

Challenge 8 – trust me I’m an AI

Mohammad Ali Agharazi Dormani - 19205451

Science fiction films like 'the Terminator' and 'I, Robot' have shown what might occur if artificial intelligence gets rogue. Many negative perspectives on AI are expressed by the Hollywood. Even some experts have some concerns. These experts think that super-intelligent AI systems can pose a considerable danger to mankind in the future. Considering the unknown possibility of AI and exponential research in this field, that may soon turn into the reality. Developers want to see common concerns at the development of AI systems. There have been some reported cases where developers ignored these warnings and made AI systems that ran rogue. For example, I heard in the news that developers made two AIs that were supposed to communicate together like humans, but they started sending unreadable messages. Such AI systems, if applied for serious applications, may pose a big danger to mankind. Thus, tech giants and developers must examine several issues with AI safe with AI and focus on establishing reliable AI systems. But corporation's only concern is money, so the governments have to regulate and monitor these AI systems. This is not easy task because it requires a lot of time and resources to do so. That is why the EU is passing some regulations and expects all its members to follow it. one of these Laws is HLEG(High-Level Expert Group on Artificial Intelligence) which will be implemented by 2020.

This part of the guideline is trying to make future and possibly present AIs as inclusive and as fair as possible. The first example that I thought of was an AI for natural language processing that can understand most accents and all languages. Most AIs can't understand Chinese, Persian or a thick Scottish accent. In addition to that, most languages in the world don't have and AI as advanced as English which has the most developed AIs. It is probably because of excessive amounts of data that are processed in English. This may be due to the fact that most of the western hemisphere speak English. I think this guideline is trying to tackle this unfair and English-centered practice and asks developers to include other languages and their accents.

Another issue that may rise is that people with disabilities may not be able to communicate to AI. For instance, mute people should have an AI dedicated to understanding sign language. Another issue that has to be solved is people that stutter cannot communicate with most major AIs like Siri, Google and Alexa. In the future we might be able to develop an AI that can understand neural signals or brain waves and allow mute people to speak like everyone else. Similarly these AIs can send neural signals to the brain and enable deaf or blind people to see and hear just like others. This will surely revolutionize the future and make disabilities irrelevant which means that AI would allow humanity to thrive and overcome its inabilities and will make a better ,fairer environment for us all. Although the ethics of these AI would also be questionable. For instance, we can never be fully sure if a person with

speaking difficulty has actually said what he meant or the AI had mistaken, which can cause a series of conflicts and problems for the person.

But even today we have many AIs that have come to our assistance when it comes to communication with disabled people. Disabled people that do not have complete control over their hands may make a lot of spelling mistakes. Most smartphone and computers can easily process language and give some suggestions for the typos that were made. In addition to that, some streaming platforms, like Youtube's automatic captioning, have come to assist people with hearing problems. Both these tools can also be used when learning another language which means the societies would be more racially diverse. Another great achievement that we have reached are voice assistants in mobile phone which allow blind people to easily use their phones.

According to this guideline the developer must use all the tools in its power to include people of any ethnicity, gender or disabilities. For this, the developer must expand and use a better data set to avoid any kind of bias. Currently we have achieved the first step in this process, by using females' voices in the data set. Developers are now trying to add accents to their natural language processing. For example, Siri now has South African, Irish, British and Australian accents, but it still doesn't have many major languages from Middle east and Asia like Arabic, Hindi, Urdu, Malay and Persian.

Another issue that may rise is unintended bias. For example, if a developer was developing an AI that predicts the future of the economy of the country and the data set collected was only from richer parts of the country like big cities and etc. the AI will predict wrong results and may ultimately result in the downfall of the economy or it might create a bigger gap between the elite class and the workers.

The guideline states the following in this matter: *In order to develop AI systems that are trustworthy, it is advisable to consult stakeholders who may directly or indirectly be affected by the system throughout its life cycle.*

This means that the AI has to have participant and/or representatives from all group of people that are going to be affected by it not just the people that hires the stakeholder. the guideline also suggests to developers to receive and obtain as much feedback as possible after the deployment of the AI.

Overall, I believe unlike the depiction in Hollywood movies AI are not dangerous and with correct regulations that European commission has implemented AI would create a better ,fairer and less discriminatory future where humans of all ethnicities and with any native language or any ability can communicate together.