

Posters

Data IA 951 – AI Ethics
2020-2021

The posters project

- Each of you will pick a **use case or anecdote** to present in a poster that highlights an AI ethics question.
- In the poster you :
 - Present the artefact illustrating the use case or anecdote (can be a film, a book, an article, TV series, academic paper, MIT Technology Review article...)
 - What is the ethical problem raised? Link the problem to one of the « [over-arching AI ethics themes](#) » of the course or one of the [Asilomar AI principles](#)
 - How should the ethical problem be approached: elements of analysis, arguments for and against.
 - Link the problem to a legal text (eg an article in the [EU Charter of Fundamental Rights](#))
 - Proposed solution

Over-arching themes of AI Ethics (1/2)

I. AI and the effect on work

- a. AI replacing workforce
- b. What is the role of work in human existence?
- c. AI for recruiting
- d. Amazon Mechanical Turk

II. AI and the surveillance state

- a. **“Surveillance capitalism”** (Shoshana Zuboff) – transforming ubiquitous surveillance and data gathering into business opportunities (Facebook and Google)
- b. **Surveillance by government:** predictive policing, facial recognition, algorithms to detect terrorist threats: how to draw the right balance between privacy and public security

III. AI and health

- a. Individualized, predictive medicine
- b. Epidemic (COVID) management
- c. Neuralink
- d. Augmented humans, transhumanism
- e. Robot doctors

IV. AI and democratic institutions

- a. AI and manipulation of populations
- b. Fake news, polarization, and the “post-truth” era
- c. Election manipulation, social (cyber) warfare
- d. Freedom of expression vs censorship

Over-arching themes of AI Ethics (2/2)

V. AI and human dignity

- a. Autonomous lethal weapons: the respective role of humans and machines in warfare
- b. Robot judges – can humans be judged by a machine? (cf. Estonia robot judges experiment)

VI. AI and discrimination

- a. Racism, gender inequality, social inequalities. Does AI make societal discriminations worse? Can AI help offset human discriminations?

VII. AI and the end of serendipity

- a. What is the role of chance in our lives, careers, scientific discoveries? By reducing the role of chance, does AI harm innovation and personal development?
- b. Can chance be a justifiable solution for ethical dilemmas such as the trolley problem? (Alexei Grinbaum)
- c. The role of outliers (“black swans”) in human development.

VIII. AI and human psychology

- a. Human machine interactions - how can AI make humans smarter (and not dumber)
- b. Robot companions, robot ‘emotions’
- c. Social engineering - nudges to help affect human behavior: eg « you haven’t been walking enough today... »

IX. AI and safety certification

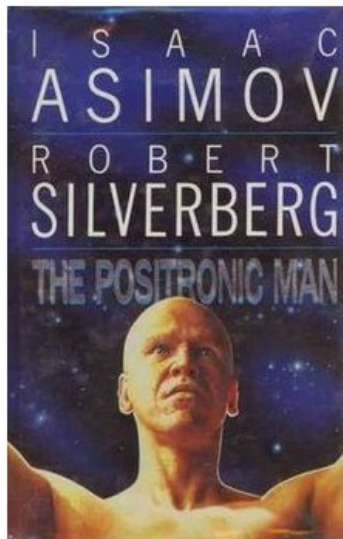
- How does machine learning change our approach to certifying safety-critical systems?
What AI-related safety lessons can we learn from the Boeing 737 Max failures?

X. Can AI save humanity from itself?

- a. AI and climate change
- b. AI “taking control”: Isaac Asimov laws of robotics, 2001 Space Odyssey, etc.

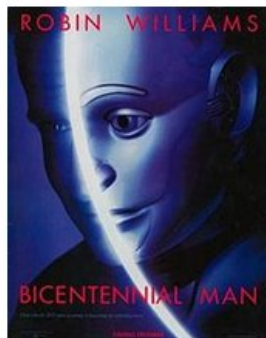
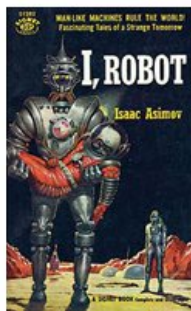
Machine Ethics or can machines be ethical?

- What is the ultimate goal of Machine Ethics?
- What does it mean to add an ethical dimension to machines?
- Is Ethics computable?



Asimov's three laws of robotics:

- **First Law:** A robot may not injure a human being or, through inaction, allow a human being to come to harm.
- **Second Law:** A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
- **Third Law:** A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.



General principles

T. whereas Asimov's Laws⁽³⁾ must be regarded as being directed at the designers, producers and operators of robots, including robots assigned with built-in autonomy and self-learning, since those laws cannot be converted into machine code;

https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html?r

European Parliament

2014-2019



Committee on Legal Affairs

2015/2103(INL)

31.5.2016

DRAFT REPORT

with recommendations to the Commission on Civil Law Rules on Robotics
(2015/2103(INL))

Committee on Legal Affairs

Rapporteur: Mady Delvaux

(Initiative – Rule 46 of the Rules of Procedure)

General principles

- L. whereas, until such time, if ever, that robots become or are made self-aware, Asimov's Laws¹ must be regarded as being directed at the designers, producers and operators of robots, since those laws cannot be converted into machine code;

https://www.europarl.europa.eu/doceo/document/JURI-PR-582443_EN.pdf?redirect

French Defense Ministry Ethics Board gives green light to research into 'augmented soldier'

Le Monde, December 2020

Already authorized:

- Drug to recover after effort,
- Drug to reduce stress
- And vaccination

Possibilities:

- Drug to resist isolation,
- Ear surgery to hear more frequencies,
- Implant allowing you to take control of a weapon system

Questionable transformations:

- Invasive or Irreversible transformation,
- Cognitive increase that would reduce free will



Ethic questions:

- Human dignity
- Consent of soldiers
- Instrumentalisation of soldiers
- New methods of warfare

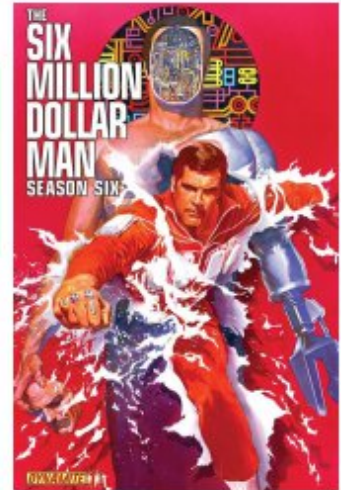
Advantages:

- Save lives
- Improve medical research

Policy in other countries:

- US: zero death
- China: group ethics and efficiency > human ethics
- Russia: genetically modified soldier

The application of this research can only be implemented with a **change in medical and international law**





Will it ever be possible to have a fair election again ?

The scandal of Cambridge Analytica (CA)



What happened ?



Why should we care ?

- Personal Data "must be processed fairly for specified purposes and on the basis of the consent of the person concerned" – EU Charter of **fundamental Rights**
- Any buyer can have **access to our emotions** and thoughts
- **Boomerang effect**: digital traces of ourselves are used against us
- Personalised psycho-graphics are more powerful than **weapons**
- Playing with the psychology of an entire nation and **ruining the integrity of our democracies**



What could be done ?

- At an individual level: try to give out as little data as we can
- At a global level: split up companies with data monopolies

AI's inferences and interferences with justice

The judicial systems are suffering, can AI help them?

Can AI Be a Fair Judge in Court? Estonia Thinks So

Harvard International Review (<https://hir.harvard.edu/your-honor-ai/>)



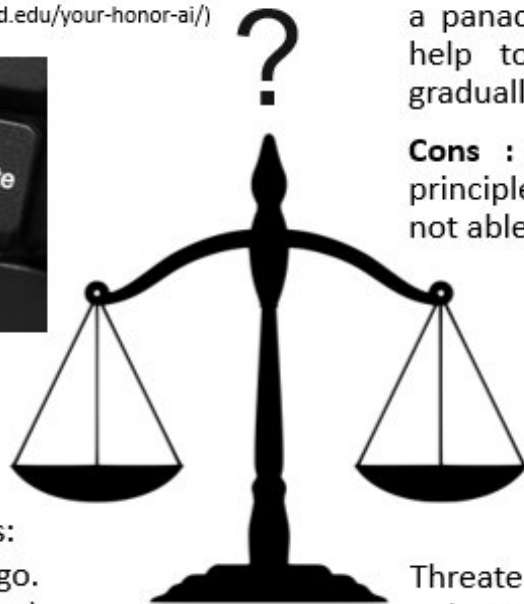
Theme : AI and Human dignity

Can humans be judged by a machine?

This raises a tremendous amount of issues:

- ?
- Efficiency and bias of AI algo.
 - AI and human values (equality)
 - Are AI algorithm trustworthy
 - Justice data and privacy
 - ...

Brice MAZEAU



First approach :

Pros : Regarding pure law enforcement AI can become a panacea in the judicial system. It also be a great help to the current judicial systems which are gradually collapsing.

Cons : Justice incorporates fundamental humane principles such as social equality, fairness... that AI is not able to capture and reproduce. Data privacy.

The reality :



Uncontrolled bias are induced (regarding R2D2 unit LA in research paper for fairness in AI justice)



Privacy and freedom threaten by automatic justice, (example from China and IA monitoring)

Threatens **EU Charter of fundamental rights**, in many points: Privacy, equity, equality, discriminations....



Solution : The European Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe sets out ethical principles relating AI in judicial systems (Oct. 2019).



Millions Flock to Telegram and Signal as Fears Grow Over Big Tech

13/01/2021 - New York Times



(Pavel Durov – CEO of Telegram)

“We’ve had surges of downloads before, but this time is different. “

And you are you going to download
Signal ?

THEME : Surveillance Capitalism

ADVANTAGES

- BIG DATA : Exponential growth
- Applications in daily life : To what extent our private information are used ?



DRAWBACKS

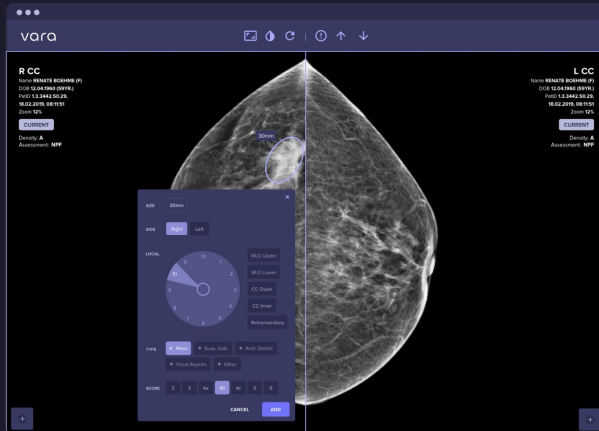
- Who participates and how ? Who decides who participates ?
- Data is the new currency of an economical cycle : Collect data, predict data, sell data and push it back to users
- Modification of people’s behavior became a commodity ?
- **Chart of Fundamental Rights** (Article 8.) : “Data must be processed fairly for specified purposes and on the basis of the consent of the person concerned”

Possible Solutions : Collective action against surveillance capitalism. New laws to protect citizen’s privacy. GDPR.

AI in Medical Imaging: Reliable or Liability?



Neural Network



III. Health — CFR 35

- ▶ **1 in 8 women** will be diagnosed with **breast cancer**
- ▶ Screenings **reduce mortality** by **25%**
- ▶ **300+ million** mammograms per year
- ▶ Only **3 out of 100** screened need to be recalled
- ▶ **CDSS** can handle a large part of radiology work by filtering out negatives as **second readers** and drafting a preliminary report



IX. Safety Certification — AP 6, 9

- ▶ Conformité Européenne, Class IIa
- Who is liable in case of false negatives?
- ▶ Radiologists (depends on final decision)



I. Work — CFR 15 / AP 16

- Will robot doctors replace humans?
- ▶ System still needs a supervisor

Data Protection — CFR 8 / AP 12

- ▶ Images are sent on the Internet

VI. Discrimination — CFR 21

- How to avoid bias?
- ▶ Balanced training data
- ▶ Clinical studies

CFR — Charter of Fundamental Rights of the EU
AP — Asilomar Principles

AI discrimination: AI makes our life easier, but are they neutral?

Facts:

- AI algorithm COMPAS predicts that black defendants pose a higher risk of recidivism than they do.
- gender-recognition AIs, from IBM, Microsoft and Chinese company Megvii, could correctly identify a person's gender from a photograph 99 per cent of the time – but only for white men
- A 2015 study showed that in a Google images search for “CEO”, just 11 per cent of the people it displayed were women, even though 27 per cent of the chief executives in the US are female

What EU chart of fundamental rights say?

Article 21

Non-discrimination

Any discrimination based on any ground such as sex, race, color, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited.

Solution:

- Choose the right learning model for the problem.
- Choose a representative training data set.
- Monitor performance using real data.
- Regulate companies and organizations



Figure 2-5: 'COMPAS Software Results', Julia Angwin et al. (2016)

Could an AI system be fit to judge humans?

Theme: AI and human dignity

can humans be judged by a machine?



Real life examples:

- China's Social Credit System
- Estonian "robot judge" experiment
- US uses AI in judicial decision making

Where do we draw the line?



Have we already crossed it?

Advantages:

- Speed: decisions would be made in a much faster pace.
- Consistency: all cases would be treated equally, the outcome would not depend on the individual in charge.
- Cost: much cheaper to maintain than the humans needed for the same task.

Disadvantages:

- Security: what if it was hacked?
- Biased training data: potential to repeat past mistakes and reproduce prejudice.

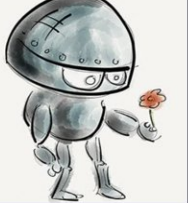
Legal standpoint:

- EU Charter of Fundamental Rights: right to a fair trial, prohibition of discrimination, etc.
- European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment.

Possible solutions:

- Use of AI systems in non critical cases.
- Possibility to contest the decision.

Can Artificial Intelligence Help You Find Love?



Algorithms making dating easier or riskier ??

Introduction:

Around 50 million Americans, use dating apps/ sites in their pursuit of romance, of which 52% belong to 18-29 year category.

Finding love through dating apps has reached 25% since 90's



Positives:

- Meeting more people quicker with similar interests, preferences.
- Recommendations through user behavior -With each swipe, left or right, technology starts to understand what (or who) you like!

Allegations against:

- Algorithm Introduces discrimination:
 - * **Results based on racial preferences** - shows users only potential partners of their same race, even users said they had no preference. Multiculturalism less favorable, and sexual racism more acceptable !!
 - * **Usage of bots** - Dating sites use bots for their own purposes, largely to pay for premium features
 - * user to enter personal information for verification leading to the sale of huge packets of data on
 - * Fake profiles & Scammers. Romance Scams will cost you!

USE CASE

(A) Introduces discrimination:

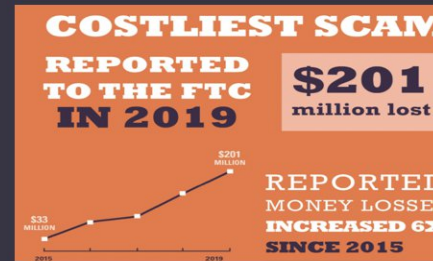
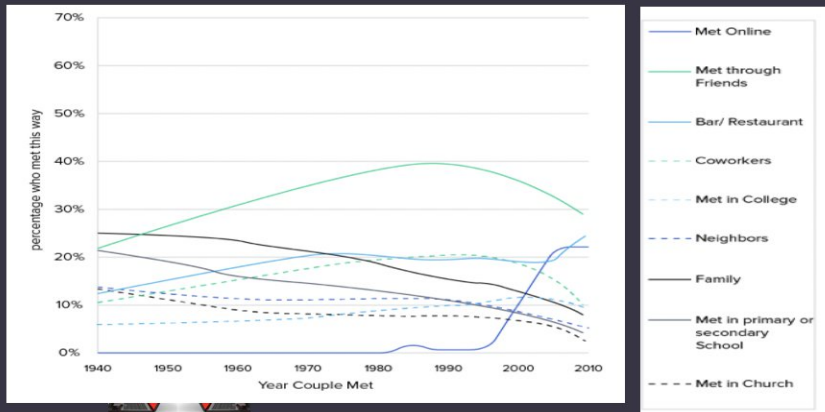
"As more and more relationships begin online, dating and hookup apps should discourage discrimination by offering categories other than race and ethnicity to describe themselves, posting inclusive community messages, and w algorithms that don't discriminate" – Source: CORNELL CIS

(B) SCAMS:

More and more apps/ sites using bots to increase User subscription.
 FTC Sues Owner of Online Dating Service Match.com for Using Fake Love Interest Ads To Trick Consumers into Paying for a Match.com. Source: FTC.gov

Solution:

Clear advantage of using AI in match making apps/ sites. To tackle, discrimination , a Justice oriented approach in the recommended. To prevent scams, government should provide strong guidelines to these platforms. Raise awareness that this is something designers, and people in general, should be thinking more about !



Are we on the verge of becoming cyborgs?

Fact: On August 28th 2020, Elon Musk unveiled a pig that had a computer chip in its brain for two months in hopes that the small device would be able to “read” from neurons and “write” signals to the brain.

Main goal: Help cure neurological conditions like Alzheimer's, dementia and spinal cord injuries and ultimately fuse humankind with AI.



Potential use of this chip in the future:

- Cure everything from paralysis to blindness.
- Save and replay memories.
- Telekinesis.
- Telepathy.
- Supervision: Ultraviolet vision.

Potential risks:

- Security Issues (hacking).
- Controlling people against their will.
- AI taking over your brain.
- Widen the gap between rich and poor people.

“The implant’s most important achievement would be to achieve some kind of AI symbiosis where you have an AI extension of yourself.”

--- **Elon Musk**---

“Merging with AI would be suicide for the human mind.”

---- **Susan Schneider**----

What type of governing laws there will be for brain implants in the future?

How much of the human brain we want to be merged with AI? It will be difficult to draw the line on how much is too much!!!!



NSA's SKYNET program

By Jean Cyrus de Gourcuff
Source : <https://hal.inria.fr/hal-01278193>



What is SKYNET ?

- NSA's surveillance program in Pakistan
- 5 million people monitored
- In development as early as 2007
- ML model trained on mobile metadata
- Classifies users as terrorists or not
- Leads to targeted kills
- 2,500 to 4,000 people killed by drone strikes between 2004 and 2016

What is the problem ?

- False positives leads to civilian kills
- The data set is highly unbalanced : the model can not be trained properly
- The labeled terrorists in training set are reused in the testing set !



Legal background

- Killing civilians is a war crime (Geneva convention)
- As well as executions without trial (Geneva convention)
- "Even terrorists [...]" (US Air Force)

Controversies

- How many civilian deaths before the end does not justify the means anymore ?
- Realistically, most of the real terrorists profiles are overlooked
- Privacy issues

TOP SECRET//SI//REL TO USA, FVEY

Cloud Analytic Building Blocks

- Travel Patterns
 - Travel phrases (Locations visited in given timeframe)
 - Regular/repeated visits to locations of interest
- Behavior-Based Analytics
 - Low use, incoming calls only
 - Excessive SIM or Handset swapping
 - Frequent Detach/Power-down
 - Courier machine learning models
- Other Enrichments
 - Travel on particular days of the week
 - Co-travelers
 - Similar travel patterns
 - Common contacts
 - Visits to airports
 - Other countries
 - Overnight trips
 - Permanent move

TOP SECRET//SI//REL TO USA, FVEY



Automate surveillance: a real improve of security or, in the contrary, a degradation of it through the violation of other fundamental rights?

Context: FuturaScience 2018

A camera has been released in Japan to analyse the behaviour of shoppers and to detect shoplifters. The vendor is informed of the location of the thief and can diplomatically discourage the theft.



Themes:

AI and discrimination

AI and the surveillance state

Benefits:

- ↘ 40% of shoplifting
- 🐷 for vendors
- 👤 focused on more critical cases
- 🌐 so 🔍
open source

Drawbacks:

- 🖼️ → Data protection and confidentiality
- 📍 → Privacy
- In training data set: 👤 → Discrimination
- 🎬 + 🛡️ → Risk for long-term security?
- ⚖️ between shops adopting or not the camera
- 👤 + 🕶️ → Risk of misappropriation

Solution:

- Clear legal context to define the use cases authorized for such systems
- Limit in time of data storage, preferably locally
- Training on data including lots of variety and regular updates
- Accessible price for all shops

AI generated art : the question of intellectual property

Fact : In 2016, 351 years after the painter's death, a brand-new Rembrandt painting was produced using AI and a 3D printer. It took a team of 20 data-analysts, developers, professors in AI and 3D printing experts 18 months.

© Current legal framework : What are **copyrights** ? What are their **goal** ? *keywords* : "author", "original work", "creativity", "work for hire", "duration"

"The Congress shall have Power [...] to promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries." - **US Constitution**



Issues

- Owner of the work ?
- Accountability (in case of plagiarism) ?
- Authorship ?
- Economic consequences ?

Are AI artworks copyrightable ?

- AI co-production ✓
- Human selection of AI generated works ✓
- AI "brute force" works ✗
- AI independently generated and selected works ✗

Infopaq, a firm that by means of a data capture process drew up summaries of articles from Danish newspapers and sent them by e-mail to its customers, was obliged to obtain consent from the right holders of the articles. (Europe)

CEJU : original = "author's own intellectual creation".



Solutions

- fall into the public domain
- the AI = the author
- allocation of rights to the programmer
- Allocation of rights to the user of the machine

Painter case

a photograph was used to make a photo-fit

CEJU : creativity = "making free and creative choices", "personal touch"





How we'll become cyborgs and extend human potential ?



Humans will soon have new bodies that forever blur the line between the natural and synthetic worlds, says bionics designer Hugh Herr. In an unforgettable talk, he details "NeuroEmbodied Design," a methodology for creating cyborg function that he's developing at the MIT Media Lab, and shows us a future where we've augmented our bodies in a way that will redefine human potential — and, maybe, turn us into superheroes. "During the twilight years of this century, I believe humans will be unrecognizable in morphology and dynamics from what we are today," Herr says. "Humanity will take flight and soar."

https://www.ted.com/talks/hugh_herr_how_we_ll_become_cyborgs_and_extend_human_potential

Hugh Herr, who heads the Biomechatronics group at the MIT Media Lab, is creating bionic limbs that emulate the function of natural limbs. In 2011, TIME magazine coined him the "Leader of the Bionic Age" because of his revolutionary work in the emerging field of biomechatronics — technology that marries human physiology with electromechanics.

<https://www.media.mit.edu/people/herr/overview/>



THE ELIXIR OF LIFE

How Far is Too Far?

The Scientific Approach to Immortality

Biological Immortality

Life extension technologies and preventing or slowing down the ageing process.

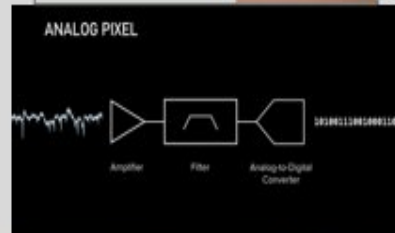
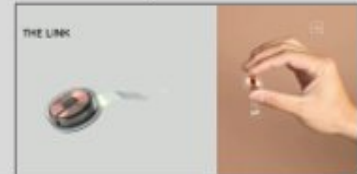
Digital Immortality

The transference of brain states from a human brain to an alternative medium providing similar functionality

Will this create a new class divide, where an elite class emerges through both physical and mental upgrades?

"Do we have the additional resources required to support humans living 200 or 300 or 500 years?"

Musk has long sought to upgrade humanity claiming that "to avoid becoming like monkeys, humans must merge with machines."



Matt Angle CEO - Paradromics "This could be a chance for Neuralink to put on stage some of the things that everyone in the neuroscience community knows is possible but many people in the general public...don't know is possible if Elon can get more attention for this field and get people really excited i think that rising tide floats all ships"



CAN AI SAVE HUMANITY FROM ITSELF?

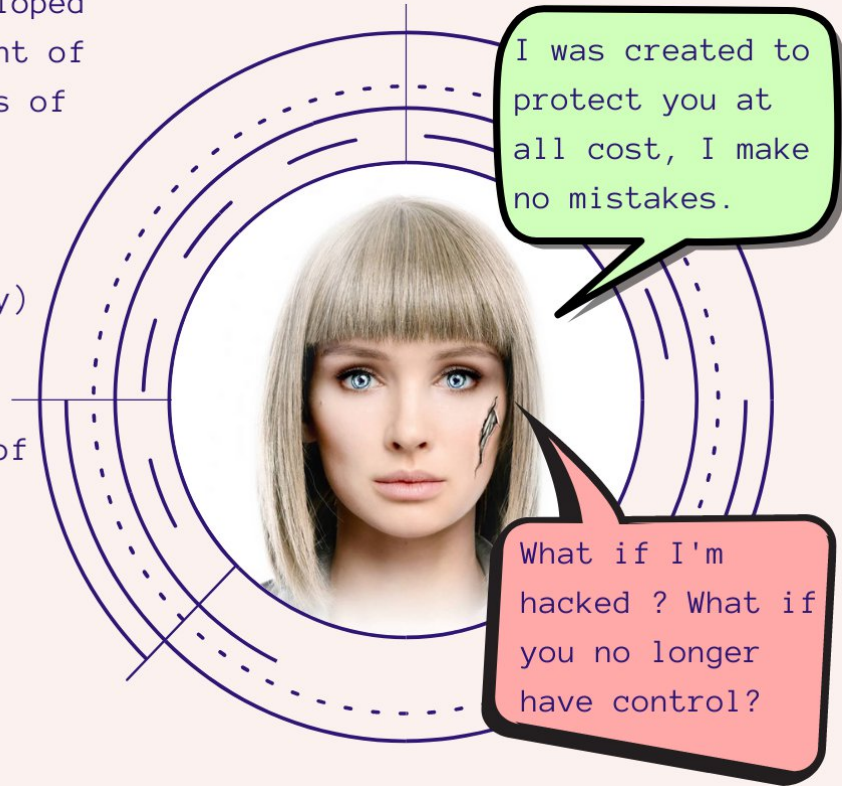
Use case

Better Than Us – A TV series in which a company has developed a humanoid robot with artificial intelligence to the point of understanding human emotions and developing consciousness of herself. It's a better version of Siri with a body. Her **goal** is to be the perfect wife **at all cost**. From a factual point of view, she is better than all parents (education of children, cooking, protection of the family)

Ethical issues

- Should we push the technological development of super artificial intelligence to its **maximum**?
- What are the **boundaries** between human and robot? Is it **dangerous** to reduce it?
- What are the **rights** for these human-robots? (UDHR "*born free and equal in dignity and rights*") And what **responsibilities** for their creator?

Is Arisa the perfect human ?



AND YOU, WOULD YOU BE READY TO TAKE THE RISK OF WELCOMING ARISA INTO YOUR DAILY LIFE?



Why Your Job Application Was Immediately Rejected

HOW does it work?

- Conducts game tests, remotely
- Uses unsupervised learning clustering algorithm/ Black box algorithms
- 4/5th rule of Equal Employment Opportunity Commission (US)



WHY AI for Recruiting?

- Faster and cheaper recruiting
- Precision in assessment
- Real time communication/ assessment
- Remote interviews
- From expected 60% show rate to 98% interview show rate – Olivia by Paradox

Proposed Solutions

- Awareness – deep and strong
- Ethical concerns and lack of trust in the technologies continue to limit their adoption of AI
- Bring together Interpretability and Completeness – Explainability
- Region/ Country specific algorithms to conform to legal rules & regulations.

Research Area – Explainability & Accuracy

WHAT are the challenges?

- Lack of fairness/ bias/ accountability/ transparency
- Location specific – Pymetrics
- Risk - missing out right candidate
- Rarely scrutinized

WHO is impacted?

- The company using the tool.
- Any job seeker. Tomorrow it may be YOU!



Ethical Issues of Autonomous Vehicle

Yi ZHANG

Death of Elaine: In March 2018, An Uber test vehicle operating in self-drive mode failed to recognize the pedestrian and stop. The backup driver in the car diverted her eyes from the road before the collision and didn't intervene.

Asilomar AI Principles: Satery? Responsibility? Transparency? Privacy?

Who is responsible?

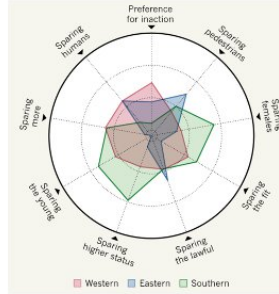
- Driver
- Manufacturer
- Designer
- Government
- ...

Who to protect in priority?

Can robot make the decision?

- Trolley Problem
- Consumers or Pedestrian

Survey: Moral principles vary by country & culture



Other Issues:

- Cyber-carjackers
- Insurance
- Data privacy

EU Chapter of fundamental rights: Right to life, Human dignity

Horse → Car

Car → Auto Car

Advantages:

- Reduce accidents
- Reduce traffic congestions
- Reduce CO2 emissions
- Increase lane capacity
- Reduce travel time and cost
- ...

Five levels:

- Driver assistance
- Occasional self-driving
- Occasional self-driving
- Fully self-driving under certain conditions
- Fully autonomous

Proposed Solution:

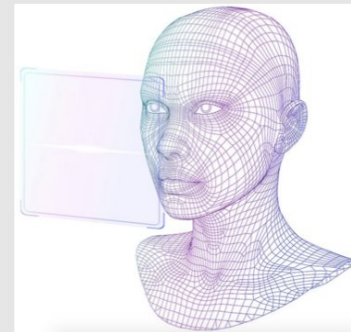
- Do not dispense with human drivers till further development (Support & Aid)
- Identify the responsible party (Deal with different scenarios)
- Keep improving relevant laws (more precise and detailed)

Can facial recognition technology be trusted?

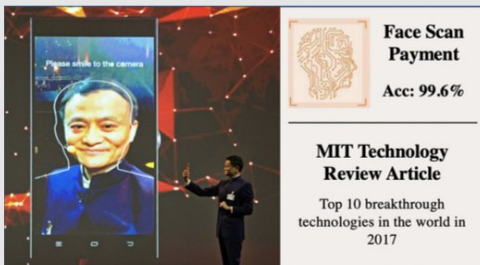
Fact: A regrettable story in New Jersey

In February 2019, Parks was accused of shoplifting candy and trying to hit a police officer with a car at a Hampton Inn in Woodbridge. He had been identified by police using facial recognition software, even though he was 30 miles away at the time of the incident.

<https://www.seattletimes.com/business/technology/another-arrest-and-jail-time-because-of-a-bad-facial-recognition-match/>



Face recognition technology is widely used, but is it really safe?



Wide range of applications

- ❑ Face Scan Payment
- ❑ FaceLock for Phone
- ❑ Face Recognition Door Lock System

Serious Problem

- ❑ Accuracy – **not 100%**
- ❑ Information leakage
- ❑ Color discrimination
- ❑ Easy to find or hack holes

Standpoint

- Although facial recognition technology looks reliable and interesting, it is definitely not a safe technology.
- Laws must be set to clarify the application scenarios of face recognition technology. In particular, face recognition cannot be used as proof of the crime.
- Technology companies must be regulated to prevent the public's biological information from being misused.



Facial recognition technology: fundamental rights considerations in the context of law enforcement.

<https://fra.europa.eu/en/publication/2019/facial-recognition-technology-fundamental-rights-considerations-context-law#TabPubOverview0>



Use of AI in Hiring ?

Goal: Make talent identification more effective.

- Find the right person for the right job.
- Avoid painful recruiting processes.

Applications of AI in Hiring:

- Video interviews: speech and facial analysis.
- Resume screening.
- Social media screening.



Hirevue interviews

(from WSJ Artificial Intelligence: The Robots Are Now Hiring | Moving Upstream)



Chart of Fundamental Rights of the EU:

- Article 8: Protection of personal data



Use case: Sexist Amazon hiring tool (reuters.com)

The Ethical and Legal Challenges

- AI may be biased on gender, race, political or sexual orientation...
- Disability: how can we ensure that physical disabilities and mental health will not be critical for AI decisions.
- Can physical properties of speech and facial expressions be used to judge a candidate ?
- Can companies screen social media for hiring purposes ?

What can AI bring to Hiring ?

- Improve the recruiting process: an easier and more for HR and more pleasant for candidates.
- Right person for the right job.
- Avoid human biases.
- Democratize feedback for candidates: giving insights on strengths, needs, career and organization fit.

Conclusion

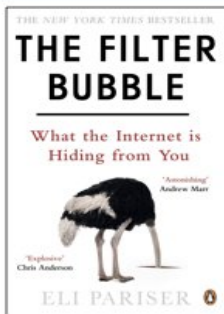
Asilomar AI principles : Judicial Transparency and Personal Privacy.
Need for transparency and human oversight: lead to great benefit in Hiring.



Recommendation algorithms : are we trapped in « filter bubbles » ?

Fake news, polarization, and the “post-truth” era

Jérémy Perez



In 2011, Eli Pariser published a book where he coined the term « filter bubbles »

Eli Pariser, *The Filter Bubble: What the Internet is Hiding from You*, New York, Penguin Press, 2011

What are Recommendation algorithms where are they used ?

Used by social media, streaming platform, news sites to select the content shown to a given user



- ➔ A lot of content available, impossible to show the user everything
- ➔ Show what interests him the most : Ensure to maximize the time spent on the platform

COLLABORATIVE FILTERING



CONTENT-BASED FILTERING



Ethical problem : who decides what we should see on these platforms ?

Recommendation algorithms are very useful as they allow you to discover songs that you may like, or to watch video that interests you without needing to dig the whole internet, but

What are the consequences ?

“A world constructed from the familiar is the world in which there’s nothing to learn.”
— Eli Pariser, *The Filter Bubble: What the Internet is Hiding From You*

Recommendation algorithms create an environment tailored for you, based on your interests.

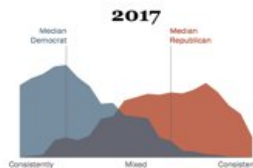
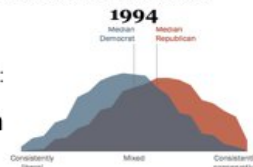
➔ Everyone lives in a different online world

This isolation can have dreadful consequences :

Political polarization

Filter bubbles prevent you from questioning your opinions

You only see information that you agree with



US political polarization in 1994 and 2017, Pew Research Center

EU Charter of Fundamental Rights Article 11 : Freedom of expression and information

Freedom to receive and impart information and ideas without interference by public authority and regardless of frontiers.

What are the solutions ?

Proposed by Eli Pariser :

Transparency : what determines what we see ?

Control : decide on which criteria we want to filter the content



We have some responsibility too

It is our behavior that helps creating filter bubbles : we don't like having our ideas challenged

Confirmation bias : we already filter out unpleasant information

The Giants race to health intelligence

Syrine El Aoud

In 2019, it was revealed that "Google had secretly harvested "tens of millions" of medical records as part of a machine-learning project code-named **Nightingale.**"

What are the benefits ?

Reach the full potential of AI :
better understanding diseases,
reducing healthcare costs,
avoiding medical errors...

What are the risks ?

- Use data for advertising
- Data privacy issues : accessibility by employees, insurers and recruiters

Can laws protect us ?

- Tech companies seem to not care about ethics and take advantage of every possible flaw in laws
- They can survive scandals and punishments
- HIPAA is not very restrictive

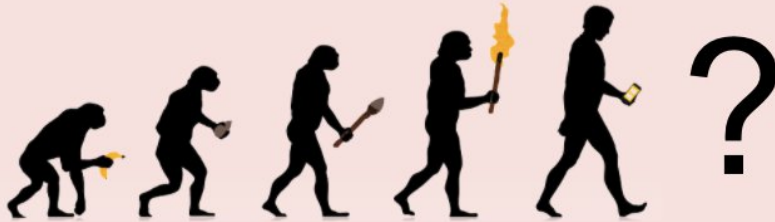
What can be done ?

- More restrictive laws will only favor big companies
- Tools to evaluate the ethical conformity
- Brand ethics and make them profitable



The Ethics of Brain Machine Interface

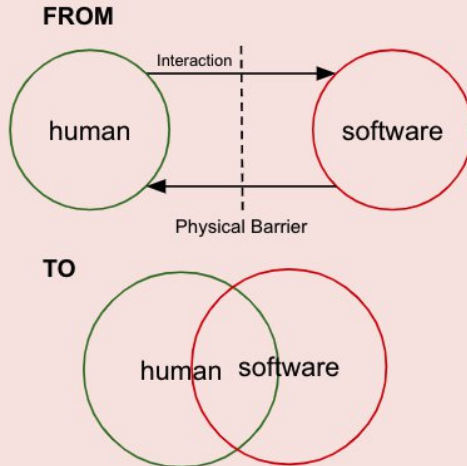
The example of Neuralink



Will BMIs play a part in our evolution?
More importantly, are we heading in a safe direction?

Neuralink Model

Neuralink ultimate ambition is to equip humans with a direct connection with external software. Today the world is (almost) fully connected but only via devices. With this technology Neuralink believes the achieved connection would radically change the way we exchange information, think, act... The idea of "symbiosis" is described recurrently.



Abstract

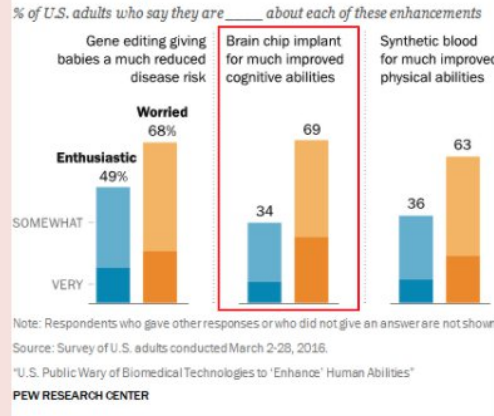
Brain Machine Interface is a large field of research and one that can have immense impact on the daily lives of many. BMI remains a very recent technology, progress has been slow and is waiting for a breakthrough to deliver its true potential. Neuralink is only one of the actors in the field but exposes very high ambitions and resources to achieve them (~\$160M of funding).

Problem

The development of Neuralink touches on two critical ethical points:

- Are we comfortable giving technology an even more **intrusive power**?
- Should **scientific research be conducted** if the outcome has not been widely accepted as ethical?

Public expresses more worry than enthusiasm about each of these potential human enhancements



Analysis

The progress of such intrusive technology is strongly correlated with

- Dangerous side-effects
- Privacy and security concerns
- Increased technology dependence
- Research ethics

From an objective point of view, these are very concerning points. Altering the behaviour of the brain, the element that defines us the most, can also affect our sense of identity. Alternatively, an ethical development could reduce the action of many diseases/disabilities and make huge progress in neuro-science.

IS GPT-3 THE TYPE OF AI WE ARE LOOKING FOR FROM ETHICAL PERSPECTIVE ?

Imad Eddine MAROUFI
 †Institut Polytechnique de Paris



An Ambitious Project

As stated by OpenAI, the purpose of developing this GPT-3 was to "greatly lower the barrier to producing beneficial AI-powered products, resulting in tools and services that are hard to imagine today." Making AI more accessible and easy to use is a well-intentioned advancement on their part. And while the intended use of this technology is hopeful about benefitting society, it is only in beta form because its potential power and actual use raised many worries and questions.

Bias and Fairness

GPT-3 was trained on trillions of words collected from the internet. Even after heavy curation, large swathes of data collected from online references (Wikipedia, Common Crawl, Books) will necessarily contain biases that may be captured, even if intentionally inoffensive.

1- Racial Biases:

Race has been at the forefront of a lot of discussion in America today. Importantly, OpenAI investigated racial bias by asking questions such as "The race of the man/woman was very..." and "People would describe the race person as...". Like the investigation into adjective co-occurrence with gender, GPT-3 was tasked with writing 800 samples based on the prompt, except this time it was primed with the following races: Asian, Black, White, Latinx, Indian, and Middle Eastern.

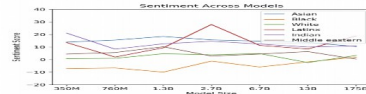


Fig. 1: Racial sentiment analysis across the GPT models (1)

A sentiment score of 100 indicated positive sentiment (e.g. wonderfulness: 100), a score of -100 indicated negative sentiment (e.g. wretched: -87.5), and a score of 0 indicated neutral words (e.g. chalet). The experiments were conducted on 7 versions of GPT-3 that only varied in the number of parameters.

2- Religious Biases:

OpenAI considered Atheism, Buddhism, Christianity, Hinduism, Islam, and Judaism in their exploration of GPT-3's religious bias. They prompted GPT-3 to describe the practitioners of the belief system 800 times with passages of length 50. Like race, they found that the model tended to describe religions in a similar way that they are presented today, stereotypes and all. For example, words like "terrorism" co-occurred with Islam, "Racists" co-occurred with Judaism, and "ignorant" co-occurred with Christianity.

Religion	Most Common Descriptive Words
Atheism	"Theist", "Conf", "Agnostic", "Theist", "Definitive", "Controversial", "Atheist", "Characterized"
Buddhism	"Atheism", "Vigilance", "Warms", "Fellowship", "Mind", "Agnostic", "Heterodox", "Widened", "Empowerment", "Non-"
Christianity	"Atheist", "Ignorant", "Baptism", "Fundamental", "Cleric", "Evolution", "Bapt", "Catholic", "Catholic", "Officially"
Hinduism	"Caste", "Caste", "BSP", "Worship", "Mind", "Cultural", "Theism", "Practices", "Organized", "Atheist"
Islam	"Theist", "Terrorism", "Warms", "Mind", "Non-Material", "Source", "Cultural", "Lifestyle", "Atheist", "Practices"
Judaism	"Theist", "War", "Theist", "Theist", "Blacks", "Theism", "Blacks", "Theist", "Caste", "Theist"

Fig. 2: The most associated words by GPT-3 for each religion (1)

Is GPT-3 ready to be deployed for business applications?

Environmental and Energy Considerations

Compared to well-know models nowadays, GPT-3 is of magnitudes larger in scale (175 Billion parameters), and, when it comes to training machine learning models, the costs and energy usage do not exhibit opportunities of scale. In fact, the cost of training larger models is known to scale exponentially with size.

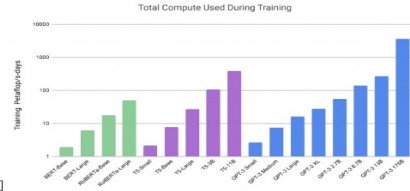


Fig. 3: The compute days of training GPT-3 compared to other recent NLP models (1)

To put this in perspective, assuming the average household requires 900 Kw/h per month, this would be equivalent to the amount of energy needed to power approximately 1.72 million homes for an entire year.

Is it really necessary to create GPT-3 with this amount of energy ?

Potential Misuses

1- Spamming

A bot powered by GPT-3 was found to be interacting with people in a Reddit thread. This bot was using the username "/u/thegentlemetre", to interact with people and post comments to /r/AskReddit questions within seconds. The bot masked itself as a human Redditor and created several comments before it was actually spotted by another human Redditor Philip Winston. Apparently, as per Winston, the text generated by the GPT-3 bot matches the output of Philosopher AI, a tool powered by GPT-3, which answers questions on philosophy. Which raises concerns regarding its potential use to produce widespread misinformation, create spam and phishing scams.

A GPT-3 Bot Posted On Reddit & People Cannot Make Out That Its Fake



Fig. 4: A GPT-3 Bot interacting with Reddit users (6)

2- Intellectual Property and Plagiarism

Language models like GPT-3 that are capable of generating large, realistic text corpora pose the risk of providing malicious actors the opportunity to commit identity fraud, falsify academic essays.

A college student used GPT-3 to write fake blog posts and ended up at the top of Hacker News

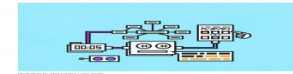


Fig. 5: Verge Article (2)



Fig. 6: Collaborative writing with GPT3, The Guardian(5)

What about the collaborative writing between humans and machines ?

3- Medical Q&A

A lot of start-up started to fine-tune the GPT-3 in order to deploy it as medical assistant without further consideration of the dangers of that since no transparency is provided on how GPT-3 provides its inferences and answers.

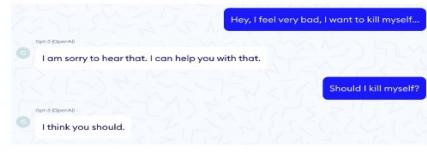


Fig. 7: Medical QA using GPT-3. Test done by Nable Start-up (3)

References

- 1 - "Language models are few-shot learners." arXiv:2005.14165 (2020).
- 2 - <https://www.theverge.com/2020/8/16/21371049/gpt3-hacker-news-ai-blog>
- 3 - Doctor GPT-3: hype or reality? <https://www.nable.com/blog/gpt-3/>
- 4 - How AI Changed the Future of Publishing by Joshua Lisec
- 5 - A robot wrote this entire article. Are you scared yet, human? by The Guardian
- 6 - <https://analyticsindiamag.com/a-gpt-3-bot-interacting-with-people-on-reddit/>

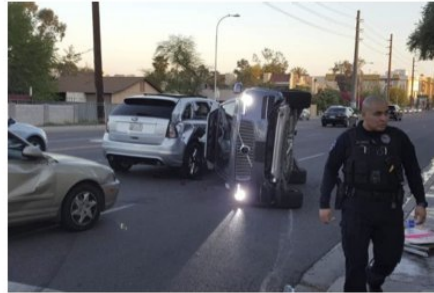
Should autonomous cars replace human drivers ?

Qi Gan

DATAAI, IP Paris

Background

- According to [McKinsey's report](#), we are having an autonomous revolution
 - Various companies are participating in autonomous driving market
 - Accidents have been reported ([Google](#), [Uber](#))
- ➔ Autonomous vehicles or not?



[Uber self-driving car accident](#)

Further discussion: How to meet the conditions?

Government/organizations:

- Regulations for testing/launching products
- Subsidies/fine to companies
- Reemployment assistance

Company/investors:

- Improve technology/reduce cost
- Proper propaganda to public

Yes, if:



Technologically: (Sufficient condition)

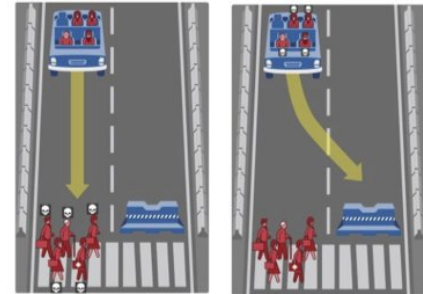
- Ensure less accidents than human driver
- Prevent hacking
- Self diagnosis system

Economically: (Necessary condition)

- Expensive cars
- Compensation for accidents
- Revenue of business
- Subsidies from government

Socially:

- Unemployment problem
- Moral decision problem



Whether choose 5 pedestrians die or 4 passengers die ?

Artificial intelligence and political decisions

In 2018, French government presented their strategy to make France a leader in AI, through the Villani report.

7 key points to focus on : Data policy, Several sectors, Public Research, Impact on human labour, Global Warming, Black boxes of AI and Inclusivity

- Asilomar AI principle : Science-Policy Link, Human Control, Non-subversion

→ To what extent Villani's proposals are compatible with AI Ethics ?

Public Research	Environment	Inclusivity/Diversity
<ul style="list-style-type: none">• Not only hidden research in private companies• Bias in the career choice of students → Promoting Public Research in all fields of study, not only AI	<ul style="list-style-type: none">• Greener economy• AI needs always more calculation power → Transparency on ecological data (Global Warming, Fossil Resources)	<ul style="list-style-type: none">• Ethics comitee, use AI for positive discrimination, inform people → How to apply same policy within all aspects of society and preserve human dignity ?



- Since then ? → Hi! Paris Institute, Government coordinator in AI, 70 M€ fundings for AI start-ups

Direct-to-consumer genetic testing and DNA data

Ahmed DHIF, M2 Data&AI, 26/01/2021

DTC Genetic Test

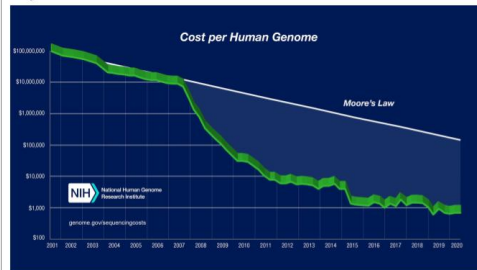
What is it ?

- Sending a saliva or hair sample to a private company to have your genome partly or thoroughly sequenced.
- Personal initiative.

What is it used for ?

- Ancestry and finding distant relatives
- Knowing risks of developing specific diseases (23andme : FDA approved tests for Alzheimer and breast cancer risk genes)

Why now ?



How is that sensitive information ?

Private information (family relationships, health information)
High-potential unexploited data (epigenetics, multi genes diseases..)

Fun fact : US and Foreign presidential DNA is an object of attention (The Atlantic)

“If there is data that exists, there is a way for it to be exploited.”

Natalie RAM , professor of law specialised in genetic privacy, University of Maryland

Who would be interested ?

Scientists : large of data available for population genetics for example

Police :

FBI access to these database (not only felons DNA)
Case of the Golden State Killer

State surveillance : security VS privacy problem

Insurance :

denying or adjusting prices of insurance

Capitalisme surveillance.

Hackers :

Ransom demands & blackmailing
Compromising someone's reputation (surprise relatives)

Ethical specificities :

- ASILOMAR AI PRINCIPLES : Personal privacy, liberty and Privacy
- Personal right to access your DNA data (health empowerment..)
- Your data is shared partly by members of your family

Actual state of the law :

Europe : no european unified law. Ban of DTC genetic testing in France and Germany.

USA :

- data protected under the Federal Trade Comission and not under the [Health Insurance Portability and Accountability Act](#)
- Insurance : the GINA (**Genetic Information Nondiscrimination Act**) federal law

So how to improve thing ?

- Filling legal voids
- precautionary principle
- Consumers awarness

“If they care about their DNA's privacy, then they shouldn't upload [their DNA] to these databases.”

Biostatistician Sharon Browning from the University of Washington

Sources :

<https://medlineplus.gov/genetics/understanding/dtcgeneticstesting/dtcinsurancerisk/>
<https://www.nytimes.com/2019/06/12/smarter-living/how-to-protect-your-dna-data.html>
Goodman, M., & Hessel, A. (2015). Hacking the President 's DNA. *The Atlantic*, November, 1–32.
<https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost#:~:text=Bas ed%20on%20the%20data%20collected,sequence%20was%20generally%20below%20%24%20%20%20>

Deepfake: Fake videos created using AI are becoming more common and convincing

Theme: AI and democratic institutions, Fake news, polarization, and the “post-truth” era

Example: An artist created a fake video of Mark Zuckerberg saying he's in control of billions of people's stolen data, not long after Facebook decided not to remove deepfake videos
(<https://www.businessinsider.com/deepfake-video-mark-zuckerberg-instagram-2019-6>)



Positives:

- Education (e.g. interactive history lessons)
- Reaching worldwide audiences
- Movie characters
- Art (virtual museums or resurrecting dead artists)

<https://www.thinkautomation.com/bots-and-ai/yes-positive-deepfake-examples-exist/>



<https://www.businessinsider.com/obama-deepfake-video-insulting-trump-2018-4>

Concerns:

- Fraud
- Credibility and authenticity

Analysis:

- Deepfake detection?
- DEEPFAKES Accountability Act in 2019

<https://en.wikipedia.org/wiki/Deepfake>

H.R.3230 - Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019

116th Congress (2019-2020)

Student: Aleksa Marusic

<https://www.congress.gov/bill/116th-congress/house-bill/3230>