



Inclusion of AI in Modern Days

17/09/2021

Mikel Garcia

Urtzi Beorlegui

Alfredo Vallejo

Index

1. [Introduction](#)
2. [Machine Manipulation](#)
3. [Machine Ethics](#)
4. [Machine Bias](#)
5. [Conclusión](#)
6. [References](#)

Introduction

Since the 1950s AI has been the holy grail of computer science, for decades the field remained in a state of theoretical possibility but technological unfeasible. But as the 21st century began to settle, some of the technological requirements were met, new and ingenious ways of processing and data were devised, and many other contingencies have led to a resurgence of the field. As a consequence AI systems have been developed and deployed in the real world, as a way of solving old problems and helping in a myriad of new ones.

But as time has passed we've started to realise that the idealistic mentality that envisioned it as a purely positive force for our societies was naive, to say the least. It's implementation in the real world has had a cascade of unintended consequences and collateral damages that weren't taken into account, let alone considered. The problems raised with it's deployment will have to be addressed if we don't want to repeat the same mistakes in the future.

With this work we'll try to give a synthesis of a couple of problems that have plagued the field, namely its potential for control and manipulation and the problem with the biases in the algorithms. But also we want to look into the future and bring some of the challenges that the field will face in the coming years, the creation of ethical agents, capable of understanding morality and ethics and applying them to real world problems in an ethical and safe manner.

Machine Manipulation

I'm sure that you've come across this situation, a friend, family member, maybe yourself, talks about a product or service, and the next thing you realize is that there's an ad on your phone about it. The most conspiratorial automatically jump to the conclusion that their phone is spying on them and that's how it knows what to show them. I wish that it was surveilling us, because that would be illegal and there would be a way to challenge it and make it stop, but that's not the case. What's really happening is way worse and worrying.

Our phone doesn't need to spy our conversations because it already has something better, our data. The trail of breadcrumbs we leave behind us, as we use the internet. Those bits of information by themselves are worthless. It's only when they are collected, classified, aggregated that they start to be useful, and become incredibly powerful tools that can be wielded against us to manipulate us.

Through data mining our data is extracted and used to profile us. Anything can go into that profile, from music tastes, to hobbies, to political preferences and ideology. Once it's in our profile it can be used to determine what we like to buy or who we are likely to vote for, and that information will be used by the platform to make a profit, be it for itself like Amazon recommending us new products related to what we've bought and show some interest. Or be it for thirdparties like Instagram and Facebook that run a massive advertising engine behind the facade of a social media platform , used to provide advertisers with the people most likely to buy their stuff.

We could dismiss it by saying that it's not anything new. Advertisement has always existed and it has always relied in some way of psychological manipulation to make us buy. But the scale and efficiency we've reached in the last few years isn't comparable to anything we've had before. Ads used to be thrown into the world, placed in newspapers, magazines, TV or the radio, with little to no control over who was going to see them. They could choose where they would appear and in what time frame or moment of the day, but nothing else, but they relied exclusively on these parameters.

Nowadays the same techniques that Spotify uses to recommend us new music, based on our tastes and on what other people similar to us listen to, are the same ones being used to show us the products that we're most likely to buy. There's no way to escape it as long as we remain in the system.

As technology improves this trend will only worsen. Sometime ago machines relied on the mathematical models devised by scientists and mathematicians, processing data in some specific way determined by them. But today with the vast quantities of data stored and the advances of Machine Learning, machines themselves have the power to infer new correlations and relationships between data points, most of them unimaginable for



humans. Those new relationships are usually rubbish, because correlation doesn't imply causation, but the lucky few useful for predicting human behaviour are the ones that make the difference. Giving birth to brand new ways of categorization that can be used to influence us in ways that we can't even imagine.

Before I continue, let me be clear about something. I'm not trying to imply that all these developments will be harmful or that technology is nefarious by itself, it's the use given to it that makes the difference. But those who control the technology are the ones that decide its use, and right now the ones who hold most of the power are big corporations, solely focused on increasing their profit for themselves and their shareholders.

So far maybe all this talk about manipulation and advertisement may seem too abstract, needlessly contrarian or ideologically motivated, but maybe with this final example I can show the real dangers that it entails. What if all this data harvesting and classification was put to use for something beyond selling stuff? What if it was used to influence elections? We won't have to put our imagination to work, it's already happened.

Maybe you'll remember the case of [Cambridge Analytica](#), a data mining company founded among others by Steve Bannon (Trump's main political strategist). The company became infamous after the 2016 presidential elections because it used its services to influence it. Enabled by a security breach in Facebook, it took advantage of the data it extracted to develop psychological profiles that then were used to influence their behaviour, through specifically targeted ads. These ads could range from purely informative pieces to straight up disinformation aimed to swing the political allegiance of the advertised, with a precision and efficiency like never seen before.

What's worse in the aftermath of this scandal, the Spanish congress tried to do the same but in an official manner. They passed a law that enabled political parties to do the same thing, elaborate psychological profiles, which could be used during the elections to inform their propagandistic strategies. Same practises, same story, but this time state sanctioned. Luckily for us the [Tribunal Constitucional banned it](#) some time later and it was never used, as far as we know. But as we've seen so far the danger still remains.

The tools I've been talking about hold the potential to revolutionise our lives, but as they're being used right now, they only serve the needs and interests of a powerful minority. Be it selling stuff that we don't need, keeping us engaged in their carefully designed attention traps, or trying to swing our political tastes.

Machine Ethics

We are facing a future, or even living a present where artificial intelligence is indeed so powerful that it could even overcome human intelligence. AI can solve impossible problems in less than seconds; but it is different when it is facing a dilemma, a human being can solve the dilemma with ethical criterias, in fact, it is difficult for an AI to take that type of decisions, this is why it is important to have the AI gifted with ethics.

Machine ethics visualize the machines as subject, rather than for the human use of machines.

machine ethics is concerned with ensuring that the behavior of machines toward human users, and perhaps other machines as well, is ethically acceptable. (Anderson and Anderson 2007: 15)

AI should be able to reasonably take into account some factors, such as: social values, moral and ethical considerations. After taking those factors into account, weigh their priorities in various multicultural contexts and act based on them.

The idea of machines following some “Laws” was famously investigated by Isaac Asimov, who proposed the “Three Laws of Robotics” (Asimov 1942):

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 Laws.

The main question comes on, ¿What behaviour should the AI follow? The algorithm should be able to identify the failure when they are working on social actions with some repercussions. The objective is clear, the code can't hurt people or companies. Due to this preoccupation, the European Parliament made a report in 2019 called “Ethics Guidelines for Trustworthy Artificial Intelligence”, following the publication of the guidelines' first draft in 2018 supervised by 52 experts focusing firstly on the human being under the defense of fundamental rights. Trustworthy Artificial Intelligence should be lawful, ethical and robust, key requirements:

- Human agency and oversight: Empower human beings, allow them to make decisions and foster their human rights.
- Technical Robustness and safety: Accurate, reliable and reproducible, avoiding unintentional harm.

- 
- Privacy and data governance: Data protection.
 - Transparency: The decisions taken should be explained to the stakeholder concerned.
 - Diversity, non-discrimination and fairness: Avoid unfair bias and fostering diversity, being accessible for anyone.
 - Societal and environmental well-being: Sustainable and environmentally friendly.
 - Accountability.

The objective is to build respectful and common frameworks for the future. It is important to agree on how to regularize ethical problems involving AI. As an example, google has its own objectives for AI applications based on the first draft of “Ethics Guidelines for Trustworthy Artificial Intelligence”:

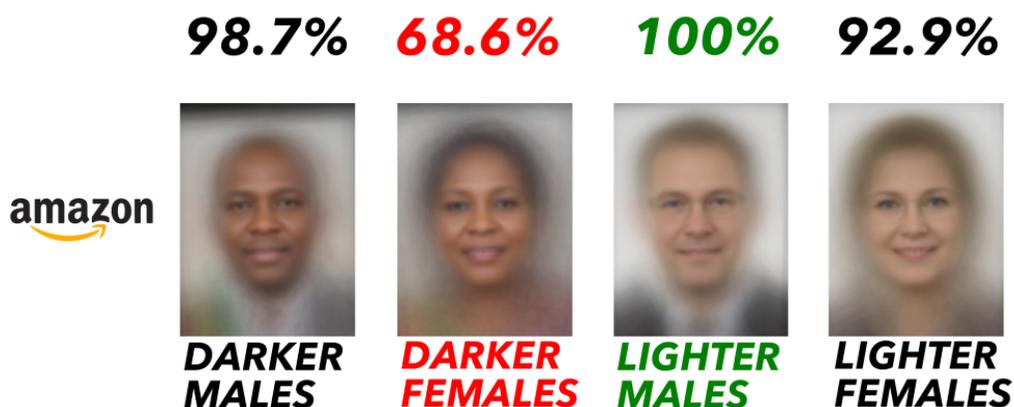
- Be socially beneficial.
- Avoid creating unfair Bias.
- Be built and tested for safety.
- Be accountable to people.
- Be made available for uses that accord with these principles.

Machine Bias

The human being is not perfect, it is a virtue that we possess, since it drives us to improve. But it can't all be pretty, we can observe these imperfections in many areas of life, humans can be selfish and evil. Considering this starting point, how do we intend to build or teach something that it works correctly?

When it comes to creating AI, the most dangerous human behaviour is prejudice. To greater or lesser extent, we all have prejudices, it is inevitable. Prejudices are part of our lives, they help us make decisions that usually are more accurate than mere chance. On the other side, there are stereotypes that severely disadvantage some groups. Such biases leave traces in the data we generate and this can be harmful for AIs that learn from that data.

For example, we have Amazon's AI, a study determined that this artificial intelligence detected faces of white men with 100% accuracy. For black women, on the other hand, this success rate dropped to 68.6%. The difference is quite large. This is because we have many more images of white men than of other people of different ethnicities and genders.





There are many more examples, AIs that linked cooking to women or labelled black men as gorillas. And so on, countless examples. This is a problem that is difficult to solve, because these intelligences learn from data that we generate and, as we have said before, we have prejudices.

This problem does not only affect image detection. It is present in areas much more sensitive to these injustices. IAs are used to decide whether to give loans or grants to applicants. It is true that when a human decides, he or she is also prejudiced. But the sad thing is that when we could eliminate this problem using artificial intelligence to be more objective, what is being achieved is to magnify these biases.

Conclusion

We see that AI has a lot to offer us if we make good use of it. It can work more efficiently than people, and is able to analyse data and create connections that we are not able to process. But like everything else, this depends on the use we make of it.

In order to do so, we must establish rules for the use of this powerful tool. Not only in its use, but also in its training.

If this is not regulated, it can have negative consequences on the daily life of the whole population. Because this has only just begun. AI will continue to develop, and there must be strong institutions to regulate the use of AI in order for this field of information technology to grow in the right direction.

As we have described, we have not yet imagined all the beneficial applications it can have. We just have to wait and see where it goes.

References

Machine Manipulation

["Big Data e Inteligencia Artificial, retos a afrontar"](#) Alfredo Vallejo Martín, Diciembre 2020

Machine Ethics

["Ethics guidelines for trustworthy AI"](#), *European Commission*, 08 April 2019.

["Objectives for AI application"](#), *Sundar Pichai, Google CEO*, 07 June 2018.

["La importancia de la ética en la inteligencia artificial"](#), Miguel Ángel Barrio, *El País*, 26 February 2019

["Ethics of Artificial Intelligence and Robotics"](#), Vincent C. Müller, *Stanford Encyclopedia of Philosophy*, 30 April 2020

Machine Bias

["Response: Racial and Gender bias in Amazon Rekognition"](#), joy Buolamwini, Jan 25, 2019

["bias in AI: Much more than a Data problem"](#), David Pereira, Jul 6, 2020

["Tackling bias in AI \(and in humans\)"](#), Jake Silberg, James Manyika; Jun 6, 2019