Practical: Artificial Intelligence (AI)

In Data Science we process a lot data through AI. With the GDPR, it is becoming increasingly important to understand the ethics behind the data that is collected, stored, processed and evaluated.

Your task is to:

• Find out what Responsible AI is?

• Find instances where AI has failed? Or been used maliciously or incorrectly.

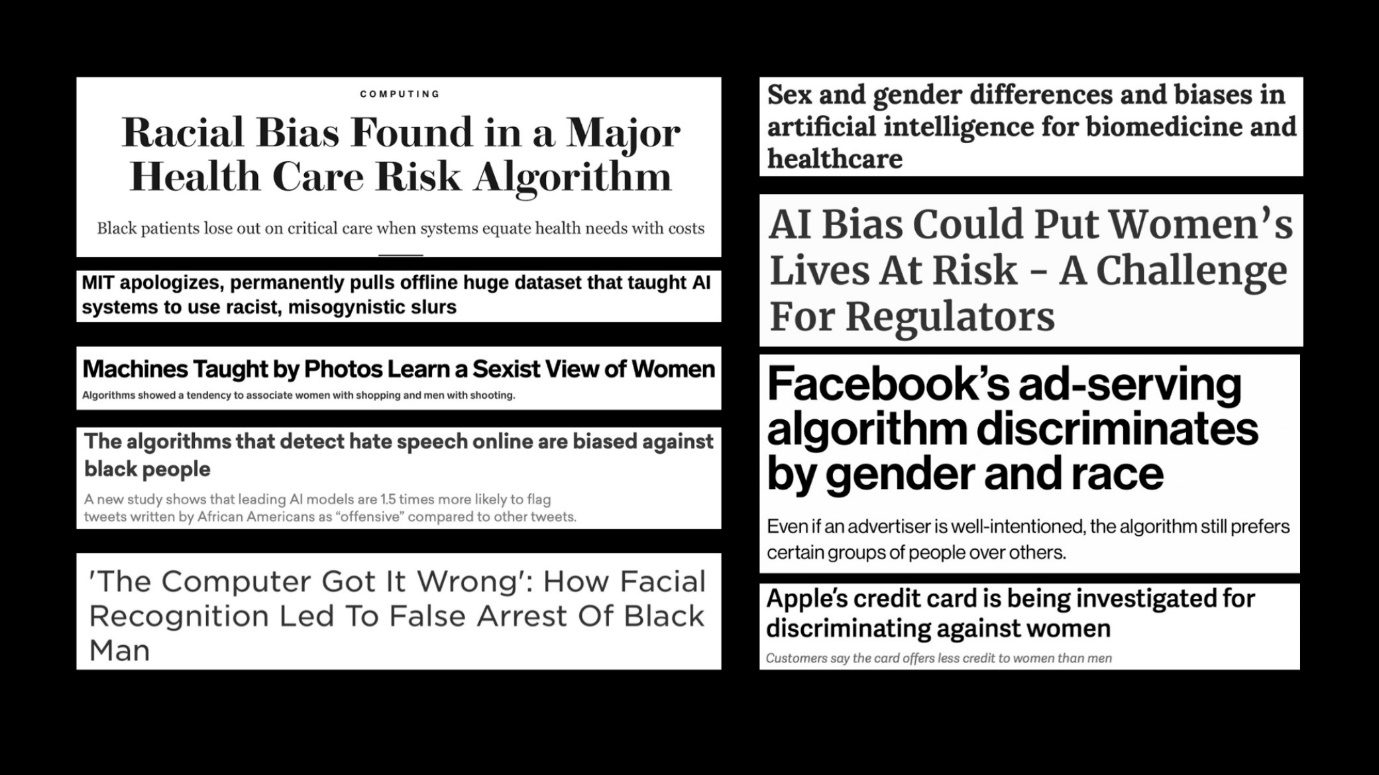
• Implications of when AI fails. There is a specific article in the GDPR Law that covers this, especially with automated decision making. (opt in and out options).

• What should organisations do to ensure that they are being responsible with AI and the wider use of data in general?

The existence of computers assisting most human activities nowadays has led to an increased usage of algorithms for consequential decision making in a growing range of contexts: from offering mortgages and credit cards to sifting job and college applications and sentencing criminals.

*“When technologies co-shape human actions, they give material answers to the ethical question of how to act” – P Verbeek4*

Numerous ethical problems have been reported as a consequence of this increased reliance on AI, from sexist hiring algorithms, racist speech detection algorithms, ageist insurance algorithms to racist and homophobic face detection algorithms. (see pic below for examples)



An explanation as to why something so seemingly objective as algorithms based on mathematical analysis executed by hardware with no emotional components can lead to such ethical violations has been given by Ayanna Howard, Prof. of AI and Ethics:

*“Humans are biased, and therefore the society is biased. And since data is collected from the society, the data is biased. This data is then used to develop AI technologies, which obviously then will be biased as well. Data is nothing but a reflection of our society”*

Not only is the biased data being used to train inherently biased models, but people developing these models also have their very own cognitive biases which can seep into AI from algorithm conception to interpretation of the results.

The consequences when AI fails are significant: deserving people losing job opportunities, access to equal medical care, access to equal resources and unfair legal issues. Bias is a cause of injustice, and it especially affects racial, gender and sexual minorities, and marginalized groups.

Therefore, rightly so, there has been a recent focus on promoting ethical AI arising from the increasing concern over its unintended negative impacts. Many governments and international organizations have released sets of ethical principles, including:

* the OECD Principles in 2019
* the Montreal Declaration in 2017,
* the U.K. House of Lords report “AI in the U.K.: ready willing and able?” in 2018
* the European Commission High-Level Expert Group (HLEG) on AI in 2018
* the Beijing AI Principles in 2019

Recent reports1,2,3 indicate that there are currently more than 70 publicly available sets of ethical principles or frameworks for AI, most of which have been released within the last six years.

Because there are so many different scenarios and use cases it seems impossible to provide ethical principles that will be specific enough to provide answers in practice, and yet broad enough to apply universally. Therefore, in order to ensure the responsible usage of personal data and AI, companies should devise their own answers to ethical dilemmas by employing systematic processes and leaving a record of these processes along the way specifying the values and rationale they employed to make decisions. This record can then provide the public with transparency regarding the rationale for a decision after the fact, as well as give the design team confidence that such a decision was made in a systematic and professional way.

Jobin et al1 reviewed 84 ethical guidelines and proposed 11 principles that characterize ethical AI:

1. transparency
2. justice and fairness
3. nonmaleficence
4. responsibility
5. privacy
6. beneficence
7. freedom and autonomy
8. trust
9. dignity
10. sustainability
11. solidarity

Bibliography

1. A. Jobin, M. Ienca, and E. Vayena, “The global landscape of AI ethics guidelines,” Nat. Mach. Intell., vol. 1, pp. 389–399, Sep. 2019.

2. J. Morley, L. Floridi, L. Kinsey, and A. Elhalal, “From what to how. An overview of Ai ethics tools, methods and research to translate principles into practices,” 2019. [Online]. Available: arXiv:1905.06876.

3. L. Floridi et al., “Ai4People—An ethical framework for a good Ai society: Opportunities, risks, principles, and recommendations,” Minds Mach., vol. 28, no. 4, pp. 689–707, 2018.

4. P. Verbeek, “Materializing morality: Design ethics and technological mediation,” Sci. Technol. Human Values, vol. 31, no. 3, pp. 361–380, 2006.