**Are we ready to embrace AI in our daily lives?**

İlhan Tanrıverdi
Data Analytics, Sabancı University

Information Law and Data Ethics

Yasin Beceni

07 September 2021

## Contents

## Introduction

Artificial Intelligence (AI) has been well known and rapidly developing with the increase in processing speeds and storage capacities. We all have seen Boston Dynamics' Atlas, Tesla's autonomous vehicles, Google's Duplex and AlphaGo. These can be considered as the best and most popular applications of AI. Except Tesla, they have no public use yet, but we all use other AI products such as navigation apps, facial recognition, text autocorrect, search and recommendation in shopping or social media, chatbots etc. Maybe you feel like you are surrounded by AI applications, but this is just the beginning. In the future, many AI applications and robots will be employed for the tasks that humans do today. Moreover, all these applications and machines will be able to communicate to each other and act autonomously.

In the following chapters, I will discuss what we fear the most about AI, what is trustworthy AI, what governments should do and what awaits us in the future.

## Fear

Do you like science fiction movies? If you have seen "The Terminator", "The Matrix", "I-Robot" or "Ex-Machina", you might fear AI. An autonomous machine sounds scary, and I think that is the part people fear. The common point of these movies is AI consciousness. All the machines in these movies are conscious and try to destroy humans for a reason that they reach by themselves after thinking. So, our first question should be: "Should AI have consciousness?"

## Consciousness

Consciousness is a complex topic that is discussed a lot for a long time. In simple terms, we can define consciousness as self-awareness. Meissner draws our attention to different areas that AI robots are employed. On one side there are robots which do repetitive work in factories or in our homes. He calls these as slave robots. On the other side there are or will be teacher, doctor, lawyer, or nurse robots. He calls these as full AI robots. Slave robots do not need to have a consciousness, but full AI robots would provide us better understanding while doing their work. From this point of view, AI that has consciousness is beneficial to humans. Conscious robots, on the other hand, are aware of themselves and their environments. Therefore, they might refuse the tasks given to them, they may request equal rights as humans, eventually they may try to take over the world. From this point of view, AI should never have consciousness (Meissner, 2020).

What about the possibility? Can AI have consciousness? There are many studies and investments about creating an AI that is self-learner and self-aware. Both companies and universities are working on it. However, there is no tangible result so far. Knight discusses whether AI can have consciousness or not in his study with a bold and assertive title: "Refuting Strong AI - Why Consciousness Cannot Be Algorithmic". In conclusion he says no, but keeps open minded (Knight, 2019).

Up to this point, we can say that it does not make sense to give consciousness to AI, and even if we try, it does not look possible anyway. Since we eliminated the possibility of Skynet taking over the world, or a war between humans and robots, can we embrace AI now? Or if we will, what criteria should AI meet to be safe and trustworthy?

## Trustworthy AI

While AI researchers have been trying to improve their work and finding more areas to employ AI, governments, organizations, and institutions have been trying to define trustworthy AI and making guidelines for this purpose. For example, European Commission has the "Ethics Guidelines for Trustworthy AI" that is prepared by "High-Level Expert Group on Artificial Intelligence". This guideline is based on a framework of three main components for a trustworthy AI. First, AI should be lawful, this means that any AI application is expected to obey laws and regulations. Second, AI should be ethical, I will mention this later in details. Third, AI should be robust, so that any unintentional harm is avoided.

The guideline then elaborates the framework with the following aspects (High-Level Expert Group on Artificial Intelligence, 2019):

### Fundamental Rights

- *Respect for human dignity*: All humans should be treated with respect as "moral subjects", rather than "objects".
- *Freedom of the individual*: Every individual should have control over his/her life and be able to make decisions about it.
- *Respect for democracy, justice, and the rule of law*: AI systems should not interfere or manipulate democratic processes and respect plurality.
- *Equality, non-discrimination, and solidarity*: AI systems should not be biased or unfair. Minority or vulnerable groups should be respected.

- *Citizens' rights*: AI can provide better governmental services for citizens. But at the same time, citizens' rights should be protected against any violation.

### Ethical Principles

- *Respect for human autonomy*: AI systems should be designed "human-centric", they should support humans and should not manipulate in anyway.
- *Prevention of harm*: AI systems must be safe and secure. They must be reliable and able to keep away malevolent attempts.
- *Fairness*: Humans should not be treated in a prejudiced way based on any differences. Any advantage, profit or cost should be distributed evenly in a fair way.
- *Explicability*: AI systems should operate with reason; this means that the results should be predictable and explainable. Especially in delicate situations, such as a cancer diagnosis or a verdict on a legal case, transparency is crucial.

### Requirements of Trustworthy AI

- *Human agency and oversight*: AI systems should support humans in their daily life or work, human freewill should not be manipulated or obstructed in any way. Therefore, humans should be involved in AI operations or have the control over AI systems.
- *Technical robustness and safety*: First, AI systems should be safe against any possible attacks from malicious actors. Second, in case of an emergency there should be a backup system or manual control interface if possible. Third, accuracy of an AI system is very significant especially in delicate areas that affect human lives. Fourth, consistency is another significant aspect of an AI system. Given the same input in the same conditions, AI systems should produce the same results.
- *Privacy and data governance*: First, any sensitive data may be used against humans, they may be treated unfairly and differently from others. Therefore, data should be protected and used in a controlled way. Second, data could contain biases, errors or inconsistencies that might lead AI systems to create unfair or incorrect results. Thus, data quality and integrity should be maintained. Third, data access, especially access to sensitive data, should be possible under clear protocols.
- *Transparency*: AI systems' design and data sets should be well documented to allow traceability and transparency. Any decision or result created by an AI system should be explainable and should make sense. Also, AI systems should introduce themselves as AI, when interacting with humans.

- *Diversity, non-discrimination, and fairness*: Biased data, that is used to train or operate AI systems, threaten diversity, plurality, and justice. Any bias in data should be carefully detected and removed. Everyone should be able to use these systems, especially people with disabilities. Moreover, stakeholders should be included and consulted in the AI systems' development process.
- *Societal and environmental wellbeing*: AI systems should be environmentally friendly, beneficial to society and human relations.
- *Accountability*: AI systems should be auditable, and their outcomes should be reportable. When a conflict occurs, trade-offs should be known and carefully evaluated regarding ethical principles and fundamental rights. In these situations, corrections should be possible if necessary.

European Commission created a guideline that is quite comprehensive. Main and common concerns are covered in fundamental rights and ethical principles, moreover, requirements for a trustworthy AI are explained in detail. The guideline describes fair, safe, robust, and human – centric AI systems that supports humans and environment (High-Level Expert Group on Artificial Intelligence, 2019).

There are more guidelines created by other governments, organizations, or institutions. In 2019, Hagendorff studied 22 guidelines and compared them from different aspects. He says, in general, %80 of them mention the key points such as accountability, privacy, or fairness (Hagendorff, 2020). It is good to have several guidelines prepared by different organizations. One day, they may be united and finally we may have one global complete guideline that merges all different perspectives.

What about practice? Do developers or researchers take these guidelines into account in their work? Hagendorff mentions a study that 63 software engineering students and 105 professional software developers participated. They were given eleven different ethical scenarios and the result is disappointing. The original result statement is: "No statistically significant difference in the responses for any vignette were found across individuals who did and did not see the code of ethics, either for students or for professionals." (Hagendorff, 2020).

Rességuier and Rodrigues underline the key difference between ethics and regulation. They indicate that AI ethics is not enough to change what developers or researchers do, there

is a vital need for regulations that force them to apply AI ethics in their work (Rességuier & Rodrigues, 2020).

Is ethics enough to create regulations? Hagendorff mentions the lack of technical explanations in AI ethics guidelines. To create concrete regulations, there is a need of concrete definitions. Hagendorff asks a few questions: “What does it mean to implement justice or transparency in AI-systems? What does a “human-centered” AI look like? How can human oversight be obtained?” and adds “the list of questions could easily be continued” (Hagendorff, 2020).

It is obvious that AI ethics should be elaborated with solid definitions and supported with technical explanations. Then, AI ethics should be the basis of regulations with sanctions.

## Dilemma

Another thing to worry about is that what should AI systems do in dilemmas? Before discussion, let me introduce you a famous dilemma from Philippa Foot, “The Trolley Problem”. There is a road which is currently repaired by five men, and you are driving a trolley. You try to stop the trolley as soon as you see the five men, but the brakes fail to stop the trolley. Suddenly, you see an alternative way on the right to escape and save the five men, but there is one man on that road and if you turn right, he will be killed. What do you do? Thomson says, everyone she asked this question said that turning the trolley is “morally permissible”. Now let’s consider another scenario where you are a doctor and have five patients waiting for organ transplant. Two patients need one lung each, other two need one kidney each and the last one needs a hearth. According to scenario, they must get the organs today, otherwise they will die, and the time is running out. Today, you get a report about a patient who visited the clinic for his yearly check-up. The report shows that he has the right blood type, and he is very healthy. You can operate the healthy man, get his organs, and transplant them to other five patients to save them. What do you do? Thomson says, everyone she asked this question said “Sorry. I deeply sympathize, but no”. Now here is the dilemma, why trolley driver may turn his trolley, but the doctor may not transfer the healthy man’s organs to other patients? In either case, one dies to save other five (Thomson, 1985). I will not go into more details here, but I wanted you to see how hard the questions could be.

There are newer and harder versions of trolley problem. For example, if on one road there is a kid and on the other road there is an elderly person, or one is a woman and the other is a man, or one is a human and the other is an animal etc. In 2018, The Moral Machine

experiment was conducted in more than 200 countries and moral preferences were measured by a multi-dimensional design. 39.61 million people participated in the experiment. People were asked to choose whom to save and whom to sacrifice in 13 different scenarios like mentioned above. The study shows that answers change due to cultural differences. For example, in some cultures people choose to save the elderly over kids (Awad et al., 2018).

What will developers do in dilemma situations? Can we foresee every scenario? Will the rulesets change from culture to culture? Awad et. al. states that: "Never in the history of humanity have we allowed a machine to autonomously decide who should live and who should die, in a fraction of a second, without real-time supervision. We are going to cross that bridge any time now, and it will not happen in a distant theatre of military operations; it will happen in that most mundane aspect of our lives, everyday transportation" (Awad et al., 2018). Therefore, we need to find answers to these questions as soon as possible.

## Conclusion

Artificial Intelligence is developing very fast, everyday new developers and researchers are joining the workforce, and new companies are coming up with new products and services. On the other side, organizations are trying to understand AI and preparing guidelines that define what is right and what is wrong. Guidelines draw frameworks about ethical principles, but they do not contain concrete definitions and technical descriptions. Therefore, it becomes a challenge to prepare regulations with sanctions based on these guidelines. Since, there are no coercive regulations, we see that developers or researchers do not consider these ethical guidelines in their work. Moreover, in addition to fundamental topics, there are tough dilemmas that are not much discussed and not found place in guidelines yet. In near future, there will be more AI applications, services, and robots around us, even in our homes. There will be much more violations and conflicts than today.

In conclusion, we need coercive regulations for developers and companies. Developers should be made aware of the dangers and apply ethical principles in their work. Dilemma scenarios should be discussed and regulated before there are too many autonomous machines around us.

## References


Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The Moral Machine experiment. *Nature*, *563*(7729), 59–64. https://doi.org/10.1038/s41586-018-0637-6

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, *30*(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8

High-Level Expert Group on Artificial Intelligence. (2019). *Ethics guidelines for trustworthy AI*. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Knight, A. (2019). *Refuting Strong AI: Why Consciousness Cannot Be Algorithmic*. 21. https://arxiv.org/abs/1906.10177

Meissner, G. (2020). Artificial intelligence: Consciousness and conscience. *AI & SOCIETY*, *35*(1), 225–235. https://doi.org/10.1007/s00146-019-00880-4

Rességuier, A., & Rodrigues, R. (2020). *AI ethics should not remain toothless!* A call to bring back the teeth of ethics. *Big Data & Society*, *7*(2), 205395172094254. https://doi.org/10.1177/2053951720942541

Thomson, J. J. (1985). *The Trolley Problem*. *94*, 22. https://doi.org/10.2307/796133