

Fundamentals of Artificial Intelligence Assignment

1- Ethical issues of AI

Craig Heptinstall Crh13- 110005643

SEM6120

Institute of Computer Science

Aberystwyth University

Abstract—The ethics of Artificial intelligence is an ever increasing debate amongst scientists and those that try to enforce new laws (e.g. privacy) or rights, with machines becoming more and more involved with the day-to-day practises and activities of people around the world. Though true AI (there is debate about that too) is something that may take many years to conceive, the ethics will always be an issue that will need addressing. So what is AI? Why are ethics important? What ethics are already considered today?

I. INTRODUCTION

Although this paper looks generally at ethics in Artificial intelligence, to understand the topic in more depth, ethics can be broken into two questions that can be asked for both humans and machines alike. Can the subject (person or machine) be ethical? And can the subject be effected by the ethics of others? The first of these questions looks at the responsibilities of AI, and whether AI should be used for such purposes as weaponry, or health care. The second question looks at AI rights, laws that could one day effect how they work, and if AI can even be considered to be under the same social circumstances as a person. This paper looks at the general area of AI ethics, and uses sources from a number of journals and news articles to discuss current (less than five years) research, and provides a critique of each.

A. What is Artificial Intelligence?

Before answering questions about the ethics of AI, it is important to define to a certain extent what the term Artificial intelligence means. To break this further, understanding the word intelligence can help define the greater meaning. Alan Turing [1] breaks intelligence into five major components, all of which should be fulfilled in order to be classed as intelligent:

- 1) Learning- The simplest form of this is trial and error, with more complicated forms such as generalisation meaning the learner can perform better in situations not encountered before.
- 2) Reasoning- Using evidence from a set of given statements to deduct a conclusion.
- 3) Problem solving- Special and general-purpose methods exists, where the special means a method of solving the problem is tailor made, while the latter means the method can be applied to a larger pool of general problems.

4) Perception- To be able to process and analyse scenes into objects, features and relationships.

5) Language- To use a system of signs, or sounds to communicate or send information to others.

Knowing the general requirements of intelligence, Artificial Intelligence should allow machines perform operations or actions that require the intelligence listed above in humans.

B. Why Ethics are important

In a world where humans are becoming more and more dependent on machines, the need for AI is exponentially increasing. Due to this circumstance, ensuring that any machines or computers used that are safety critical, are depended on (e.g. healthcare), or cases where sensible decisions are required should be ethically correct. This relates back to the reasoning component of the AI definition. One example from the Cambridge Handbook of AI [2], could be that in the near future where for instance a machine that decides on mortgage decisions is found to be being unfairly handing out successful applications based on discriminative terms, who is to blame? Why is the machine coming to such results? It could be said that as long as humans are imperfect, so machines will be. Alongside these kind of ethics, there is also the other side to be considered, the ethical rights of the machines themselves. Though this topic requires more imagination about the future, machines that have feelings should still be considered. For instance, if a machine was working in bad conditions, or being treated badly, should they have rights?

C. AI ethics being considered today, and tomorrow

Although laws and rules have been around for misuses of technology for some time now, no concrete laws regarding the use of Artificial Intelligence are under use at this time, meaning uses of AI for weaponry or wrongdoing is not currently monitored. Because at this time AI is not fully developed, and may have some time to go, creating laws that manage unethical behaviour might not fit well with future and more advanced AI. Aside from laws which are maintained by the courts, AI and machines used for public purposes are still judged by society, and general census around misusing AI can affect how they are used. Drones for example, which are now becoming more and more popular with the general public have meant they are used for many uses, from capturing images to performing deliveries. There are however circumstances where

privacy is breached, or accidents from using autonomous drones have taken place. The use of Artificially Intelligent drones will be looked at in the following section of this paper. Alongside current examples where ethics have had to be taken into careful consideration, fictional ethics and rules have been created, which have then become a part of current research. The Author and professor Isaac Asimov created three basic laws of robotics (made famous by the fictional story of I, Robot) in [4] 1942 which were:

- 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 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 Law.

Although fictional, the laws have been deeply discussed in current AI advances, such as in an edition of the IEEE Intelligent Systems [5], where the laws were modified to try reflect today's current AI. Relating this back to current and future uses of intelligent machines, the three laws are not always a necessity to engineers, for instance military uses of autonomous vehicles designed to kill enemies would not have such laws built into their designs. Creating a standard applicable set of rules for future AI could be integral to ensuring that future technologies are morally and ethically correct.

II. RECENT RESEARCH INTO ETHICS OF AI

Because ethics is such a broad topic, a great amount of research is already available. This part of the paper considers three papers from the past five years, looking at ethical uses of AI, through to considering the ethics of social and emotional AI. With each, a discussion of the research undertaken and findings will be referenced.

A. *The Ethics of Artificial Intelligence*

A paper already mentioned briefly earlier, this piece of research from the Cambridge Handbook of Artificial intelligence [2] looks at the ethical use of AI for a number of scenarios, discussing moral status of machines, the transparency of AI workings, and the ideas of superintelligence, referencing various examples of AI in good and bad circumstances. As discussed earlier, if AI is to be used more often for day-to-day activities, the ethical decisions taken by the intelligence must be justified by the engineers, creators of the intelligence, and the intelligence itself. The paper spoken about here starts by mentioning

B. *Emotion, Artificial Intelligence, and Ethics*

C. *Considering Social and Emotional Artificial Intelligence*

III. OVERALL FINDINGS AND CONCLUSION

REFERENCES

- [1] What is AI?, B.J. Copeland, 2000.
www.alanturing.net/turing_archive/pages/referencearticles/whatisai.html
- [2] Nick Bostrom and Eliezer Yudkowsky. *The Ethics of Artificial Intelligence*. Cambridge Handbook of Artificial Intelligence, 2011.
- [3] Robots Are People Too, J.F. Weaver 2014.
www.slate.com/articles/technology/future_tense/2014/07/ai_drones_ethics_and_laws_if_corporations_are_people_so_are_robots.single.html
- [4] Isaac Asimov's "Three Laws of Robotics", I. Asimov 1942.
<http://www.auburn.edu/~vestmon/robotics.html>
- [5] Beyond Asimov: The three laws of responsible robotics, R. Murphy, D. Woods, 2009.
<http://www.computer.org/csdl/mags/ex/2009/04/mex2009040014-abs.html>
- [6] Kevin LaGrandeur. 2010 *Emotion, Artificial Intelligence, and Ethics*. New York Institute of Technology.
- [7] M. Schroeder, G. McKeown. 2010 *Considering Social and Emotional Artificial Intelligence*.