Ethics in the Age of AI

Prof. Lokke Moerel

Morrison & Foerster

Data Science & Data Ethics
e-Conference by EAA

29/30 June 2020

Did not have proper brakes…

Where electricity extended man's

physical power

Information technology will extend man's

thinking power

Nicolas Carr, 'The Big Switch' (2013)

Will *AI* replace you?

Microsoft sacks journalists to replace them with robots

Users of the homepages of the MSN website and Edge browser will now see news stories generated by AI

▲ From July, the MSN homepage will no longer feature news stories produced by journalists at PA Media, formerly the Press Association. Photograph: Alamy

$$AI = \text{Cheap Predictions}$$

Jul 10, 2018, 12:28am

The Economics Of Artificial Intelligence – How Cheaper Predictions Will Change The World

Bernard Marr Contributor ⓘ

non-predictions → *predictions*

Complementary goods

AI ⟷ *Data + Judgement*

Judgement

Input

output

AI

is neither good nor bad, nor is it neutral.

It is as good as the data fed to the algorithms.

1st Law of Technology of Melvin Kranzberg, 2015

BUSINESS NEWS   OCTOBER 10, 2018 / 5:12 AM / UPDATED 15 HOURS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

That is because Amazon's computer models were trained to vet applicants by observing patterns in resumes submitted to the company over a 10-year period. Most came from men, a reflection of male dominance across the tech industry.

NEWS · 24 OCTOBER 2019 · UPDATE 26 OCTOBER 2019

Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals – and highlights ways to correct it.

The researchers found that the algorithm assigned risk scores to patients on the basis of total health-care costs accrued in one year. They say that this assumption might have seemed reasonable because higher health-care costs are generally associated with greater health needs. The average black person in the data set that the scientists used had similar overall health-care costs to the average white person.

european actuarial academy

Data - Ethics - Actuary

become a supporter | subscribe | search

find a job | US edition

theguardian

world | opinion | sports | soccer | tech | arts | lifestyle | fashion | business | travel | environment

browse all sections

Women less likely to be shown ads for high-paid jobs on Google, study shows

Automated testing and analysis of company's advertising system reveals male job seekers are shown far more adverts for high-paying executive jobs

If recruiters click more on white male candidates…

Judgement

output

AI in healthcare

Objective: predict probability of death for pneumonia patients to discover priorities in treatment

Output: patients suffering both pneumonia and asthma had lower risk (surprising!)

Why: patients with a history of asthma were admitted directly to the IC

The data set does not always tell the whole story

You are a data scientist with pharmaceutical company

- One of your prescription medications is in very short supply
- The business wants to develop a tool to predict how to best prioritize distribution
- The tool seems to disfavor certain ZIP codes
- The algorithm is a black box
- Because of the tool:
  - Distribution has improved
  - Complaints about non-availability have decreased

What should I do?

You are commercial director of a bank

- Divorce is large indicator for defaulting on mortgages and loans
- You are offered a tool predict likelihood of divorce based on analytics of public social media posts
- All your competitors have this tool

Do you implement?

AI Risks

Data is our biggest *asset* &

our biggest *liability*

CEO Google

The AI did it

is not an acceptable excuse.

Algorithmic Accountability implies an obligation to report and justify algorithmic decision making and to mitigate any negative social impacts or potential harms.

How to Hold Algorithms Accountable, MIT, Diakopulos and Friedler, 2017

- If you have data you can use it
  - No IP rights
  - No "ownership" of data

- Data protection rules are organizing principles

- Portal function for other fundamental rights

- **Privacy includes ethics!**

- 118 countries have "GDPR style" legislation
  - India, Australia, Brazil…

- EU "exports" GDPR in freetrade agreements
  - Japan, Israel, New Zealand

- Outliers
  - China, Russia, U.S., however…

GDPR matchup: The California Consumer Privacy Act 2018

Jul 31, 2018   Save This

Privacy is a human right, we need a GDPR for the world: Microsoft CEO

weforum.org

Privacy no longer a social norm, says Facebook founder

▲ People have become more comfortable sharing private information online, says Facebook founder Mark Zuckerberg. Photograph: Eric Risberg/AP

FACEBOOK

Zuckerberg promises 'complete overhaul' of Facebook geared towards user privacy at F8

POLICY \ BUSINESS \ TECH

IBM will no longer offer, develop, or research facial recognition technology

IBM's CEO says we should reevaluate selling the technology to law enforcement

By Jay Peters | @jaypeters | Jun 8, 2020, 8:49pm EDT
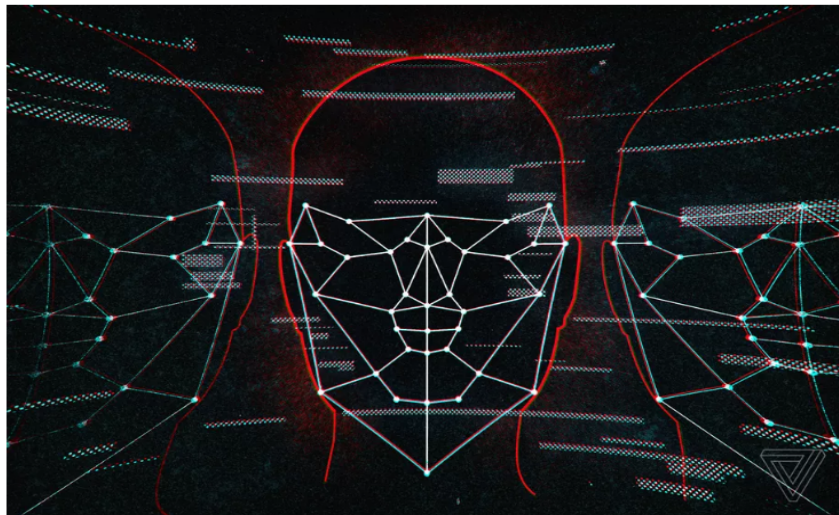
f 🐦 ↗ SHARE

Illustration by Alex Castro / The Verge

What are the challenges?

- *Discrimination & unfair bias*

- *Unforeseen applications*

- *Rendition of the Self*

- *Is consent possible?*

What are the challenges?

2. Unforeseen applications

AI analyses vast amounts of data

Reveals correlations and patterns
(descriptive)

Prediction of likely behavior of individuals

(predictive)

Used to direct or influence the future actions of individuals
(prescriptive)

sign in   become a supporter   subscribe   search

jobs   US edition

theguardian

US   politics   world   opinion   sports   soccer   tech   arts   lifestyle   fashion   business   travel   environment

browse all sections

home › tech

Artificial intelligence (AI)

New AI can guess whether you're gay or straight from a photograph

An algorithm deduced the sexuality of people on a dating site with up to 91% accuracy, raising tricky ethical questions
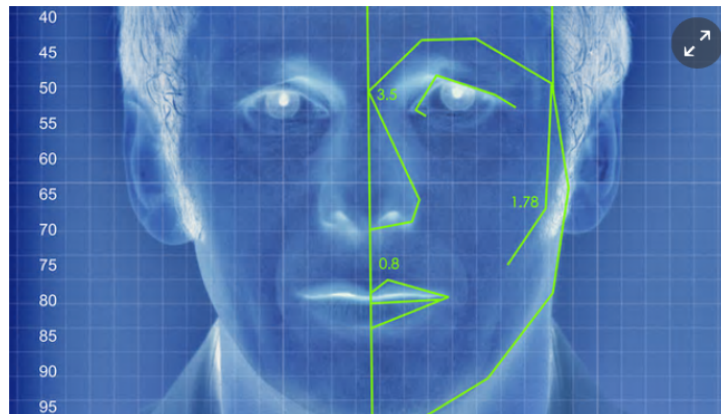
This article is 1 month old

59,104

Sam Levin in San Francisco

@SamTLevin

email

Thursday 7 September 2017 19.46 EDT

Most popular in US

What are the challenges?

3. The Rendition of the Self

Stanford | News

Home   Find Stories   For Journalists   Contact

Search Stanford news... 🔍

JANUARY 12, 2015

New Stanford research finds computers are better judges of personality than friends and family

Stanford researchers have found that computers can judge personality traits more accurately than one's friends and colleagues. In fact, artificial intelligence can draw inferences about a person as accurately as a spouse, according to Stanford postdoctoral fellow Michal Kosinski.

The results showed that a computer could more accurately predict the subject's personality than a work colleague by analyzing just 10 likes; more than a friend or a roommate with 70; a family member with 150; and a spouse with 300 likes.

"Given that an average Facebook user has about 227 likes (and this number is growing steadily), artificial intelligence has a potential to know us better than our closest companions do," wrote Kosinski and his colleagues.

Facebook targets 'insecure' teens

Daily News Article — Posted on May 2, 2017

Jump to...

- Questions
- Background

Print

- Entire Page
- Just Article
- Just Questions

(by Darren Davidson, The Australian) — Facebook is using sophisticated algorithms to identify and exploit Australians as young as 14, by allowing advertisers to target them at their most vulnerable, including when they feel "worthless" and "insecure," secret internal documents reveal.
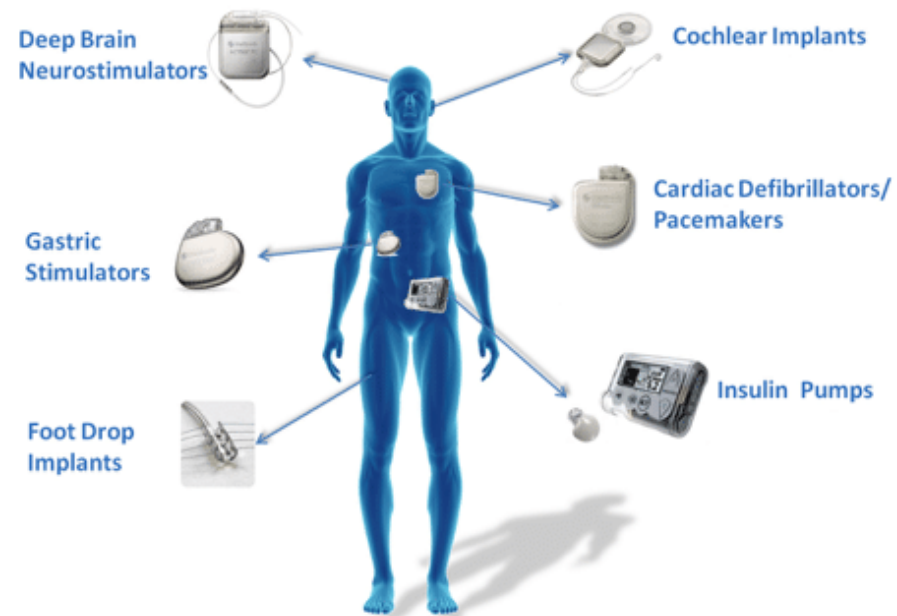
A 23-page Facebook document seen by *The Australian* marked **"Confidential: Internal Only"** and dated 2017, **outlines how the social** network can target *"moments when young people need a confidence boost"* in pinpoint detail.

By monitoring posts, pictures, interactions and internet activity in real-time, Facebook can work out when young people feel "stressed" "defeated" "overwhelmed" "anxious" "nervous" "stupid" "silly" "useless" and a "failure," the document states. [Such information gathered through a system dubbed sentiment analysis could be used by advertisers to target young Facebook users when they are potentially more vulnerable.]

WIRELESS IMPLANTABLE MEDICAL DEVICES

Deep Brain Neurostimulators

Cochlear Implants

Cardiac Defibrillators/ Pacemakers

Gastric Stimulators

Insulin Pumps

Foot Drop Implants

We have 4 - 5.000 data points
on every adult in the U.S.

CEO Cambridge Analytica

Did WE agree to this?

How to regulate all this?

The effects of new technology cannot be easily predicted until the technology is extensively deployed

yet once deployed become entrenched and are then difficult to change

David Collingridge, 'The Social Control of Technology' (1980)

- Require organizations that implement a new technology to evaluate the impact on privacy already in the design stage
  - *Privacy by Design*

- There are numerous ethical issues that are less visible and that we do not yet have good answers for
  - *Ethics by Design*

Data Protection Impact Assessment

The 10 Commandments List, Short Form

1. You shall have no other gods before Me.

2. You shall not make idols.

3. You shall not take the name of the LORD your God in vain.

4. Remember the Sabbath day, to keep it holy.

5. Honor your father and your mother.

6. You shall not murder.

7. You shall not commit adultery.

8. You shall not steal.

9. You shall not bear false witness against your neighbor.

10. You shall not covet.

Human Centric – deploy AI only if clear benefit for people and society

(no unfair bias)

Human Control – our processes within human direction and control

(no black box)

Human Accountability – we will remain accountable

(no blaming the algorithm)

Decision-making – transparency about decisions & underlying reasoning

(review & appeal)

Fairness – no discrimination

(no unfair bias)

Transparency – upfront & explanation after the fact

(no black box)

Automated decision-making – right to obtain human intervention

(review & appeal)

Accountability – demonstrate compliance

(burden of proof is on controller)

So how to achieve

Algorithmic Accountability?

1. Design for development at the outset

2. Select the training dataset (representative, accurate, no bias)

3. Verification of algorithmic outcomes in set stages of the development

4. Ensure transparency and explainability of outcomes to individuals

5. Can algorithm be used to prevent unlawful discrimination?

6. Ensure auditability of the algorithm

7. Document in DPIA

- You rarely have the right data

- You rarely have enough data

- Historical data is rarely unbiased

- Removing group-indicators does not work
  - For any group indicator there is a near-proxy...

- How to verify for bias if you do not have the **group indicators** in the first place?
  - You need a Master Data Set for verification purposes
  - What if you do not have the sensitive categories? Can you collect these for these purposes?

- With deep learning, development in stages is not possible

- You will have to verify ex-post output (rather than rely on ex-ante measures)
  - New training methods: adversarial training

- Explainability: the gap between the technical meaning and societal meaning

- We need to re-think explainability as a means to actually increase autonomy

Google Brain Built a Translator so AI Can Explain Itself

By Udit Agarwal | January 12, 2019 | 👁 23 | 💬 1

- Actuaries & privacy & ethics experts part of design teams
- Indepth technical knowledge
- Training in ethical dilemmas
- Open culture

3-line of defense - compliance model

hinders innovation

Morrison & Foerster LLP

You are welcome to join my network!

Contact

Data Science & Data Ethics
e-Conference by EAA

29/30 June 2020