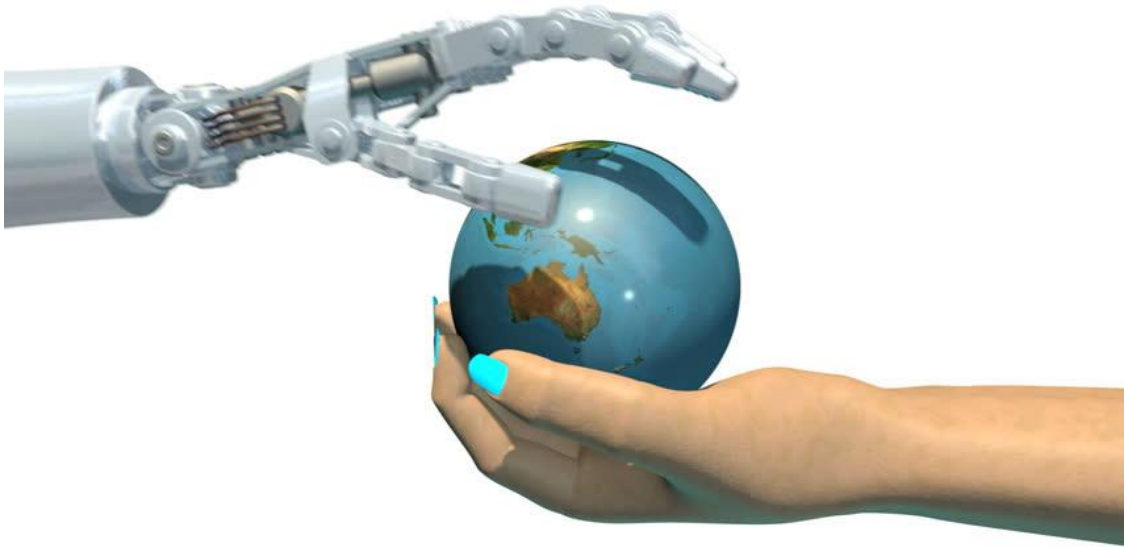


eman ta zabal zazu



Universidad  
del País Vasco

Euskal Herriko  
Unibertsitatea



## **Ethical issues of Artificial Intelligence**

Ander Munarriz Kristupas Mišutis

Advanced Techniques in Artificial Intelligence

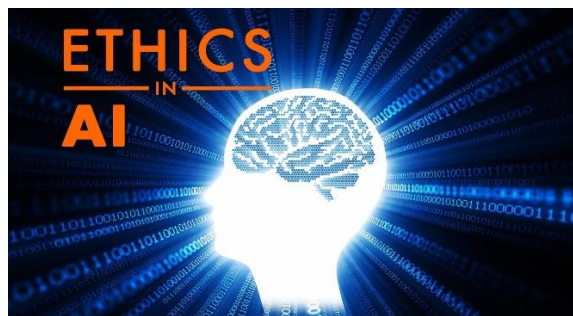
# Contents

1.	Introduction.....	3
2.	Ethics in Artificial Intelligence .....	3
3.	Future challenges in Artificial Intelligence Ethics.....	4
4.	Examples.....	6
5.	Artificial Intelligence companies, groups and countries.....	10
6.	Conclusion .....	12
7.	References.....	13

# 1. Introduction

Artificial Intelligence are integrating in our society extremely rapidly: more and more functions can be done not by people, but robots and it is only a matter of time till they be smarter than humans and actions need to be taken before it happens. People like Bill Gates, Elon Musk, Stephen Hawking and others already expressed their concerns with the AI and said that it can cause danger for humanity. Our society needs to ensure that robots and machines will act ethically and figure out who should provide the morality norms that AI should be aware of. The main goal of this report is:

1. To address what is ethics in AI, what challenges it faces;
2. Ways that AI is used now ethically (or unethically) today in big companies work;
3. Summarize the importance of ethics in Artificial Intelligence.



## 2. Ethics in Artificial Intelligence

The ethics of Artificial Intelligence is part of ethics of technology that specifies on robots and other AI beings. These ethics are usually contained of two in some sense similar types of ethics: Robot Ethics and Machine Ethics:

- Robot Ethics(roboethics) – explores how people should design, deploy and treat robots and how robots can be used either to harm or benefit human race.
- Machine Ethics – considers the prospects for creating computers and robots capable of making moral decisions (behaving morally).

Both disciplines: Robot Ethics and Machine Ethics are emergent and just beginning to take form. In this report Robot and Machine ethics expressed just as Artificial Intelligence ethics.

### 3. Future challenges in Artificial Intelligence Ethics

There are many different possibilities that can happen if the ethics of AI would not be created and formed understandably. One of them is **danger to humanity**: Joseph Weizenbaum argued in 1976 that AI technology should not be used to replace people in positions that require respect and care, such as any of these:

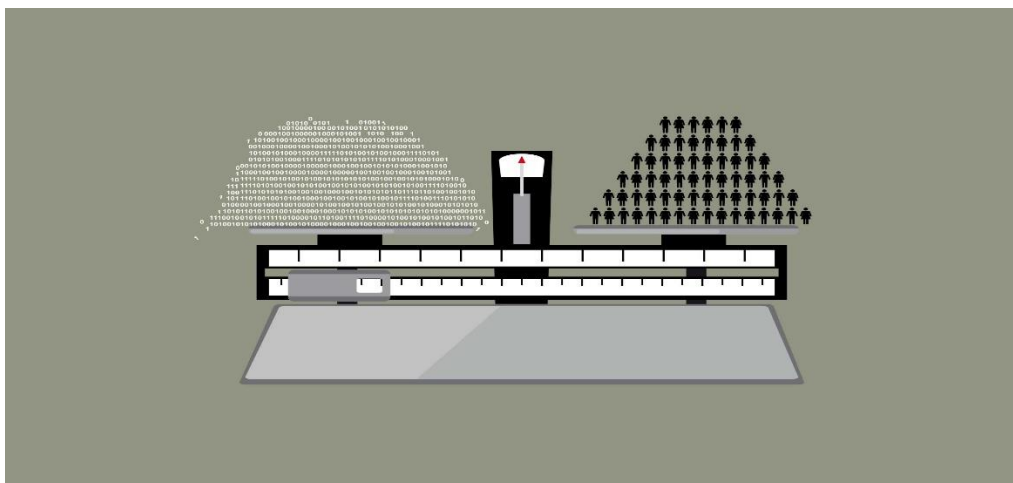
- A customer service representative;
- A therapist;
- A nursemaid for the elderly;
- A soldier;
- A judge;
- A police officer.

These jobs need to be left for human beings, because it requires an authentic feeling of empathy, especially those that will use weapons (soldier, police officer). In 2018 July 18<sup>th</sup> more than 2,400 AI researchers (Demis Hassabis, Elon Musk, Stuart Russel) signed a pledge promising not to build autonomous weapons. The pledge started with sentence:

- **“Artificial intelligence (AI) is poised to play an increasing role in military systems. There is an urgent opportunity and necessity for citizens, policymakers, and leaders to distinguish between acceptable and unacceptable uses of AI.”**

Although there is mutual consent that there cannot be AI that can cause human death there is still a lot of AI used in military weapons: Many governments have begun to fund programs to develop AI weaponry. The United States Navy recently announced plans to develop autonomous drone weapons, paralleling similar announcements by Russia and Korea respectively. The future ethics of AI should make it clear what can and cannot be done in military activities.

Another problem that needs AI ethics involvement: **unemployment**. According PwC international jobs automation study, which analyzed the job market in 29 countries, just 1% of jobs will disappear thanks to automation. Later in a future the job market will feel bigger impact, because AI is growing so fast we can only predict how it will look. Research firm Gartner predicted that by 2020 AI will create 2.3 million jobs and eliminate 1.8 million and by 2022 it can generate over \$4 trillion in business value for and industry. Companies are invested in a profit, so the job for AI ethics is to organize a correct balance between robot and human workers.



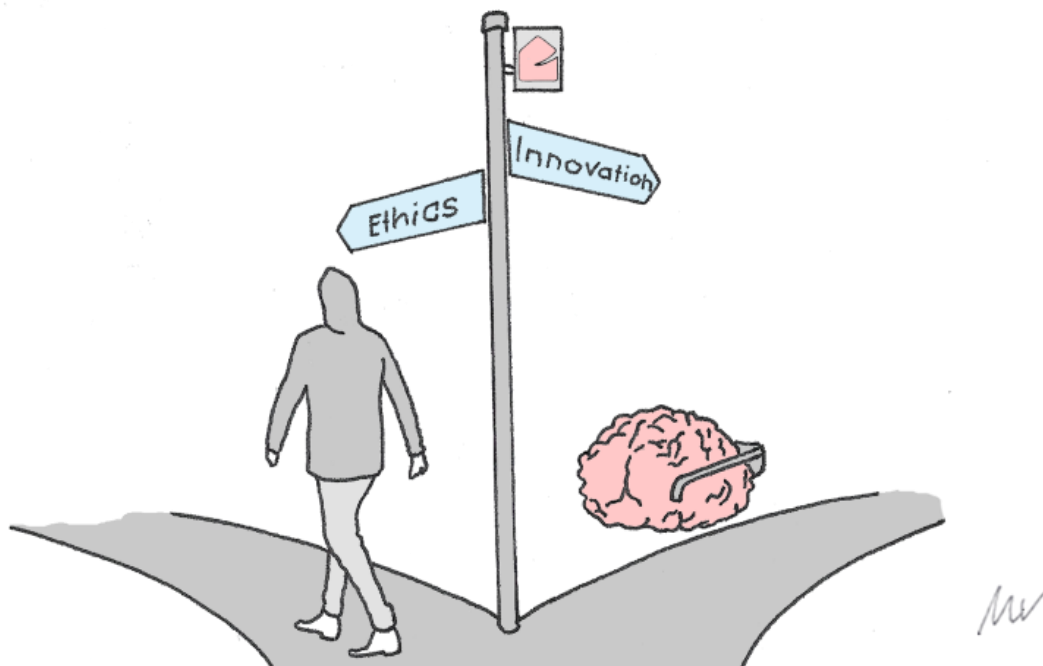
**Robot rights** are important AI topic that needs rules. In 2018 European Commission received an open letter made by more than 150 experts in robotics, AI, law, medical science and ethics that warned EC about how robot rights violate human rights because earlier EC received proposal that requires a special legal status of “electronic persons” for autonomous robots. The experts wrote: “From an ethical and legal perspective, creating a legal personality for a robot is inappropriate. Humankind stands on the threshold of an era when ever more sophisticated robots, [...], seem poised to unleash a new industrial revolution, which is likely to leave no stratum of society untouched, it is vitally important for the legislature to consider all its implications.”. How to do that is one of the hardest tasks for AI ethics to solve.

## 4. Examples

Unemployment, weaponizing AI and robots rights are more abstract problems may not seem important now for people that have nothing to do with it, but ethics of AI also needs be active in internet used by billions of people every day. There are real-life examples that demonstrates the importance of groups of experts to prevent a poor use of artificial intelligence.

For instance, Facebook’s motto is to go and develop fast, and if they make mistakes they will fix them quickly. Due to this reckless company policy, it has recently faced several controversies regarding its data privacy and security policies and fake news, but a major consequential mistake has been Myanmar. A country which refers to the internet as Facebook has been dealing with the worst kind of hate speech when the Rohingya, a mostly Muslim group, had suffered a brutal campaign of ethnic cleansing. In 2017 over 725.000 of the Rohingya people fled to Bangladesh and around 10 thousand people were killed because of the hate speech caused crimes

Even if Facebook did not cause it, it has been a decisive tool for those seeking to spread hate as many reports had shown. Posted by politicians, military leaders and prominent Buddhist monks. Facebook's actions have been sloppy and slow, that also applies for its artificial intelligence, which detects and deletes hateful comments. The AI has been unable to detect any of that in Myanmar because it was not compatible with the different language fonts. The company ignored repeated warnings as far back as 2013. This situation shows that "developing fast, even if we make mistakes" is not a valid one with companies that operate in so many places around the world.



Another company giant with AI problems is YouTube. Its well-known artificial intelligence, the YouTube algorithm, as informed by an academic paper that analyzed the algorithm's deep neural networks, has been described as one of the largest and most sophisticated industrial recommendation systems in existence. Being this huge and complex AI naturally it comes with many controversies.

There are 1.5 billion YouTube users in the world, which is more than the number of households that own televisions. It is the biggest platform and has more diverse content than any show on TV, and unlike TV, users decide which content they want to watch. Or at least that would be the case with informed and experienced users. The reality is much more different.

What they watch is shaped by YouTube's algorithm, which ranks billions of videos to identify 20 "up next" clips that are both relevant to a previous video and most likely, statistically speaking, to keep a person hooked on their screen. Sometimes, these clips, as we will see, show why the recommendations are something to be concerned with. For example, the algorithm has been found to be promoting violent videos for toddlers in YouTube Kids, a dedicated app for children, spreading conspiracy-theory videos and, most importantly, fueling disinformation during the presidential election. Zeynep Tufekci, a widely respected sociologist and technology critic, said that search and recommender algorithms are misinformation engines.

The reasons for this behavior are unknown, as the algorithm is a private property, but a computer programmer who worked on the recommendation system said that the priorities YouTube gives its algorithms are dangerously skewed. They say it distorts reality, its focus is on watch time and ad-revenue for the companies instead of optimizing for what is truthful.





Chaslot, a computer programmer that worked on the algorithm, was concerned with the distortions of content, and how it only shows that which reinforces their existing view of the world. He proposed many changes to suppress fake news but all of them got rejected. After some research, he suggests that YouTube amplifies videos that are divisive, sensational and conspiratorial.

Luciano Floridi, a professor at the University of Oxford's Digital Ethics Lab, says that "algorithms that shape the content we see can have a lot of impact, particularly on people who have not made up their mind. Gentle, implicit, quiet nudging can over time edge us toward choices we might not have otherwise made".

Tufekci states that once this gets normalized, the recommended content will get more and more bizarre and hateful. "The question before us is the ethics of leading people down hateful rabbit holes full of misinformation and lies at scale just because it works to increase the time people spend on the site – and it does work."

But YouTube is not only drowning on alt-right hate speech and violence, it is also targeting a particular demographic of content creators on its platform: LGBTQ+ people. YouTube's AI has been demonetizing, blocking and age-gating anything that could have any word referring to this group.

YouTube has defended itself from these claims saying that their system sometimes makes "mistakes" in understanding context and nuances. "That the algorithm should not filter out content based on gender, race or political viewpoints and promised to fix the system", but no real progress has been made yet and reality stays the same.

As the trans creator Chase Ross said, many videos of him have been restricted and/or deleted completely, many of them with the word trans in it (the algorithm even ran anti-LGBTQ+ ads in his videos), and he is just one of many that have faced the same issues. Some creators uploaded the same video twice but taking the word "trans" or "LGBT" off, and the restriction was gone.

YouTube wrote an email explaining that they do not have trigger words for the algorithm and that the system is bias-free. They said they use machine learning to evaluate content against their advertisers guidelines and ended by encouraging creators to appeal so that the algorithm could learn and get better. However, this practice puts the burden on the channel holder and may damage their channel.

It is troubling to see how YouTube is giving preference to ad revenue coming from homophobic and transphobic companies, divisive videos and conspiracy content instead of supporting diverse creators and providing a platform for marginalized people as they have hypocritically stated before. But, how are other companies (and countries) going to tackle future problems with AI when they come?

## 5. Artificial Intelligence companies, groups and countries

As seen before, AI will disrupt society as never seen before, and companies should be making the first steps on ethical AI. Fortunately, many organizations and governments are aware of the problems, and are creating projects and groups of experts focused on AI ethics.

Google's ethics group, DeepMind, is the most advanced ethical research unit in the world, working in collaboration with experts in the field and independent advisors, as well as people that would be affected by the rise of the AI. However, they are not the only ones concerned with developing ethical AI.

AI-Ethics, a web-page dedicated to foster dialogue between researchers and experts, published a list of several organizations which they consider leaders on the field of AI ethics. From the list of 20, it is interesting to mention three of them.

The first one was an initiative conducted by Reid Hoffman, founder of LinkedIn, Knight Foundation, and Omidyar Network. They have made an investment of \$27 million and are working with the MIT Media Lab and the Berkman Klein Center for Internet & Society at Harvard University. Their main objective is to support AI ethics projects and activities in the connectivity between computer sciences, the social sciences and the humanities, apart from convening and supporting a network of people and institutions working to maximize the benefits of AI.

Also, the global law firm K&L Gates LLP has made a \$10 million contribution at Carnegie Mellon University to establish the K&L Gates Endowment for Ethics and Computational Technologies. This will be used to support the university in advancing research about the ethical and political issues that may arise and will create many new conferences also faculty positions and doctoral fellowships.

Then, the IEEE Standards Association has developed an initiative on ethics of autonomous and intelligent systems, a way to understand and define the basic rules of ethics for the future to come. To achieve that objective, they launched “Ethically Aligned Design”, a crowd-sourced global treatise regarding this field, divided in sections that go from specific aspects of AI development, to broader subjects like economics, law, weapons systems...

Along with companies, some governments have stated their interest in not only increasing development in AI, but also on regulating it.

Germany, for example, is still developing its national AI strategy, but they already have passed a law about the ethics that programmers must follow when creating artificial intelligence. They have also formed an Ethics Commission on Automated Driving and the country’s cabinet has adopted the guidelines, making it the first government in the world to do so. Also, 7 German firms are implementing ethics into their research.

The United Kingdom's AI strategy, called the AI Sector Deal, has a large focus on the development of data trusts, with clearly defined rules and responsibilities, as a promising example of how the public and private sectors could collaborate to advance AI research without sacrificing privacy or security. They also established the new Centre for Data Ethics and Innovation, which will develop methods for measuring and improving trust in data-driven technologies like AI, as well as recommend ways for addressing the novel ethical challenges AI could pose.

On the other hand, India does not have yet an AI ethics organization, but it has an excellent education system that trains highly qualified software engineers and will be a major source of talent for big companies. India can thus play a major positive role in key areas such as AI ethics, bringing traditions of democracy, sensitivity to diversity, and other areas that will be critical to developing ethical AI algorithms and ensuring that AI is beneficial to broad numbers of people.

And how is Spain doing in the field? This same year the Acuilae company has announced ETHYKA, the first ethical module in the country for artificial intelligence on the automotive sector, but they would also like to carry it a bit further and study, for example, the consciousness of machines and the impact it will have on society.

## 6. Conclusion

As seen in this report, AI is already being used in ways that influence our society, and it will play a much bigger role in the future. From social media to military use, as well as economy, there is still many questions about the use we will make of AI and how it will impact our lives. For that, companies must support the creation of groups that could answer those ethical questions, and governments should invest in researching the field too. Time is not on Artificial Intelligence ethics side and it must improve drastically.

## 7. References

[https://en.wikipedia.org/wiki/Ethics\\_of\\_artificial\\_intelligence](https://en.wikipedia.org/wiki/Ethics_of_artificial_intelligence)

<http://peterasaro.org/writing/Asaro%20Wallach,%20Machine%20Ethics%20Intro.pdf>

<https://intelligence.org/files/EthicsofAI.pdf>

<https://techcrunch.com/2018/06/21/the-future-of-ai-relies-on-a-code-of-ethics/>

<https://aiethics.princeton.edu/case-studies/>

<https://ai-ethics.com/2017/08/12/other-ai-ethics-groups/>

<https://www.datainnovation.org/2018/08/event-recap-house-countries-are-preparing-for-the-global-ai-race/>

<http://www.vdu.lt/lt/kaip-uztikrinti-kad-robotai-laikytusi-morales-normu/>

<http://peterasaro.org/writing/Asaro%20Wallach,%20Machine%20Ethics%20Intro.pdf>

<https://futureoflife.org/2018/07/18/ai-companies-researchers-engineers-scientists-entrepreneurs-and-others-sign-pledge-promising-not-to-develop-lethal-autonomous-weapons/?cn-reloaded=1>

<https://www.cnn.com/2018/08/01/five-of-the-scariest-predictions-for-ai.html>

<https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/>

<https://deepmind.com/applied/deepmind-ethics-society/>

<https://www.datainnovation.org/2018/08/event-recap-house-countries-are-preparing-for-the-global-ai-race/>

<http://www.ticpymes.es/tecnologia/noticias/1105169049504/etica-ia-espanola.1.html>

<https://www.knightfoundation.org/press/releases/knight-foundation-omidyar-network-and-linkedin-founder-reid-hoffman-create-27-million-fund-to-research-artificial-intelligence-for-the-public-interest>

<https://standards.ieee.org/industry-connections/ec/autonomous-systems.html>

<https://www.cmu.edu/news/stories/archives/2016/november/endowment.html>