image ab. Sie schreibt und lehrt zu den Themen digitale Kultur und postkritische Museologie und arbeitet als freiberufliche Kuratorin.

Magda Tyżlik-Carver
Magda Tyżlik-Carver erforscht relationale Beziehungsnetze zwischen Menschen und dem Nicht-Menschlichen. Sie setzt diese Begriffe in Beziehung zu Konzepten des Posthumanen als kuratorische Entitäten, d. h. zu Algorithmen, Bots, Software und Computerinfrastrukturen. Tyżlik-Carver lehrt derzeit an der Universität Aarhus im Studiengang Digital Design and Information Studies und veröffentlichte 2016 ihre Dissertation unter dem Titel *Curating in/as Common/s. Posthuman Curating and Computational Culture*.

Yvonne Zindel
Yvonne Zindel arbeitet als Kuratorin und Vermittlerin insbesondere zu Möglichkeiten der Digitalität und Nachhaltigkeit. Sie setzt sich mit NFTs, Blockchain und Degrowth sowie den Möglichkeiten für ein dekoloniales, antirassistisches, feministisches und inklusives Kuratieren und Vermitteln auseinander.

Acknowledgements

First and foremost, we would like to express our gratitude to all the wonderful persons who made *Training the Archive* possible.

Many thanks to Hannes Bajohr, Nick Couldry, Elisa Giardina Papa, Adam Harvey, Mar Hicks, Mél Hogan, Moritz Ibing, Maya Indira Ganesh, Leif Kobbelt, Isaak Lim, Ulises Mejias, Matteo Pasquinelli, Gabriel Pereira, Anna Ridler, Alexa Steinbrück, Giulia Taurino, Magda Tyżlik-Carver for contributing to this publication, as well as to all peer reviewers for their helpful comments. We extend our thanks to the curators who we interviewed about their practice, as well as to the colleagues who tested the Curator's Machine: Galina Dekova, Raffael Dörig, Severin Dünser, Anna Fricke, Joasia Krysa, Marie Lechner, Janice Mitchell, Daniel Muzyczuk, Lisa Oord, Kasia Redzisz, Nora Riediger, Ana Sophie Salazar, Tina Sauerländer, Sabine Maria Schmidt, and Xiaoyu Weng. Furthermore, we thank Eva Cetinić, Vincent Christlein, Geoff Cox, Katrin Glinka, Sybille Krämer, Mattis Kuhn, Roland Meyer, Fabian Offert, Tillmann Ohm, Gaia Tedone, and Yvonne Zindel for their contributions to the 'Art & Algorithms' conference, as well as Mads Pankow for moderating, Anna Burst for the music, and Off Office for their sensitive approach to the visual design of the project.

We gratefully acknowledge all the (current and former) employees of Ludwig Forum Aachen who supported the realization of this project in numerous ways: Sonja Benzner, Esther Boehle, Marie Gentges, Mailin Haberland, Fanny Hauser, Nadine Henn, Feodora Heupel, Maren Hoch, Miriam Kroll, Annette Lagler, Holger Otten, Stefanie Wagner, and Julia Zeh. We would also like to thank the City of Aachen and the Department of Cultural Affairs; in particular Sabine Gerhards, Jaroslaw Gussmann, Jana Härter, Markus Hartmeier, Dieter Haubrich, Barbara Jakubowski, Nico Kaczmarczyk, Yvonne Knauff, Peter Marbaise, Olaf Müller, Claudia Nadenau, Ulli Nellessen, Dagmar Neuner, Nico Rüttgers, Jana Schampera, Irit Tirtey, Werner Wassenberg, Michael Weihmann, and Werner Wosch.

From among our partners HMKV Hartware MedienKunstVerein, Dortmund and the Visual Computing Institute of RWTH Aachen University we would like to thank Simone Czech, Ann-Kathrin Drews, Tim Elsner, Sabrina Fiedler, Moritz Ibing, David Kleinekottmann, Johanna Knott, Leif Kobbelt, Isaak Lim, Mathias Meis, Katharina Priestley, Marian Schneider, and all the collaborators, scientists and innovators we had the pleasure of meeting at conferences, on Zoom calls, and in exciting conversations within the 'Digital Culture' program, such as Bálint Alpári, Marina Bauernfeind, Johannes Bernhardt, Dominik Busch, Jonas Carstens, Lena tom Dieck, Alina Fuchte, Tabea Golgath, Margarethe Grad-Hamburg, Joachim Haenicke, Miriam Hausner, Yannick Hofmann, Leonardo Impett, Harald Klinke, Martin Lätzel, MIREVI, Clemens Neudecker, Hedda Roman, Thomas Rost, Ben Scheffler, Jan Sölter, Sonja Thiel, Dusan Totovic, Maximilian Westphal, and Sonja Wunderlich.

Finally, *Training the Archive* could not have been realized without the generous funding of the Kulturstiftung des Bundes (German Federal Cultural Foundation); special thanks to Sara Holstein, Juliane Köber, Christopher Krause, Julia Mai, Fabian Märtin, Marie-Kristin Meier, Stephanie Regenbrecht, Sandra Rutke, and Julian Stahl for the collaboration.

Dominik Bönisch, research project manager of *Training the Archive,* extends his gratitude to Inke Arns, Eva Birkenstock, and Francis Hunger for their trust, advice, and support at every step of this project.

Danksagung

Zuallererst möchten wir uns bei allen großartigen Personen bedanken, die *Training the Archive* möglich gemacht haben.

Vielen Dank an Hannes Bajohr, Nick Couldry, Elisa Giardina Papa, Adam Harvey, Mar Hicks, Mél Hogan, Moritz Ibing, Maya Indira Ganesh, Leif Kobbelt, Isaak Lim, Ulises Mejias, Matteo Pasquinelli, Gabriel Pereira, Anna Ridler, Alexa Steinbrück, Giulia Taurino, Magda Tyżlik-Carver für ihren hervorragenden Beitrag zu dieser Publikation sowie an alle Peer-Reviewer*innen für die hilfreichen Kommentare. Wir danken den Kurator*innen, die wir zu ihrer individuellen Arbeitsweise befragen konnten, sowie den Kolleg*innen, die mit uns die Curator's Machine getestet haben: Galina Dekova, Raffael Dörig, Severin Dünser, Anna Fricke, Joasia Krysa, Marie Lechner, Janice Mitchell, Daniel Muzyczuk, Lisa Oord, Kasia Redzisz, Nora Riediger, Ana Sophie Salazar, Tina Sauerländer, Sabine Maria Schmidt und Xiaoyu Weng. Darüber hinaus richten wir unseren Dank an Eva Cetinić, Vincent Christlein, Geoff Cox, Katrin Glinka, Sybille Krämer, Mattis Kuhn, Roland Meyer, Fabian Offert, Tillmann Ohm, Gaia Tedone und Yvonne Zindel für ihre spannenden Vorträge auf der Konferenz zu ‚Kunst & Algorithmen' sowie Mads Pankow für die Moderation, Anna Burst für die Musik und Off Office für ihren feinfühligen Umgang mit der visuellen Gestaltung des Projekts.

Wir danken allen (aktuellen und ehemaligen) Mitarbeiter*innen des Ludwig Forum Aachen, die zur Umsetzung von *Training the Archive* in vielfältiger Weise beigetragen haben: Sonja Benzner, Esther Boehle, Marie Gentges, Mailin Haberland, Fanny Hauser, Nadine Henn, Feodora Heupel, Maren Hoch, Miriam Kroll, Annette Lagler, Holger Otten, Stefanie Wagner und Julia Zeh. Genauso danken wir der Stadt Aachen und dem Kulturbetrieb, insbesondere Sabine Gerhards, Jaroslaw Gussmann, Jana Härter, Markus Hartmeier, Dieter Haubrich, Barbara Jakubowski, Nico Kaczmarczyk, Yvonne Knauff, Peter Marbaise, Olaf Müller, Claudia Nadenau, Ulli Nellessen, Dagmar Neuner, Nico Rüttgers, Jana Schampera, Irit Tirtey, Werner Wassenberg, Michael Weihmann und Werner Wosch.

Von unseren Partner*innen – dem HMKV Hartware MedienKunstVerein, Dortmund und dem Visual Computing Institute der RWTH Aachen University – danken wir herzlich Simone Czech, Ann-Kathrin Drews, Tim Elsner, Sabrina Fiedler, Moritz Ibing, David Kleinekottmann, Johanna Knott, Leif Kobbelt, Isaak Lim, Mathias Meis, Katharina Priestley, Marian Schneider sowie all den Mitarbeiter*innen, Wissenschaftler*innen und Innovator*innen, die wir auf Konferenzen, bei Zoom-Calls oder in spannenden Gesprächen im Rahmen des Programms ‚Kultur Digital' kennenlernen durften, u. a. senden wir Dank an Bálint Alpári, Marina Bauernfeind, Johannes Bernhardt, Dominik Busch, Jonas Carstens, Lena tom Dieck, Alina Fuchte, Tabea Golgath, Margarethe Grad-Hamburg, Joachim Haenicke, Miriam Hausner, Yannick Hofmann, Leonardo Impett, Harald Klinke, Martin Lätzel, MIREVI, Clemens Neudecker, Hedda Roman, Thomas Rost, Ben Scheffler, Jan Sölter, Sonja Thiel, Dusan Totovic, Maximilian Westphal und Sonja Wunderlich.

Schließlich wäre *Training the Archive* ohne die großzügige Förderung durch die Kulturstiftung des Bundes nicht zu realisieren gewesen; ein besonderer Dank gilt Sara Holstein, Juliane Köber, Christopher Krause, Julia Mai, Fabian Märtin, Marie-Kristin Meier, Stephanie Regenbrecht, Sandra Rutke und Julian Stahl für die freundliche Zusammenarbeit.

Einen ganz besonderen Dank richtet Dominik Bönisch, wissenschaftlicher Projektleiter von *Training the Archive,* an Inke Arns, Eva Birkenstock und Francis Hunger für ihr Vertrauen, ihren Rat und ihre Unterstützung bei jedem Schritt dieses Projekts.

Colophon / Impressum

This book is published on the occasion of the research project / Dieses Buch erscheint anlässlich des Forschungsprojekts: *Training the Archive*, 01.01.2020–31.12.2023. A project of Ludwig Forum Aachen in cooperation with HMKV Hartware MedienKunstVerein, Dortmund, and the Visual Computing Institute of RWTH Aachen University / Ein Projekt des Ludwig Forum Aachen im Verbund mit dem HMKV Hartware Medien KunstVerein, Dortmund und dem Visual Computing Institute der RWTH Aachen University.

Ludwig Forum Aachen
Director / Direktorin: Eva Birkenstock
Deputy Director / Stellv. Direktorin: Fanny Hauser, Annette Lagler (Emerita)
Executive Secretary / Geschäftsführende Sekretärin: Nadine Henn
Administration and Registrar / Administration und Registrarin: Feodora Heupel
Curators / Kurator*innen: Esther Boehle, Holger Otten
Curatorial Trainees / Kuratorische Volontärinnen: Mailin Haberland, Lisa Oord
Research Project Manager / Wissenschaftliche Projektleitung *Training the Archive:* Dominik Bönisch
Research Trainee / Forschungsvolontärin: Galina Dekova
Documentation and Inventory / Dokumentation und Inventarisierung: Nora Riediger
Library and Archive / Bibliothek und Archiv: Sonja Benzner
Conservation / Konservatorinnen: Julia Rief, Miyon Schultka, Christina Sodermanns-Janßen
Visitor Service / Besucher*innenservice: Birgit Makowski
Department of Cultural Affairs, Aachen / Kulturbetrieb der Stadt Aachen: Olaf Müller (Head / Betriebsleitung), Irit Tirtey (Financial Manager / Kaufmännische Geschäftsführung)
Museum Services of the Department of Cultural Affairs, Aachen / Museumsdienst des Kulturbetriebs der Stadt Aachen: Pia vom Dorp (Head / Leitung), Marie Gentges, Petra Kather, Manuela Mitrolidis, Karoline Schröder
Technical Services of the Department of Cultural Affairs, Aachen / Technischer Dienst des Kulturbetriebs der Stadt Aachen: Werner Wosch (Head / Leitung), Joachim Gabriel, Jaroslav Gussmann, Stefan Hansen, Markus Hartmeier, Daniel Hensel, Wolfgang Meehsen, Werner Wassenberg, Michael Weihmann

Publication / Publikation
Editors / Herausgeber*innen: Inke Arns, Eva Birkenstock, Dominik Bönisch, Francis Hunger
Editorial Manager / Redaktionsleiter: Dominik Bönisch
With contributions by / Mit Beiträgen von: Inke Arns, Hannes Bajohr, Eva Birkenstock, Dominik Bönisch, Nick Couldry, Elisa Giardina Papa, Adam Harvey, Mar Hicks, Mél Hogan, Francis Hunger, Moritz Ibing, Maya Indira Ganesh, Leif Kobbelt, Isaak Lim, Ulises Mejias, Matteo Pasquinelli, Gabriel Pereira, Anna Ridler, Alexa Steinbrück, Giulia Taurino, Magda Tyżlik-Carver
Translations / Übersetzungen: Erika Rubinstein, Düsseldorf
Copyediting and proofreading / Lektorat und Korrektorat: Nancy Chapple, Berlin (EN), Freies Lektorat Anna Lina Dux, Kassel (DE)
Interview transcription / Transkription: Maria Akingunsade, Hamburg
Graphic design / Grafikdesign: Off Office, München (Johannes von Gross, Robin Körner, Markus Lingemann, Leonie Seitz)
Paper / Papier: Gardamatt Eleven, Munken Lynx Rough, Arto Gloss
Print / Druck: DZA Druckerei zu Altenburg GmbH

Published by / Erschienen im: Verlag der Buchhandlung Walther und Franz König
Ehrenstr. 4, D–50672 Köln

Bibliographic information by the Deutsche Nationalbibliothek (German National Library): The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbiografie (German National Bibliography); detailed bibliographic data are available on the Internet at / Bibliografische Informationen der Deutschen Nationalbibliothek: Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie, detaillierte bibliografische Daten sind im Internet abrufbar unter: http://dnb.d-nb.de.

Distribution / Vertrieb
Buchhandlung Walther König
Ehrenstr. 4, D–50672 Köln
Tel: +49 (0) 221 / 20 59 6 53
verlag@buchhandlung-walther-koenig.de

ISBN 978-3-7533-0566-0

Training the Archive is funded by the program / wird gefördert im Programm

Ku/tur Digital
KULTURSTIFTUNG DES BUNDES

Funded by / Gefördert von

Die Beauftragte der Bundesregierung für Kultur und Medien

Collaborative Partner / Verbundpartner

HMKV
Hartware MedienKunstVerein

Digital Partner / Digitaler Partner

RWTH AACHEN UNIVERSITY

Ludwig Forum Aachen is supported by / wird unterstützt von

Peter und Irene Ludwig Stiftung

FREUNDE DES LUDWIG FORUMS FÜR INTERNATIONALE KUNST E.V.

Educational Partner / Bildungspartner

STAWAG

Mobility Partner / Mobilitätspartner

ASEAG

Cultural Partner / Kulturpartner

WDR 3

Ludwig Forum Aachen
Jülicher Straße 97–109
D–52070 Aachen
www.ludwigforum.de

Ein Museum der stadt aachen

Ludwig
Forum
Aachen

Verlag der Buchhandlung Walther und Franz König