Learning Journal 2

Mohammad Ali Agharazi Dormani – 19205451

Artificial intelligence is a part of our everyday lives today, and it poses an unprecedented combination of potential good or potential damage. We are all well acquitted with Hollywood depiction of future AI where the robots will try to wipe out humanity. on the other hand, we have Artificial Intelligences that we use to help us in our daily tasks even in this learning journal I used artificial intelligence countless times to error check my grammar and spelling. Even simple things like Youtube's auto-caption or Google translate is Al-driven. Those systems helped me a lot with learning English. I am not sure if I could have been able get to this level or proficiency in English if it wasn't for Al. so as a matter of fact I am a huge fan of AI, but I still believe that Artificial intelligence has to be controlled in some sense and I think European Union's guidelines for trust-worthy AI is on the right path.

The purpose of this set of guidelines is to help Europe member countries and their tech companies steer the way toward moral and comprehensive AI that is fair transparent and under humans' direct control. The EU is not the first ruling substance in this world to put out guidelines for the moral growth of artificial intelligence, but it had one of most significant ones. On the other hand, China and Russia have a different strategy. They are solely concerned with the growth of AI in any aspect possible. The US has taken a similar approach to the EU, but it is much slower.

The first guideline is "Human agency and oversight". This is the first and maybe the most important requirement for future AIs. When built, carefully and deployed with the right human oversight, AI has the potential to do significant greater good for the world than harm. However ,as depicted in the movie terminator, without this oversight AI can go rouge and may even try to destroy us ,its creators, or it may be like Chappie (2015) which was good at some parts of the movie and evil when it was getting input from criminals but because the creator had initially taught Chappie not to kill, Chappie chose the right path at the end. To me both scenarios could be possible, but the most realistic scenario was from Isaac Asimov many sci-fi depictions of robots. Robots are able to understand anything and do anything unless it will break one of the three laws that humans have embedded in them which he called "Laws of Robotic". They are the following :

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.

3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

And an additional zeroth law which was introduced in one of the last books in the series called Foundation and Earth.

0. A robot may not injure humanity, or, by inaction, allow humanity to come to harm.

I think this nicely resembles what HLEG intended.

Another issue that the developers have to prevent, according to the guideline, is the cyber-attacks carried out on these artificial intelligences, which can be devastating. These attacks may result in loss of colossal amounts of data or even worse, they can turn the AI rogue. Another issue that my rise with these kinds of attacks is if the attacker may manipulate the data and ultimately cause harm. For example, if an AI is used in a weather forecasts and the data is manipulated, and the AI doesn't predict a hurricane or etc. there may be major casualties.