With a Human Face? - Aesthetics and Politics of Datafication

Jasmin Schädler

Dutch Art Institute (DAI) Art Praxis

Graduate School ArtEZ University of the Arts Master of Arts Thesis

Supervisor: Antonia Majaca

July 2020

Table of Contents

Abstract

"With a Human Face? - Aesthetics and Politics of Datafication" investigates artistic strategies that recenter the humanness of data. By following the works of three contemporary artists, data collection and dataset formation are critically discussed. Mimi Ọnụọha's work enfolds the discussion of missing datasets and the consequences; this is complimented with a discussion of search engine data voids. Adam Harvey's work cautions about the overcollection of biometric data, especially in the context of law enforcement and in relation to the discussion around camera surveillance. Caroline Sinders' work stresses the humanness of data and the consequential importance of human rights with respect to collection and processing. A discussion of design, technology and science unfolds through her work. The aim of the paper is to bring further attention to perspectives reaching beyond the practical. This is done by questioning the inevitability of technology and by identifying its ideological components. Finally, complexity and diversity in perspectives are called upon as the outlook to further development.

Acknowledgements

My gratitude goes to: Antonia Majaca as my thesis supervisor who guided this process patiently. Antonia Majaca, Basam El Baroni and Ana Teixeira Pinto for their guidance during my time at the DAI. They helped me greatly in further shaping my interests and finding new inspirations. Bethany Crawford and Joannie Baumgärtner for re-reading my writing and helping me form my ideas during the final steps. Brooke Perry for giving my writing the final touch. Carloalberto Treccani for exchanging ideas and new inspirations.

Introduction – The Birth of Data Collection

The practice of data collection is old, ancient, archaic. The oldest relicts of cuneiform scripts (ca. 3300 BCE) are data records, records of economic activities as well as lexical lists, grouping and categorizing words.[1] The latter organize words under generic categories like “animal species” as explained by archeologist Kristina Sauer, curator of the Uruk-Warka-Collection of the German Archaeological Institute at the University of Heidelberg. These lists form some of the first simple datasets. More systematic data collection with respect to population is also a long-established practice. Already highly sophisticated when the Roman Provincial Census started around 11/0 BCE,[2] it is a substantial aspect of the Judeo-Christian narrative due to its prominent part in the Gospel of Luke.[3] The predominantly Christian West starts its cultural history with a census and the dominant position of data collection never fades from there. The abstraction of humans into countable and economic entities continues throughout Western history and goes hand in hand with the development of capitalism. The media scholar Jonathan Beller summarizes the major developments in modern history well when he writes “this earlier period of digitization had many names, most tellingly if also disavowingly perhaps, 'Humanism,' but its overarching operation was the (uneven) commodification of life.”[4] The enslavement and consequential abstraction of human beings in the transatlantic slave trade as well as the “management” of indigenous populations leading to subsequent genocides are peaks in the violence of datafication. Beller therefore already coins this period in time Digital Culture I.[5] Data collection has always been a way of “consolidating power” over people's lives as summarized in *Data Feminism* by artist Catherine D’Ignazio and media scholar Lauren F. Klein. Many instances of violent data collection are referenced in their footnotes, like colonial counting and statistics on minoritized ethnicities.[6]

Fast forwarding some centuries from the Roman census, another census takes a prominent position in the progress of data collection and evaluation. The U.S. census of 1890 was endangered because “census data from the year 1880 had not yet been analyzed.”[7] A point had been reached where the amount of data was exceeding the capacities to process data, a so-called “control crisis” as media scholar Felix Stalder quotes James Beninger.[8] Subsequently, punch card machines

[1] Kristina Sauer, “Ordnung ist das halbe Leben!,” in *5300 Jahre Schrift,* ed. Michaela Böttner, Ludger Lieb, Christian Vater, Christian Witschel (Heidelberg: Verlag das Wunderhorn, 2017), 15.

[2] W. Graham Claytor and Roger S. Bagnall, *“*The Beginnings of the Roman Provincial Census: A New Declaration from 3 BCE,” *Greek, Roman, and Byzantine Studies* 55, no. 3 (2015): 642.

[3] Lk 2:1: “In those days Caesar Augustus issued a decree that a census should be taken of the entire Roman world.”

[4] Jonathan Beller, *The Message is Murder* (London:Pluto Press, 2018), 6.

[5] Beller, *The Message is Murder*, 5–6.
“Global commodification, settler colonialism, the mercantile system, the middle passage, slavery, plantations, and industrial capitalism instantiated a first order digital culture[...].”

[6] Catherine D’Ignazio and Lauren F. Klein, *Data Feminism (*Cambridge, MA: The MIT Press, 2020)*,* 12.

[7] Felix Stalder, *The digital condition*, trans. by Valentine Pakis (Newark: Polity Press, 2018), 41.

[8] Stalder, *The digital condition*, 41.

invented by Hermann Hollerith were bought and put to use to speed up efficiency of data processing. These machines are the predecessors of the contemporary computer and Hollerith's company would later become IBM (International Business Machines Corporation).[9] Beller notes, the census is "media of racialization" and IBM subsequently provided the "infrastructural support of the Holocaust."[10] From this point in time, data processing machines started to grow in capacity and relevance. Since the first punch card systems had helped to overcome the problem of too much data to process, data collection was once again allowed to increase, and it did.

This very brief look into the history of datafication is not an attempt to reveal it as a human condition but to nevertheless pledge for its subversion and reclaiming. The writing of the history of progress is an agenda in the disguise of objectivity. As the cuneiform scripts show, classification is a way of making sense of the world. It is a method which humans also apply mentally to navigate their surroundings and filter uncountable inputs. If one collects too much data, has weak filters or over-categorizes the surroundings, this is considered pathological. These conditions are referred to with terms like neurosis, paranoia, apophenia, autism, schizophrenia and conspiracy theory. In recent years, the same pathological terminology has been applied to datafication and computation based data evaluation.[11] Already, philosopher Gilles Deleuze and psychiatrist Felix Guattari establish this conflation of vocabulary in *Anti-Oedipus*, although they do not consider the physical machine itself but a concept of the machine as an extension of human desire.[12] Their approach gains new relevance today with data being viewed as an extension of human desire.

Today we live in the era of Big Data, a term which has been around in its present meaning since the 1990s.[13] Big Data is not only the accumulation of massive amounts of data ("In 2014, 2.5 quintillion bytes of data were produced every day.")[14] but is also a way of reasoning. It pits causation against correlation, favors what over why and questions hierarchies of logic.[15] Apart from

[9] Stalder, *The digital condition*, 41–42.

[10] Beller, *The Message is Murder*, 105.

[11] Hito Steyerl,. "A Sea of Data: Apophenia and Pattern (Mis-)Recognition," *e–flux Journal*, no. 72 (April 2016), https://www.e-flux.com/journal/72/60480/a-sea-of-data-apophenia-and-pattern-mis-recognition/.
Clemens Apprich et al., *Pattern Discrimination (Pattern Discrimination*. Lüneburg: Meson Press, 2018).
Matteo Pasquinelli, "How a Machine Learns and Fails – A Grammar of Error for Artificial Intelligence," *spheres Journal for Digital Culture*, no. 17 (November 2019), http://spheres-journal.org/how-a-machine-learns-and-fails-a-grammar-of-error-for-artificial-intelligence/.
Matteo Pasquinelli, "Machines that Morph Logic: Neural Networks and the Distorted Automation of Intelligence as Statistical Inference," *Glass Bead Journal,* no. 1 (2017), https://www.glass-bead.org/article/machines-that-morph-logic/?lang=enview.
Ben Klemens, "a Library for Scientific Computing," Apophenia, last modified April 26, 2016, accessed June 3, 2020, http://apophenia.info/.
Peter Krieg, *Die Paranoide Maschine (*Leipzig: E.A. Seemann, 2005).

[12] Gilles Deleuze and Félix Guattari, *Anti-Ödipus,* trans. Bernd Schwibs (Frankfurt am Main: Suhrkamp, 2019), 497.

[13] Steve Lohr, "The Origins of 'Big Data': An Etymological Detective Story," *The New York Times BITS*, February 1, 2013, https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/.

[14] Wendy Hui Kyong Chun, "Big Data as Dram," *ELH* 83, no. 2 (Summer 2016): 372.

[15] Chun, "Big Data as Drama," 372.

these discussions around reasoning and the investigation of (artificial) intelligence, its true presence is in identity politics. Big Data, in a way, is the present-day equivalent of psychoanalysis, the method of reading out presumed subconscious desires of subjects and drawing conclusions from those findings which subsequently define the world(s) we live in. Just as Deleuze and Guattari thought it to be necessary to criticize the methods of psychoanalysis as an over-mythologization (a mythological apophenia) of a (bourgeois) identity, many voices today criticize Big Data and the technologies surrounding it as an over-determination of identities. With the means of Big Data, identities are fixed through statistical correlations to increase economic efficiency as well as control.

This paper will investigate means of data collection in the world of Big Data. Surveillance and governance will be the main considerations. Central to the discussion will be the approaches of three artist researchers who critically engage with present practices of data collection and datafication. Art and artists are often the first to critically investigate and subvert new technologies. Deleuze and Guattari mention Buster Keaton and Charlie Chaplin numerous times in *Anti-Oedipus* as important positions in critiquing and interpreting the machine as a cultural phenomenon beyond pure technical aspects.[16] Stalder also stresses the importance of artists like Nam June Paik and Bruce Nauman when he discusses early critical engagements with electronic media: “[The electronic image] engendered in the moment of its appearance an autonomous reality beyond and independent of its representational function. A whole generation of artists began to explore forms of existence in electronic media, which they no longer understood as pure media of information.”[17]

In the present discussion around datafication and related algorithmic technologies the artists leading the way are Mimi Ọnụọha, Adam Harvey and Caroline Sinders, in order of appearance. The first chapter structured around Ọnụọha's work focuses on the concept of datasets as a social and political force and who or what is represented in them and their absence. Apart from Ọnụọha's work, the concept of data voids in relation to search engines as discussed by Microsoft researcher Micheal Golebiewski and media scholar danah boyd, and data failures as discussed by media scholar Safiya Umoja Noble will be central. The second chapter takes off from Harvey's work with respect to data collection and surveillance, and specifically face image data. Facial recognition technology is compared with the narrative of CCTV surveillance, and both are discussed within the

Judea Pearl, *The Book of Why: the New Science of Cause and Effect* (NEW YORK: BASIC BOOKS, 2020). Luciana Parisi, *“*Instrumental Reason, Algorithmic Capitalism, and the Incomputable,*” i*n *Alleys of Your Mind: Augmented Intelligence and Its Traumas*, ed. Matteo Pasquinelli, 25–137 (Lüneburg: Meson Press, 2015), 128. Luciana Parisi, “XENO-PATTERNING: predictive intuition and automated imagination,” *Journal of the theoretical Humanities* 24, no. 1 (February 2019): 82–97.

[16] Deleuze and Guattari. *Anti-Ödipus*, 409, 514.

[17] Stalder, *The digital condition*, 44.

context of law enforcement. The third chapter engages with Sinders' work and the debate around data ownership and technology design. The narrative of determinism versus the consideration for human rights is a main focus. An additional example is found in the graphical user interface as discussed by media scholar Nishant Shah.

The aim of this paper is to investigate current practices and discussions around data collection and processing. Following the lead of the discussed examples, data is being reconnected with its human origins. This is a first step in questioning the narratives around current algorithmic technologies. An important further step is to question the terminology which will only be briefly mentioned in this paper. However, the use of intelligence and learning as a way of anthropomorphizing technology and therefore placing it in a potential position of power is questioned. The assumptions accompanying this practice cause the continuation and enforcement of hierarchies, oppression and often violent stereotypes. This happens because neutrality is assigned to machines and algorithmic processes because they are thought to be human independent and inevitable. The human labor, as for example in the case of crowd workers, these technologies are based on is devaluated and becomes invisible. This paper does not (yet) omit the use of all the terminology of this narrative.

Chapter 1 – Missing datasets and malicious data voids

The work of artist Mimi Ọnụọha will be the guiding principal for this chapter. Departing from two of her projects the discussion will engage with aspects of data collection and dataset creation. Lead by the artist's work, attention will be payed especially to gaps in a world of superfluous data. The discussion will also show how not only the collection of data but the formation of datasets is a powerful tool, especially with respect to search engines. Another dominating theme throughout the chapter is the question of authorship, power and governance with respect to data collection and dataset creation. The concepts of missing datasets and the threat of data voids will unfold with respective examples. Claims of neutrality with respect to data tools and search engines will be indirectly contested.

1.1 Introduction to the data-focused work of Mimi Ọnụọha

Ọnụọha's work is investigating and critiquing the meaning and power of data and more specifically, data collection in society. At the same time, Ọnụọha also makes propositions and educates the general public on data under algorithmic governance. She brings the people back into the picture after it seemed like they had long dispersed into an abundance of data points, disintegrated and fragmented.

Ọnụọha is an artist, researcher and programmer based in New York City. She studied at Tish NYU in the Interactive Telecommunications Program. Ọnụọha has been researching in the context of different residency programs over the last years, among others the Eyebeam Center for Arts & Technology, the Data & Society Research Institute as well as the National Geographic Digital Storytelling Fellowship.

As she stated in a talk she gave at Kikk Festival 2019 in Belgium, Ọnụọha's favorite definition of data is by Mitchell Whitela, who writes “data are measurements extracted from the flux of the real.”[18] But what Ọnụọha's work actually focuses on is not data itself but data collection. In the article “Points of Collection” she published on the Data & Society website in 2016, Ọnụọha states “the conceptual, practical, and ethical issues surrounding 'big data' and data in general begin at the very moment of data collection.”[19] Exactly this moment, the moment of data collection, is the main focus of Ọnụọha's artistic work. The reason for this interest is a project from 2014 where Ọnụọha herself unintentionally became the collector of data. The project is called “*69” and, as stated by the artist at Kikk Festival, is an intervention more than an art project. As a reaction to

[18] Mimi Ọnụọha, “What Is Missing Is Still There,” filmed November 1, 2019 at Kikk Festival, YouTube, 39:08, March 2, 2020, https://youtu.be/57Lgztk62uY.

[19] Mimi Ọnụọha, “The Point of Collection,” *Medium. Data & Society: Points*, October 31, 2016, https://points.datasociety.net/the-point-of-collection-8ee44ad7c2fa.

being cat called frequently in her neighborhood and being annoyed by never thinking of an appropriate response, the artist started to hand out papers with a phone number to the perpetrators. They texted the number expecting to message her, but the number actually led to a server with scripted replies. A side product of this was a data basis with all the numbers of the cat callers. Ọnụọha was suddenly in the powerful position of owning a very specific dataset of Brooklyn cat callers, although it all started with her being in the vulnerable position of being publicly objectified. Through this experience, she realized there is a relationship behind every data set, a relationship between the collector and the collected; human relationships.[20] While it helped her to process the unpleasant experiences she had made in public space, it also lead her onto a trajectory she is still following today.

1.2 The impact of missing datasets

One of Ọnụọha's long term projects is called "The Library of Missing Datasets." This project, first presented in 2016, is a collection of data sets which have not (yet) been collected. She shows this work in exhibitions as a file cabinet with folders inside, annotated with the names of the sets – the missing datasets are chosen in relationship to the place and context of the exhibition. When the folders are opened, they are of course empty. As she also recounts during the Kikk talk, to watch people open up the folders to make this realization is always exciting.[21] The word "missing" in the title of the work she contextualizes in the GitHub online repository as "both a lack and an ought: something does not exist, but it should. That which should be somewhere is not in its expected place; an established system is disrupted by distinct absence."[22] While the critical debate with respect to Big Data is often only concerned with the abundance of data collection, Ọnụọha looks at the blind spots of data and specifically at the meaning of those blind spots. Why do these blind spots exist? What are the consequences? Who profits from their absence? D'Ignazio and Klein define this method with the term "examining power" in their book *Data Feminism* and postulate it as "the first principle of data feminism." They state that the term "means naming and explaining the forces of oppression that are so baked into our daily lives – and into our datasets, our databases, and our algorithms – that we often don't even see them."[23]

The answers to the questions Ọnụọha is asking are not always obvious or simple. When engaging with this work, the most prominent aspect at first glance is the everlasting incompleteness of a list of missing datasets. And as the artist states herself, "this list will always be incomplete, and

[20] Ọnụọha, "What Is Missing Is Still There."

[21] Ọnụọha, "What Is Missing Is Still There."

[22] Mimi Ọnụọha, "On Missing Data Sets," *GitHub*, January 25, 2018, https://github.com/MimiOnuoha/missing-datasets.

[23] D'Ignazio and Klein, *Data Feminism,* 24.

is designed to be illustrative rather than comprehensive."[24] Ọnụọha creates a hands-on experience of datasets for the viewer/user. The work demystifies it while making the power behind it graspable. It becomes obvious what a dataset really is: a list of items with a common denominator. The items of the list often already exist but must be collected under a specific title (the common denominator) to become a dataset. This common denominator is a pattern, the first pattern, empirically identified by noticing a common trait and the incentive to create the dataset in the first place. It is the discovery of the resource that induces the mining process of the ore which will be purified into further patterns/correlations.

Ọnụọha poses three specific questions with this work: "What is a Missing Data Set? [...] Why do they matter? [...] Why are they missing?"[25] The essence of her answer to the first question is the correlation between missing data and issues that affect vulnerable groups who are not in charge of data collection. One example for this instance from the artists exemplary list (that was recently crossed off because it is no longer missing) is the number of people killed through law enforcement in the US. The vulnerable group consists of the potential victims who could be protected by stronger observation of these instances of killing and serious prosecution thereof. Examples of similar cases are also discussed by mathematician Cathy O'Neil in her 2016 book *Weapons of Math Destruction*. O'Neil starts a thought experiment of how she would close a gap of missing datasets to improve the situation in American prisons: "If I had a chance to be a data scientist for the justice system, I would do my best to dig deeply to learn what goes on inside those prisons and what impact those experiences might have on prisoners' behavior."[26] However, since this kind of investigation would hold up superficial economic efficiency and stir up public attention on misconduct, and because prisoners do not have a strong and influential lobby, even when these datasets are collected, they do not necessarily impact the status quo. O'Neil even uses the term "black box" when she discusses prisons since only the input and the output are considered but what happens inside stays invisible and intangible.

The second question targets hidden social biases and indifferences and asks if what "we ignore reveals more than what we give our attention to."[27] The fact that certain aspects of life are never considered to be worth a dataset tells a lot about their assumed relevance. However, this does not mean they are not important or impactful for anyone, it only means the ones affected are not in power, nor does their affectedness profit anyone in power. An example for this is a subproject of the library. Ọnụọha was invited to support a group of Asian-American Broadway actors in 2016 to

[24] Ọnụọha, "On Missing Data Sets."

[25] Ọnụọha, "On Missing Data Sets."

[26] Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy (*New York: Crown Publishing Group, 2016), chap. 5, Apple Books.

[27] Ọnụọha, "On Missing Data Sets."

understand the reasons why they are not hired more often. They created an overview of the cast's demographics of Broadway season 2014/2015. The reason why the theaters had never collected this information "according to a representative at the Broadway League, is because it relies on 'self-identification of ethnicity and therefore don't have accurate data.' That hasn't prevented the League, however, from collecting the same type of demographic information about its audiences."[28] It becomes quite obvious how the latter is directly useful to the theaters whereas the former would only mean additional work and self-reflection instead of profit. Only once the performers themselves have questioned the hiring politics of the theaters, was it taken up and made to matter.

The answer to the third question is partially also visible in the first two question-answer pairs. However, as a critical researcher, Ọnụọha has also pointed out less obvious reasons for the missing status of certain data sets. As shown above, visible power dynamics are often the reason why certain data is not collected. A further reason why data is missing can be the impossibility of its collection because it resists quantification. Ọnụọha gives the example of US currency outside of the US. As she explains in her Kikk talk, cash is too anonymous to be properly accounted for with the available tools. Also, the ratio of work and benefit plays an important role. The most pressing example is the reporting of sexual assault. Unfortunately, the victim has to produce enough evidence by revisiting the assault in often very public settings without any guarantee the perpetrator will be convicted. The decision to rather not report incidents at all often seems healthier.

The last answer Ọnụọha includes looks at the term "missing" from another perspective. The definition of "not existing although it should" does not apply for all cases. There is data which is more dangerous to the ones affected by its collection and therefore efforts are made to prevent the act of collection for as long as possible. The example linked in the text repository is the case of ID cards which undocumented immigrants can apply for in some American cities to be able to participate in quotidian life. To receive this ID card, they must hand in personal information such as name and address. The administrators of these ID cards have announced to destroy the data "in case a conservative Republican wins the White House and demands the data".[29] The presence of the data would make it easy to track down and deport immigrants and therefore it is important to keep the data set unavailable. In contrast, this also applies to many missing datasets that would affect wealthy parts of society. O'Neil discusses the datasets provided to create predictive crime models which are used by many police departments to statistically evaluate where their presence is most needed in order to prevent or register crime early. She stresses the problem data causes on nuisance

[28] Mimi Ọnụọha, "Broadway won't document its dramatic race problem, so a group of actors spent five years quietly gathering this data themselves," *Quartz, December 4, 2016,* https://qz.com/842610/broadways-race-problem-is-unmasked-by-data-but-the-theater-industry-is-still-stuck-in-neutral/.

[29] Tara Palmeri, "Municipal ID law has 'delete in case of Tea Party' clause," *New York Post*, February 16, 2015, https://nypost.com/2015/02/16/municipal-id-law-has-delete-in-case-of-tea-party-clause/.

crimes; the focus will stay on the poorer neighborhoods, new data is collected only there and “this creates a pernicious feedback loop.”[30] However, data on specific crimes of the wealthy with respect to finance are not recorded or even noticed with a similar investment. If these datasets existed and were included in such predictive crime models, police would certainly have to take different patrolling routes. But they remain missing due to a strong interest in their non-existence.

“The Library of Missing Data Sets” is a meaningful artwork reaching much further than what it is made of. It is almost difficult to tell where it ends. Not only is it continuously developed by the artist, it also guides the viewer into understanding the impact of data on our lives. The artwork once more makes data visible as human after heavy technomorphization through the tech industry (quite in contrary to the anthropomorphizing of technology). Ọnụọha's project is also a reminder for the necessity of data diversity to achieve equality in representation. Beller summarizes the consequences of the long-standing missing diversity of training data very well when he discusses one of the first machine learning examples. He points out Claude E. Shannon's 1948 paper “A Mathematical Theory of Communication” which was the groundwork for all text processing algorithms.[31] Shannon's system is supposed to be content indifferent, and therefore neutral. However, the text Shannon chose as the “ur-text” for the development of a statistical method of text creation was one far from being neutral: the historian Dumas Malone’s biography *Jefferson the Virginian* (1948).[32] Beller refers to historian Annette Gordon-Reed who investigated the sources for this biography and uncovered the omission of accounts stating Jefferson's fatherhood in two cases from a relationship with Sally Hemings, an enslaved women owned by Jefferson.[33] One very specific world view was already at the heart of machine learning data long before text generation and recognition technology had become a mainstream application, automatically including a long list of missing datasets. The aspect of text generation and suggestion is also what Beller points out as critical with respect to statistical plausibility of scenarios. Since a dataset that violently omits the representation of enslaved people as important figures of American history is the starting point of text production technology, this omission consequently reproduces itself and “thus #Blacklivesmatter becomes a statistically improbable event.”[34] This shows once more the importance of Ọnụọha's project to point out the implications of missing datasets with respect to participation in political and social life.

[30] O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, chap. 5.

[31] Beller, *The Message is Murder*, 57.

[32] Beller, *The Message is Murder*, 72.

[33] Beller, *The Message is Murder*, 73.

[34] Beller, *The Message is Murder*, 74.

1.3 Data voids – hijacking missing datasets

"The Library of Missing Data Sets" is indirectly linked with another work by Ọnụọha, called "We Are Searching For" (2015).[35] The latter is a collection of search terms the artist recorded on 39 computers at the Royal College of Art in London during a period of four months. The collection is presented on a large poster in several columns. It is a curious method of collection and there is no mention of privacy and consent on the artists website. This therefore triggers two questions: should these semi-public computers grant privacy to its users? Did the users know their search terms were being collected? Data traces become visible.

The search terms themselves also hint at suggestions made by auto-suggest functions of search engines. They help the user to find things they only have partial information about. In Ọnụọha's presentation of the dataset, the impact of the auto-suggest function becomes visible especially in cases where people did not know how to spell their inquiry. For example, trying to spell the artist Yves Klein with a C leads in very different directions. However, auto-suggest goes further than curious search terms entered by art students. To create an awareness for how we search and how our searches are guided is fundamental in times when the internet has become one of the most important information resources in many parts of the world. Also here, missing data plays a crucial role. In 2018, Microsoft researcher Michael Golebiewski coined the term "data void." Together with danah boyd, head of the New York based research institute Data&Society (Ọnụọha is a former fellow of this institution), Golebiewski has since published two reports on this phenomenon. In the most recent publication from 2019, the phenomenon is explained as "search terms for which the available relevant data is limited, nonexistent, or deeply problematic. Recommender systems also struggle when there's little available data to recommend."[36] This "low-quality data situations" is called "data voids."[37]

Data voids are loop-holes of missing datasets which are easily hijacked for the spreading of conspiracy theories and misinformation, often with ideological causes. This is a moment where missing datasets actually pose an imminent threat to society – a guerrilla information warfare. Especially in times of machine learning, all data is a dataset in some way or another. The main aspect for data to become a dataset is to have a common denominator as mentioned before. Therefore, every search in a search engine creates a dataset. The term dataset is never used in the data void report, the authors solely refer to data. After all, a dataset is a grouping of data points, yet

[35] Mimi Ọnụọha, "We Are Searching For," MIMI ỌNỤỌHA, last modified February 13, 2015, accessed April 25, 2020, http://mimionuoha.com/we-are-searching-for.

[36] Michael Golebiewski and danah boyd, "DATA VOIDS WHERE MISSING DATA CAN EASILY BE EXPLOITED," *DATA & SOCIETY*, November 2019, 5, https://datasociety.net/wp-content/uploads/2019/11/Data-Voids-2.0-Final.pdf.

[37] Golebiewski and boyd, "DATA VOIDS WHERE MISSING DATA CAN EASILY BE EXPLOITED," 5.

it does not mean they have only been collected to form a specific dataset. A lot of data is out there in the form of posts, news reports, private image uploads, blogs, etc. Data is much more than numbers in a table collected for a scientific cause as has become apparent in the public debate of the last two decades. In the case of data voids, there is no or very little high-quality data available on certain topics or datasets are specifically created with bad intentions (popularizing search terms). The types discussed in the report are the following five: Breaking News, Strategic New Terms, Outdated Terms, Fragmented Concepts and Problematic Queries. The report is looking specifically into instances of misinformation and hate speech.

In the case of "Breaking News," data void refers to missing data on a formerly rare term (referring, for example, to a city or a name or a concept) suddenly in the focus of the public eye due to an incident. Since it takes time to produce information, it is a moment often affected by abuse. A very prominent example is the 2017 Sutherland Springs Murders. The deadly shooting happened inside and in front of a church in Sutherland, Texas. Leads about the shooter being allegedly "associated with left-leaning groups" were seeded in different outlets like Twitter and Reddit by users asking questions.[38] Although it was completely made up, it was the only initial information available and therefore showed up first when the incident was searched for until further information had been published. The authors of the report point out one very important aspect with respect to data production in such cases. Referring to an article with the headline "Antifa Responsible for Sutherland Springs Murders, According to Far-Right Media," they explain how search engines cut off parts of headlines to display objects on the search page, in this case, "Antifa' Responsible for Sutherland Springs Murders... ." Therefore, even though this article tried to clarify the source of the wrong allegations, it continued to feed them. Especially people who limit their searches to superficially skimming through the search page were prone to be further drawn into wrong beliefs. What can be learned from this example is the importance of understanding how things and events are registered in a specific frame of datasets – in this case, headlines on search engine result pages.

For "Strategic New Terms," content is produced without a peak in attention. These cases are long-term projects that produce a variety of resources. Such data voids are not necessarily ever brought to general attention but can have major impact on people either way. If the term does rise to wider attention, it is far more difficult to clear up because the content around it has been there for much longer than all the potentially new content trying to clear up the data void. This leads to high ranking search results as in the case of "crisis actors."[39] An example in which a term actually did not gain wide attention but radicalized at least one individual, is "black on white crime." The young

[38] Golebiewski and boyd, "DATA VOIDS WHERE MISSING DATA CAN EASILY BE EXPLOITED," 20.

[39] Golebiewski and boyd, "DATA VOIDS WHERE MISSING DATA CAN EASILY BE EXPLOITED," 24–25.

white man who killed nine people and injured one more at the Emanuel African Methodist Episcopal Church in Charleston retells his unquestioned story of radicalization in his manifesto: “This prompted me to type in the words 'black on white crime' into Google, and I have never been the same since that day.”[40] He directly entered a data void which displayed an asymmetrical portrayal of reality and fostered him to further engage with ideologies supporting his racist prejudices and narrowing his perception of the world further. In the German context, a term a journalist warned about on Twitter in reference to the first data voids report is “Bürgerkrieg Deutschland” (civil war Germany).[41] He shows results coming up for this search term and points out this data void is a potential threat that should be resolved with the production of quality content before it explodes.

‘‘Outdated Terms” are cases for which no more quality content is being produced because they were replaced by new terms or debates just stopped being relevant. According to the authors, outdated terms mainly affect platforms that only access the content published within their infrastructure, for example YouTube and Facebook. These platforms favor new content which gives access to manipulation in the case of outdated terms. Whatever is newly added is likely to show up with a higher rank, even comments to older content can be affective in this case. General search engines like Bing and Google, the report states, are less affected because they can access a lot of authoritative content associated with the terms. However, one issue they do face is the auto-suggest function which can propose additions to outdated terms that lead users into data voids.

In the case of “Fragmented Concepts,” resolving data voids poses the greatest challenge. The same topics can be searched for by different key words. The used key words also often point towards different political inclinations. Theoretically, search engine algorithms will start to converge the different terms referring to the same matters over time, especially for general search engines. The top listed results will then be the results considered to be of highest quality, although different wording to the search query might be used primarily. However, in the U.S., accusations have been raised by “political figures” towards this practice of content convergence, calling it “anti-conservative bias.”[42] Therefore, “search engine designers are especially loathe to help associate synonymous terms that cross political lines, which leads to 'distinct clusters' pointing towards the same matters.”[43] These clusters bear potential for media manipulators and are often referred to as filter bubbles.

The last category the report presents is called “Problematic Queries.” In this case, users are

[40] Golebiewski and boyd, “DATA VOIDS WHERE MISSING DATA CAN EASILY BE EXPLOITED,” 25.

[41] Gerret von Nordheim, “Suchergebnisse zu 'Bürgerkrieg Deutschland',” Twitter, September 28, 2018, https://twitter.com/gvnordheim/status/1045647563058302976.

[42] Golebiewski and boyd, “DATA VOIDS WHERE MISSING DATA CAN EASILY BE EXPLOITED,” 32.

[43] Golebiewski and boyd, “DATA VOIDS WHERE MISSING DATA CAN EASILY BE EXPLOITED,” 32.

already starting their search with highly problematic key words. A prominent example the authors point out is “Did the Holocaust exist?” The query already implies a conspiracy theory and for a long time, only content creators with this inclination would produce for this query. After it had been identified as a data void, quality content producers were asked to include counter statements that would directly connect with the query. This can of course only happen once a problematic query is known and at the same time there are uncountable possibilities to consider.

Although all the points brought forward in the data voids reports are valid and important, the publications have to be read with a grain of salt because they are not completely honest with respect to the power of search engine providers, specifically large general search engines like Google and Bing. Although it is certainly true that they have no influence on content production in general, their infrastructure is far from powerless and the list of data voids has some gaps where the only entity to blame is the search engines. One of the authors, Golebiewski, is principal program manager at Microsoft Bing and therefore directly involved in the structuring of a large search engine. The apologetic tone the reports takes on, with respect to what search engines are capable of, has an air of hypocrisy knowing his position: “While Bing and Google can – and must – work to identify and remedy data voids, many of the vulnerabilities lie at the heart of what these search engines do. [...] Without new content being created, there are certain data voids that cannot be easily cleaned up.”[44]

Safiya Umoja Noble has investigated data voids, or as she terms it “data failures” caused by search engines and/or their vulnerabilities in her 2018 book *Algorithms of Oppression*. Her focus lies on Google as the market leader. She does not mention Bing or other search engines specifically. Noble brings forward issues related to internal corporate culture at Google, proprietary algorithms, business relations to other companies, sexist and racist page ranking and Google bombing.

In 2017, the Google employee James Damore authored the “anti-diversity manifesto” which was supported by other employees.[45] This manifesto, triggered by an investigation about gender related pay gaps, denied women success in software engineering due to “psychological inferiority.” Noble points out the importance of remembering the humans behind every technology, specifically in the case of search engines which are often promoted as neutral information providers. In comparison to the data voids report, it is one matter if there is no “quality data” available but it is another matter if the people behind a search engine are willing to promote other kinds of data. What Noble calls “algorithmic oppression” in most cases affects people of color and women. This

[44] Golebiewski and boyd, “DATA VOIDS WHERE MISSING DATA CAN EASILY BE EXPLOITED,” 43.

[45] Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism (*New York: New York University Press, 2018), chap. 1, Apple Books.

becomes apparent when looking at her example of searching for “black girl.”[46] The previous top hits on Google sexualized the search query. In the case of Bing, this is still true as a search done on May 7, 2020 has shown. It is also still true for “Asian girl” on both search engines. However, it is not the case for “white girl.” Numerous such examples exist, such as who is displayed when searching “CEO,” “professor style” or “three black teenagers” versus “three white teenagers.”[47] These data voids or data failures make the priorities of search engine providers apparent and questionable. Noble has been pointing out issues of sexism and racism with respect to search engines since 2010 but still many cases prevail. Her examples also put the concept of quality data, the measure of standard used in the data voids reports, up for debate – who defines what quality data is?

To fully grasp the reasons for these data failures is difficult since “it is impossible to know when and what influences proprietary algorithmic design [...].”[48] Whereas Noble tackles the proprietary issue from a user rights perspective, questioning the opacity of page rank systems and its part in reproducing racist and sexist biases, Golebiewski and boyd are focusing on the abuse of the latter through outside attackers. They do not consider bad or careless intentions from within the companies. Without an ethical standard and a system of evaluation with respect to the algorithms behind search engines, it seems unlikely to uncover or resolve these data failures. Due to the opaqueness of the mechanisms behind search engines, the users are not able to evaluate search results other than if they were helpful in solving a concrete problem. Felix Stalder stresses the power this unleashes because “this reality lends an enormous amount of influence to the institutions and processes that provide the solutions and answers.”[49]

1.4 Conclusion – Data responsibility

The conclusion of this detour into the data voids report is a question: do search engines (platform specific or not) have an educational mandate and where does curation of search engines start and end? Creating datasets brings great responsibility as Ọnụọha shows in her work. However, this responsibility is frequently overlooked for convenience or economic profit. The instances of dataset creation are often not directly visible, and the example of search engines is only one specific example out of many. Platforms responsible for making these datasets accessible have their own interests while at the same time the responsibility is often overwhelming. Malicious manipulators are fast in finding new blind spots and tricking search engine optimization. While the search engine providers are trying to offer a diverse palette without specific political inclinations, they also

[46] Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*, chap. 1.
[47] Safiya Umoja Noble, “Google Has a Striking History of Bias Against Black Girls,” *TIME*, March 26, 2018, https://time.com/5209144/google-search-engine-algorithm-bias-racism/.
[48] Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* , chap. 1.
[49] Stalder, *The digital condition*, 124.

provide a platform to manipulators, discrimination and conspiracy theories. Data collection and dataset compilation are places of power. Ọnụọha initiates the consideration of who is not represented as a guiding principle of dataset engagement.

Chapter 2 – Facial recognition data and biometric surveillance

This chapter is initiated by the work of Adam Harvey. It will be looking into dataset creation and measures of surveillance with a focus on facial recognition. Facial recognition technology will be contextualized looking back at camera surveillance. A main focus will be applications in law enforcement.

2.1 Introduction to the data-focused work of Adam Harvey

Harvey is an artist and researcher who, like Ọnụọha, has graduated from the Interactive Telecommunications Program at Tish New York University.[50] He is now based in Berlin. His work over the years has been engaged with creating awareness for data collection, especially in the field of biometric data as well as subversive re-appropriation of and protection against data collection. He has created different camouflage concepts including an LED anti-photo gadget, hair and face make-up and a false face generator to avert phase recognition, a vest to divert thermal drone targeting, as well as a microcontroller device for geolocation spoofing.[51] For the present discussion, however, the focus will be on his "MegaPixels" project.

2.2 The "MegaPixels" project and the politics of facial recognition

"MegaPixels" was initiated in 2017 together with the programmer and artist Jules LaPlace. This project investigates public datasets compiled by different entities to help train face recognition machine learning algorithms. In the beginning "MegaPixels" was an interactive installation piece which allowed visitors/users to engage with "the largest publicly available facial recognition training dataset in the world, called MegaFace," through a physical interface.[52] The installation was produced for the exhibition "Glass Room," curated by Tactical Tech. The custom-made algorithm compared the user to the available faces in the MegaFace dataset and printed a receipt including the closest match, a look-alike percentage and the user's face. The installation piece did not collect any data on those interactions as stated by the artist. However, out of ca. 15.000 users, Harvey states two people reported finding their exact image without prior knowledge of being part of the dataset. The work developed into a larger research project beyond the MegaFace dataset and is accessible through a webpage.[53] So far seven major image training datasets have been discussed with larger

[50] Adam Harvey, "About," Adam Harvey, last modified August 1, 2019, accessed March 28, 2020, https://ahprojects.com/about/.

[51] Adam Harvey, "Projects," Adam Harvey, last modified February 1, 2019, accessed March 28, 2020, https://ahprojects.com/projects/.

[52] Adam Harvey, "MegaPixels: Faces," Adam Harvey, last modified November 1, 2017, accessed March 28, 2020, https://ahprojects.com/megapixels-glassroom/.

[53] Adam Harvey and Jules LaPlace, "Art and research publication investigating the ethics, origins, and individual privacy implications of face recognition datasets," MegaPixels, last modified March 11, 2020, accessed May 2,

contributions on the website but many more have been investigated and featured in news articles. All seven data sets presented in depth are annotated and show human faces. One aim of the project is to show how face recognition datasets are compiled usually without the consent or knowledge of the individuals visible on the images, nor of, in most cases, the photographers. The project questions the ethics of the practice as well as the implications of facial recognition software in relation to biometric surveillance. Major newspapers have been reporting on the findings and collaborating in the research, like the *Financial Times* and the *New York Times*. The project is a valuable contribution to the ongoing discussion of how to proceed with the technology of facial recognition on a legal and governance level that gained new speed in 2019 after some cities in the U.S. (e.g. San Francisco and Somerville) banned the use of facial recognition technology "by the police and other agencies."[54] Following those bans, several initiatives and petitions like www.banfacialrecognition.com have been started to support a nationwide ban in the U.S.. The European Union published new regulations on "processing of personal data through video devices" in January 2020 without a ban, but voiced concern with respect to facial recognition technology because "researchers report that software used for facial identification, recognition, or analysis performs differently based on the age, gender, and ethnicity of the person it's identifying. Algorithms would perform based on different demographics, thus, bias in facial recognition threatens to reinforce the prejudices of society."[55]

Currently, two critical perspectives are circulating. The first is specifically singling out facial recognition technology from all biometric technologies and advocating for its ban, like philosopher Evan Selinger and computer scientist Woodrow Hartzog in their 2019 *New York Times* article.[56] They have been warning about the danger of face recognition technology since 2018 and the risks of oppression it bears. The other critical take on the topic is warning against focusing on facial recognition technology alone. Instead, privacy specialist Bruce Schneier argues for discussing the use of "all technologies of identification, correlation and discrimination, and decide how much we as a society want to be spied on by governments and corporations — and what sorts of influence we want them to have over our lives."[57] Harvey reacted to a Twitter post by Selinger who posted

2020, https://megapixels.cc/.

[54] Kate Conger, Richard Fausset, and Serge F. Kovaleski, "San Francisco Bans Facial Recognition Technology," *The New York Times*, May 14, 2019, https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html.

[55] "Guidelines 3/2019 on processing of personal data through video devices, Version 2.0," The European Data Protection Board, adopted on January 29, 2020, https://edpb.europa.eu/sites/edpb/files/consultation/edpb_guidelines_201903_videosurveillance.pdf.

[56] Evan Selinger and Woodrow Hartzog, "Read Your Facial Expressions? The benefits do not come close to outweighing the risks," *The New York Times*, October 17, 2019, https://www.nytimes.com/2019/10/17/opinion/facial-recognition-ban.html.

[57] Bruce Schneier, "We're Banning Facial Recognition. We're Missing the Point," *The New York Times*, January 20, 2020, https://www.nytimes.com/2020/01/20/opinion/facial-recognition-ban-privacy.html.

Schneier’s article, stressing the urgent need for tech regulations with respect to facial recognition technologies. Harvey points out that a definition of facial recognition is yet to be made and that “face recognition is just one of many forms of remote spectral reconnaissance.”[58] With this comment, Harvey stresses that regulations would be easily circumvented without a clear definition of what they are regulating and at the same time cut off the debate, consequently allowing for the use of these technologies. The economy scholar Shoshana Zuboff specifically warns in her book *The Age of Surveillance Capitalism* about the investment of large data owning companies, like Google and especially Facebook, in lobbying against laws that would restrict the collection and use of peoples biometric information, including images, without their consent.[59] The question between fast or thorough action arises considering regulations of biometric technology.

Not only is this technology available to larger companies, state security and law enforcement and has the potential to create an Orwellian state which in 2020 suddenly became a reality much faster than expected due to the COVID-19 crisis and the policing against Black Lives Matter protesters following the death of George Floyd. These technologies are also available to and used by private individuals increasingly and for security in semi-private sectors like supermarkets. The scandals and discussions around Clearview have shown how the technology is easily abused as a fancy gimmick. The private company offering facial recognition services has granted private accounts to potential or actual investors and friends as reported in March in a *New York Times* article.[60] The individuals with access to the software often use the tool in questionable ways without much reflection on the implications, similar to a fun toy, like finding out the name of someone’s date or celebrity look-alikes as discussed in the same article. The Clearview dataset is not publicly available but the company's CEO Hoan Ton-That has stated during many public occasions that the data is scraped from publicly available sources and the company saves the images until they are specifically asked to take down a picture by the owner of a face.[61]

In contrast to the Clearview dataset, many other datasets that also scraped their data from online sources are accessible to literally everyone. To have a vast and annotated dataset available for training a machine vision algorithm is essential. It will define the reliability of any image recognition software. Harvey helps to widen the public awareness for the presence of these datasets and how they are compiled. As will become apparent, it is not necessary for publishing platforms to

[58] Evan Selinger, “Facial recognition technology has *unique* affordances,” Twitter, January 20, 2020, https://twitter.com/EvanSelinger/status/1219320097678004229.

[59] Shoshana Zuboff, *The Age of Surveillance Capitalism: the Fight for Human Future at the New Frontier of Power (*New York: PublicAffairs, 2019), chap. 8, Apple Books.

[60] Kashmir Hill, “Before Clearview Became a Police Tool, It Was a Secret Plaything of the Rich,” *The New York Times*, March 5, 2020, https://www.nytimes.com/2020/03/05/technology/clearview-investors.html.

[61] “Clearview AI's founder Hoan Ton-That speaks out,” CNN Business, YouTube, 32:53, March 6, 2020, https://youtu.be/q-1bR3P9RAw.

sell their users' data to third party companies or to grant specific access to data scrapers. The internet is a gold mine, but the potential data miners need the right tools to purify the ore and refine it for the market. Maybe fracking is an even better comparison: what just looked like mud at first is actually a source of wealth but will leave behind major damage to the exploited habitat. The former news correspondent Patricia Ticineto Glough argues, "data that is most typically bracketed out as noise in sociological methods – such as affect, or the dynamism of nonconscious or even nonhuman capacity [...] is central to the datalogical turn."[62] In the case of images this also includes blurry and low resolution images rich in noise.

2.3 The datasets of the "MegaPixels" project

The first news article initiated through Harvey's research for the "MegaPixels" project is from 2017. While researching public datasets he came across one particular case he brought to wider attention through a post on Twitter. The headline of the article published by *THE VERGE* was "Transgender YouTubers had their videos grabbed to train facial recognition software."[63] The computer scientist Karl Ricanek had compiled a dataset of people undergoing hormone replacement therapy to transition genders. The data was scraped from the internet and consisted of links to transitioning diaries and time lapse videos. Ricanek had already removed the dataset due to personal concerns before it became a news story. However, on the webpage of his research group "The Face Aging Group," part of the University of North Carolina Wilmington Institute for Interdisciplinary Identity Sciences, the title of a conference paper is reminiscent of its existence. The paper with the title "Is the Eye Region More Reliable Than the Face? – A Preliminary Study of Face-based Recognition on Transgender Dataset" (2013) also hints at the wider potential of this research.[64] Today, as showcased by Clearview and the Russian AI company N-TechLab, facial recognition also appears to work when only the eyes are visible.[65] The conference paper of Ricanek's group contributed to make this progress in facial recognition technology. When contacted for the article, Ricanek explained he was thinking of potential terrorist threats where hormonal transitioning could have helped a terrorist to cross borders unrecognized even by facial recognition. After being told

[62] Patricia Ticineto Clough, Karen Gregory, Benjamin Haber, and R. Joshua Scannell, "The Datalogical Turn," in *The User Unconscious: On Affect, Media, and Measure,* ed. Patricia Ticineto Clough (Minneapolis: University of Minnesota Press, 2018), 103.

[63] James Vincent, "Transgender YouTubers had their videos grabbed to train facial recognition software," *THE VERGE*, August 22, 2017, https://www.theverge.com/2017/8/22/16180080/transgender-youtubers-ai-facial-recognition-dataset.

[64] "Publications & Conferences," Face Aging Group, last modified October, 9, 2019, accessed April 30, 2020, http://www.faceaginggroup.com/?page_id=21.

[65] CNN Business, "Clearview AI's founder Hoan Ton-That speaks out," 11:31.
"Moscow's Facial Recognition Tech will Outlast the Coronavirus," VICE News, YouTube, 07:43, April 16, 2020, 01:01, https://youtu.be/pbGq3REp4PI?t=61.

how threatening it can be for a transgender person to be revealed unwillingly, the researcher apologized.[66]

The seven large data sets to which Harvey and LaPlace have dedicated reports on the "MegaPixels" website are Brainwash, Duke MTMC, MegaFace, MS-CELEB-1M, Oxford Town Centre, UnConstrained College Students and WILDTRACK. Six of them have been removed, moved or temporarily deactivated for research since the reports came out. Each report gives insights into the sources of the data, the areas of application and the users of the dataset, compiled through academic citations.

The dataset that was part of the project from the beginning is called MegaFace.[67] The before mentioned installation also used it as a resource. It consists of "4,753,320 faces of 672,057 identities from 3,311,471 photos downloaded from 48,383 Flickr users' photo albums."[68] The dataset was compiled by the University of Washington and published in 2016. The images were taken from the Yahoo Flickr Creative Commons 100 Million Dataset which was released in 2014. This original dataset consists only of links to the images' locations on Flickr to pay tribute to the creators.[69] The researchers of Washington University were not sufficed with solely the links, but downloaded the images and therefore also lost the creators' attribution. As analyzed by the "MegaPixels" project, two thirds of the images included in MegaFace do not allow commercial use. However, research also happens within commercial companies (e.g. Google, N-TechLab (Russia), senseTime (China) have used MegaFace for their research) and although the dataset was never published within a commercial product, it helped to improve them – so the line between research and commercial use is extremely thin and fragile. Even the researchers of Washington University who compiled the dataset are active entrepreneurs. Prof. Ira Kemelmacher-Shlizerman is a popular public speaker for research turned business in relation to facial recognition. She is the developer of Dreambit, an app that allows you to change your hairstyle which she sold to Facebook in 2016.[70] Apart from her professorship at Washington University she holds a position at Google.[71]

[66] James Vincent, "Transgender YouTubers had their videos grabbed to train facial recognition software," *THE VERGE*, August 22, 2017, https://www.theverge.com/2017/8/22/16180080/transgender-youtubers-ai-facial-recognition-dataset.

[67] Adam Harvey and Jules LaPlace, "MegaFace Dataset," MegaPixels, last modified October 22, 2019, accessed May 2, 2020, https://megapixels.cc/megaface/.

[68] Harvey and LaPlace, "MegaFace dataset."

[69] Kashmir Hill, and Aaron Krolik, "How Photos of Your Kids Are Powering Surveillance Technology," The *New York Times*, corrected October 11, 2019, https://www.nytimes.com/interactive/2019/10/11/technology/flickr-facial-recognition.html.

[70] Ira Kemelmacher-Shlizerman, "How and Why Did University of Washington Professor Ira Kemelmacher-Shlizerman Build Dreambit and Sell To Facebook," filmed May 24, 2017 at LDV Vision Summit, Vimeo, 11:07, August 1, 2017, https://vimeo.com/227942118.
She shares this hair app idea with the CEO of Clearview Hoan Ton-That who invented an app that allowed people to wear Trumps hair on photos. In contrast to Ira Kemelmacher-Shlizerman he never successfully sold his app.

[71] Ira Kemelmacher-Shlizerman, "About," Ira Kemelmacher-Shlizerman, last modified May 22, 2020, accessed June

Since its release in 2016, MegaFace has been used to create several other derivative datasets with more specific applications like age or low resolution. The original dataset has been widely used in academic publications as well as patents. Harvey and LaPlace also show the geographic origins of users which includes an increasing number of citations from Chinese institutions, including the National University of Defense Technology. And although a *New York Times* article from October 2019 in collaboration with Harvey's research has pointed out Illinois residents who are included in the dataset have been treated unlawfully (Illinois has "the Biometric Information Privacy Act, a 2008 measure that imposes financial penalties for using an Illinoisan's fingerprints or face scans without consent") and are able to take legal measures against the use of their images, the scientists responsible for its publication did not react or take actions.[72] After inquiry of the *New York Times* a spokesperson of Washington University has stated "[a]ll uses of photos in the researchers' database are lawful. The U.W. is a public research university, not a private entity, and the Illinois law targets private entities."[73] However, as mentioned before, the University of Washington did allow others to use the dataset including research teams of private entities. Since March 2020 the so-called "MegaFace challenge" is concluded and the dataset is no longer available.[74]

The Microsoft MS-Celeb-1M dataset, published also in 2016, is a unique dataset because it not only consists of annotated images but also includes a list of suggested names for which no images have been scraped and included.[75] All people included either in the image dataset (100.000 individuals and 10.000.000 images, approximately 100 per person) or the list of suggestions (900.000 names) are people with public appearances due to their occupations, this is why the dataset is called "Celeb" as in celebrity. Some of them, as stated by the artists on the project website, are data activists, journalists and/or artists who publicly oppose and criticize face recognition technology and biometric surveillance. Prominent examples are Hito Steyerl, Trevor Paglen and the before mentioned danah boyd, Shoshana Zuboff and Bruce Schneier. MS-Celeb-1M was used to create the dataset "Racial Faces in the Wild."[76] After an article published in the *Financial Times* in collaboration with the "MegaPixels" project, the MS-Celeb-1M website was taken down and Microsoft announced the termination of the project in 2019. Harvey continuously

4, 2020, https://sites.google.com/view/irakemelmacher/about?authuser=0.

[72] Hill and Krolik, "How Photos of Your Kids Are Powering Surveillance Technology."

[73] Hill and Krolik, "How Photos of Your Kids Are Powering Surveillance Technology."

[74] Ira Kemelmacher-Shlizerman and Steve Seitz. "MegaFace and MF2: Million-Scale Face Recognition," MegaFace, last modified March 31, 2020, accessed June 17, 2020, http://megaface.cs.washington.edu/.

[75] Adam Harvey and Jules LaPlace, "Microsoft Celeb Dataset," MegaPixels, last modified May 2, 2020, accessed May 2, 2020, https://megapixels.cc/msceleb/.

[76] Harvey and LaPlace, "Microsoft Celeb Dataset."

finds active copies of the dataset.[77] The “MegaPixels” page also includes geolocated usage, again showing a large number of citations from China, including the National University for Defense Technology and the Chinese company SenseTimes. The latter was involved in providing face recognition technology to Chinese authorities to monitor Uighur Muslims in the Xinjiang province as reported by the New York Times in May 2019.[78]

The other five datasets discussed on the “MegaPixels” webpage have something in common which also differentiates them from the two mentioned above. The images of those five datasets were not scraped from the internet but specifically recorded in the so called “wild” or uncontrolled, unrehearsed surroundings without or only partial knowledge of the recorded subjects in public or semi-public spaces. Three of them were recorded on university campuses in the U.S. and Switzerland.

The Duke Multi-Target Multi-Camera dataset has particularly gained a lot of attention from researchers due to its set up; multi-camera (eight cameras) and multi-target.[79] This setup which allows observation of the formation and dispersion of groups is specifically interesting to study for state security research projects, among them Homeland Security, the National University for Defense Technology, a company working for the UK Ministry of Defense and a company providing law enforcement software to Scotland Yards and Queensland police as stated by the “MegaPixels” report. The images were recorded in 2014, the dataset was published in 2016 by the computer science department of Duke University but has been removed in reaction to the MegaPixels research. None of the targets were asked for consent, they were only informed by posters in the area of surveillance.[80] There is no information available if people noticed these signs. The cameras, specifically put in place for the data collection, were not particularly hidden or secluded.

A similar dataset is the UnConstrained College Students dataset recorded and published by the University of Colorado.[81] This dataset has two specificities which make it stand out, although it was not very widely used. The first aspect is related to the funding of the research which includes several intelligence and national security agencies like ODNI (Office of Director of National Intelligence) and IARPA (Intelligence Advance Research Projects Activity). The second aspect which distinguishes this dataset is the way the data was recorded. The researchers installed a digital

[77] Adam Harvey, “Is distributing the torrent for the Microsoft-Celeb face recognition training dataset illegal?,” Twitter, June 12, 2020, https://twitter.com/adamhrv/status/1271219064866852864?s=20.

[78] Chris Buckley, and Paul Mozur, “How China Uses High-Tech Surveillance to Subdue Minorities,” *The New York Times*, May 22, 2020, https://www.nytimes.com/2019/05/22/world/asia/china-surveillance-xinjiang.html.

[79] Adam Harvey and Jules LaPlace, “Duke MTMC Dataset,” MegaPixels, last modified May 2, 2020, accessed May 2, 2020, https://megapixels.cc/duke_mtmc/.

[80] Madhumita Murgia, “Who’s using your face? The ugly truth about facial recognition,” *Financial Times*, last modified September 18, 2019, https://www.ft.com/content/cf19b956-60a2-11e9-b285-3acd5d43599e.

[81] Adam Harvey and Jules LaPlace, “Unconstraint College Students Dataset,” MegaPixels, last modified April 18, 2019, accessed May 2, 2020, https://megapixels.cc/uccs/.

camera that recorded through an open window from an angle invisible to all targets. Pictures, recorded in 2012 and 2013, were strategically taken during breaks and in between classes when students change locations and therefore an increased number of individuals are visible on the images. The whole setup was only constructed to create this dataset, in a manner that is closer to methods a stalker would use than a researcher should apply. However, the research was planned and approved beforehand. Due to the MegaPixels website, the dataset was temporarily removed for revision. The involved researchers did not avoid conversations with the artists and newspapers involved and informed them that no data had been provided to government agencies. Nevertheless, government agencies can still benefit from public research results based on the data even if they would not gain direct access to the dataset.

The third campus dataset called WILDTRACK was recorded at ETH Zürich, Switzerland.[82] This dataset, published in 2017, was meticulously annotated through the crowdsourcing platform Amazon Mechanical Turk. The annotation was time intensive because each photo can potentially show dozens of individuals. Again, consent was not specifically asked, but signs were placed next to the cameras.

Another dataset, this time consisting of a video, was recorded on Oxford (U.K.) town square by a state owned CCTV camera for public safety in public space.[83] It is also the oldest of all the datasets discussed, most likely recorded in 2007 and published in 2009 by the University of Oxford. Harvey has pointed out the removal of this dataset on Twitter on June 12, 2020.[84]

The last dataset discussed by the project was recorded inside a San Francisco café, ironically called Brainwash, the only one from a privately owned space.[85] The dataset recorded through the shop's CCTV camera consists of 11.917 images each showing a number of people and was published in 2015. The dataset was created by Stanford University and the German Max Planck Institute for Informatics. In response to the reporting around the "MegaPixels" project, the dataset is no longer available and also the citations in two papers by the National University for Defense Technology have been removed.

Only through the thorough and continued research of facial recognition datasets Harvey has been engaged with since 2010, the general public is slowly growing aware of the scale this field

[82] Adam Harvey and Jules LaPlace, "Wildtrack Dataset," MegaPixels, last modified May 2, 2020, accessed May 2, 2020, https://megapixels.cc/wildtrack/.

[83] Adam Harvey and Jules LaPlace, "Oxford Town Centre Dataset," MegaPixels, last modified May 2, 2020, accessed May 2, 2020, https://megapixels.cc/oxford_town_centre/.

[84] Adam Harvey, "Another surveillance training dataset take down," Twiter, June 12, 2020, https://twitter.com/adamhrv/status/1271370337087885313?s=20.

[85] Adam Harvey and Jules LaPlace, "Brainwash Dataset," MegaPixels, last modified May 2, 2020, accessed May 2, 2020, https://megapixels.cc/brainwash/.

has.[86] The “MegaPixels” project summarized here only shows a small excerpt of datasets and focuses on their collection more than their application. The focus on data collection is crucial because without data there would be no facial recognition. For algorithms to recognize faces and identify people they need training data. The unreliability caused by missing data diversity has been prominently discussed with respect to malfunctioning of facial recognition, especially when systems are confronted with people of color and especially women of color since a majority of training data consists of white male faces.[87] For a long time training data was artificially produced in photo studios where, for example, military employees were asked to pose.[88] The aim is, however, to apply facial recognition in uncontrolled settings. To develop reliable facial recognition systems in “the wild,” diversity is also necessary in angle, position, lighting, size and resolution. Through the wide availability of images on the internet, especially with the mentioned diversity, the practice of finding instead of producing images has become common without consideration for privacy or copyright. Even public and semi-public spaces in real life are implicitly scraped for images as the college and café datasets show.

2.4 CCTV surveillance and the security argument

The debate around facial recognition technology is extremely similar to, and also directly connected to, the debate around camera surveillance in public space. Visual surveillance has slowly grown into a common aspect of public life. In 2004, a very similar project to “MegaPixels” was published, the “Urbaneye Project.” Just like “MegaPixels,” it is also a website presenting critical reports on a technology involved in biometric surveillance. The “Urbaneye Project” is focused on the presence of CCTV cameras in Europe, especially in its capitals. It was coordinated by the Technical University Berlin and supported by the European Commission.[89] Its purpose, as stated on the website of the latter, was to “contribute to the formulation of guidelines for a regulation of CCTV meeting both the demands for efficient and suitable employment and the claims for effective and accountable control.”[90] However, the researchers clearly state on their own website, “the content of this website does not necessarily reflect the views of the European Commission regarding these issues.”[91] The duration of the project was limited to 30 months and it has not been

[86] Murgia, “Who’s using your face? The ugly truth about facial recognition.”

[87] Steve Lohr, “Facial Recognition Is Accurate, if You’re a White Guy,” *The New York Times,* February 9, 2018, https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html.

[88] Murgia, “Who’s using your face? The ugly truth about facial recognition.”

[89] “Welcome to the Urbaneye Project,” Urbaneye Project. last modified March 16, 2006, accessed April 3, 2020, http://www.urbaneye.net.

[90] “On the threshold to urban panopticon?.” European Commission, last modified April 12, 2005, accessed May 15, 2020, https://cordis.europa.eu/project/id/HPSE-CT-2001-00094/.

[91] Urbaneye Project, “Welcome to the Urbaneye Project.”

updated since 2004. The inception report of the “Urbaneye Project,” written by Leon Hempel und Eric Töpfer, already warns about algorithmic surveillance including facial recognition as a prospect and the “need for further action and regulation” in reference to the European Union.[92] On a side note, quite ironically, the European Commission currently supports a startup called “Urban Data Eye” through its Horizon 2020 program.[93] This small enterprise is doing the opposite of what the “Urbaneye Project” did. It monetizes camera surveillance through AI enforced movement tracking, specifically targeted at public areas to enhance business.

The "Urbaneye Project” inception report maps out the spread of state lead camera surveillance in Europe. What is particularly interesting with respect to the discussion of regulation, is the introduction of new laws or the loosening of former regulations protecting privacy to make camera surveillance possible in the first place as referred to by the researchers. Some examples include the revision of the police act in North Rhine-Westphalia, Germany in 2000 and a security act from 1995 allowing video surveillance of public space in France.[94] Neither loosening nor increasing regulations hindered the push to increase the introduction of a surveillance technology which had questionable effectiveness with respect to crime reduction but high efficiency of policing. This seems absurd in comparison to the current debate of regulating technologies like facial recognition and other biometric surveillance technologies. The fear of regulation, in the end, leads to more regulation through policing which is made possible by those technologies in the first place. Freedom is not preserved by restraining regulation of these technologies. Meanwhile, the main arguments for security and crime reduction which help surveillance technologies to be implemented on state levels also never hold true. This can be well observed in the example of camera surveillance. The promise was to prevent crimes through the fear of being caught easily and quickly clearing up crimes that took place. The camera recording would be the objective witness police investigation had always dreamed about. Real world application does not turn out to be so straight forward nor objective as summarized by sociologist Clive Norris in his paper “THERE’S NO SUCCESS LIKE FAILURE AND FAILURE’S NO SUCCESS AT ALL – Some critical reflections on understanding the global growth of CCTV surveillance” written as a chapter of the 2012 book *The Global Growth of Camera Surveillance*. In the case of camera surveillance, many problems were caused by data storage limitations and time-consuming data classification. These are, of course, both aspects which seem to be resolved today as data storage has become abundantly available and machine vision is supposed to automate classification. But is this really the case?

[92] Leon Hempel and Eric Töpfer, “Working Paper No.1: Inception Report,” Centre for Technology and Society Technical University Berlin, 2002, 8, http://www.urbaneye.net/results/ue_wp1.pdf.

[93] “Actionable Data to optimize,” Urban Data Eye. last modified May 9, 2019, accessed April 26, 2020, http://urbandataeye.com/.

[94] Hempel and Töpfer, “Working Paper No.1: Inception Report,” 2–3, 10.

Machine vision can help in finding a particular face, but can it identify a crime? Will it ever be able to judge the difference between an embrace of love and sexual assault when this is, at times, even difficult for human observers? The part camera surveillance played in solving crimes is surprisingly low in comparison to what arguments for its implementation promised.

A wider survey of CCTV efficiency evaluations as conducted by Norris showed the impact ranges from positive (41%) to indifferent (43%) to negative or undesirable (15,9%).[95] The main effect in higher clear-up rates was shown in property crimes, especially in car parks, whereas in violent crimes there was no significant effect. One study from Australia showed that most incidents that took place in surveilled areas were not registered through the surveillance process but through other investigations. And in San Francisco it was observed that footage from public surveillance programs has helped in actually charging someone with a crime six times within three years.[96]

A study from the British Home Office also showed fear of crime among the population was not reduced due to the implementation of camera surveillance in public spaces but in some cases increased for people aware of the cameras.[97] And most importantly, interviews with offenders showed they were not stopped by camera surveillance but only saw it as an obstacle to "work around."[98]

Another aspect already present with camera surveillance is the discrimination of minorities who are disproportionally targeted due to prejudices.[99] This problem has already been identified for algorithmic law enforcement where prevailing discrimination was continued by the algorithmic evaluation because it was inscribed in the training data. A widely discussed example is the risk management algorithm COMPAS, a commercial product from a company called Equivant, employed by several criminal justice departments in the U.S.[100]

2.5 Facial recognition and law enforcement

Facial recognition systems are also already employed in crime investigations as became apparent with the reporting on Clearview in the beginning of 2020 by journalist Kashmir Hill in the *New York Times*. Before Clearview, it was not possible to use facial recognition in law enforcement outside of datasets directly available to the agencies like mug shots and drivers licenses.[101]

[95] CliveNorris, "The Global Growth of Camera Surveillance," in *Eyes Everywhere: the Global Growth of Camera Surveillance*, ed. Aaron Doyle, Randy K. Lippert, and David Lyon (London: Routledge, 2012), 32.

[96] Norris, "The Global Growth of Camera Surveillance," 32–33.

[97] Norris, "The Global Growth of Camera Surveillance," 33.

[98] Norris, "The Global Growth of Camera Surveillance," 37.

[99] Norris, "The Global Growth of Camera Surveillance," 38–39.

[100] "Algorithms in the Criminal Justice System: Risk Assessment Tools," Electronic Privacy Information Center, last modified 2020, accessed June 4, 2020, https://epic.org/algorithmic-transparency/crim-justice/.

[101] Kashmir Hill, "The Secretive Company That Might End Privacy as We Know It," *The New York Times*, January 18, 2020, https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html.

Clearview, as already mentioned above, only uses publicly available images similar to the MegaFace dataset but collects them from a variety of platforms. Many of those platforms have sent cease-and-desist letters to Clearview because the company is violating their terms of conditions with its data scraping practices. Wether the intention behind this reaction is due to ethical or instead economic reasons is up for debate. Facebook already claimed in 2014 to have reached "97.35 percent accuracy" for facial recognition on "the Labeled Faces in the Wild (LFW) dataset" in a post titled "DeepFace: Closing the Gap to Human-Level Performance in Face Verification" and many large tech companies have a history of working with military and law enforcement.[102] Extrapolating from this, it is likely they are not pleased to find out someone else is exploiting their data for commercial use while also endangering long term investments in lobbying to prevent regulations that would make commercial use entirely impossible or at least much more difficult.

When the news broke about the services Clearview offers to law enforcement it was not the fact that it is technically possible, but the fact it had already been in use for several months without public notice and without any proof of reliability, that was shocking. The company and the few officers who have publicly talked about the company, stress the successful cases where the software helped to identify offenders quickly. The same examples are named in various articles and interviews, but it remains unknown how many this really includes and if there have been false matches that were overlooked. As with camera surveillance, which is now an important source for material fed to Clearview (in its function as "an after-the-fact research tool"[103]), it seems to primarily help with property crimes. Apart from the obviously dangerous implications this technology can have, once it becomes available to the wider public, like stalking or even the preparation of a crime by knowing peoples habits, it continues to carry forward prejudice profiling which can lead to exclusion of people from public spaces or even decrease their safety based on their physical appearance. On top of this, it remains unclear how reliable facial recognition really is. So far, there is no proof it will always lead to the correct person or what the chances of doppelgänger will be, as feared by the law scholar Clare Garvie "who has studied the government's use of facial recognition."[104] In her study on the government and law enforcement applications of facial recognition since 2001, Gravie points out the creative uses of the technology when it comes to the probes being fed to facial recognition systems for identification of suspects.[105] She reports

[102] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," FACEBOOK Research, June 24, 2014, https://research.fb.com/wp-content/uploads/2016/11/deepface-closing-the-gap-to-human-level-performance-in-face-verification.pdf.

[103] "How Clearview AI Works," Clearview, last modified March 24, 2020, accessed June 4, 2020, https://clearview.ai/.

[104] Hill, "The Secretive Company That Might End Privacy as We Know It."

[105] Clare Garvie, "Garbage in, Garbage out: Face recognition on flawed data," Georgetown Law. Centre on Privacy & Technology, May 16, 2020, https://www.flawedfacedata.com/.

how images are enhanced and how forensic sketches and even celebrity and art historical look-alikes are used to search image datasets for potential matches. On top of the diverse practice of generating input data, she also summarizes the missing standards of how to apply the search results as evidence. In June 2020 the first case of misidenitification due to missing standards was publicly discussed, this certainly means many more have remained unnoticed.[106]

But apart from misconduct with respect to facial recognition technology, there is also no data on actual doppelgänger, especially when only partial information is available through low quality images or covered parts of the face. Just like fingerprints[107] and DNA,[108] which for a long time were thought to be unquestionable evidence, also facial recognition will have its flaws, maybe in unexpected and unforeseen ways. Facial recognition tools have shown their flaws with a lower rate of recognizing people of color, especially women, as identified by computer scientist Joy Adowa Buolamwini in her 2017 thesis titled "Gender Shades." This topic was again reviewed by Buolamwini and her colleague, computer scientist Inioluwa Deborah Raji in an audit of the impact her previous research had on these tools to reduce error rates – "darker females are the most improved subgroup (17.7% - 30.4% reduction in error)."[109] Buolamwini has also founded the Algorithmic Justice League to create awareness and offer tools to work on problems of exclusion and misrepresentation inside artificial intelligence applications. Her approach is an object of debate. Media scholar Ramon Amaro has put forward doubt in his 2019 e-flux essay "As if" by stating, "to merely include a representational object into a computational milieu that has already positioned the white object as the prototypical characteristic catalyzes disruption on the level superficiality. From this view, the white object remains whole, while the object of difference is seen as alienated, fragmented, and lacking in comparison." Today's technologies continue the violent classifications introduced by the fathers of biological and scientific racism as found in Carl Linnaeus (taxonomy) and Francis Galton (statistical heredity / modern eugenics). Amaro raises the question if it is necessary to subordinate to technologies with such violent pasts. In her paper on biometric

[106] Kashmir Hill, "Wrongfully Accused by an Algorithm," *The New York Times*, June 24, 2020, https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html.

[107] Adrian Lobe, "Wer keine Biometrie hat, ist kein Bürger," *Süddeutsche Zeitung*, October 27, 2018, https://www.sueddeutsche.de/digital/biometrie-gesichtserkennung-fingerabdruck-spracherkennung-1.4183394. "Die meist nach kaukasischem Aussehen modellierten Registraturen haben in der Praxis (etwa beim Frequent-flyer-Programm) dazu geführt, dass bei bestimmten Gruppen, etwa Frauen asiatischer Herkunft, deren Fingerlinien nur wenig ausgeprägt sind, die biometrischen Merkmale nicht "lesbar" waren."

[108] Matthew Shaer, "The False Promise of DNA Testing," *The Atlantic*, 15 June. 2016, https://www.theatlantic.com/magazine/archive/2016/06/a-reasonable-doubt/480747/. ""Ironically, you have a technology that was meant to help eliminate subjectivity in forensics," Erin Murphy, a law professor at NYU, told me recently. "But when you start to drill down deeper into the way crime laboratories operate today, you see that the subjectivity is still there: Standards vary, training levels vary, quality varies.""

[109] Inioluwa Deborah Raji and Joy Buolamwini, "Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products," Conference on Artificial Intelligence, Ethics, and Society, 2019, 4, https://dam-prod.media.mit.edu/x/2019/01/24/AIES-19_paper_223.pdf.

capitalism prominently featuring Galton as the inventor of fingerprint technology, media scholar Ariana Dongus states, “embedded categories such as race, class, gender, disability, and age order people into segmented groups within a population. In presenting the social order as natural, biometric systems maintain the race, gender, and class (b)orders within states.”[110] Dongus identifies biometric methods as “a colonial practice” always first tested in places of asymmetric power relations like colonies or war-zones before they are returned to the West as “higher-order knowledge.”[111] Seeing these two examples juxtaposed brings up the question of privileged access versus tools of oppression. In the examples Dongus investigates, biometric information is collected on a population controlled under exceptional circumstances like war and/or incarceration. Biometric identification is solely seen and used as means of surveillance and securitization. In the cases of omission Buolamwini is stressing, biometric identification is also a way of participating in social activities that do not only cover aspects of surveillance. They are ways of personal securitization as, for example, unlocking one's smart phone. Or, they are meant to entertain like face filters. When people are excluded from such activities this creates first of all the impression of disadvantage and discrimination instead of relief of not having to participate in an oppressive system. Is it a privilege to stay anonymous? Only if it is chosen anonymity. Unchosen anonymity bares dangerous cases of discrimination. This includes also the possibility for a person to be wrongfully recognized as the offender in a crime due to imprecision caused by underrepresentation.

2.6 Conclusion – Educating about facial recognition technology

The work of Adam Harvey initiated a broader discussion of biometric surveillance and governance and its false entanglement with scientific determinism. The current debate around the application of face recognition technology is a déjà-vu as the comparison to camera surveillance has shown. Also, the promised accuracy will crumble as the history of fingerprints and DNA as forensic tools forecasts. However, in neither case did the obvious failures and fallacies lead to their complete abandonment. To prevent further misconduct, that is already taking place in the case of face recognition as Garvie's report has shown, it is important to create awareness for new biometric technologies and how they affect everyone in the long run. “MegaPixels” contributes enormously to this awareness by connecting different aspects. The selection of datasets the project website presents focuses not only on data scrapped from the internet, but also from CCTV cameras and data specifically recorded by researchers. Educating the general public about the practices around face

[110] Ariana Dongus, “Galton’s Utopia – Data Accumulation In Biometric Capitalism,” *spheres Journal for Digital Culture*, no. 17 (November 2019), 12, http://spheres-journal.org/galtons-utopia-data-accumulation-in-biometric-capitalism/.

[111] Dongus, “Galton’s Utopia – Data Accumulation In Biometric Capitalism,” 6–7.

recognition is currently one of the most important tasks to generate informed and wide criticism. The 2020 Black Lives Matter protests have sparked a new wave of debate considering the dangers of facial recognition technology, especially with respect to discrimination of black people and people of color. In June 2020, IBM and Amazon have publicly announced the restriction of their services with respect to facial recognition. Amazon is calling for more regulations and will therefore grant a one-year moratorium to their law enforcement services in the US.[112] IBM is suggesting a national dialogue.[113] Neither company completely bans facial recognition technology. Adam Harvey has posted his skepticism on Twitter following these statements.[114] Clearview continues to believe in its services.[115]

[112] "We are implementing a one-year moratorium on police use of Rekognition" Amazon, published June 20, 2020, Accessed June 17, 2020, https://blog.aboutamazon.com/policy/we-are-implementing-a-one-year-moratorium-on-police-use-of-rekognition.

[113] Arvind Krishna, "IBM CEO's Letter to Congress on Racial Justice Reform," IBM, published June 8, 2020, accessed June 17, 2020, https://www.ibm.com/blogs/policy/facial-recognition-susset-racial-justice-reforms/.

[114] Adam Harvey, "The skeptics corner," Twitter, June 11, 2020, https://twitter.com/adamhrv/status/1270999156660875264?s=20.

[115] Nick Statt, "Amazon bans police from using its facial recognition technology for the next year," *THE VERGE,* June 10, 2020, https://www.theverge.com/2020/6/10/21287101/amazon-rekognition-facial-recognition-police-ban-one-year-ai-racial-bias.

Chapter 3 – Designing for transparency: data ownership and collection

This chapter will look into data ownership as a right of the people. Data has become an extension of human life and therefore, data collection should happen while regarding human rights. Zuboff describes the status quo in opposition to these values when she defines surveillance capitalism as "an expropriation of critical human rights that is best understood as a coup from above: an overthrow of the people's sovereignty."[116] Caroline Sinders' work pushes for human rights centred design, especially in the field of data collection. Her guiding concepts are transparency and community as will be shown. The relationship between technology, design and science will be investigated with respect to claims of inevitability. The graphical user interface as discussed by Nishant Shah will be used as an example.

3.1 Caroline Sinders: Human Rights Centered Design

Sinders is an artist and machine-learning design researcher particularly focused on user experience and interface design who investigates human interaction with digital products through an anthropological lens.[117] She is the founder of Convocation Design + Research, "an agency focusing on the intersections of machine learning, user research, designing for public good, and solving difficult communication problems."[118] Sinders is also a graduate of New York University Tish School's Interactive Telecommunications Program and a Mozilla Foundation fellow.[119] At Mozilla she researched Human Rights Centered Design, a follow up to Human Centered Design as propagated by the design firm IDEO, a "design and management framework that develops solutions to problems by involving the human perspective in all steps of the problem solving process."[120] When designing in the framework for artificial intelligence, Sinders claims it is necessary to shift the focus to Human Rights Centered Design, "one that focuses on data accountability and creation," and in accordance to the UN Universal Declaration of Human Rights which grants basic rights to all people.[121] Noble presumes "artificial intelligence will become a major human rights issue in the twenty-first century."[122] This especially affects data collection and data ownership. To be part of a dataset should be a choice and should require consent. Sinders calls her approach "designing for

[116] Zuboff, "The Age of Surveillance Capitalism," chap.THE DEFINITION.

[117] Caroline Sinders, "Feminist Data Set," Caroline Sinders, published May 26, 2020, 13, *https://carolinesinders.com/wp-content/uploads/2020/05/Feminist-Data-Set-Final-Draft-2020-0526.pdf*.

[118] Caroline Sinders, "About," Caroline Sinders, last modified April 25, 2017, accessed April 28, 2020, https://carolinesinders.com/about/.

[119] Sinders, "About."
Caroline Sinders, "AI is more than math: using art and design to interrogate bias in AI," filmed May 6, 2019 at re:publica, YouTube, 44:56, May 6, 2019, https://youtu.be/e0wyEnuRi3U.

[120] Sinders, "AI is more than math: using art and design to interrogate bias in AI," 17:05.

[121] Sinders, "AI is more than math: using art and design to interrogate bias in AI," 17:44.

[122] Noble, *Algorithms of Oppression*, 25.

transparency" where transparency is aimed towards the product and not the user. Transparency serves to answer questions of how algorithms affect the user, what data will be collected and why. This resembles Audrey Tang's approach of governance in Taiwan. As minister without portfolio with the background of a software engineer, her work focuses on integrating technology into democratic processes. Hereby, she proclaims to make the government radically transparent to the people instead of the people transparent to the state.[123]

Following the user experience designer Amber Case's suggestions in "Designing Calm Technology," Sinders also advocates in her 2019 re:publica talk for the possibility to download information a company has collected on a user as well as having it permanently deleted. The option to opt out of a dataset has gained momentum in 2020 through Clearview. Since the company was criticized for their scraping practices, they now offer a form on their website through which everyone can apply for the removal of their face from the Clearview dataset, not just retrospectively but into the future as well. According to a *VICE* article from February 2020, this also affects images posted by others than oneself.[124] However, one must send in a copy of government issued ID which will certainly stop many from performing this action.

Overall, Sinders' approach demands giving options to users. Since the introduction of the General Data Protection Regulation (GDPR) in Europe in 2018, many websites give options to configure what data they are allowed to collect, and many set the smallest amount necessary as default. However, often the "enable all purposes" button is placed strategically to be mistaken for the "save choices" button (Sinders would call this a "Dark Pattern" as coined by user experience designer Harry Brignull)[125] and sometimes websites cannot be visited at all when data tracking is not granted. This weakens the idea of "articulate privacy policies" as desired by Sinders and Case. In May 2020, the European Data Protection Board has issued guideline updates to further minimize loopholes that allow tricks to avoid direct consent and data extortion by denying access.[126]

Sinders summarizes her idea of designing for transparency under three captions: legibility, the ability to audit and space for impact interaction. Overall, she stresses the ethical focus of design with respect to artificial intelligence and more specifically machine learning and areas of data collection. Sinders views data as inherently human and participatory. Without people, there is no

[123] Audrey Tang, "Digital Social Innovation," filmed May 6, 2019 at re:publica, YouTube, 53:59, May 12, 2019, https://youtu.be/jl9mt5OEH0c.

[124] Anna Merlan, "Here's the File Clearview AI Has Been Keeping on Me, and Probably on You Too," *Vice*, February 28, 2020, https://www.vice.com/en_us/article/5dmkyq/heres-the-file-clearview-ai-has-been-keeping-on-me-and-probably-on-you-too.

[125] Caroline Sinders, "Dark Patterns and Design Policy," *Medium. Data & Society: Points*, May 20, 2020, *https://points.datasociety.net/dark-patterns-and-design-policy-75d1a71fbda5*.

[126] "Guidelines 05/2020 on consent under Regulation 2016/679, Version 1.1," The European Data Protection Board, adopted on May 4, 2020, https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf.

data she says, and people must be treated with dignity.[127]

To center AI design around human rights and dignity is directly connected with Ọnụọha's practice of creating visibility for human relationships with respect to data. In Sinders' approach, people are granted agency, and artificial intelligence is not considered to be human independent. Important here is the design aspect. To design something already implies human agency. After all, it means to devise meaning, taking its Latin roots into consideration. It differs from a simplified generic scientific approach where meaning is discovered as in natural laws like gravity. Technology is sandwiched between these two approaches. The deterministic mindset of the natural sciences is often adapted by technology when a certain standard has been reached and established. This goes even as far as transferring the concept of biological evolution to technological advancement excluding human influence and decisions as relevant factors. The journalist and technology podcast producer Rose Eveleth summarized this development in the 2019 *VOX* article "The biggest lie tech people tell themselves — and the rest of us."[128] She refers to inventor Ray Kurzweil's "law of accelerating returns" through which he states "technological evolution as a continuation of biological evolution."[129] Eveleth criticizes this argument as a way to avoid responsibility and regulations that are completely normal in other areas of society like the food, drug or mining industry. Sinders also brings forward this comparison in her re:publica talk promoting the concept of datasheets developed by the AI NOW Institute.[130] These datasheets would label datasets similar to labeling of food or drug ingredients and therefore make their content and purpose more accessible.

Sinders stands for ethical design in the context of AI. The idea of ethical AI has gained a lot of attention in recent years, especially from technology companies. AI NOW, for example, as well as the before mentioned Data & Society Research Institute were founded and funded with the help of Microsoft, and the former is also funded through Google and DeepMind. Also, many other topic related research groups at universities receive funding from the industry.[131] In Germany, Facebook invested in seed funding for the "TUM Institute for Ethics in Artificial Intelligence" in 2019.[132] As discussed by Rodrigo Ochigame in his *The Intercept* article "The invention of ethical AI," it is

[127] Sinders, "AI is more than math: using art and design to interrogate bias in AI."

[128] Rose Eveleth, "The biggest lie tech people tell themselves — and the rest of us," *Vox*, October 8, 2019, https://www.vox.com/the-highlight/2019/10/1/20887003/tech-technology-evolution-natural-inevitable-ethics.

[129] Ray Kurzweil, *The Singularity Is near: When Humans Transcend Biology (*London: Duckworth, 2008), 16.

[130] Timnit Gebru et al., "Datasheets for Datasets," Proceedings of the 5th Workshop on Fairness, Accountability, and Transparency in Machine Learning: Stockholm, Sweden, 2018, https://arxiv.org/abs/1803.09010.

[131] Rodrigo Ochigame, "THE INVENTION OF ETHICAL AI: How Big Tech Manipulates Academia to Avoid Regulations," *The Intercept*, December 20, 2019, https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/.

[132] Joaquin Quiñonero Candela, "Facebook and the Technical University of Munich Announce New Independent TUM Institute for Ethics in Artificial Intelligence," About Facebook, January 20, 2019, https://about.fb.com/news/2019/01/tum-institute-for-ethics-in-ai/.

always dubious when tech companies are directly involved in the research that is supposed to advice regulations.[133] He recounts instances he witnessed as a researcher at the MIT Media Lab in which research was directly influenced by corporate or military agendas.[134] This kind of involvement can also grow into a source of false scientific inevitability or "evolution" due to practicability or profitability as represented by corporations. The importance of ethical AI should not be diminished by these corporate efforts. However, it is necessary to keep up the ethical debate independently of corporations and create awareness for the rights to consent as well as educate and spread technological literacy to nourish critical thought.

3.2 The determinism of the GUI and the invisible human

Sinders is working at the intercept of independent criticism, corporations and research institutions. She asks questions that interrogate the relationship between design and policy with the user first in mind, and specifically looks at interfaces – the places of human interaction with technology as well as policy through design. A specific example of user interfaces, namely the graphical user interface (GUI), is also a place of inevitability claims towards technology. To interact with computers graphically seems obvious and is well established as the most common form of human-machine interaction as of today. For his IMPAKT Festival talk in 2019, Shah questions this hegemony under the title "From GUI to No UI."[135] In his argument, Shah brings together different aspects of deterministic assumptions towards technology. He starts with the assumption of rationality – we assume machines are rational and since the GUI is a machine it is rational, whereas the user is human and therefore irrational or, in other words, neurotic. Using media artist Kelly Dobson's "OMO" project (a robot with changing breathing patterns) as an example, Shah points out that machines are only repetitive (which implies rational and predictive) by design and not by default. He proposes to question what else computation is able to do by questioning the GUI as the default interface but also by questioning it as inevitable and neutral entity. Just like Sinders is creating awareness for the power of design and how dignity can be granted or denied to users, Shah is steering the attention to the regulatory effect interface design has on user behavior. Exactly this kind of regulation is currently promoted as endangered freedom by potential state directives that would determine user regulation. The intentional influencing that is acted out through hardware as much as through software is obscured into fake neutrality by calling upon the inevitability of scientific findings. In all places, human involvement is erased. This affects

[133] Ochigame, "THE INVENTION OF ETHICAL AI: How Big Tech Manipulates Academia to Avoid Regulations."
[134] Ochigame, "THE INVENTION OF ETHICAL AI: How Big Tech Manipulates Academia to Avoid Regulations."
[135] Nishant Shah, "From GUI to no UI," filmed November 2, 2019 at IMPAKT Festival, YouTube, 01:10:23, November 11, 2019, https://youtu.be/vaeoAeEBNcI.

the design/engineering level as much as the repetitive labor of the (digital) proletariat (labeling data and content moderation as much as device assembly). Although the former are invisible by choice, the latter are invisible by design. And as Sinders points out, it all starts and ends with granting or denying data its humanness.[136] The granting of humanness applies to all fields of technology. Shah therefore proclaims the necessity to always look for the people who have become invisible through an interface, "the hidden and subsumed bodies."[137] He observed that people working in precarious digital maintenance often have more rights when they are considered part of the machinery instead of being seen as (replaceable) humans.

3.3 The "Feminist Dataset"

In contrast to the previously mentioned scientific framing of technology, the speculative approach practiced in design where the agency of change is within the idea and not the discovery, is usually found in early stages of development. Artificial Intelligence is treated by some as a scientific discovery that has just not fully manifested itself. By others, it is seen as a speculation that guides a process. Sinders is engaging with the speculation that can be mended and changed to achieve a preferred goal with an accompanying path leading up to this goal. Part of this path is not only her work on user sovereignty through Human Rights Centered Design but also her "Feminist Dataset" project. Sinders has been hugely involved in research around online harassment. Her project at the Wikimedia Foundation was to research hate speech vocabulary of the alt-right.[138] The fact that she researched this area and collected data, in this case language of a particular subculture, again contributed to the representation of hate speech in datasets even in an instance where it is necessary to counteract a negative phenomenon - currently the hate speech dictionary that resulted from it is used by the Southern Poverty Law Center to support cases against hate speech.[139] Sinders consequently thought about strategies which could counteract hate speech and "how inequity and bias manifest inside a dataset."[140] This lead her to the idea for the "Feminist Dataset". Currently, the "Feminist Dataset" is considered an art project by Sinders. She tours different places like art galleries and feminist bookshops worldwide to give lectures and most importantly workshops.[141] At every stop, the "Feminist Dataset" grows through new contributions. It does not only consist of vocabulary and concepts from the intersectional feminist realm but also includes images, interviews

[136] Sinders, "AI is more than math: using art and design to interrogate bias in AI."
[137] Shah, "From GUI to no UI."
[138] Sinders, "AI is more than math: using art and design to interrogate bias in AI."
[139] Sinders, "AI is more than math: using art and design to interrogate bias in AI."
[140] Sinders, "AI is more than math: using art and design to interrogate bias in AI," 14:58.
[141] Katharine Schwab, "This Designer Is Fighting Back Against Bad Data – With Feminism," *Fast Company*, April 16, 2018, https://www.fastcompany.com/90168266/the-designer-fighting-back-against-bad-data-with-feminism.

and complete books, fiction as well as theory. Sinders' idea behind this dataset is "that to remove bias within machine learning, the 'removal of bias' itself has to be manifested into a 'thing' to teach or sway the algorithms."[142] For algorithms to execute the removal of biases, it is necessary to closely define what bias is and what bias removal is. In the context of machine learning, this is achieved by collecting examples for an algorithm to analyze and reenact. The project raises questions if such an algorithm is realisable or desirable and therefore redirects importance to human decisions and assesment. With her way of data collection, Sinders is aiming to establish data collection as a feminist and especially collective practice of intervention as well as a form of protest.

In the long run, Sinders plans to use the dataset to train a feminist chatbot.[143] It is a direct reaction not only to her personal experiences in online harassment but also to chatbot experiments like Microsoft's Tay (thinking about you) chatbot that was left to learn from Twitter interactions without any safety net against racism in place, because it was not part of the developers reality.[144] Within 16 hours the experiment was stopped because the chatbot had turned into a Nazi. This development led Sinders to ask if it is possible to "create an AI that doesn't harm? Is it possible to create an Anti-Fa AI, and what would that data set look like?"[145] The basis of these questions and of the entire project is, however, another: Is ethical data collection possible? Just as with design in general, Sinders stresses the ethical aspects of data collections: "Unlike machine learning projects where you'll scrape a bunch of data, and just throw it at something, I'm interested in how you can ethically grow data."[146] To reach this goal she does not impose an approach but develops it together with a community, in this case the intersectional feminist community. It is an attempt to "digitize [a] form of equality" to influence machine learning systems.[147] The community-based approach is again a parallel to Audrey Tang's way of creating community engagement in policy making. Sinders includes every aspect of machine learning systems and therefore also data labeling, usually provided through crowd working platforms. These platforms in their current structure do not fulfill feminist and intersectional requirements as they are highly unethical in the way workers are treated and payed. Therefore, Sinders has developed a new crowd sourcing platform with a wage calculator that only accepts a reasonable and realistic relationship between data units, time spent per unit and payment per unit in consideration of living wage ("Washington State, where Amazon is

[142] Caroline Sinders, "Building a Feminist Data Set for a Feminist AI," Schloss Post, October 20, 2017, https://schloss-post.com/building-feminist-data-set-feminist-ai/.

[143] Sinders, "Feminist Data Set," 13.

[144] Tahani Nadim, "c u soon humans need to sleep now so many conversations today thx," in *The influencing Machine*, ed. Tahani Nadim (Berlin: nGbK, 2018), 72.

[145] Sinders, "Building a Feminist Data Set for a Feminist AI."

[146] Schwab, "This Designer Is Fighting Back Against Bad Data–With Feminism."

[147] Sinders, "Building a Feminist Data Set for a Feminist AI."

headquartered").[148] The project becomes an all-encompassing example of the real possibility for change and the potential for equality in the realm of data design. In this spirit, it already follows the suggested principles of data feminism as proposed by D'Ignazio and Klein: Examine Power, Challenge Power, Elevation Emotion and Embodiment, Rethink Binaries and Hierachies, Embrace Pluralism, Consider Context, Make Labor Visible.[149]

In reaction to Sinders' talk at re:publica, one member of the audience who stated he was "into ethics" questioned this approach. He thought a feminist dataset would exclude many people and therefore should not become the default for platforms like Facebook and Twitter because "they include [...] everybody's ethics."[150] He asked Sinders if she is opting for splitting platforms depending on peoples ethics or if she would really want to make the "Feminist Dataset" the general default. Sinders responded by saying that using a single dataset for anything would be very irresponsible. But the question leaves behind a worrying afterthought, especially with Sinders' personal involvement in mind, creating awareness for Gamergate, an online movement that harassed female game developers and journalists for speaking out on the misogynist practices of the industry, with serious repercussions for Sinders and her family.[151] What the question implies without actually stating it is the "right to be a misogynist/racist/..." online. Certainly, this is an extrapolation, but the person did not state which ethics would be infringed by the "Feminist Dataset" in comparison to exclusion caused by currently dominating datasets. Again, a vague fear of regulation is opposing the sole idea of equality taking unevenly spread privileges into consideration.

3.4 Conclusion – The future of data-based networking culture

Finally, a conclusion to the discussion of Sinders' work could be the imbalance between fear of regulation and the possibility of creating a digitized world that treats users with dignity and respect. The latter often is seen as the direct enemy of the former. To put it naively: does this mean a digitized world is not for everyone or at least not for everyone in the same way? Certainly, this is true for the offline world / IRL and certainly IRL has merged already with the digital, virtual and online, especially with respect to data collection. But the fear of regulation as it is enacted now directly opposes the utopian idea of equality cyber space once was thought to promise in its early days. Currently, two different concepts of regulations and rights clash: the right to stay anonymous

[148] Caroline Sinders and Cade Diehm, "Calculator," Technically Responsible Knowledge, last updated April 24, 2020, accessed May 31, 2020, http://trk.network/#calculator.

[149] D'Ignazio and Klein, *Data Feminism*, vii.

[150] Sinders, "AI is more than math: using art and design to interrogate bias in AI," 44:55.

[151] Caroline Sinders, "That time the Internet sent a SWAT team to my mom's house," B*oing Boing*, July 24, 2015, https://boingboing.net/2015/07/24/that-time-the-internet-sent-a.html.

and the right to be safe. The latter is used as an argument against the former. However, the implementation of safety has become an excuse for invasive collection of personal data for the creation of trust. As media scholar Wendy Hui Kyong Chun argues with reference to information scholar Helen Nissenbaum, "the reduction of trust to security assumes that danger stems from outsiders, rather than 'sanctioned, established, powerful individuals and organizations'."[152] Sinders argues for trust through transparency but transparency on the side of the service provider.[153]

Shah even goes as far as to reject the internet completely because "what we have is broken."[154] For this reason, he is collaborating with the Feminist Internet Research Network (FIRN). The network is a subproject of the Association for Progressive Communications (APC). The APC's "vision is for people to use and shape the internet and digital technologies to create a just and sustainable world, leading to greater care for ourselves, each other and the earth."[155] What is crucial in their approach is the focus on creating accessibility to the internet in places where currently it is not accessible. According to Internet World Stats, this is true for around 40.4% of the world's population as of May 2020.[156] However, certainly 100% of the world's population are part of a dataset which is already shown by the before mentioned statistics. Only if everyone has equal access to the internet and is considered equally in technology is there even the slightest chance for freedom and equality which is currently considered at risk through lurking regulations. However, when considering Beller's ontology of digital culture it seems completely impossible to separate it from racism and discrimination because it has been formed by them for much longer than expected: "Built on an axiomatics of racial inequality and gender inequality, today's codifications, abstractions and machines, far from being value-neutral emergences intelligible in some degree-zero history of technology, are rather racial formations, sex-gender formations, and national formations—in short, formations of violence."[157] With this in mind, the "Feminist Dataset" seems like an impossible undertaking since it will always return to structures of violence. Is it even worth engaging then? Discussions like "Re-imagining the Internet: Roadmaps to digital Equality" that took place between political scientist Nanjira Sambuli, lawyer and activist Renata Ávila Pinto, activist Esra'a Al Shafei and curator Mi You in June 2020 as part of the Goethe Institute Latitude Festival are extremely necessary to push other approaches and open up new ways of imagining the

[152] Chun, "Big Data as Drama", 375.

[153] Caroline Sinders, "How UX Design Creates Trust," Adobe, October 8, 2019, https://xd.adobe.com/ideas/perspectives/social-impact/building-trust-through-user-experience-design/.

[154] Shah, "From GUI to no UI."

[155] "About," Association for Progressive Communications, accessed June 4, 2020, https://www.apc.org/en/about.

[156] "World Internet Users and 2020 Population Stats," Internet World Stats, last modified June 2, 2020, accessed June 5, 2020, https://www.internetworldstats.com/stats.htm.

[157] Beller, *The Message is Murder*, 2.

digital.[158] During this discussion, the speakers stressed the fact that they were not there to offer solutions but to start complicating and adding complexity. Yet, there were also some graspable proposals and even existing examples. Ávila Pinto suggested a minimal way of connecting people free of charge and control and You mentioned the project ReUnion by designer Yin Aiwen, a social network of care focusing on relationships between people as the smallest unit instead of the individual users.[159] The debate also stressed the importance to look elsewhere than only North America and Europe and illustrated how many indigenous languages have vocabulary that conveys concepts of networking culture much better than most colonial languages. This is an important reminder to finally stop the overwriting of the Global South through the arrogant argument of progress by the Global North. The reminiscence to the Haitian Revolution instantly comes to mind as the most progressive revolution of the 18th and 19th century. Yet, it was never fully acknowledged by the Global North and even actively diminished and occluded.

[158] Nanjira Sambuli, Renata Ávila Pinto, Esra'a Al Shafei and Mi You, "RE-IMAGINING THE INTERNET: ROADMAPS TO DIGITAL EQUALITY," filmed June 4, 2020, at Goethe Institut Latitude Digital Festival, YouTube, 04:14:17 - 05:59:14, June 4, 2020, https://youtu.be/LW7UXFjgbcI.

[159] Yin Aiwen, "About," ReUnion, last modified January 31, 2020, accessed June 5, 2020, https://www.reunionnetwork.org/#about.

Conclusion – Complexity instead of ideology

The question whether certain technologies like facial recognition could be banned altogether because the harm they cause is larger than the gain has become apparent. Maybe instead, it is necessary to understand where technology ends and ideology starts. Returning to the example of camera surveillance, what part is technology and what part is ideology? Filming and recording can be considered part of technology. However, even those applications were developed within a context as stressed repeatedly by Beller: "Ideas operating 'in the silence of technologies' are without a doubt ideological in the sense that they are divorced from their material conditions of production and dissemination as Regis Debray taught us, but technologies operating in the silence of social difference are solely techno-logical only in the sense that their emergence as a social formation that sediments those social relations into an apparatus is suppressed."[160] The ideological aspects of technology cannot be ignored, but the relevance given to technology can be adjusted as discussed by media scholars Matteo Pasquinelli and Vladan Joler in their paper "The Nooscope Manifested – Artificial Intelligence as Instrument of Knowledge Extractivism." They state it is necessary to "secularize AI from the ideological status of 'intelligent machine' to one of knowledge instrument."[161]

The computer scientists Glen Weyl and Jaron Lanier have published a text on *WIRED* in March 2020 titled "AI is an ideology, not a technology."[162] The two authors, who are also researchers at Microsoft and refer to themselves as enthusiasts of deep / convolution networks, are not writing about bans but about rethinking of certain technologies. One aspect they stress especially is the relevance of human labour to advance technology, especially in the case of what is summarized under the headline of artificial intelligence today. With human labour they are not just thinking about the engineers and programmers but about the large labour forces needed to annotate or produce data in the first place. To put this in dimension: during a 2015 TED Talk Fei-Fei Li, the Stanford professor and founder of ImageNet proudly stated "ImageNet was one of the biggest employers of the Amazon Mechanical Turk workers: together, almost 50,000 workers from 167 countries around the world helped us to clean, sort and label nearly a billion candidate images [...]" for the purpose of training image recognition algorithms.[163] People like Weyl and Lanier (as well as Shah and Sinders) are assigning agency to this labour force by changing the language used to speak

[160] Beller, *The Message is Murder*, 99.

[161] Matteo Pasquinelli and Vladan Joler, "The Nooscope Manifested: Artificial Intelligence as Instrument of Knowledge Extractivism," KIM research group (Karlsruhe University of Arts and Design) and Share Lab (Novi Sad), May 1, 2020 (preprint forthcoming for *AI and Society*), https://nooscope.ai.

[162] Jaron Lanier and Glen Weyl, "AI is an Ideology, Not a Technology," *WIRED*, March 15, 2020, https://www.wired.com/story/opinion-ai-is-an-ideology-not-a-technology/.

[163] Fei-Fei Li, "How We Teach Computers To Understand Pictures," TED 15, March 17, 2015. New World AI, May 23, 2020, https://www.newworldai.com/how-we-teach-computers-to-understand-pictures/.

about certain advancements like image recognition. Instead of saying "a machine is learning to see" when a program distinguishes cats from dogs, they suggest instead that "people contributed examples in order to define the visual qualities distinguishing 'cats' from 'dogs' in a rigorous way for the first time."[164] Along those lines, the already mentioned Taiwanese minister Tang refers to AI as "assistive intelligence" to stress that it is never human independent.[165]

As stressed by all three artists discussed in the chapters above, a major aspect of technology currently referred to as "AI" is data, specifically human data. To effectively influence the further progression of technology, which is always a means of enhancing human life, is to bring attention to the humanness of data. The media scholar Luciana Parisis argues that "gathering data and quantifying behaviors, attitudes, and beliefs, the neoliberal world of financial derivatives and big data also provides a calculus for judging human actions, and a mechanism for inciting and directing those actions."[166] This will in return shape how the world is structured and perceived for human interaction and where human agency begins and ends. The notion of the incomputable is a relevant aspect to be considered here. The incomputable aspects of every scientific theorem define the borders of its scope. The assigned determinism, also referred to in Chapter 3, is a simplification that does not hold true infinitely. In the context of computation, this is represented by Turing's Halting Problem. In a simplified summary, this proof shows no computing machine can predict if any given computing machine (including itself) will be able to halt (determine) or not for any given problem.[167] With the Halting Problem, Turing showed that mathematics is not absolutely deterministic, and neither are computing machines. Therefore, they are not capable of all-encompassing predictions. Quantum mechanics has shown this in the world of particles outside the philosophical realm of pure mathematics by embracing uncertainty. Even before quantum mechanics, thermo dynamics already hinted at a non-deterministic interpretation of the physical world. However, to accept this as a state of mind, it is first necessary to accept the ideological component of classical physics**.** With this in mind, the physicist Flavio Del Santo explains "due to the tremendous predictive success of Newtonian physics (in particular in celestial mechanics), it became customary to conceive an in principle limitless predictability of the physical phenomena that would faithfully reflect the fact that our Universe is governed by determinism."[168] The grounds

[164] Lanier and Weyl, "AI is an Ideology, Not a Technology."

[165] Tang, "Digital Social Innovation."

[166] Parisi, *"*Instrumental Reason, Algorithmic Capitalism, and the Incomputable,*"* 127.

[167] Alan Turing, "On Computable Numbers, with an Application to the Entscheidungsproblem," in *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life: Plus The Secrets of Enigma*, ed. B. Jack Copeland (Oxford: Oxford University Press, 2004), 72–74.

[168] Flavio Del Santo, "Indeterminism, causality and information: Has physics ever been deterministic?," FQXi Community, March 11, 2020, 1, https://fqxi.org/community/forum/topic/3436.

were laid by major philosophers of the enlightenment like Hume, Kant and Leibniz.[169] The questioning of determinism in physics is also a deeply philosophical question since it is directly connected with the concept of free will or the possibility to determine human action. Turing's Halting Problem is mainly a mathematical problem, yet the Turing machine is the blueprint for contemporary computing machines. However, for a long time computing machines were tools for deterministic solutions, representatives of deductive reasoning. Although this perception prevails in the common understanding, the actual applications have surpassed this way of reasoning. Current advancements in pattern based algorithmic approximations in the form of correlations are reaching beyond the concept of deterministic answers. Computational reasoning has taken the form of abductive reasoning based on datasets that are inevitably incomplete because they could never be anything other than that. This way of reasoning allows to reach beyond expected premises. Yet, this also entails a very high possibility of error due to incompleteness. Although trial-and-error is part of scientific research, it is only reasonable to a certain extent. To become meaningful in the long run, seemingly contingent causalities have to become transparent. If they stay opaque, any application or generalization deduced from them stays contingent. This is a serious problem in current machine-learning applications, especially when they affect human behavior. As the investigations of all three artists discussed in this paper show, this cannot be taken lightly and needs to be discussed thoroughly to prevent power asymmetries from growing further. All three artists make these issues accessible to a wider non-expert public. Technology shapes communities and affects human perception and behavior. Technological advancements should never be taken as inevitable and they are never neutral. As Li said, legitimately, “AI will change the world. Who is going to change AI?”[170] She already did her part in shaping AI in a controversial way, done with the typical scientific curiosity that only thinks about repercussions in a second step, if at all. Li prominently used the comparison between the learning process of a small child and the “learning process” of an image recognition algorithm.[171] Introducing this parallel is also part of an ideology that trivializes image recognition algorithms. It is time to introduce complexity instead and use this opportunity to revisit hierarchies of knowledge. Luciana Parisi writes the “predictive intuition in neural networks can explain the formation of patterns beyond deductive premises and demarcate the advance of artificial imagination in the dynamic architecture of machine thinking.”[172] She builds her argument through the concept of xeno-patterning and alienness. This is an interesting approach that opens new perspectives; however it also adds layers of mystification. It is a good start but what comes

[169] Del Santo, “Indeterminism, causality and information: Has physics ever been deterministic?,” 6.

[170] Fei-Fei Li, “On AI and Machine Learning.” Filmed April 10, 2017 at Grace Hopper Celebration. YouTube, 28:20, February 6, 2018, https://youtu.be/XlnbNFW2tX8.

[171] Li, “How We Teach Computers To Understand Pictures.”

[172] Parisi, “XENO-PATTERNING: predictive intuition and automated imagination,” 83.

next? As this paper has shown, artists will play a relevant and crucial role in this process.

Bibliography

Aiwen, Yin. “About.” ReUnion. Last modified January 31, 2020. Accessed June 5, 2020. https://www.reunionnetwork.org/#about.

Amazon. “We are implementing a one-year moratorium on police use of Rekognition.” Published June 20, 2020. Accessed June 17, 2020. https://blog.aboutamazon.com/policy/we-are-implementing-a-one-year-moratorium-on-police-use-of-rekognition.

Apprich, Clemens, Wendy Hui Kyong Chun, Florian Cramer, and Hito Steyerl. *Pattern Discrimination*. Lüneburg: Meson Press, 2018.

Association for Progressive Communications. “About.” Accessed June 4, 2020. https://www.apc.org/en/about.

Beller, Jonathan. *The Message Is Murder*. Longon: Pluto Press, 2018.

Buckley, Chris, and Paul Mozur. “How China Uses High-Tech Surveillance to Subdue Minorities.” *The New York Times*, May 22, 2020. https://www.nytimes.com/2019/05/22/world/asia/china-surveillance-xinjiang.html.

Chun, Wendy Hui Kyong. “Big Data as Drama.” *ELH* 83, no. 2 (Summer 2016): 363–382.

Claytor, W. Graham and Roger S. Bagnall,. *“*The Beginnings of the Roman Provincial Census: A New Declaration from 3 BCE.” *Greek, Roman, and Byzantine Studies* 55, no. 3 (2015): 637–653.

Clearview. “How Clearview AI Works.” Last modified March 24, 2020. Accessed June 4, 2020. https://clearview.ai/.

Clough, Patricia Ticineto, Karen Gregory, Benjamin Haber, and R. Joshua Scannell. *“*The Datalogical Turn.” In *The User Unconscious: On Affect, Media, and Measure,* edited by Patricia Ticineto Clough, 94–114. Minneapolis: University of Minnesota Press, 2018.

CNN Business. “Clearview AI's founder Hoan Ton-That speaks out.” YouTube, 32:53, March 6, 2020. https://youtu.be/q-1bR3P9RAw.

Conger, Kate, Richard Fausset, and Serge F. Kovaleski. “San Francisco Bans Facial Recognition Technology.” *The New York Times*, May 14, 2019. https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html.

Deleuze, Gilles, and Félix Guattari. *Anti-Ödipus*. Translated by Bernd Schwibs. Frankfurt am Main: Suhrkamp, 2019.

Del Santo, Flavio. "Indeterminism, causality and information: Has physics ever been deterministic?." FQXi Community, March 11, 2020. https://fqxi.org/community/forum/topic/3436.

D'Ignazio, Catherin and Lauren F. Klein. *Data Feminism*. Cambridge, MA: The MIT Press, 2020.

Dongus, Ariana. "Galton's Utopia – Data Accumulation In Biometric Capitalism." *spheres Journal for Digital Culture*, no. 17 (November 2019). http://spheres-journal.org/galtons-utopia-data-accumulation-in-biometric-capitalism/.

Electronic Privacy Information Center. "Algorithms in the Criminal Justice System: Risk Assessment Tools." Last modified 2020. Accessed June 4, 2020. https://epic.org/algorithmic-transparency/crim-justice/.

European Commission. "On the threshold to urban panopticon?." Last modified April 12, 2005. Accessed May 15, 2020. https://cordis.europa.eu/project/id/HPSE-CT-2001-00094/.

The European Data Protection Board. "Guidelines 3/2019 on processing of personal data through video devices, Version 2.0." Adopted on January 29, 2020. https://edpb.europa.eu/sites/edpb/files/consultation/edpb_guidelines_201903_videosurveillance.pdf.

---. "Guidelines 05/2020 on consent under Regulation 2016/679, Version 1.1." Adopted on May 4, 2020. https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf.

Eveleth, Rose. "The biggest lie tech people tell themselves — and the rest of us." *Vox*, October 8, 2019. https://www.vox.com/the-highlight/2019/10/1/20887003/tech-technology-evolution-natural-inevitable-ethics.

Face Aging Group. "Publications & Conferences." Last modified October, 9, 2019. Accessed April 30, 2020. http://www.faceaginggroup.com/?page_id=21.

Garvie, Clare. "Garbage in, Garbage out: Face recognition on flawed data." Georgetown Law. Centre on Privacy & Technology, May 16, 2020. https://www.flawedfacedata.com/.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, Kate Crawford. "Datasheets for Datasets." Proceedings of the 5th Workshop on Fairness, Accountability, and Transparency in Machine Learning: Stockholm, Sweden, 2018. https://arxiv.org/abs/1803.09010.

Golebiewski, Michael, and danah boyd. “DATA VOIDS WHERE MISSING DATA CAN EASILY BE EXPLOITED.” *DATA & SOCIETY*, November 2019. https://datasociety.net/wp-content/uploads/2019/11/Data-Voids-2.0-Final.pdf.

Harvey, Adam. “About Adam Harvey.” Adam Harvey, last modified August 1, 2019. Accessed March 28, 2020. https://ahprojects.com/about/.

---. “Another surveillance training dataset take down.” Twiter, June 12, 2020. https://twitter.com/adamhrv/status/1271370337087885313?s=20.

---. “Is distributing the torrent for the Microsoft-Celeb face recognition training dataset illegal?” Twitter, June 12, 2020. https://twitter.com/adamhrv/status/1271219064866852864?s=20.

---. “MegaPixels: Faces.” Adam Harvey, last modified November 1, 2017. Accessed March 28, 2020. https://ahprojects.com/megapixels-glassroom/.

---. “Projects.” Adam Harvey, last modified February 1, 2019. Accessed March 28, 2020. https://ahprojects.com/projects/.

---. “The skeptics corner.” Twitter, June 11, 2020. https://twitter.com/adamhrv/status/1270999156660875264?s=20.

Harvey, Adam, and Jules LaPlace. “Art and research publication investigating the ethics, origins, and individual privacy implications of face recognition datasets.” MegaPixels, last modified March 11, 2020. Accessed May 2, 2020. https://megapixels.cc/.

---. “Brainwash Dataset.” MegaPixels, last modified May 2, 2020. Accessed May 2, 2020. https://megapixels.cc/brainwash/.

---. “Duke MTMC Dataset.” MegaPixels, last modified May 2, 2020. Accessed May 2, 2020. https://megapixels.cc/duke_mtmc/.

---. “MegaFace Dataset.” MegaPixels, last modified October 22, 2019. Accessed May 2, 2020. https://megapixels.cc/megaface/.

---. “Microsoft Celeb Dataset.” MegaPixels, last modified May 2, 2020. Accessed May 2, 2020. https://megapixels.cc/msceleb/.

---. “Oxford Town Centre Dataset.” MegaPixels, last modified May 2, 2020. Accessed May 2, 2020. https://megapixels.cc/oxford_town_centre/.

---. “Unconstraint College Students Dataset.” MegaPixels, last modified April 18, 2019. Accessed May 2, 2020. https://megapixels.cc/uccs/.

---. “Wildtrack Dataset.” MegaPixels, last modified May 2, 2020. Accessed May 2, 2020. https://megapixels.cc/wildtrack/.

Hempel, Leon, and Eric Töpfer. “Working Paper No.1: Inception Report.” Centre for Technology and Society Technical University Berlin, 2002.

http://www.urbaneye.net/results/ue_wp1.pdf.

Hill, Kashmir. “Before Clearview Became a Police Tool, It Was a Secret Plaything of the Rich.” *The New York Times*, March 5, 2020. https://www.nytimes.com/2020/03/05/technology/clearview-investors.html.

---. “The Secretive Company That Might End Privacy as We Know It.” *The New York Times*, January 18, 2020. https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html.

---. “Wrongfully Accused by an Algorithm.” *The New York Times*, June 24, 2020. https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html.

Hill, Kashmir, and Aaron Krolik. “How Photos of Your Kids Are Powering Surveillance Technology.” The *New York Times*, corrected October 11, 2019. https://www.nytimes.com/interactive/2019/10/11/technology/flickr-facial-recognition.html.

Internet World Stats. “World Internet Users and 2020 Population Stats.” Last modified June 2, 2020. Accessed June 5, 2020. https://www.internetworldstats.com/stats.htm.

Kemelmacher-Shlizerman, Ira. “About.” Ira Kemelmacher-Shlizerman, last modified May 22, 2020. Accessed June 4, 2020. https://sites.google.com/view/irakemelmacher/about?authuser=0.

--- “How and Why Did University of Washington Professor Ira Kemelmacher-Shlizerman Build Dreambit and Sell To Facebook.” Filmed May 24, 2017 at LDV Vision Summit. Vimeo, 11:07, August 1, 2017. https://vimeo.com/227942118.

Kemelmacher-Shlizerman, Ira, and Steve Seitz. “MegaFace and MF2: Million-Scale Face Recognition.” MegaFace, last modified March 31, 2020. Accessed June 17, 2020. http://megaface.cs.washington.edu/.

Klemens, Ben.“Apophenia a Library for Scientific Computing.” Apophenia. Accessed June 3, 2020. http://apophenia.info/.

Krieg, Peter. *Die Paranoide Maschine*. Leipzig: E.A. Seemann, 2005.

Krishna, Arvind. “IBM CEO’s Letter to Congress on Racial Justice Reform.” IBM, published June 8, 2020. Accessed June 17, 2020. https://www.ibm.com/blogs/policy/facial-recognition-susset-racial-justice-reforms/.

Kurzweil, Ray. *The Singularity Is near: When Humans Transcend Biology*. London: Duckworth, 2008.

Lanier, Jaron, and Glen Weyl. “AI is an Ideology, Not a Technology.” *WIRED*, March 15, 2020. https://www.wired.com/story/opinion-ai-is-an-ideology-not-a-technology/.

Li, Fei-Fei. “How We Teach Computers To Understand Pictures.” TED 15, March 17, 2015. New World AI, May 23, 2020. https://www.newworldai.com/how-we-teach-computers-to-understand-pictures/.

---. “On AI and Machine Learning.” Filmed April 10, 2017 at Grace Hopper Celebration. YouTube, 28:20, February 6, 2018. https://youtu.be/XlnbNFW2tX8.

Lobe, Adrian. “Wer keine Biometrie hat, ist kein Bürger.” *Süddeutsche Zeitung*, October 27, 2018. https://www.sueddeutsche.de/digital/biometrie-gesichtserkennung-fingerabdruck-spracherkennung-1.4183394.

Lohr, Steve. “Facial Recognition Is Accurate, if You’re a White Guy.” *The New York Times,* February 9, 2018. https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html.

---. “The Origins of ‘Big Data': An Etymological Detective Story.” *The New York Times BITS*, February 1, 2013. https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/.

Merlan, Anna. “Here’s the File Clearview AI Has Been Keeping on Me, and Probably on You Too.” *Vice*, February 28, 2020. https://www.vice.com/en_us/article/5dmkyq/heres-the-file-clearview-ai-has-been-keeping-on-me-and-probably-on-you-too.

Murgia, Madhumita. “Who’s using your face? The ugly truth about facial recognition.” *Financial Times*, last modified September 18, 2019. https://www.ft.com/content/cf19b956-60a2-11e9-b285-3acd5d43599e.

Nadim, Tahani. “c u soon humans need to sleep now so many conversations today thx.” In *The influencing Machine*, edited by Tahani Nadim, 72–81. Berlin: nGbK, 2018.

Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press, 2018. Apple Books.

---. “Google Has a Striking History of Bias Against Black Girls.” *TIME*, March 26, 2018. https://time.com/5209144/google-search-engine-algorithm-bias-racism/.

von Nordheim, Gerret. “Suchergebnisse zu 'Bürgerkrieg Deutschland'.” Twitter, September 28, 2018. https://twitter.com/gvnordheim/status/1045647563058302976.

Norris, Clive. “The Global Growth of Camera Surveillance.” In *Eyes Everywhere: the Global Growth of Camera Surveillance*, edited by Aaron Doyle, Randy K. Lippert, and David Lyon, 23–45. London: Routledge, 2012.

Ochigame, Rodrigo. “THE INVENTION OF ETHICAL AI: How Big Tech Manipulates Academia to Avoid Regulations.” *The Intercept*, December 20, 2019. https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/.

O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group, 2016. Apple Books.

Ọnụọha, Mimi. “Broadway won’t document its dramatic race problem, so a group of actors spent five years quietly gathering this data themselves.” *Quartz, December 4, 2016.* https://qz.com/842610/broadways-race-problem-is-unmasked-by-data-but-the-theater-industry-is-still-stuck-in-neutral/.

---. “On Missing Data Sets.” *GitHub*, January 25, 2018. https://github.com/MimiOnuoha/missing-datasets.

---. “The Point of Collection.” *Medium. Data & Society: Points*, October 31, 2016. https://points.datasociety.net/the-point-of-collection-8ee44ad7c2fa.

---. “We Are Searching For.” MIMI ỌNỤỌHA. Last modified February 13, 2015. Accessed April 25, 2020. http://mimionuoha.com/we-are-searching-for.

---. “What Is Missing Is Still There.” Filmed November 1, 2019 at Kikk Festival. YouTube Video, 39:08, March 2, 2020. https://youtu.be/57Lgztk62uY.

Palmeri, Tara. “Municipal ID law has ‘delete in case of Tea Party’ clause.” *New York Post*, February 16, 2015. https://nypost.com/2015/02/16/municipal-id-law-has-delete-in-case-of-tea-party-clause/.

Parisi, Luciana. *“*Instrumental Reason, Algorithmic Capitalism, and the Incomputable.*”* In *Alleys of Your Mind: Augmented Intelligence and Its Traumas*, edited by Matteo Pasquinelli, 25–137. Lüneburg: Meson Press, 2015.

---. “XENO-PATTERNING: predictive intuition and automated imagination.” *Journal of the theoretical Humanities* 24, no. 1 (February 2019): 82–97.

Pasquinelli, Matteo. “How a Machine Learns and Fails – A Grammar of Error for Artificial Intelligence.” *spheres Journal for Digital Culture*, no. 17 (November 2019). http://spheres-journal.org/how-a-machine-learns-and-fails-a-grammar-of-error-for-artificial-intelligence/.

---. “Machines that Morph Logic: Neural Networks and the Distorted Automation of Intelligence as Statistical Inference.” *Glass Bead Journal,* no. 1 (2017). https://www.glass-bead.org/article/machines-that-morph-logic/?lang=enview.

Pasquinelli, Matteo, and Vladan Joler. “The Nooscope Manifested: Artificial Intelligence as Instrument of Knowledge Extractivism.” KIM research group (Karlsruhe University of Arts and Design) and Share Lab (Novi Sad), May 1, 2020 (preprint forthcoming for *AI and Society*). https://nooscope.ai.

Pearl, Judea. *The Book of Why: the New Science of Cause and Effect*. NEW YORK: BASIC BOOKS, 2020.

Quiñonero Candela, Joaquin. “Facebook and the Technical University of Munich Announce New Independent TUM Institute for Ethics in Artificial Intelligence.” About Facebook, January 20, 2019. https://about.fb.com/news/2019/01/tum-institute-for-ethics-in-ai/.

Raji, Inioluwa Deborah , and Joy Buolamwini. “Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products.” Conference on Artificial Intelligence, Ethics, and Society, 2019. https://dam-prod.media.mit.edu/x/2019/01/24/AIES-19_paper_223.pdf.

Sambuli, Nanjira, Renata Ávila Pinto, Esra'a Al Shafei and Mi You. “RE-IMAGINING THE INTERNET: ROADMAPS TO DIGITAL EQUALITY.” Filmed June 4, 2020, at Goethe Institut Latitude Digital Festival. YouTube, 04:14:17 - 05:59:14, June 4, 2020. https://youtu.be/LW7UXFjgbcI.

Sauer, Kristina. “Ordnung ist das halbe Leben!.” In *5300 Jahre Schrift,* edited by Michaela Böttner, Ludger Lieb, Christian Vater, Christian Witschel, 14–17. Heidelberg: Verlag das Wunderhorn, 2017.

Schneier, Bruce. “We’re Banning Facial Recognition. We’re Missing the Point.” *The New York Times*, January 20, 2020. https://www.nytimes.com/2020/01/20/opinion/facial-recognition-ban-privacy.html.

Schwab, Katharine. “This Designer Is Fighting Back Against Bad Data – With Feminism.” *Fast Company*, April 16, 2018. https://www.fastcompany.com/90168266/the-designer-fighting-back-against-bad-data-with-feminism.

Selinger, Evan. “Facial recognition technology has *unique* affordances.” Twitter, January 20, 2020. https://twitter.com/EvanSelinger/status/1219320097678004229.

Selinger, Evan, and Woodrow Hartzog. “Read Your Facial Expressions? The benefits do not come close to outweighing the risks.” *The New York Times*, October 17, 2019. https://www.nytimes.com/2019/10/17/opinion/facial-recognition-ban.html.

Shaer, Matthew. “The False Promise of DNA Testing.” *The Atlantic*, 15 June. 2016. https://www.theatlantic.com/magazine/archive/2016/06/a-reasonable-doubt/480747/.

Shah, Nishant. “From GUI to no UI.” Filmed November 2, 2019 at IMPAKT Festival. YouTube, 01:10:23, November 11, 2019.
https://youtu.be/vaeoAeEBNcI.

Sinders, Caroline. “About.” Caroline Sinders, last modified April 25, 2017. Accessed April 28, 2020.
https://carolinesinders.com/about/
---. “AI is more than math: using art and design to interrogate bias in AI.” Filmed May 6, 2019 at re:publica. YouTube, 44:56, May 6, 2019.
https://youtu.be/e0wyEnuRi3U.
---. “Building a Feminist Data Set for a Feminist AI.” Schloss Post, October 20, 2017.
https://schloss-post.com/building-feminist-data-set-feminist-ai/.
---. “Dark Patterns and Design Policy.” *Medium. Data & Society: Points*, May 20, 2020.
https://points.datasociety.net/dark-patterns-and-design-policy-75d1a71fbda5.
---. “Feminist Data Set.” Caroline Sinders, published May 26, 2020.
https://carolinesinders.com/wp-content/uploads/2020/05/Feminist-Data-Set-Final-Draft-2020-0526.pdf.
---. “How UX Design Creates Trust.” Adobe, October 8, 2019.
https://xd.adobe.com/ideas/perspectives/social-impact/building-trust-through-user-experience-design/.
---. “That time the Internet sent a SWAT team to my mom's house.” B*oing Boing*, July 24, 2015.
https://boingboing.net/2015/07/24/that-time-the-internet-sent-a.html.

Sinders, Caroline, and Cade Diehm. “Calculator.” Technically Responsible Knowledge, last updated April 24, 2020. Accessed May 31, 2020.
http://trk.network/#calculator.

Stalder, Felix. *The Digital Condition*. Translated by Valentine Pakis. Newark: Polity Press, 2018.

Statt, Nick, “Amazon bans police from using its facial recognition technology for the next year.” *THE VERGE,* June 10, 2020.
https://www.theverge.com/2020/6/10/21287101/amazon-rekognition-facial-recognition-police-ban-one-year-ai-racial-bias.

Steyerl, Hito. “A Sea of Data: Apophenia and Pattern (Mis-)Recognition.” *e–flux Journal*, no. 72 (April 2016). https://www.e-flux.com/journal/72/60480/a-sea-of-data-apophenia-and-pattern-mis-recognition/.

Taigman, Yaniv, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. “DeepFace: Closing the Gap to Human-Level Performance in Face Verification.” FACEBOOK Research, June 24, 2014.
https://research.fb.com/wp-content/uploads/2016/11/deepface-closing-the-gap-to-human-level-performance-in-

face-verification.pdf.

Tang, Audrey. “Digital Social Innovation.” Filmed May 6, 2019 at re:publica. YouTube, 53:59, May 12, 2019. https://youtu.be/jl9mt5OEH0c.

Turing, Alan. “On Computable Numbers, with an Application to the Entscheidungsproblem.” In *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life: Plus The Secrets of Enigma*, edited by B. Jack Copeland, 58–90.Oxford: Oxford University Press, 2004.

Urban Data Eye. “Actionable Data to optimize.” Last modified May 9, 2019. Accessed April 26, 2020. http://urbandataeye.com/.

Urbaneye Project. “Welcome to the Urbaneye Project.” Last modified March 16, 2006. Accessed April 3, 2020. http://www.urbaneye.net.

VICE News. “Moscow's Facial Recognition Tech will Outlast the Coronavirus.” YouTube, 07:43, April 16, 2020. https://youtu.be/pbGq3REp4PI?t=61.

Vincent, James. “Transgender YouTubers had their videos grabbed to train facial recognition software.” *THE VERGE*, August 22, 2017. https://www.theverge.com/2017/8/22/16180080/transgender-youtubers-ai-facial-recognition-dataset.

Zuboff, Shoshana. *The Age of Surveillance Capitalism: the Fight for Human Future at the New Frontier of Power*. New York: PublicAffairs, 2019. Apple Books.