Zewail City of Science and Technology

Machine learning course

Instructor: Mohamed Elshenawy

Ahmed Adel 201901464

Omar El-Sakka 201900773

Abdulla Sabry 201701484

University of Science and Technology January,

Building an Artificial intelligence (AI) system was not releasing any concerns from the majority of people until it started to intermediate in every aspect of our lives; especially when it became responsible for recruiting people and being responsible for making decisions in governments that shall affect a whole population, consequently that led many companies to put many efforts in making rules and regulations that should control AI systems.

Building an AI system is always related to many risks that should be thought of and addressed first. According to Mike Thomasa, the first immediate risk is job automation, and to what degree machines will substitute humans. The second most intriguing concern, as AI learns from human behavior, it's very likely that it can learn to be biased to a certain gender or certain color. And above all of them, it's really concerning when thinking about AI for weapons, for example, drones, is it possible that the AI system will decide to make a more huge mess than specified? And to what degree it will cause inequality between countries in wars? These questions led many big companies to have second thoughts while developing AI systems, Like Google, Microsoft, and many others.

There are many ethical frameworks that have been developed to state some rules, which should be deployed in every institution that funds AI research. For example; Universal Guidelines for Artificial Intelligence, which have been

developed in 2018, has 12 guidelines. Most importantly they ensure the right to transparency, as all individuals have the right to know the basics of the AI system. And define the right of humans for the final determination. And it states that institutions must make sure that AI systems don't reflect bias. And AI systems cant be deployed to society, only after adequate assessment of its purports and considering public safety. And most importantly, humans must be able to terminate the system if it's out of control. All of these guidelines are made specifically to control the AI system and make it under human control.

Many companies like Microsoft have spent a great deal of time putting some guidelines for their AI systems. They always state that they are committed to developing AI that puts people first. They have six principles, first of them is fairness, as they ensure to make an AI system that treats all people equally, as it's very important in criminal justice for example. The second one is Reliability & Safety; as AI systems must be extremely reliable while assessing human health, the wrong decisions might lead to catastrophes, and physical systems like self-driving cars. The third principle is that AI should respect the privacy of humans. And it adds that it should mainly empower humanity, and respect diversity. And to make that possible systems must be transparent and understandable for all individuals. And these principles will allow us to account

for AI systems. Moreover, they try to achieve these principles by establishing strategies and empowering other organizations as well.

Google as well is trying to set important rules to run their AI systems. For example, fairness and Interpretability to be able to thrush AI systems. In addition to respecting privacy and security.

In the end, I think it's very comforting thinking that big companies like Google and Microsoft are taking responsible AI problems seriously, And trying to address new challenges that are concerning people. I strongly believe that we should have more control over AI systems and regularize them, and let the details and basics of AI systems be publicly available to enable much more investigation to stop any unanticipated problems in the future.

References

1. **Thomas, M. (2021, July 28).** ***7 Dangerous Risks of Artificial Intelligence*****. Built In.** https://builtin.com/artificial-intelligence/risks-of-artificial-intelligence
2. ***Benefits & Risks of Artificial Intelligence*****. (2021, November 29). Future of Life Institute.** https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/
3. **Microsoft. (2018).** ***Responsible AI principles from*****.** https://www.microsoft.com/en-us/ai/responsible-ai
4. **Google AI. (2017).** ***Responsible AI practices –*****.** https://ai.google/responsibilities/responsible-ai-practices/