CON101: Ethics in AI Systems

Apar Ahuja
2019CS10465

October 2020

All scientific development in human history has been a double edged sword. All advancement has increased quality and longevity of human life and at the same time created the resources to destroy the earth many times over. Such ethical issues crop up when we look at the future of Artificial Intelligence also.

Human Ethics is a highly complex code defining right or wrong over the ages but when it comes to AI this code may get compromised by the creator of the AI Program. The developer of AI defines and Teaches the AI system what is acceptable/right and should be preserved and what is unacceptable/wrong and should be destroyed. An armed drone can be programmed to kill a terrorist or a soldier or an unarmed baby. AI systems may turn racist or biased by profiling human beings selectively. They may be fooled by fake data to make errors and irrational conclusions. All such backdoors and vulnerabilities need to be addressed to prevent unintended consequences.

Modern AI systems such as the GPT-3 can write articles, computer code and answer questions based on a reference. The risks include fake news articles, misinformation, spams and fraudulent academic writing. Imagine, students putting up their assignments to an AI system and making their way through college. As a matter of fact with the current AI technologies available, one might not even be able to detect if this assignment was written by me or GPT-3!

Finally, the biggest ethical issue in creating AI systems will be in fixing the responsibility of errors, accidents and unintended consequences. Who is responsible for an accident done by a driverless vehicle? AI systems might divide the society into two; those who gain by the advent of advanced AI systems replacing costlier and less efficient human labour and those who lose their jobs to AI systems making them unemployed. AI systems replacing human beings will be more impersonal and would be less empathetic to their client's needs and problems.

AI systems might take some time to catch up to our expectations, they still just understand the rules of the language (task) and correlations, not the context itself. But with the present rate of development, our systems will get better at even more complex problems faster than one can imagine.