

## AI TRAPS: Automating Discrimination

---

14 & 15 June 2019

Press Coverage



## Media Overview

### Articles

furtherfield - 02.08.2019, *AI TRAPS: Automating Discrimination*

heise online - 15.06.2019, *Gegen die Fallen der KI: Die feministische Alexa kommt*

heise online - 16.06.2019, *Datenforscherin. "Künstliche Intelligenz ist kaputt"*

taz.de - Event Insertion, 06.06.2019, 13.04.2014, 14.06.2019

### Interviews

Exberliner - 13.06.2019, *It's the structures, not the tech: Os Keyes*

Spiegel Online - 15.06.2019, *Künstliche Intelligenz Wie viel Rassismus steckt in Algorithmen?*

---

# Furtherfield

---



## AI TRAPS: Automating Discrimination

02/08/2019

Lorin Decarli

On June the 16th Tatiana Bazzichelli and Lieke Ploeger presented a new Disruption Network Lab conference entitled "AI TRAPS" to scrutinize Artificial Intelligence and automatic discrimination. The conference touched several topics from biometric surveillance to diversity in data, giving a closer look at how AI and algorithms reinforce prejudices and biases of its human creators and societies, to find solutions and countermeasures.

A focus on facial recognition technologies opened the first panel "THE TRACKED & THE INVISIBLE: From Biometric Surveillance to Diversity in Data Science" discussing how massive sets of images have been used by academic, commercial, defence and intelligence agencies around the world for their research and development. The artist and researcher Adam Harvey addressed this tech as the focal point of an emerging authoritarian logic, based on probabilistic determinations and the assumption that identities are static and reality is made through absolute norms. The artist considered two recent reports about the UK and China showing how this technology is yet unreliable and dangerous. According to data released under the UK's Freedom of Information Law, 98% of "matches" made by the English Met police using facial recognition were mistakes. Meanwhile, over 200 million cameras are active in China and – although only 15% are supposed to be technically implemented for effective face recognition – Chinese authorities are deploying a new system of this tech to racial profile, track and control the Uighurs Muslim minority.

Big companies like Google and Facebook hold a collection of billions of images, most of which are available inside search engines (63%), on Flickr (25%) and on IMDb (11 %). Biometric companies around the world are implementing facial recognition algorithms on the pictures of common people, collected in unsuspected places like dating-apps and social media, to be used for private profit purposes and governmental mass-surveillance. They end up mostly in China (37%), US (34%), UK (21%) and Australia (4%), as Harvey reported.

Metis Senior Data Scientist Sophie Searcy, technical expert who has also extensively researched on the subject of diversity in tech, contributed to the discussion on such a crucial issue underlying the design and implementation of AI, enforcing the description of a technology that tends to be defective, unable to contextualise and consider the complexity of the reality it interacts with. This generates a lot of false predictions and mistakes. To maximise their results and reduce mistakes tech companies and research institutions that develop algorithms for AI use the Stochastic Gradient Descent (SGD) technique. This enables to pick a few samples selected randomly from a dataset instead of analysing the whole of it for each iteration, saving a considerable amount of time. As Searcy explained during the talk with the panel moderator, Adriana Groh, this technique needs huge amount of data and tech companies are therefore becoming increasingly hungry for them.



In order to have a closer look at the relation between governments and AI-tech, the researcher and writer Crofton Black presented the study conducted with Cansu Safak at The Bureau of Investigative Journalism on the UK government's use of big data. They used publicly available data to build a picture of companies, services and projects in the area of AI and machine learning, to map what IT systems the British government has been buying. To do so they interviewed experts and academics, analysed official transparency data and scraped governmental websites. Transparency and accountability over the way in which public money is spent are a requirement for public administrations and they relied on this principle, filing dozens of requests under the Freedom of Information Act to public authorities to get audit trails. Thus they mapped an ecosystem of the corporate nexus between UK public sector and corporate entities. More than 1,800 IT companies, from big ones like BEA System and IBM to small ones within a constellation of start-ups.

As Black explained in the talk with the moderator of the keynote Daniel Eriksson, Transparency International Head of Technology, this investigation faced systemic problems with disclosure from authorities, that do not keep transparent and accessible records. Indeed just 25% of the UK-government departments provided some form of info. Therefore details of the assignments are still unknown, but it is at least possible to list the services those companies deploying AI and machine learning can offer governments: connect data and identify links between people, objects, locations; set up automated alerts in the context of border and immigration control, spotting out changes in data and events of interest; work on passports application programs, implementing the risk-based approaches to passports application assessments; work on identity verification services using smartphones, gathering real time biometric authentications. These are just few examples.



Maya Indira Ganesh opened the panel "AI FOR THE PEOPLE: AI Bias, Ethics & the Common Good" questioning how tech and research have historically been almost always developed and conducted on prejudiced parameters, falsifying results and distorting reality. For instance, data about women's heart attacks hadn't been taken in consideration for decades, until doctors and scientists determined that ECG-machines calibrated on the data collected from early '60s could neither predict heart attacks in women, nor give reliable data for therapeutic purposes, because they were trained only on male population. Just from 2007 ECG-machines were recalibrated on parameters based on data collected from female individuals. It is not possible to calculate the impact this gender inequality had on the development of modern cardiovascular medicine and on the lives of millions of women.

As the issue of algorithmic bias in tech and specifically in AI grows, all big tech firms and research institutions are writing ethics charters and establishing ethics boards sponsoring research in these topics. Detractors often refer to it as ethics-washing, which Ganesh finds a trick to mask ethics and morality as something definable in universal terms or scale: though it cannot be computed by machines, corporations need us to believe that ethics is something measurable. The researcher suggested that in such a way the abstraction and the complexity of the machine get easy to process as ethics becomes the interface used to obfuscate what is going on inside the black box and represent its abstractions. "But these abstractions are us and our way to build relations" she objected.

Ganesh wonders consequently according to what principle it shall be acceptable to train a facial recognition system, basing it on video of transgender people, as it happened in the alarming "Robust transgender face recognition" research, based on data from people undergoing hormone replacement therapy, Youtube videos, diaries and time-lapse documentation of the transition process. The HRT Transgender Dataset used to train AI to recognize transgender people worsens the harassment and the targeting that trans-people already experience daily, targeting and harming them as a group. However, it was partly financed by FBI and US-Army, confirming that law enforcement and national security agencies appear to be very interested in these kinds of datasets and look for private companies and researchers able to provide it.

In this same panel professor of Data Science and Public Policy Slava Jankin reflected on how machine learning can be used for common good in the public sector. As it was objected during the discussion moderated by Nicole Shephard, Researcher on Gender, Technology and Politics of Data, the "common good" isn't easy to define, and like ethics it is not universally given. It could be identified with those goods that are relevant to guarantee and determine the respect of human rights and their practice. The project that Jankin presented was developed inside the Essex Centre for Data analytics in a synergic effort of developers, researches, universities and local authorities. Together, they tried to build an AI able to predict within reliability where children lacking *school readiness* are more likely to be found geographically, to support them in their transition and gaining competencies, considering social, economic and environmental conditions.

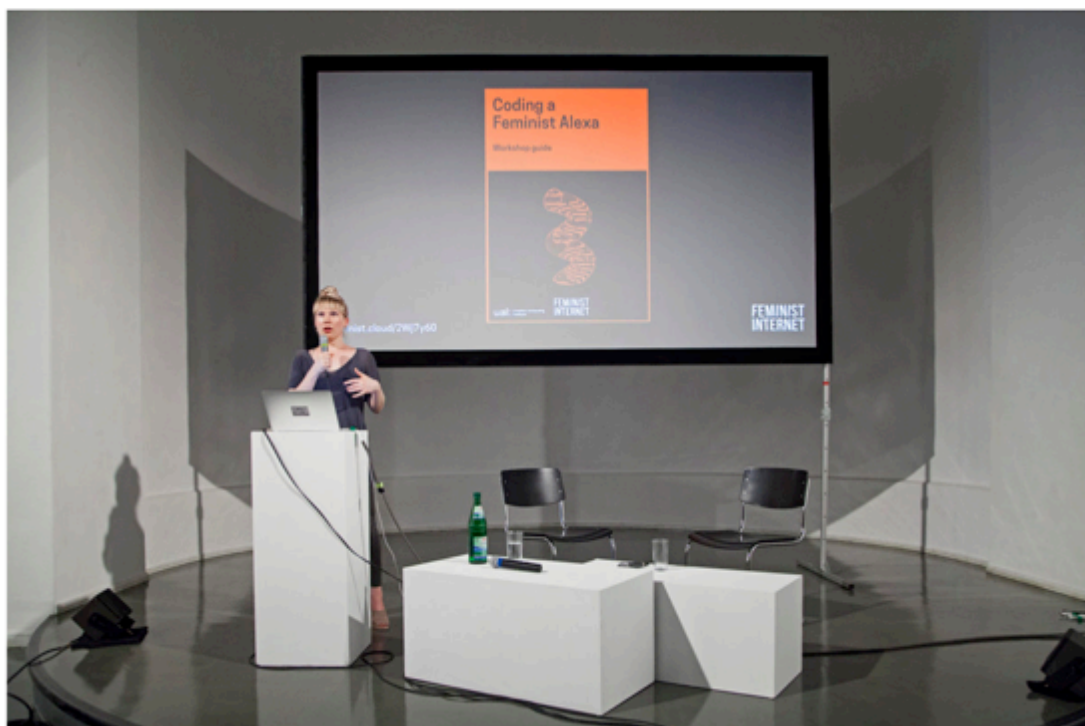


The first keynote of the conference was the researcher and activist Charlotte Webb, who presented her project Feminist Internet in the talk "WHAT IS A FEMINIST AI?"

<<There is not just one possible internet and there is not just one possible feminism, but only possible feminisms and possible internets>>. Starting from this assumption Webb talked about Feminist Human Computer Interaction, a discipline born to improve understandings about how gender identities and relations shape the design and use of interactive technologies. Her Feminist Internet is a no profit organisation funded to make internet a more equal space for women and other marginalized groups. Its approach combines art, design, critical thinking, creative technology development and feminism, seeking to build more responsible and bias-free AI able to empower people considering the causes of marginalization and discrimination. In her words, a feminist AI is not an algorithm and is not a system built to evangelize about a certain political or ideological cause. It is a tool that aims at recognizing differences without minimizing them for the sake of universality, meeting human needs with the awareness of the entire ecosystem in which it sits.

Tech adapts plastically to pre-existing discriminations and gender stereotypes. In a recent UN report, the 'female' obsequiousness and the servility expressed by digital assistants like Alexa, the Google Assistant, are defined as example of gender biases coded into tech products, since they are often projected as young women. They are programmed to be submissive and accept abuses. As stated by Feldman (2016) by encouraging consumers to understand the objects that serve them as women, technologists abet the prejudice by which women are considered objects. With her projects, Webb pushes to create alternatives that educate to shift this systemic problem – rather than complying with market demands – first considering that there is a diversity crisis in the AI sector and in the Silicon Valley. Between 2.5 and 4% of Google, Facebook and Microsoft employees are black, whilst there are no public data on transgender workers within these companies. Moreover, as Webb pointed out, just 22% of the people building AI right now are female, only 18% of authors at major AI-conferences are women, whilst over 80% of AI-professors are men. Considering companies with decisive impact on society women comprise only 15% of AI research staff at Facebook and 10% in Google.

Women, people of colour, minorities, LGBTQ and marginalized groups are substantially not deciding about designing and implementing AI and algorithms. They are excluded from the processes of coding and programming. As a result the work of engineers and designers is not inherently neutral and the automated systems that they build reflect their perspectives, preferences, priorities and eventually their bias.



Washington Tech Policy Advisor Mutale Nkonde focused on this issue in her keynote "RACIAL DISCRIMINATION IN THE AGE OF AI." She opened her dissertation reporting that Google's facial intelligence team is composed by 893 people, and just one is a black woman, an intern. Questions, answers and predictions in their technological work will always reflect a political and socioeconomic point of view, consciously or unconsciously. A lot of the tech-people confronted with this wide-ranging problem seem to undermine it, showing colour-blindness tendencies about what impacts their tech have on minorities and specifically black people. Historically credit scores are correlated with racist segregated neighbourhoods and risk analyses and predictive policing data are corrupted by racist prejudice, leading to biased data collection reinforcing privileges. Without a conscious effort to address racism in technology, new technologies will replicate old divisions and conflicts. By instituting policies like facial recognition we just replicate rooted behaviours based on racial lines and gender stereotypes mediated by algorithms. Nkonde warns that civil liberties need an update for the era of AI, *advancing racial literacy in Tech.*

In a talk with the moderator, the writer Rhianna Ilube, the keynote Nkonde recalled that in New York's poor and black neighbourhood with historically high crime and violence rates, Brownsville, a private landlord in social housing wanted to exchange keys for facial recognition software, so that either people accept surveillance, or they lose their homes. The finding echoes wider concerns about the lack of awareness of racism. Nkonde thinks that white people must be able to cope with the inconvenience of talking about race, with the countervailing pressures and their lack of cultural preparation, or simply the risk to get it wrong. Acting ethically isn't easy if you do not work on it and many big tech companies just like to crow about their diversity and inclusion efforts, disclosing diversity goals and offering courses that reduce bias. However, there is a high level of racial discrimination in tech sector and specifically in the Silicon Valley, at best colour-blindness – said Nkonde – since many believe that racial classification does not limit a person's opportunities within the society, ignoring that there are instead economic and social obstacles that prevent full individual development and participation, limiting freedom and equality, excluding marginalized and disadvantaged groups from the political, economic, and social organization. Nkonde concluded her keynote stressing that we need to empower minorities, providing tools that allow overcoming autonomously socio-economic obstacles, to fully participate in society. It is about sharing power, taking in consideration the unconscious biases of people, for example starting from those designing the technology.



The closing panel “ON THE POLITICS OF AI: Fighting Injustice & Automatic Supremacism” discussed the effect of a tool shown to be not neutral, but just the product of the prevailing social economical model.

Dia Kayyali, Leader of the Tech and Advocacy program at WITNESS, described how AI is facilitating white supremacy, nationalism, racism and transphobia, recalling the dramatic case of the Rohingya persecution in Myanmar and the oppressive Chinese social score and surveillance systems. Pointing out critical aspects the researcher reported the case of the Youtube anti-extremism-algorithm, which removed thousands of videos documenting atrocities in Syria in an effort to purge hate speech and propaganda from its platform. The algorithm was trained to automatically flag and eliminate content that potentially breached its guidelines and ended up cancelling documents relevant to prosecute war crimes. Once again, the absence of the ability to contextualize leads to severe risks in the way machines operate and make decisions. Likewise, applying general parameters without considering specificities and the complex concept of identity, Facebook imposed in 2015 new policies and arbitrarily exposed drag queens, trans people and other users at risk, who were not using their legal names for safety and privacy reasons, including domestic violence and stalking.

Researcher on gender, tech and (counter) power Os Keyes considered that AI is not the problem, but the symptom. The problem are the structures creating AI. We live in an environment where few highly wealthy people and companies are ruling all. We have bias in AI and tech because their development is driven by exactly those same individuals. To fix AI we have to change requirements and expectations around it; we can fight to have AI based on explainability and transparency, but eventually if we strive to fix AI and do not look at the wider picture, in 10 years the same debate over another technology will arise. Keyes considered that since its very beginning AI-tech was discriminatory, racialized and gendered, because society is capitalist, racist, homo-transphobic and misogynistic. The question to pose is how we start building spaces that are prefigurative and constructed on values that we want a wider society to embrace.

As the funder and curator of the Disruption Network Lab Tatiana Bazzichelli pointed out during the moderation of this panel, the problem of bias in algorithms is related to several major “bias traps” that algorithm-based prediction systems fail to win. The fact that AI is political – not just because of the question of what is to be done with it, but because of the political tendencies of the technology itself – is the real aspect to discuss.

In his analysis of the political effects of AI, Dan McQuillan, Lecturer in Creative and Social Computing from the London University, underlined that while the reform of AI is endlessly discussed, there seems to be no attempt to seriously question whether we should be using it at all. We need to think collectively about ways out, learning from and with each other rather than relying on machine learning. Countering thoughtlessness of AI with practices of solidarity, self-management and collective care is what he suggests because bringing the perspective of marginalised groups at the core of AI practice, it is possible to build a new society within the old, based on social autonomy.

What McQuillan calls the AI realism appears to be close to the far-right perspective, as it trivialises complexity and naturalises inequalities. The character of learning through AI implicates indeed reductive simplifications, and simplifying social problems to matters of exclusion is the politics of populist and Fascist right. McQuillan suggests taking some guidance from the feminist and decolonial technology studies that have cast doubt on our ideas about objectivity and neutrality. An antifascist AI, he explains, shall involve some kinds of people’s councils, to put the perspective of marginalised groups at the core of AI practice and to transform machine learning into a form of critical pedagogy.

*Pic 7: Dia Kayyali, Os Keyes, Dan McQuillan and Tatiana Bazzichelli during the panel “ON THE POLITICS OF AI: Fighting Injustice & Automatic Supremacism”*





We see increasing investment on AI, machine learning and robots. Automated decision-making informed by algorithms is already a predominant reality, whose range of applications has broadened to almost all aspects of life. Current ethical debates about the consequences of automation focus on the rights of individuals and marginalized groups. However, algorithmic processes generate a collective impact too, that can only be addressed partially at the level of individual rights, as it is the result of a collective cultural legacy. A society that is soaked in racial and sexual discriminations will replicate them inside technology.

Moreover, when referring to surveillance technology and face recognition software, existing ethical and legal criteria appear to be ineffective and a lack of standards around their use and sharing just benefit its intrusive and discriminatory nature.

Whilst building alternatives we need to consider inclusion and diversity: If more brown and black people would be involved in the building and making of these systems, there would be less bias. But this is not enough. Automated systems are mostly trying to identify and predict risk, and risk is defined according to cultural parameters that reflect the historical, social and political milieu, to give answers able to fit a certain point of view and make decisions. What we are and where we are as a collective, what we have achieved and what we still lack culturally is what is put in software to make those same decisions in the future. In such a context a diverse team within a discriminatory conflictual society might find ways to flash the problem of bias away, but it will get somewhere else.

The truth is that automated discrimination, racism and sexism are integrated in tech-infrastructures. New generation of start-ups are fulfilling authoritarian needs, commercialising AI-technologies, automating biases based on skin colour and ethnicity, sexual orientation and identity. They develop censored search engine and platforms for authoritarian governments and dictators, refine high-tech military weapons training them using facial recognition on millions of people without their knowledge. Governments and corporations are developing technology in ways that threaten civil liberties and human rights. It is not hard to imagine the impact of the implementation of tools for robotic gender recognition, within countries were non-white, non-male and non-binary individuals are discriminated. Bathrooms and changing rooms that open just by AI gender-detection, or cars that start the engine just if a man is driving, are to be expected. Those not gender conforming, who do not fit traditional gender structures, will end up being systematically blocked and discriminated.

Open source, transparency and diversity alone will not defeat colour-blinded attitudes, reactionary backlashes, monopolies, other-directed homologation and cultural oppression by design. As it was discussed in the conference, using algorithms to label people based on sexual identity or ethnicity has become easy and common. If you build a technology able to catalogue people by ethnicity or sexual identity, someone will exploit it to repress genders or ethnicities, China shows.

In this sense, no better facial recognition is possible, no mass-surveillance tech is safe and attempts at building *good tech* will continue to fail. To tackle bias, discrimination and harm in AI we have to integrate research on and development of technology with all of the humanities and social sciences, deciding to consciously create a society where everybody could participate to the organisation of our common future.

Curated by Tatiana Bazzichelli and developed in cooperation with Transparency International, this Disruption Network Lab-conference was the second of the 2019 series *The Art of Exposing Injustice*. More info, all its speakers and thematic could be found here: <https://www.disruptionlab.org/ai-traps>. The videos of the conference are on [Youtube](#) and the Disruption Network Lab is also on [Twitter](#) and [Facebook](#).

To follow the Disruption Network Lab sign up for its [Newsletter](#) and get informed about its conferences, ongoing researches and projects. The next Disruption Network Lab event "Citizen of evidence" is planned for [September 20-21](#) in Kunstquartier Bethanien Berlin. Make sure you don't miss it!

Photocredits: Maria Silvano for Disruption Network Lab

## Gegen die Fallen der KI: Die feministische Alexa kommt

Aktivistinnen haben Prototypen für androgyne feministische Sprachassistenten entwickelt. Damit wollen sie über KI-Monopole und Vorurteile gegen Frauen aufklären.

Lesezeit: 3 Min.  In Pocket speichern

   382



Charlotte Webb – Mitgründerin des Zirkels Feminist Internet auf der Konferenz "AI Traps" in Berlin, (Bild: Stefan Krempf)

15.06.2019 12:05 Uhr

Von Stefan Krempf

Sexismus und weithin gepflegte Vorurteile sind eine zunehmende Herausforderung für Entwickler von Anwendungen Künstlicher Intelligenz (KI). Selbst der Unesco ist inzwischen aufgefallen, dass digitale Assistenten oder Chatbots wie Amazons Alexa, Apples Siri, Googles Assistant oder Microsofts Cortana meist weibliche Stimmen zumindest voreingestellt haben und ziemlich unterwürfig sind. Und wenn Verbraucher in den Interaktionen mit den digitalen Helfern ermuntert werden, die ihnen dienenden Objekte als weiblich zu verstehen, unterstützen die Hersteller zugleich die Ansicht, dass Frauen Objekte sind.

Sprachassistenten würden oft als digitale Variante des Einkaufszettels oder der Aufgabenliste in Heim und Küche verwendet, erklärte die Mitgründerin des Zirkels Feminist Internet, Charlotte Webb, am Freitag auf der zweitägigen Konferenz "AI Traps" in Berlin, mit der das Disruption Network Lab gegen Diskriminierung durch KI vorgehen will. Die Verbraucher bevorzugten dafür offenbar weibliche Stimmen, sodass die an möglichst hohen Verkaufszahlen interessierten Produzenten diesem Wunsch nachkämen.

### Standards für feministische KI

Webb und ihre Mitstreiter haben sich so das Ziel gesetzt, eine "feministische Alexa" und weitere Sprachassistenten zu entwickeln, die vom gesamten Konversationsdesign her über Vorurteile gegen Frauen aufklären. Mit dem Projekt F'xa ist der Prototyp eines solchen Chatbots bereits fertig, fast zehn weitere Modelle unterschiedlichster Couleur haben die Aktivistinnen im Rahmen eines gemeinsam mit dem Creative Computing-Institut der University of the Arts London durchgeführten Workshops entworfen und Grundgerüste dafür programmiert.

Nach Standards für feministische KI mussten die Macher dabei nicht lange suchen, da die Aktivistinnen und Forscherinnen Josie Young sowie Shaowen Bardzell auf diesem Feld schon Vorlagen produziert hatten. Demnach geht es auf einer übergeordneten Ebene etwa darum, Nutzern algorithmische Entscheidungsfindungen der Software bewusst zu machen, pluralistische Ansichten und die Selbstreflexion zu stärken, systemische Zusammenhänge aufzuzeigen und eine echte Teilhabe schon im Designprozess zu erlauben.

## **Sprachassistent für Teenager**

Doch wie könnte eine feministische Konversation mit einer angemessenen Stimme konkret aussehen, wie wird das Gerät gegenüber dem Nutzer in diesem Sinne präsentiert? F'xa etwa "sagt niemals 'Ich'", brachte Webb ein Beispiel. Ein einschlägiger Bot sollte nicht zu menschlich wirken, um die Nutzer davon abzuhalten, zu intensive Gefühle zu dem KI-System aufzubauen. Bei den Baumstrukturen für die Antworten auf zu erwartende Fragen sei dann jedes Detail wichtig, um sexistische Nuancen weitestgehend auszuschließen.

B(o)(o) gehört mit zu den schon ausgearbeiteten Assistenten, dessen Motto es ist, nichts für tabu zu erklären. Ausgerichtet ist dieser auf Teenager, die ihre Identität während der Pubertät weiter ausformen. Gefragt gewesen sei ein "sicherer Raum für alle als peinlich empfundenen Fragen" etwa über sexuelle oder emotionale Normen, berichtete Webb. Das System höre vor allem zu und gebe manchmal in umgangssprachlichen, keinem Geschlecht eindeutig zurechenbaren Ton Tipps und betreibe vor allem Sexualaufklärung. So würden etwa Jungs beruhigt, die erstmals Haarwuchs an Beinen oder unter den Achseln feststellen, dass sie damit nicht zu einem Pelzmonster würden.

## **Feministische Zusatzfunktionen für Alexa**

Am besten sei es, menschliche Assoziationen in Bezug auf die Sprachassistenten weitgehend zu vermeiden, erläuterte die Feministin. Viel hänge im weiteren davon ab, wer überhaupt zu welchem Zeitpunkt sprechen dürfe und welche Passagen einer Unterhaltung aufgegriffen würden. Um tiefer in diese Designentscheidungen einzusteigen, führe der Zusammenschluss mittlerweile auch Workshops mit einer früheren Amazon-Programmiererin durch, um spezielle feministische Zusatzfunktionen ("Skills") für Alexa selbst zu bauen und in Umlauf zu bringen.

Dieser Schritt habe zunächst zu internen Diskussionen geführt, ob man überhaupt eine solche Plattform eines "Monopolisten" nutzen solle, räumte Webb ein. Auch dies verstoße streng genommen bereits gegen die feministische Linie, da der E-Commerce-Riese etwa auf Gratisleistungen von Nutzern setze und fragliche interne Arbeitsverhältnisse habe. Man sei aber übereingekommen, dass es besser sei, dieses System mit eigenen Anwendungen zu unterwandern und dessen Strukturen offen zu problematisieren, als sich ganz davon fernzuhalten.

## Zugehörigkeitsgefühle, die Nutzer zu der Technik entwickeln könnten

Generell umriss die Hauptrednerin der Konferenz die Grundzüge feministischer KI als ein System, in das jenseits der prinzipiellen Offenheit und Geschlechterlosigkeit nicht von vornherein ideologische oder politische Zwecke einkodiert werden dürften. Unterschiede müssten dabei anerkannt und die Zugehörigkeitsgefühle vorweggenommen werden, die Nutzer zu der Technik entwickeln könnten. Wichtig sei es auch, "feministische", vorurteilsfreie Daten in die Modelle einzubringen und Algorithmen mit einem Gespür für Diversität mit einem diversen Team zu entwickeln. Diskriminierungen jeglicher Art müssten unterbleiben.

Die Technologieforscherin Maya Indira Ganesh schätzte zumindest die kurzfristigen Erfolgsaussichten eines feministisch geprägten Systemwandels bei Künstlicher Intelligenz skeptisch ein. Es bräuchte sehr lange, um bessere Trainingssets für KI-Algorithmen zu produzieren, gab sie zu bedenken. Zumindest Firmen würden darauf nicht warten oder hinarbeiten, da sie im Kapitalismus rasch weiter Produkte verkaufen müssten. Um aus Vorurteilen herauszukommen, gebe es keine vorgefertigte Lösung.

## Ethische Regeln finden in der Praxis keine Anwendung

Schier "jeder" gebe sich zwar inzwischen ethische Regeln für Künstliche Intelligenz, verwies Ganesh etwa auf die EU, das Weltwirtschaftsforum oder China. In der Praxis werde in der Regel aber weiter "unethisch" gearbeitet. Google habe den eigens einberufenen Ethikrat sogar kurzfristig wieder auflösen müssen, nachdem publik geworden sei, dass ein Rassist dort eine führende Position innehaben sollte. Interner Druck von Mitarbeitern sei immer hilfreich, um Auswüchse wie den Einsatz von KI für militärische Systeme zu verhindern. Auch Standardisierungsbemühungen oder der Austausch auf Konferenzen unter Einschluss von Humanwissenschaftlern könnten mittelfristig einen Weg ebnen für Taxonomien gegen die Voreingenommenheit von Algorithmen, die sich aus den unterschiedlichsten, auch gesellschaftlichen Quellen nähre. (bme)

 Kommentare lesen (382)

# Datenforscherin: "Künstliche Intelligenz ist kaputt"

In KI-Modellen könnten immer mehr Trainingsdaten verarbeitet werden, was die Monopolbildung unterstütze, warnt die New Yorker Praktikerin Sophie Searcy.

Lesezeit: 3 Min.  In Pocket speichern

   376



(Bild: whiteMocca/Shutterstock.com)

16.06.2019 17:58 Uhr

Von Stefan Krempf

Firmen und Forschungseinrichtungen, die Algorithmen für Künstliche Intelligenz (KI) entwickeln, werden immer datenhungriger. Dies verdeutlichte Sophie Searcy, Data Scientist beim "KI-Bootcamp" Metis in New York, auf der am Samstag zu Ende gegangenen Konferenz "AI Traps" in Berlin. "Die Leute in Unternehmen reden nur noch darüber, wie sie Daten bekommen." Mit immer umfangreicheren Trainingsdaten-Sets wollten sie immer bessere KI-Lösungsmodelle entwickeln.

Leistungsfähigere Algorithmen an sich seien nicht schlecht, erläuterte Searcy. Sie halfen dabei, dass Maschinen die "echte Welt" besser und schneller einschätzen könnten. Jedes KI-Modell sei letztlich eine kleine Funktion, um "etwas Größeres einzufangen". Es gehe dabei um Lernprozesse in der Form, dass Erfahrungen verarbeitet würden, um bestehende Modelle so zu aktualisieren, dass sie auch für künftige Entwicklungen nützlich sind.



Sophie Searcy, Data Scientist beim "KI-Bootcamp" Metis in New York  
(Bild: Stefan Krempf)

## Daten, die man nicht hat

Bei der Erkennung von Fröschen anhand von Frosch-Fotos beispielsweise könnte es immer neue Spezies geben, die Aufnahmen zudem in einem immer wieder anderen Licht oder von einem anderen Winkel aus gemacht sein. Je mehr dieser Variablen berücksichtigt und verarbeitet werden könnten, desto genauer arbeite der Algorithmus. Bisläng ist es dabei laut der Praktikerin am schwierigsten, anhand verfügbarer Trainingsdaten einzuschätzen, "wie ein Modell mit Daten arbeitet, die man nicht hat".

Bisher habe es gute Gründe gegeben, nicht alle existierenden Daten zu sammeln, verdeutlichte Searcy anhand des Frosch-Beispiels. Weil man immer wieder das gesamte Datenset auf einmal habe beobachten müssen, sei die Trainingszeit durch das Einfügen neuer Bilder exponentiell gewachsen, was die Komplexität enorm erhöht habe.

Searcy forderte daher eine deutlich strengere Regulierung der datengetriebenen Branche. "Die Industrie muss als inhärentes Risiko behandelt werden" – so wie Pharmafirmen, die ein neues Arzneimittel entwickelten, unterstrich sie. Angesichts des "hohen Schadenspotenzial" reiche auch die Datenschutz-Grundverordnung (DSGVO) nicht aus, wonach etwa biometrische Daten nur zur eindeutigen Identifizierung einer natürlichen Person verwendet werden dürfen und Systeme transparent sein müssen.

## Monopole auf Daten und Inhalte

Diese systembedingte Bremswirkung wird der Expertin zufolge durch Verwendung stochastischer Gradientenabstiegsverfahren nun zunehmend aufgehoben. Stochastic Gradient Descent (SGD) verringere die Menge an Testdaten durch eine zufällige Auswahl. Sie können so einfacher weiterverarbeitet werden. "Deep Fakes" etwa lebten von diesem Prozess. Die Folge sei der rasant wachsende Datenhunger der Entwickler von KI-Modellen. Jeder drehe nur noch an den virtuellen Stellschrauben für Algorithmen und schaue darauf, wie sich durch das Sammeln von immer mehr Trainingsdaten Genauigkeitsverluste reduzieren und höhere Erkennungsraten erzielen ließen.

Der große Nachteil dieses Prozesses sei, dass damit Monopole auf Daten und Inhalte gefördert würden, monierte Searcy. Die den Markt beherrschenden Techkonzerne wie Amazon, Facebook, Google oder Microsoft könnten damit auch den Arbeitsmarkt kontrollieren und KI-Entwickler an sich binden. Damit hätten sie auch eine gewisse Machtposition selbst gegenüber der Politik und den Regulierern, die schwer zu brechen sei. Generell würden KI-Modelle eingesetzt, um menschliche Arbeit zu ersetzen oder sonst allenfalls neue Tätigkeitsformen zu ermöglichen. Diese Mechanismen seien "schlecht für jeden", die Allgemeinheit profitiere nicht davon. Searcys Resümee: "Künstliche Intelligenz ist kaputt."

## Strengere Regulierung der datengetriebenen Branche – DSGVO nicht ausreichend

Vorgaben für Ethik, Transparenz oder Open Source für KI reichen Searcy zufolge nicht aus, um die Datenmonopole in die Schranken zu weisen. Der Anspruch der Google-Mutter Alphabet etwa, die Technik nur einzusetzen, wenn die Vorteile für die Allgemeinheit die Schäden für Einzelpersonen überwiegen, sei die verhasste Lachnummer in der Branche, meinte die Insiderin: "Die Firmen werden sich nie von sich aus dafür entscheiden, das Richtige zu tun." Auf Open Source setzten sie auch nur, "wenn es für sie nützlich ist" und darüber Dritte leichter eingespannt werden könnten. Etwa bei der Gesichtserkennung oder Regierungsaufträgen gebe es aber keinerlei Anreiz dafür, Modelle frei zu veröffentlichen.

## Trainingsdaten ohne Einwilligung der Betroffenen

Der Künstler Adam Harvey berichtete auf der Tagung anhand von Erkenntnissen aus seinem Projekt "Megapixels", dass Firmen genauso wie Forscher ihre Trainingsdaten-Sets vielfach "wild", also ohne Einwilligung der Betroffenen aufbauten. Das sei beispielsweise bei Daten für die Gesichtserkennung der Fall. Da derlei Systeme nach wie vor hohe Fehlerraten aufwiesen, sei in diesem Bereich die Jagd auf die Daten voll im Gange. Microsoft etwa habe sich für die Datenbank Celeb einfach bei Bildern von Promis sowie weniger bekannten Personen im Web bedient, die Duke University habe für das "Multi-Tracking-Register" DukeMTMC aus einem Fenster des Instituts mit Teleobjektiv Aufnahmen von Studenten gemacht.

Diese Datenbanken seien zwar inzwischen größtenteils offiziell abgeschaltet worden, führte Harvey aus: "Aber man kann sie nicht wirklich aus dem Netz bekommen." So kursierten die Inhalte nach wie vor in "akademischen Torrents" in Peer-to-Peer-Netzwerken "rund um die Welt". Der Aktivist hält es für erwiesen, dass Teile davon etwa von der chinesischen Armee übernommen und nun auch für die Unterdrückung der muslimischen Minderheit in der autonomen Region Xinjiang verwendet werden und fordert, dass die beteiligten Firmen und Hochschulen "dafür haften". Für KI-Versuche sollten die Betreiber zudem allenfalls die biometrischen Daten ihrer eigenen Belegschaft nutzen.

## "Social Scoring" von Kindern und Jugendlichen

Nicht gut an kamen beim Publikum die Schilderungen des Datenwissenschaftlers Slava Jankin von der Hertie School of Governance über Analyseprojekte in englischen Gemeinden. Dort seien mit Methoden des Social Scoring "Risikoprofile" von Kindern und Jugendlichen erstellt worden, um möglichst frühzeitig zu erkennen, ob diese Missbrauch zum Opfer fallen oder nicht altersgerecht schulreif werden könnten. Ausgewertet habe man dafür etwa auch Akten von Sozialarbeitern. Da aufgrund von Datenschutzbeschränkungen Verknüpfungen mit Registern von Polizei oder Klinken nicht hätten hergestellt werden dürfen, liege die Vorhersagegenauigkeit bei dem System bei 74 Prozent.

Informatiker wüssten offenbar oft gar nicht, was sie tun, kritisierte ein Gastforscher dieses Projekt. Die Folgen für die Betroffenen vor Ort könnten überaus gefährlich sein. "Die Eltern haben das gefordert", verteidigte Jankin den Ansatz. Das technische System solle Sozialarbeiter nicht ersetzen, sondern ihnen die Arbeit erleichtern. Der Kreis spare durch das Projekt Geld. Man habe zudem einen Ethikrat mit Abgesandten von den Eltern und anderen Gemeindevertretern eingerichtet. (bme)

 [Kommentare lesen \(376\)](#)





Mittler und Mutter sind Pilsch bei der Premiere

### Die Familie gewinnt immer

Was heute das eine, morgen cabarettgeschrien werden sollte, entsteht gewöhnlich erst dann, wenn es wirklich ernst wird. Im Fall der Handlung vom Überleben der Familie, die von...

Das Fest – Studientheater des 3. Studienjahres (Jugendtheater Puppentheater der Hochschule für Schauspielkunst Ernst Busch, Schlossteichstraße 100, 10178 Berlin, 12. Juni 19, 13.00 Uhr, 12,50 €)

Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)

Inszeniert von Christiane G. ... 2019 69 GRIPS ... MORMOLIS | MUGNOG-KINDER! ... 8.6. | 18 Uhr ... Akademie der Künste, Hansaroseweg

Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)

13. Jahre Deutscher ... Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)

NELE STUHLER KEINE AHNUNG ... PERFORMANCE ...

Motiv ... Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)

13. Jahre Deutscher ... Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)

knotipp ... club 7 ...

14-16 JUNI ... AI TRAPS ... AUTOMATING DISCRIMINATION ... THE POLITICS OF AI: Fighting Injustice & Automatic Supremacy ... WHAT IS FEMINIST AI? ... RACIAL DISCRIMINATION IN THE AGE OF AI ... THE 16TH CONFERENCE OF THE DISRUPTION NETWORK LEAD ... DISRUPTIONLAB.ORG/AI-TRAPS

Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)

13. Jahre Deutscher ... Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)
Mutter (19-20:00 Uhr)
Hinter der Bühne (19-20:00 Uhr)



Walter Drogus

### Der Walter aus dem Lützowviertel

Der Lützowviertler wohnt am Ende der Straße, wo es immer noch einen schönen Blick gibt auf den Lützowpark und den Lützowviertler Wald. Walter Drogus hat sich hier ein Nest gebaut. Er ist ein Mann mit viel Lebenserfahrung und viel Können.

Walter Drogus ist in der Lützowviertler Straße 102, 13. 6., 10419 Berlin zu Hause. Er ist ein Mann mit viel Lebenserfahrung und viel Können.

#### Maltege

- Biergarten - 13.00 Uhr
13.00 Uhr Biergarten
13.00 Uhr Biergarten
13.00 Uhr Biergarten

#### Kommuni

- Mitgliedschaft - 13.00 Uhr
13.00 Uhr Mitgliedschaft
13.00 Uhr Mitgliedschaft
13.00 Uhr Mitgliedschaft

AI TRAPS: AUTOMATING DISCRIMINATION. THE POLITICS OF AI: fighting Injustice & Algorithmic Supremacy. WHAT IS FEMINIST AI? RACIAL DISCRIMINATION IN THE AGE OF AI. THE 16TH CONFERENCE OF DISRUPTION NETWORK LAB.

KUNSTLEBEN IM BÜRO: 11.04.2019 bis 19.06.2019

#### Stell

- Stellungsangebot - 13.00 Uhr
13.00 Uhr Stellungsangebot
13.00 Uhr Stellungsangebot
13.00 Uhr Stellungsangebot

#### Ver

- Veranstaltung - 13.00 Uhr
13.00 Uhr Veranstaltung
13.00 Uhr Veranstaltung
13.00 Uhr Veranstaltung

#### Interdisziplinäre

- Interdisziplinäre - 13.00 Uhr
13.00 Uhr Interdisziplinäre
13.00 Uhr Interdisziplinäre
13.00 Uhr Interdisziplinäre

PUGS IN LOVE QUEER WEEK 12 - 15 JUNI STUDIO R

Interdisziplinäre - 13.00 Uhr
13.00 Uhr Interdisziplinäre
13.00 Uhr Interdisziplinäre
13.00 Uhr Interdisziplinäre



Die vierfeldige Ausstellung war ein Erfolg. Sie hat viel Aufmerksamkeit erregt und ist ein wichtiger Beitrag zur Queer-Kunstszene.

Kontakt: 13.00 Uhr
13.00 Uhr Kontakt
13.00 Uhr Kontakt
13.00 Uhr Kontakt

MIRIAM SOBER SPEAKING VOLUMES



Künstliche Intelligenz

## Wie viel Rassismus steckt in Algorithmen?

Mutale Nkonde ist Forscherin und kämpft gegen Rassismus im Zeitalter künstlicher Intelligenz. Ihre bisherige Bilanz für die Tech-Branche ist vernichtend.

Ein Interview von [Sonja Peteranderl](#)



Mutale Nkonde erforscht, wie Technologie Rassismus reflektiert - und was sich dagegen tun lässt

#womenofcolortechchat

Samstag, 15.06.2019 13:47 Uhr

Drucken Nutzungsrechte Feedback Kommentieren

Die Forscherin Mutale Nkonde, derzeit Fellow am New Yorker Forschungsinstitut Data & Society, kämpft mit ihrem Projekt "Racial Literacy in Tech" zusammen mit anderen Wissenschaftlern gegen [Rassismus](#), der durch [Algorithmen](#) und andere Technologien verstärkt wird.

"Gesichtserkennungsalgorithmen sind darauf trainiert, weiße Gesichter zu erkennen, Flughafen-Scanner verstehen schwarze Frisuren nicht, Kreditwürdigkeits-Scores sind mit rassistisch getrennten Nachbarschaften korreliert und Daten für das Predictive Policing (vorausschauende Polizeiarbeit) werden von rassistischen Vorurteilen geplagt", heißt es in ihrem [Forschungsbericht "Advancing Racial Literacy in Tech"](#). "Ohne eine bewusste Anstrengung, Rassismus in Technologie anzugehen, ist es unvermeidlich, dass neue Technologien alte Trennungen replizieren."

ANZEIGE



**Hauswert-Rechner: Wie viel ist Ihr Haus wert?**

Die Immobilienpreise sind auf Rekordhoch. Jetzt Preis ermitteln und zum Mega-Preis verkaufen!

Auf der [Disruption Network Lab Konferenz](#) "AI Traps - Automating Discrimination" an diesem Wochenende in Berlin wird Nkonde über Rassismus im Zeitalter künstlicher Intelligenz sprechen. SPIEGEL ONLINE hat sie vorab interviewt.

---

## Zur Person

**Mutale Nkonde** forscht als Fellow des Data and Society Research Institute in New York City zu Rassismus, Technologie und Politik. Sie arbeitet seit 2016 als Tech-Beraterin für die demokratische Kongressabgeordnete Yvette Clarke und entwickelt Strategien, um Algorithmen besser zu kontrollieren.

---

**SPIEGEL ONLINE:** Sie untersuchen, wie Rassismus in neuen Technologien zutage tritt. Was erkennen Sie da und was muss sich ändern?

**Mutale Nkonde:** Die Bürgerrechte brauchen ein Update für die Ära der künstlichen Intelligenz. In New York will ein privater Vermieter in Sozialwohnungen zum Beispiel [↗ Schlüssel gegen Gesichtserkennungssoftware austauschen](#). Davon sind arme, schwarze Menschen betroffen, die in Brownsville leben, einem Viertel mit historisch hohen Kriminalitäts- und Gewalttaten. Sie müssen die Überwachung akzeptieren - oder sie verlieren ihre Wohnung. Das ist nicht nur eine Verletzung von New Yorker Mietgesetzen, es werden auch ohne jegliche Kontrolle biometrische Daten gesammelt.

**SPIEGEL ONLINE:** Gesichtserkennung verbreitet sich nicht nur in schwarzen Communities.

**Nkonde:** In diesem Fall sollen die Gesichtserkennungssysteme für Wohnungen nur in schwarzen Nachbarschaften eingesetzt werden, nicht auf der Upper East Side, wo Menschen wie [Donald Trump](#) wohnen. Von den Negativ-Effekten von Technologie sind oft Communitys betroffen, die schon immer Ziel von Überwachung waren und unter Generalverdacht standen. Das reicht in den USA bis in die Sklavenszeit zurück. Algorithmische Entscheidungsfindungen bei Jobbewerbungen oder in der Strafjustiz oder auch Gesichtserkennung sind die neuen Varianten dieses Systems.

**SPIEGEL ONLINE:** Wie viel Rassismus steckt in Algorithmen oder den ihnen zugrunde liegenden Daten?

**Nkonde:** "ProPublica" hat etwa nachgewiesen, dass Afro-Amerikaner höhere Risikobewertungen erhalten als weiße Kriminelle. Es wird also eher davon ausgegangen, dass sie erneut straffällig werden, was sich wiederum auf die Strafe auswirkt. In der New Yorker Datenbank zu kriminellen Gangs sind vor allem junge schwarze Männer oder Latinos erfasst. Es reicht schon aus, auf demselben Foto wie ein Gangmitglied aufzutauchen, sich im selben Gebäude aufzuhalten oder eine bestimmte Farbe zu tragen.

**SPIEGEL ONLINE:** Weil die Farben mit Gangs assoziiert werden?

**Nkonde:** Menschen haben in anderen Vierteln das Recht, Rot, Blau, Grün zu tragen, ganz wie sie wollen. Aber wer in einer armen Nachbarschaft wohnt und eine bestimmte Farbe trägt, wird einer Gang zugerechnet, selbst wenn er nicht kriminell ist. Und wer in der Gangdatenbank erfasst wurde, erfährt es nicht einmal. Du hast keine Chance, deinen Namen wieder aus der Datenbank entfernen zu lassen und wirst schlimmstenfalls einfach verhaftet, ohne zu wissen, warum.

**SPIEGEL ONLINE:** Die Daten werden dann in anderen Systemen wie Predictive Policing weiterverwertet.

**Nkonde:** Die Betroffenen wissen nicht, wer alles Daten aufnimmt, was mit den Daten passiert und ob sie mit Polizei- oder Immigrationsbehörden geteilt werden. Die New Yorker Polizei hat etwa Mugshots, Fotos von Verdächtigen, an IBM verkauft, damit die ihre Datenbanken erweitern konnten.

**SPIEGEL ONLINE:** Beobachten Sie, dass sich die betroffenen Communitys inzwischen stärker organisieren?

**Nkonde:** Was sich geändert hat: Überwachungstechnologie wird heute zunehmend auch gegen weiße Menschen eingesetzt. Sie wurden durch Skandale wie Cambridge Analytica aufgerüttelt und wissen jetzt, dass sich Technologie auch gegen sie richten kann und dass ihre Daten missbraucht werden können. Dadurch wird nun auch die Stimme von Aktivisten gehört, die sich seit Jahren für diese Themen stark gemacht haben - meistens weiße, liberale Personen. Der schwarzen Community hört dagegen niemand zu.

**SPIEGEL ONLINE:** Was muss geschehen, damit die Perspektiven schwarzer Menschen sichtbar werden?

**Nkonde:** Wir müssen "Racial Literacy" erreichen. Es muss ein Bewusstsein für Rassismus entstehen - und weiße Menschen müssen damit klarkommen, dass es unbequem ist, über Rasse zu sprechen. Ich verstehe Literacy als Fähigkeit, die sich weiterentwickelt und die man üben muss, wenn man nicht gut darin ist.

**SPIEGEL ONLINE:** Welche Note würden Sie an Tech-Firmen in punkto "Racial Literacy" vergeben?

**Nkonde:** Für unseren Bericht haben wir acht Monate lang geforscht und dabei auch rund 20 Angestellte interviewt, die in allen möglichen großen Tech-Firmen aus dem Silicon Valley arbeiten. Manche hatten das Gefühl, dass Rassismus nichts mit ihrer Arbeit zu tun habe, weil sie nur Entwickler seien. Sie argumentierten etwa, sie seien farbenblind. Das Problem: Wer das behauptet, hat auch nicht den Anspruch, etwas zu verbessern.

**SPIEGEL ONLINE:** Gab es auch positive Tendenzen?

**Nkonde:** Manche wollen anders über das Problem reden, weil sie glauben, dass der Begriff "Diversity" nicht spezifisch genug ist. Sie wollen die Kultur verändern, so dass über "Rasse" gesprochen werden kann und Menschen tatsächlich als "weiß" oder "schwarz" gelabelt werden. Es herrscht große Angst, Klartext zu reden.

**SPIEGEL ONLINE:** Wollen sich die Firmen auch von innen verändern - indem sie etwa schwarzen Nachwuchs stärker fördern?

**Nkonde:** Was das Recruiting betrifft, bin ich sehr enttäuscht. Ich bin Googles Machine Learning Team durchgegangen und habe nur einen schwarzen Mann und eine schwarze Frau gefunden - unter 893 Mitarbeitern. Es sind kaum schwarze Menschen an der Entwicklung und Erforschung künstlicher Intelligenz beteiligt, die Quoten schwarzer Mitarbeiter bei Tech-Unternehmen sind gering.

**SPIEGEL ONLINE:** Was würden Sie Tech-Konzernen raten?

**Nkonde:** Die Unternehmen sollten über Racial Literacy nachdenken, aber sie brauchen auch eine greifbare Politik. Wir entwickeln gerade ein Assessment-Tool, das messen soll, ob es eine konkrete Strategie gibt. Gibt es einen Plan, um bestimmte Communitys zu erreichen und Rassismus-Probleme anzugehen? Wie sehr steht eine Firma dahinter? Ist der Plan finanziert, und wie lange? Es ist ein Langzeitprojekt.

*Berlin in English since 2002*

# EXBERLINER

EXBERLINER.com



WHAT'S ON

FEATURES

IN PRINT

## It's the structures, not the tech: Os Keyes

BY FREYA WERNER    JUNE 13, 2019

[RSS](#)    [PRINT](#)



© Dorothy Edwards/Crosscut

Disruption Network Lab returns to Kunstquartier Bethanien June 14-15 with another critical reflection on the use of technology – **AI Traps**. This time DNL looks at artificial intelligence and asks: How does AI reinforce prejudice and influence political discourse/action? Ph.D. student Os Keyes focuses their studies on data, gender and infrastructures of control – with a special focus on facial recognition – and joins the panel *The Politics of AI* (June 15, 6:30pm) to explore how the bias in AI can be countered.

### **How did you develop an interest in AI?**

Before starting my Ph.D. programme I actually worked as a data scientist for three years at a couple of different companies, one that runs Wikipedia and the other in information security. Hanging around in those spaces, working with those people and just dealing with data on such a massive scale in ways that were extremely sensitive made me aware of the ethical implications of all of this. At the information security company, I started building machine-learning models for a living... and becoming aware of the fact that I am building the system to run when I am not paying attention – like it will be going off making decisions based on random rules that I programmed into it and data I gave to it.

**You once said it was a university paper titled “White, Man, and Highly Followed: Gender and Race Inequalities in Twitter” that initially got you interested in facial recognition... because it made you so mad. Why?**

This paper used facial recognition on Twitter profile pictures to analyze race and gender on social media. Anyone whose picture they could not analyze was discarded from the data set.

And these people went off and built a big data-set on how people of different races and genders interact. My problems were of a purely scientific nature: in the study there were only three races, two genders and no awareness of the fact that information you have about a user can be more than just a profile picture. For example, the text in their profile itself could lead to a completely different gendered conclusion.

But my greatest annoyance was actually that no one else noticed this flaw in the study and that it was fundamentally incapable of handling the existence of trans people. So that sort of kicked off my interest and I wrote a big paper that analyzed all the literature on gender recognition. And I inevitably found that facial recognition was trans-exclusionary.

**Do you have any other examples where a transgender person directly experienced gender bias because facial recognition?**

So one example might be illustrative, as small as it may seem. There are these gender recognition apps. The idea is that you can do like gender-flipped photos, which everyone thinks is hilarious, and gender detection and all the rest. A while ago I met a group of trans people who had run into these systems and had been playing around with them. How unhappy some of them were, that the system had told them that they were the wrong gender was absolutely heartbreaking and entirely understandable, right? Because its literal physical infrastructures saying, ‘No, you are not real, you don’t count’ and claiming that it is scientific and drawing on scientific authority. And as small as that might seem if you think about the ways in which it is going to be built into other forms of existing systems where gendering takes place you can see the potential for their heartbreak to become a lot more ubiquitous... just imagine human-robot interactions where the robot determines whether it should address you as miss or sir.



### **So where is facial recognition mostly being used and what are the tangible risks?**

The answer is we don't know where and that's incredibly worrying. It should worry us that it's happening and we don't know where it's happening, but it should also worry us that just because it's being done by random private companies and not the state, does not mean that it is not incredibly risky. First because of how much of the world is dominated by private companies. Second because data from private companies doesn't stay in private companies hands for long. Any time a national security crisis might come up or a policing issue, companies tend to defer to the state. And when they do that all the data they were gathering, which you did not know they were gathering, suddenly ends up in the hands of the government. It is everywhere, it is increasingly common and I am very worried by the fact that we don't recognize how ubiquitous it is and that who has access to that data can change at the whim of the state.

---

### **So is AI going in the wrong direction?**

I can see it going in one of two directions, if we actually make structural changes we need, AI is a useful tool. It is a thing that can be used with consent, with permission and with contextual awareness in a way that frees human beings up to concentrate on the work and activities that can't be automated.

In a world where we don't make the structural changes we need, it looks much the same in theory, but there are two major differences. The first is we automate more stuff than AI is actually good for. Second, we become more comfortable and numb to the ubiquitous data collection, the ubiquitous surveillance, the idea that systems should tell us how best to run our lives. That is a society which is very heavily controlled which does not have room for multiple ways of doing things.

### **So how do we ensure that we don't go that second route and how do we keep AI beneficial?**

The problem is not AI. It is the symptom. The problem is the structures that are creating AI. We live in an environment where there are a few highly wealthy people and companies who have all of the power. Our problem is that we have state infrastructures which are designed to preserve themselves first and for most. The reason why we are seeing these issues in AI is that its development is driven by exactly the same people.

So to fix AI, one of the things we can do is to change requirements and expectations around it, like building in the expectation that it will make the world actively better. We can insist on consent to having AI deployed, based on explainability and transparency. But ultimately if all we do is fix AI, then in 10 years we will have exactly the same debate over another technology.

**Disruption Network Lab: AI Traps, Jun 14-15 | Kunstquartier Bethanien, Kreuzberg. See [website](#) for complete programme.**

disruption network lab | artificial intelligence | Interviews