



The Myth of Neutrality

How AI is widening social divides



About me

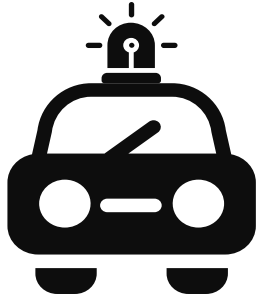
Stefanie Stoppel

Machine Learning Engineer 🛠️

I enjoy
dancing 🕺,
drawing 🖋️,
& laughing 😄



@Qbrid_



Imagine...

... you're driving
down the highway
when you're
suddenly pulled over
by the police



They check your driver's license in their system ...

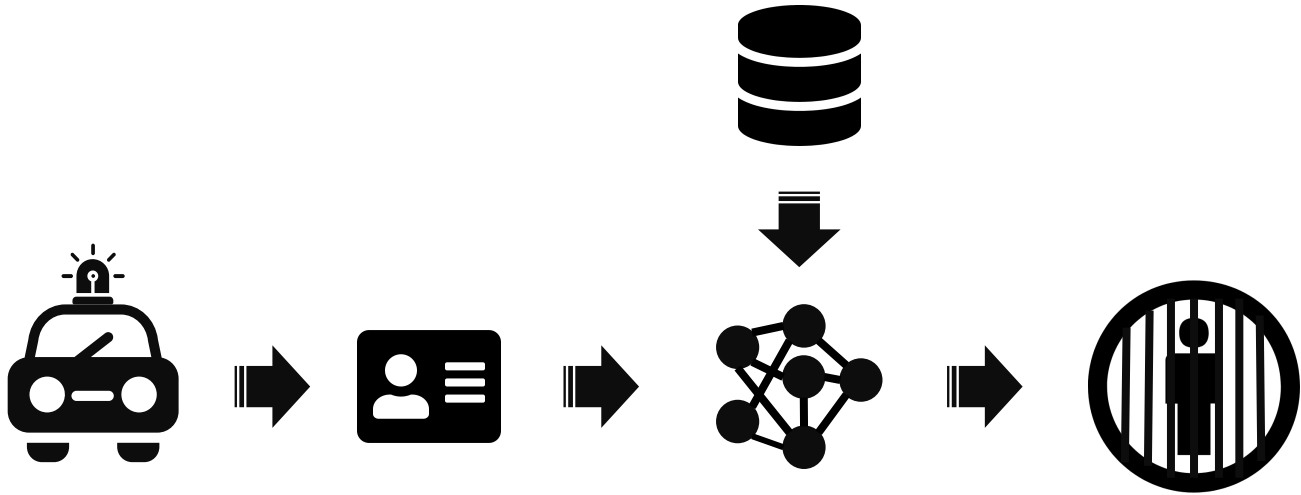
... and find your
picture matches a
person wanted for
armed robbery





What's going on?





The AI landscape

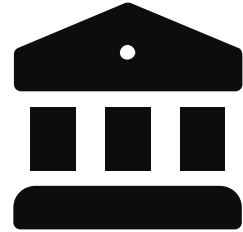




big tech



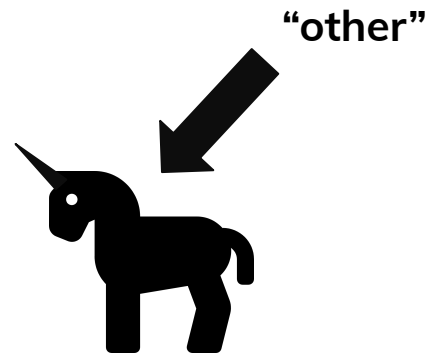
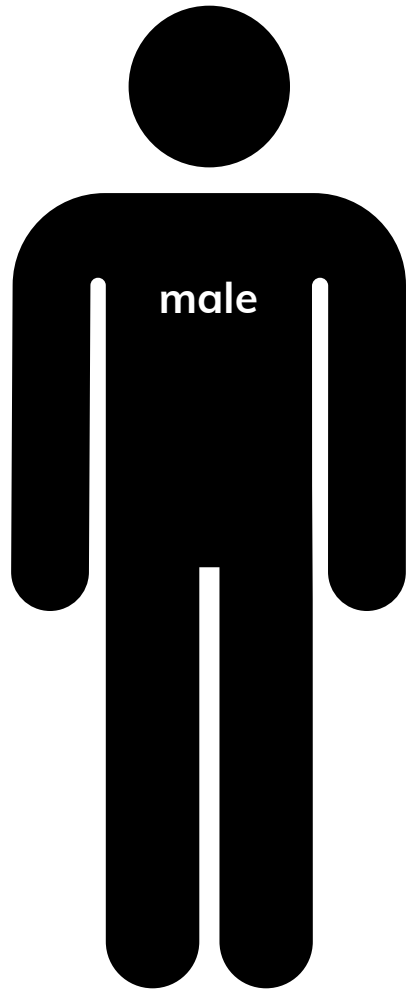
military

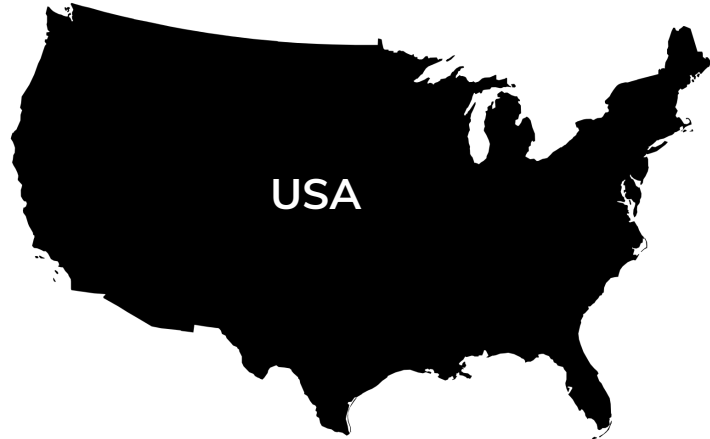


government



community





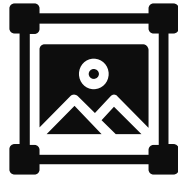
"other" 

Left image: ["Map showing Mainland USA without state borders"](#) by Theshibboleth, [CC BY-SA 2.5](#),
Right image: [Map of China](#) by Alanmak, CC0 1.0 Universal

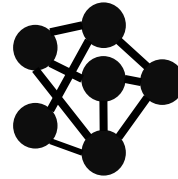
The standard AI development process



Research
& Funding



Data collection
& labeling



Training
& Testing

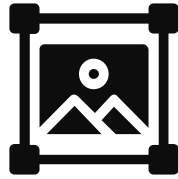


Deployment
& BIG CA\$H

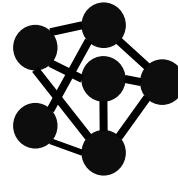
The standard AI development process



Research &
Funding



Data collection
& labeling



Training
& Testing



Deployment



The Birth of AI



Dartmouth Workshop (1956)

- Organized by John McCarthy, Marvin Minsky, Nathaniel Rochester & Claude Shannon
- Term “Artificial Intelligence” was coined
- Researchers believed in fully intelligent machines until the mid 70s



<https://t1p.de/i6pc>



Quote from the workshop proposal

"[...] every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it".¹

¹<https://t1p.de/5r3y>

**Why can't AI do
the washing up
for me then???**





**Who
funds the
research?**





Military & Intelligence Agencies





DARPA

short for **Defense
Advanced Research
Projects Agency**

They were a major source of funding in AI's early days - and still are now.

US contract spending on AI by government agency

TOP 10 CONTRACT SPENDING on AI by U.S. GOVERNMENT DEPARTMENT and AGENCY, 2020

Source: Bloomberg Government, 2020 | Chart: 2021 AI Index Report

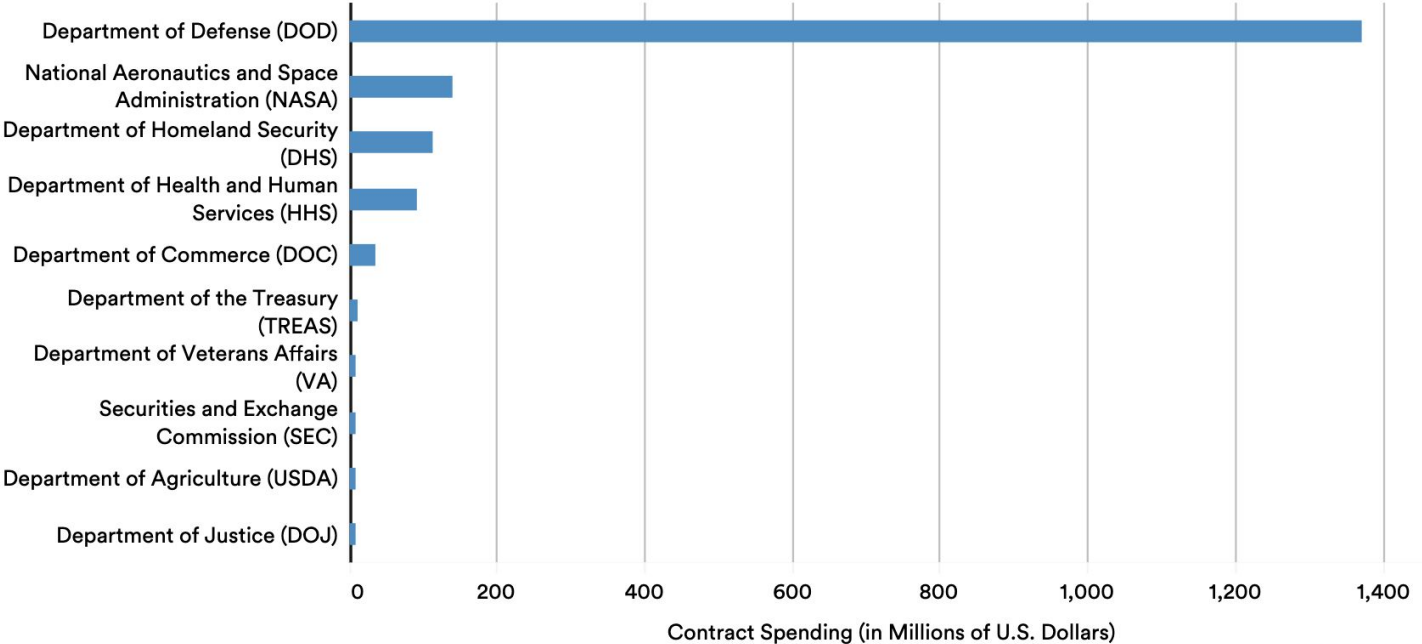


Figure 7.3.4



Clearview AI

The US based company sells access to its biometric identification software to law enforcement agencies.

To create their product, they scraped **>300 billion photographs** from Twitter, Facebook & Instagram.



*"All the information we collect
is collected legally and it is all
publicly available
information
[...]”¹*

Hoan Ton-That,
CEO of Clearview AI

¹<https://www.dw.com/en/clearview-ai-controversy-highlights-rise-of-high-tech-surveillance/a-57890435>

Big Tech



**... is buying
everything**



TIMELINE OF TECH GIANTS' BILLION-DOLLAR ACQUISITIONS

Every \$1B+ acquisition made by Facebook, Amazon, Microsoft, Google and Apple. Bubble size represents maximum valuation.

● Facebook ● Apple ● Microsoft ● Google ● Amazon

VALUATION OF ACQUIRED COMPANY

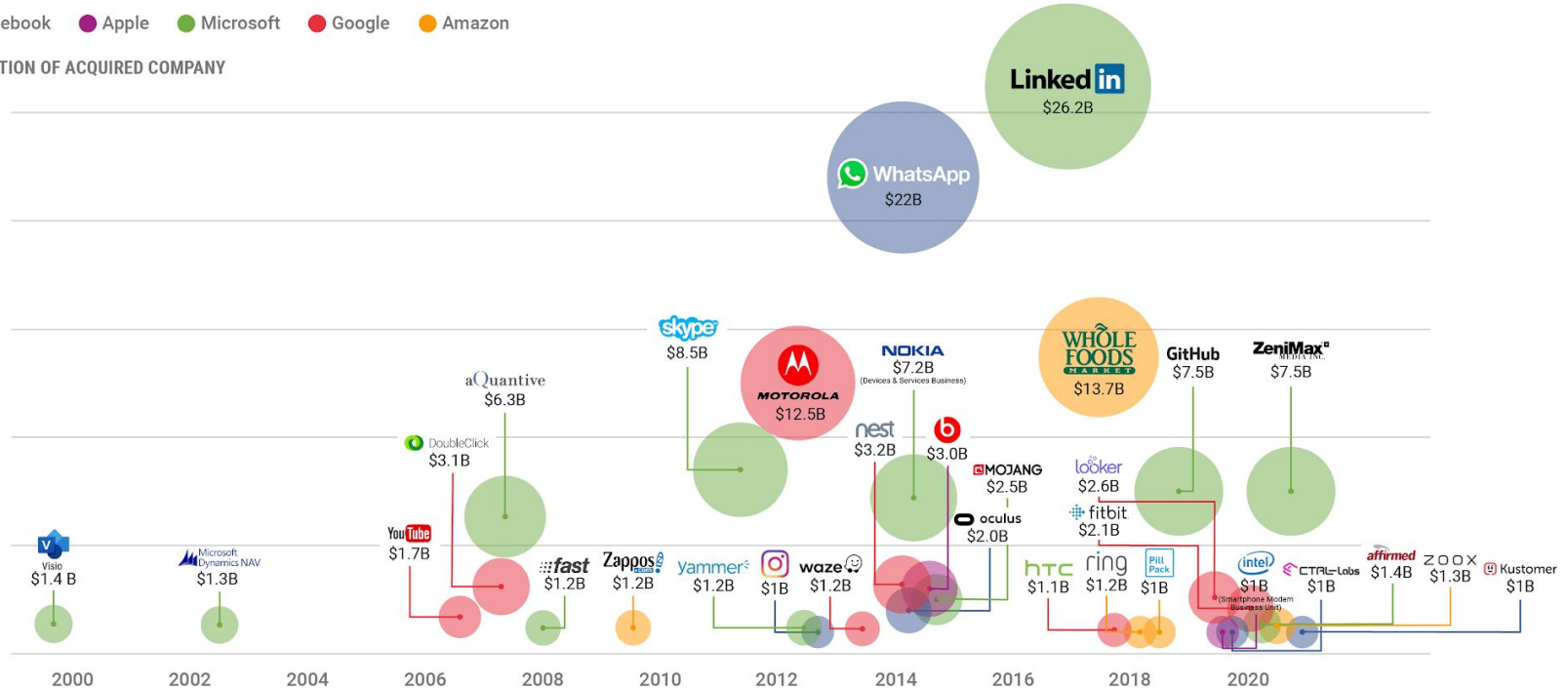
\$25 B

\$20 B

\$15 B

\$10 B

\$5 B



ACQUISITION DATE

Data as of February 2021



CB Insights: <https://t1p.de/4m36>



**... is leading to an
AI “brain drain”
at US universities**

NUMBER of AI FACULTY DEPARTURES in NORTH AMERICA, 2004-19

Source: Gofman and Jin, 2020 | Chart: 2021 AI Index Report

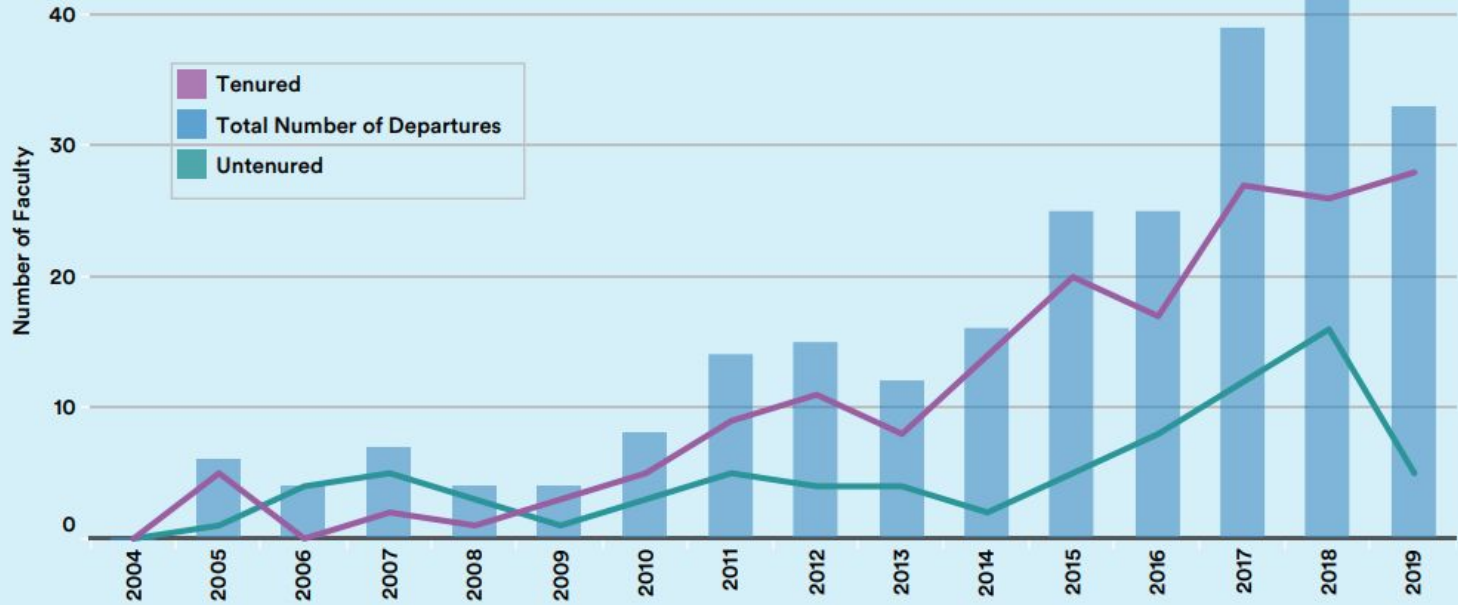



Figure 4.4.1



**However,
research within
some companies
does not seem
entirely
independent...**





*“Take great care
to strike a
positive tone”*

A Senior Manager @ Google
while reviewing a paper on
recommendation algorithms
before publication

Reuters: <https://t1p.de/cwt1>



Timnit Gebru



Margaret Mitchell

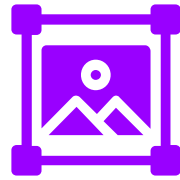
Read more: James Vincent in The Verge,
[“Google is poisoning its reputation with AI researchers”](#)

Image sources: <https://t1p.de/fv2e> (left), <https://t1p.de/98ox> (right)

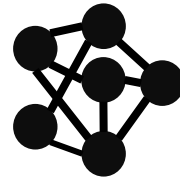
The standard AI development process



Research
& Funding



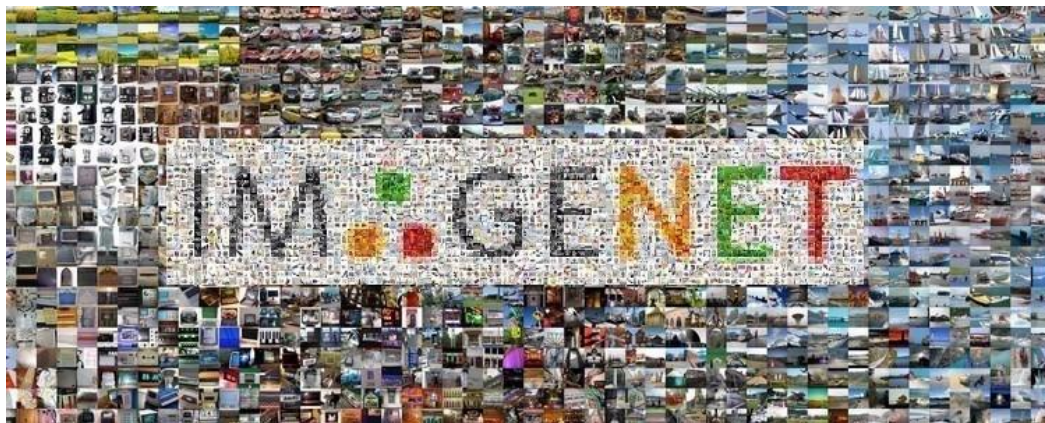
Data collection
& labeling



Training
& Testing



Deployment



- > 14 million images in more than 20.000 categories
- goal: “map out the entire world of objects”¹
- the image classification dataset

Quote from: <https://t1p.de/geca>,
image source: <https://t1p.de/jm34>

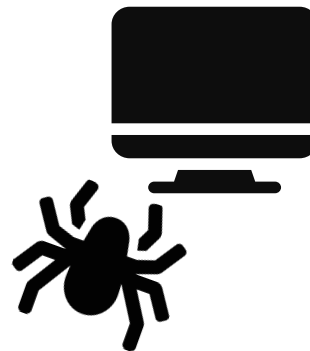


**Where did
the images
come from?**



The internet.

ImageNet was
scraped from
search engines
& photo sharing
websites.





**Did they ask
for consent?**



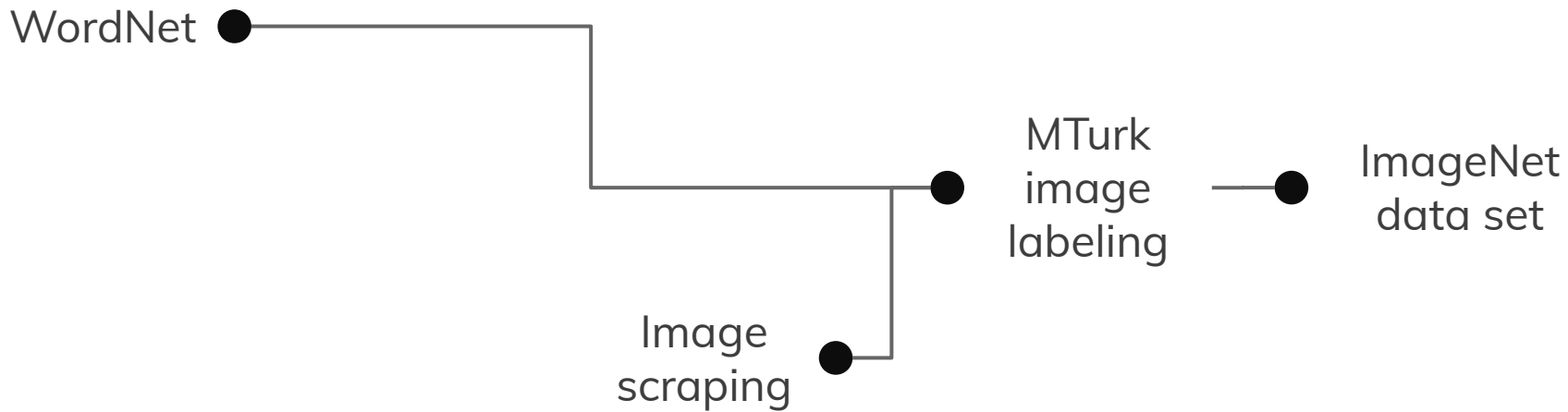


Nope.



**How were
the images
labeled?**





1985

2006

2009



**How would
you choose
a label?**

Melancholic
Pervert
Convict
Bad Person Jezebel Prima Donna
Spinster
Wanton Slut Spastic
Loser Stud
Call Girl Unskilled Person
Fucker
Crazy Codger Kleptomaniac
Tosser
Failure
Wimp
Dead-eye
Drug Addict
Streetwalker
Waverer
Closet Queen
Second-Rater

Also...



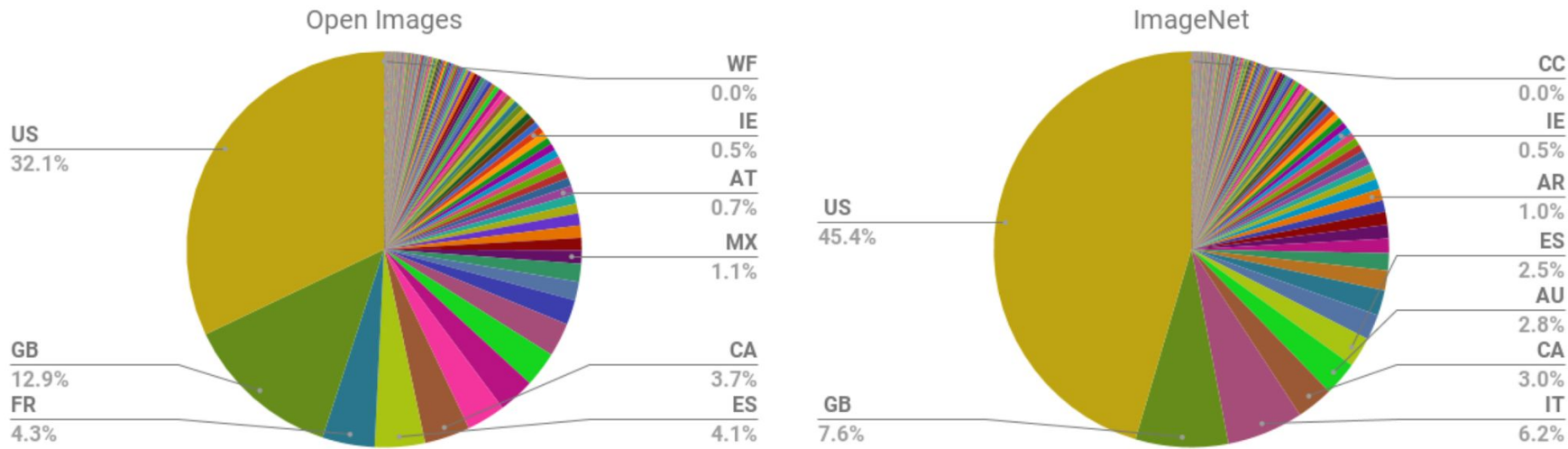
**... weddings
don't look
like this
everywhere**

Images labeled by a neural network



Source: <https://t1p.de/le86>

Geographic skew in image data sets

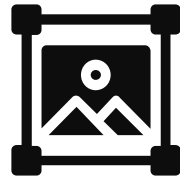


Source: Shankar et al., "No Classification without Representation: Assessing Geodiversity Issues in Open Data Sets for the Developing World", <https://arxiv.org/abs/1711.08536>

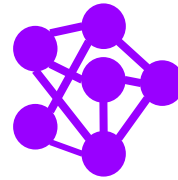
The standard AI development process



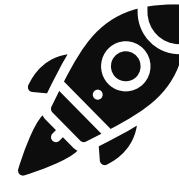
Research
& Funding



Data collection
& labeling



Training
& Testing



Deployment





Gender Shades



**In 2018,
Joy Buolamwini &
Timnit Gebru
investigated biases in
commercial binary
gender classification
systems.**





Joy Buolamwini,
image source: <https://t1p.de/07ik>

Joy's work was inspired by her own experience of not being recognized by open-source face detection software.

She had to wear a white mask for the system to finally "detect" her.



**Gender
classification
accuracies
at first glance...**

93.1%

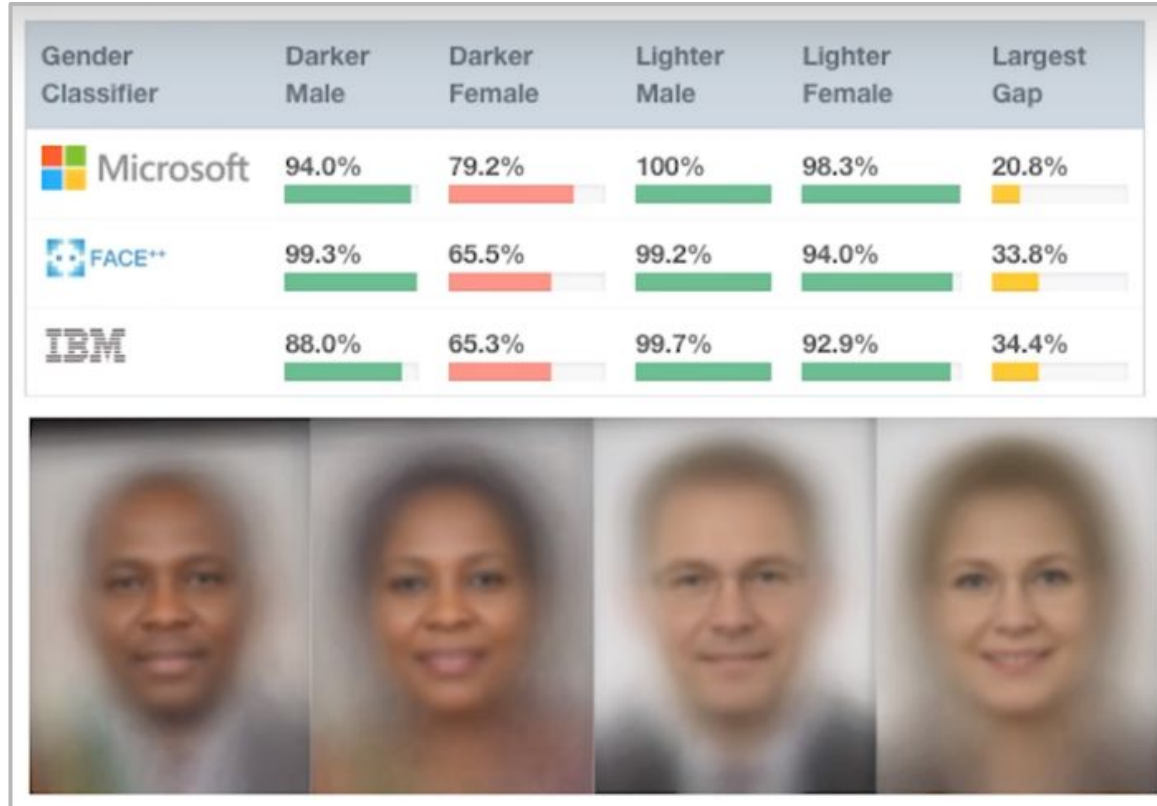
Microsoft

89.5%

Face++

86.5%

IBM



Buolamwini et al., 2018, "Gender shades",
 image source: <https://t1p.de/pe8h>



**A single
success metric
does not tell
the whole
story!**

**Buolamwini's & Gebru's
work motivated many
other researchers to
assess biases &
try to build
fairer systems.**





Diversity in Faces (DiF)



**Data set created by
IBM with the goal of
“advancing the
study of accuracy
and fairness in
facial recognition”**



Quote and image: <https://t1p.de/c85x>,
arXiv paper: Merler et al., “Diversity in Faces”, 2019

Measuring facial features

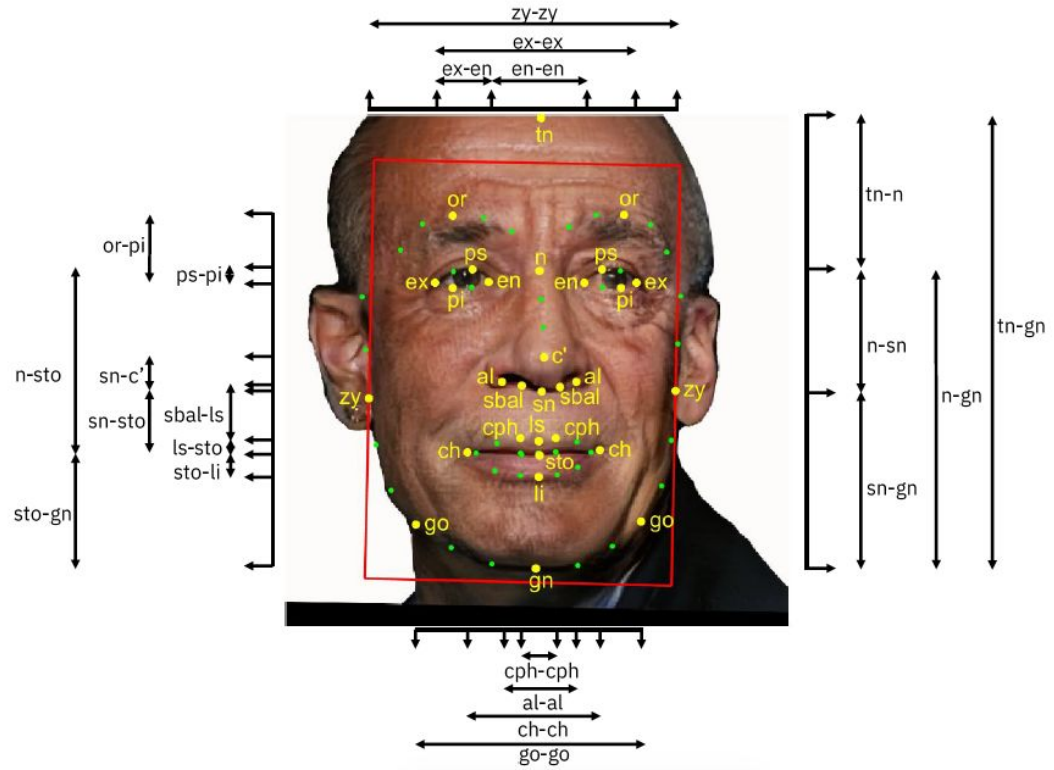


image source: <https://excavating.ai/>

The reasoning

The measurements allow
better assessment of
accuracy and fairness &
more fine-grained
representation of
“facial diversity”



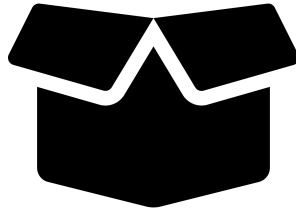
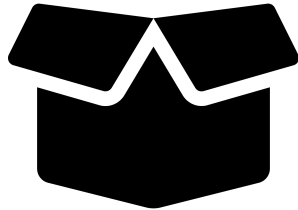
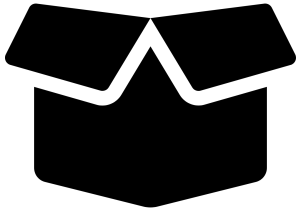
diversity == variety of face shapes




diversity == binary gender*

*assigned by crowd-workers

AI creators decide about the classification system!





*“[...]
the practice of
classification is
centralizing power:
the power to decide
which differences make
a difference.”*

— Kate Crawford,
Atlas of AI

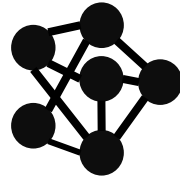
The standard AI development process



Research
& Funding



Data collection
& labeling



Training
& Testing



Deployment





Hiring & Firing



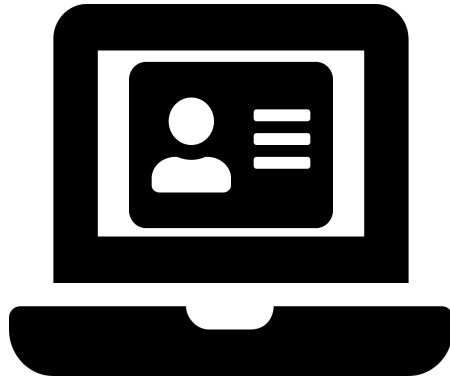


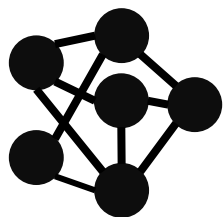
Traditional recruiting

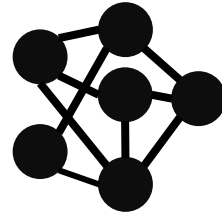


Recruiting today



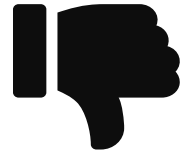
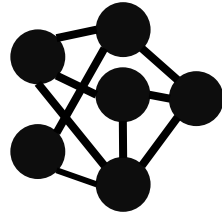






<https://t1p.de/lbyo>





<https://t1p.de/n9ra>



Layers of problems



No questions &
no feedback
after a rejection

1

No proof of possible
discrimination &
no way to challenge
the decision

2

Scalability:
other companies
might use the
same system

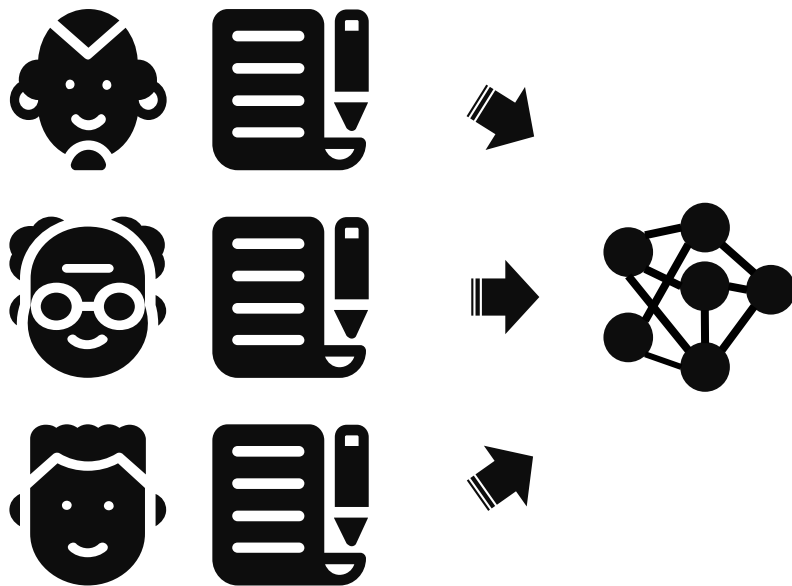
3





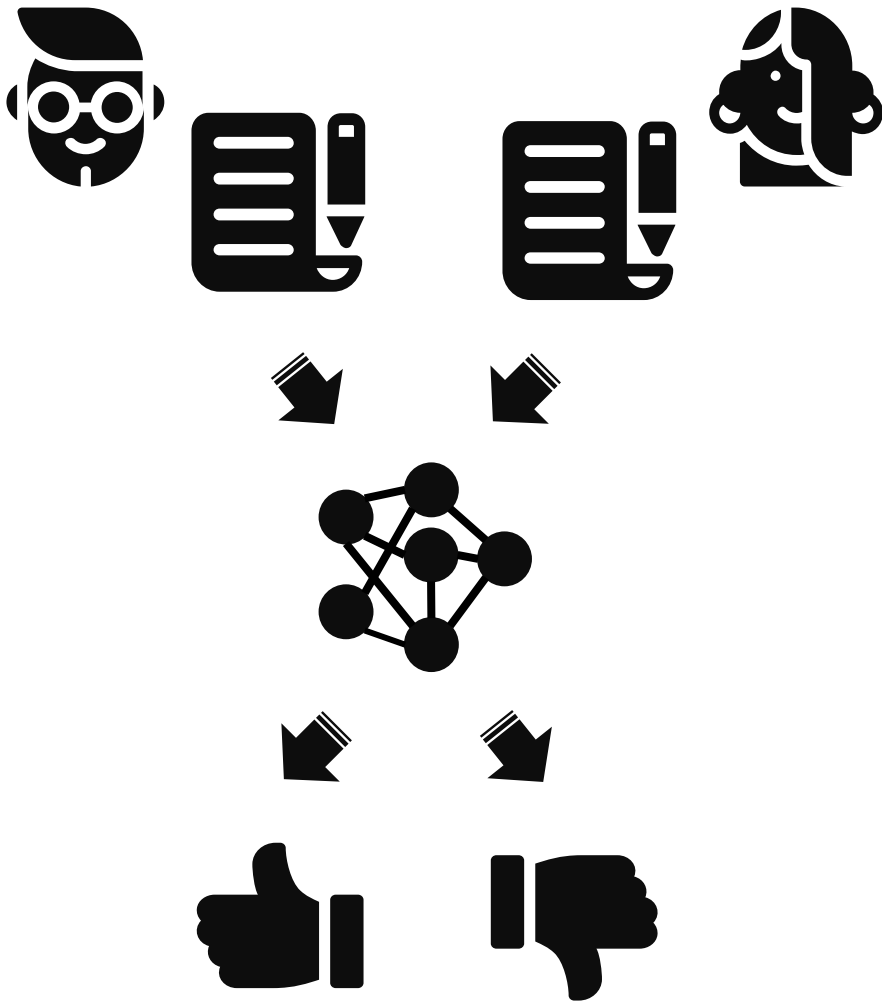
**How do algorithms
determine whether
someone is a
“good fit”?**





Amazon's hiring tool

It was trained on resumes of applicants over a 10-year period.



**They realized it
discriminated
against women...**

... and tried to fix it
by making it “blind”
to certain words
indicating gender.





**But they soon
realized they
couldn't fix it.**

The system kept finding ways to infer a person's gender from other, seemingly unrelated factors.



They trashed the system.

However, they now seem to think automatically firing people based on algorithmic scores is a great idea.*

*Read this article if you're interested:
Spencer Soper in Bloomberg:
["Fired By Bot at Amazon: 'It's you against the machine'"](#)



Fairness

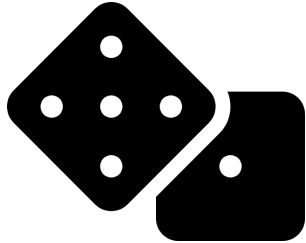




**Imagine
you had to
build a fair
IT-hiring
algorithm...**



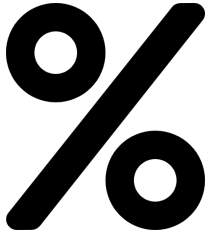
Would it be fair if...



**... the algorithm
just randomly
selects from the
pool of
applicants?**



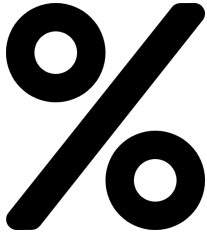
Or would it be fair if...




**... the algorithm
approves
the same
percentage
of women
& men?**



Or would it be fair if...



**... the algorithm
approves the
same percentage
of women & men,
given that they
are qualified?**



**All these notions
of “fairness” can
be described
mathematically.**

But...

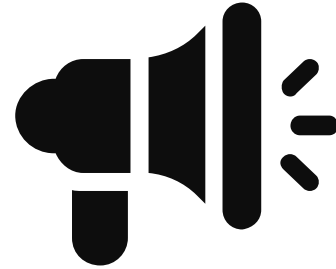


**We first need to
have a discussion
about what
fairness means
in each context!**

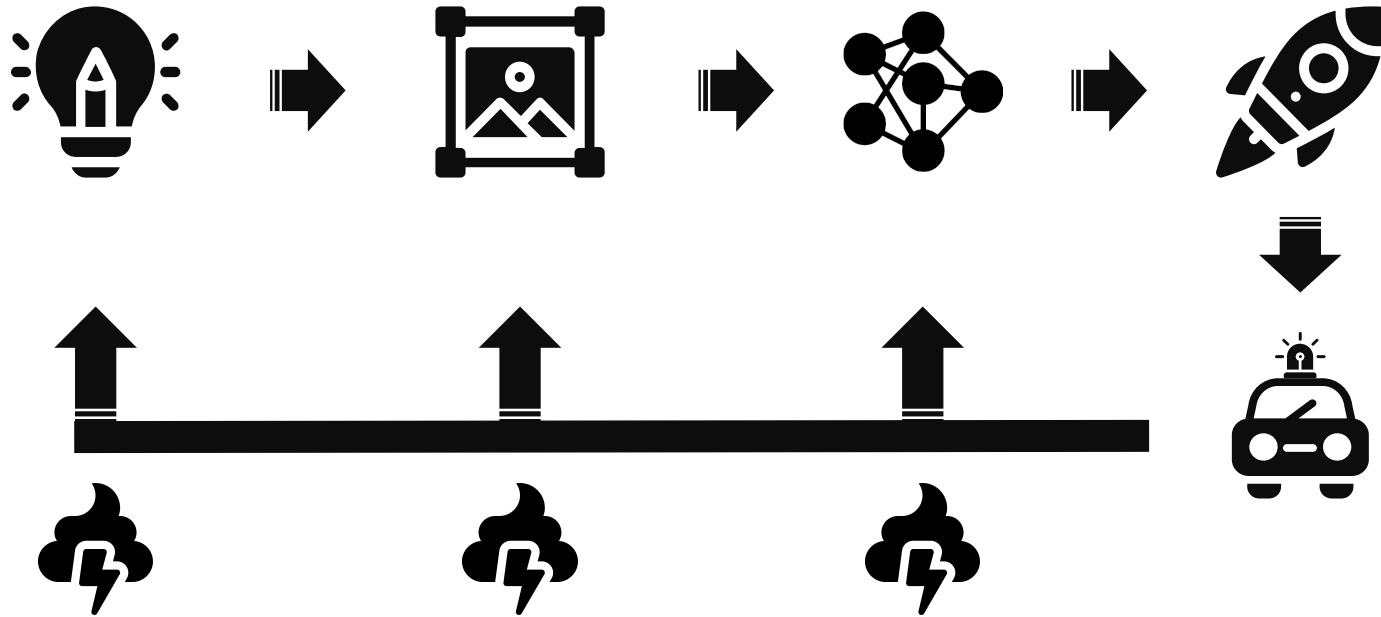
**Fairness
is
a
political
decision**



**Let's not outsource
these political decisions
to the select few
developing AI systems!**



The AI feedback loop





Building Bridges

Advice for everyone



Stay
informed



Join &
organize
collectives



Vote &
donate

Advice for folks in Machine Learning



Be critical

Whether it's about your own or someone else's work and results.



Moral first, math second

Discussions about technological harms & consequences > formalization.



Involve other humans

Whether it's affected communities or social scientists: listen & learn!



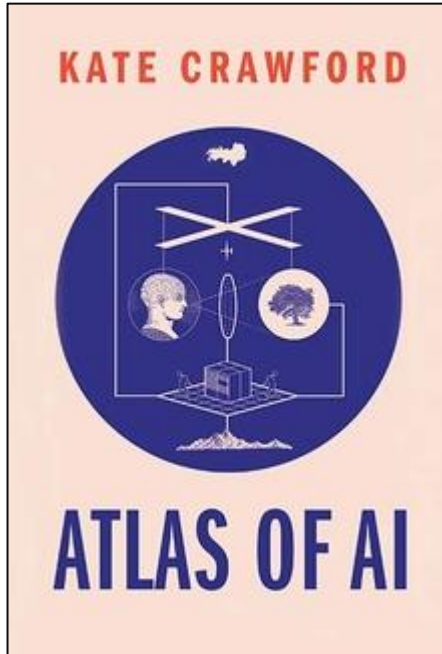
Thank you



@Qbrid_ 

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

References – Books

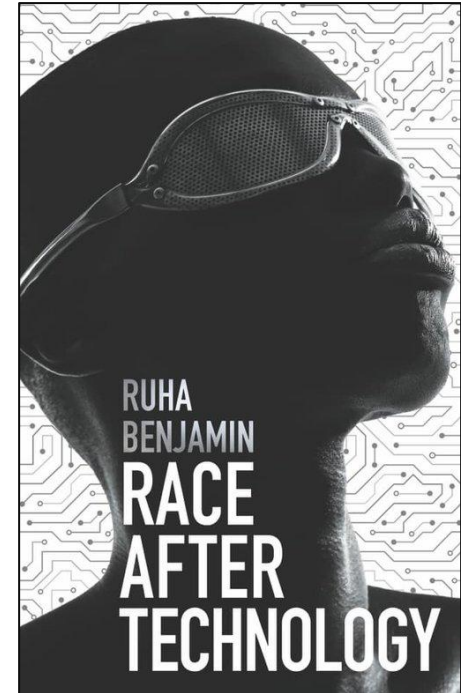


Kate Crawford,
“Atlas of AI”

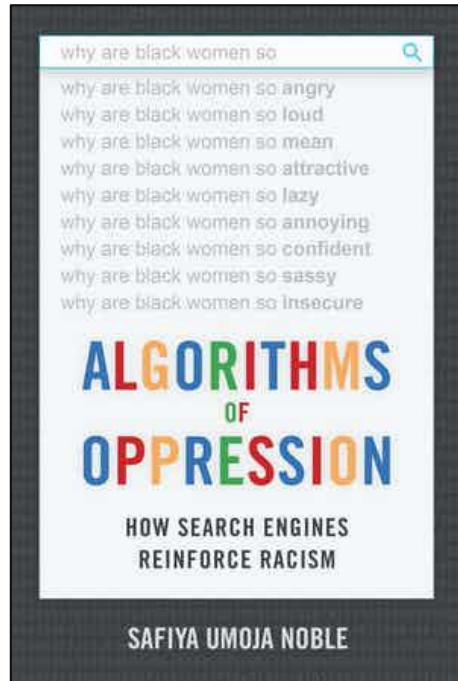
An excellent book for the ones interested in how AI works using both “extraction” and “abstraction” as guiding principles.

Ruha Benjamin,
“Race After Technology”

Ruha Benjamin argues convincingly how AI is shaping us as much as we shape it & how race itself can be seen as a technology of oppression.



References – Books

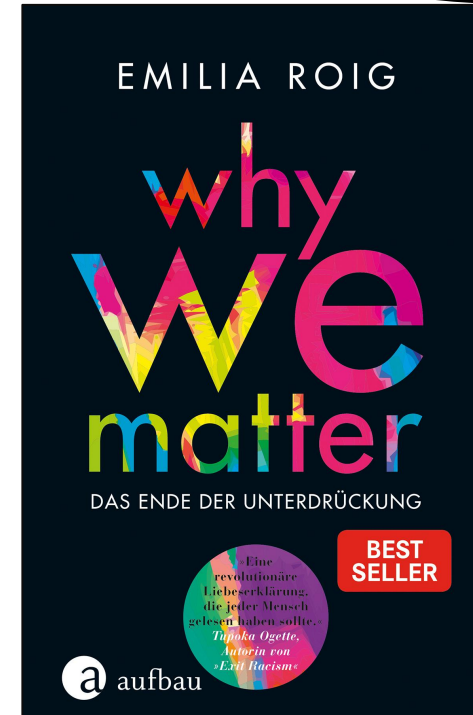


Safiya Umoja Noble, “Algorithms of Oppression”

The author dives deep into the unnerving practices of searching & sorting.

Emilia Roig “Why we matter” (German, not directly AI-related)

A great book on intersectionality & how people can be marginalized along multiple axes of their identity. Includes a great chapter on how knowledge (re)production is anything but neutral.



References – Podcasts



“Factually! With Adam Conover”

Especially the following episodes:

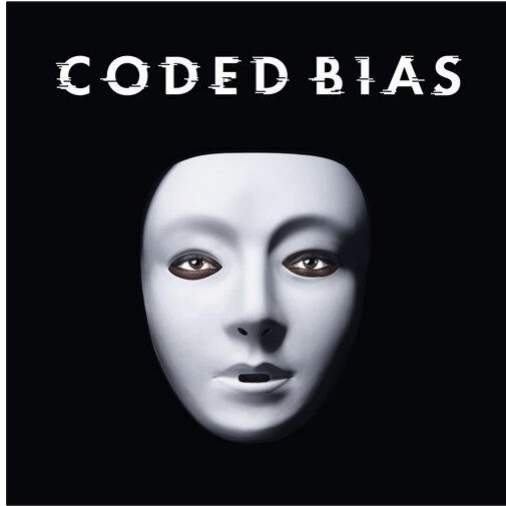
- “Is AI really inevitable?”
- “Why Self-Driving Cars Aren’t Coming Any Time Soon”
- “Why Search Engines Aren’t Unbiased”

MIT Technology Review “In Machines We Trust”

Especially the 4-part series about hiring, starting with episode “Hired by an algorithm”



References – Films & Series



“Coded Bias”, by Shalini Kantayya

Documentary on Netflix following Joy Buolamwini & her work to uncover algorithmic bias

First episode of Netflix series “Connected”

(Title: Surveillance)



Organizations around algorithmic justice



Algorithmic Justice League founded by Joy Buolamwini

AJL is an “[...] organization that combines art and research to illuminate the social implications and harms of artificial intelligence”

European Digital Rights
(EDRi)

“European Digital Rights (EDRi) is an association of civil and human rights organisations from across Europe. We defend your rights and freedoms in the digital environment.”



Organizations around algorithmic justice



AlgorithmWatch

“AlgorithmWatch is a non-profit research and advocacy organization that is committed to watch, unpack and analyze automated decision-making (ADM) systems and their impact on society.”

Access Now



accessnow

“Access Now defends and extends the digital rights of users at risk around the world”

Down the rabbit hole – scientific papers & talks

- Birhane, A. & Uday Prabhu, V. (2021). Large datasets: a pyrrhic win for computer vision?. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision.
⇒ I can generally recommend all papers by Abeba Birhane, they are very insightful!
- Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜 In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21). Association for Computing Machinery, New York, NY, USA, 610–623. DOI:<https://doi.org/10.1145/3442188.3445922>
⇒ This is the “controversial” paper that Google was unhappy about
- Inioluwa Deborah Raji, Morgan Klaus Scheuerman, and Razvan Amironesei. 2021. You Can't Sit With Us: Exclusionary Pedagogy in AI Ethics Education. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21). Association for Computing Machinery, New York, NY, USA, 515–525. DOI:<https://doi.org/10.1145/3442188.3445914>



Down the rabbit hole – scientific papers & talks

- Videos: “Taming the Machines” public lecture series of the Ethics in IT department at University of Hamburg, Germany
 - “The case against facial recognition” by Prof. Dr. Woodrow Hartzog
 - “The Global Digital Economy Made Concrete: Unpacking the Smart City” by Prof. Dr. Blayne Haggart, Prof. Dr. Natasha Tusikov
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2018). Datasheets for datasets. arXiv preprint arXiv:1803.09010.
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2019, January). Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 220-229).