Ethics of Artificial Intelligence

Introduction to Computing

Taimoor ul islam (14031)

Date: 10/2/2022

ETHICS OF ARTIFICIAL INTELLIGENCE

# SUMMARY

The possibility of creating thinking machines host to tons of ethical issues. The first section discuss issue in near future of AI. The second section outlines the challenges in ensuring that AI will operate safely as it approaches to human intelligence.

## ISSUES IN NEAR FUTURE OF AI

**Transparency:** If the algorithm is based on complex neural network than it is completely impossible for the programmers to know why or even how algorithm taking any specific decision but if it is based on the decision trees then they could find the reason behind behavior of algorithm by inspecting its code. It is important to develop AI algorithms that are not only powerful but transparent for inspection in case of any misuse.

**Predictability:** Transparency is not the only desirable feature of AI. It is also important that decisions of AI algorithm be predictable but it is not the case. The developers of the AI algorithms only knows that the algorithm will reach to its final goal but they doesn’t know the step by step plan through which the algorithm will reach to its final goal. The algorithm will compute every possible step-by-step plan available to it and then compare the compare the consequences of each plan and execute the best plan according to its experience to reach its final goal. Hence decision of AI algorithms are unpredictable.

**Robust against manipulation:** An AI based machine vision system to scan airline luggage for bombs must be robust against human adversaries who are finding flaws in algorithm for example placing a specific shape near pistol in luggage which will neutralize that effect and the system won’t detect it. Robustness against human manipulation is the ordinary criterion in information security.

**Responsibility:** When AI system will perform cognitive task previously performed by humans who will be blamed in case it fails at its assigned task. The programmer or the end user? Identifying a person who should be directly responsible for any problem is major social criterion.

## CHALLANCES FOR ENSURING SAFE OPERATION OF AI

**Predictability of AI:** The behavior of an AI is not predictable to its developer that is because a final goal is set to the algorithm to where it has to reach at the end and there are multiple steps that algorithm can perform to reach that final point. The algorithm will calculate the number of steps or paths available to it to reach that final goal. It will then compute the possible distant consequences of each step it can take and then compare that results that which result is near to its main goal. Now the path it will chose depend upon the algorithm.

**Predictability of AGI:** It is easy to envision the kind of safety we need in Artificial narrow intelligence although we cannot predict the local behavior of AI but AI is working only in a restricted domain that’s why it is easy to understand what kind of safety we need. But handling an AGI is a different class of problem because in this case we cannot envision the kind of safety we need. Think an AGI can foresee the distant consequences of thousands of different actions across thousands of different domains and could adapt and learn operating in a new domain faster than human which ever no human may even encounter. Here there is no compact specifications of good behavior for domains that human never encounter.

## CONCLUSION

The Conclusion is that although AI technologies are serving humans and beneficial to mankind in many aspects but we should also consider the issues of hand-overing cognitive work to AI in social dimensions. We should ensure the safety of an AGI during its development.