

## Professional Computing Ethics Essay 2017

### Limits of AI Systems

Since "intelligent" systems are always very complex, and since we are not able to prove program correctness for any but very simple programs, we can never be sure that artificial intelligence systems (including self-driving cars) have no errors in them. Even if we could be sure our programs are correct, we can never be sure that we have not overlooked some combination of circumstances, such as has happened in "flash crashes" in the stock exchange resulting from "black box trading".

AI systems are not only already complex, but when combined with "deep learning" systems and self-improving software, we are faced with systems that can become quite incomprehensible to humans, and we can never be absolutely sure that they will not at some stage exhibit "rogue behaviour".

This has led computer scientists to be divided on whether we should continue developing such "intelligent" systems, some thinking because of the dangers that it is unethical to do so (e.g. Hawking <http://www.independent.co.uk/life-style/gadgets-and-tech/news/stephen-hawking-artificial-intelligence-could-wipe-out-humanity-when-it-gets-too-0.html>); while others believe that it would be unethical not to do so in view of the potential benefits to humanity, and that we can build in sufficient safeguards (e.g. Kurzweil, <http://www.kurzweilai.net/dont-fear-artificial-intelligence-by-ray-kurzweil>).

What are the ethical issues here? Identify the different cases put forward, analysing them in the light of the ACS Code of Ethics, and offering your opinion (with evidence) as to which approach you believe is "most ethical".

### Some Resources

A simple, popular introduction (by Kuipers):

<https://theconversation.com/beyond-asimov-how-to-plan-for-ethical-robots-59725>

A more scholarly introduction (also by Kuipers) is:

<https://web.eecs.umich.edu/~kuipers/papers/Kuipers-sss-16.pdf>

AAAI Spring Symposium on Ethical and Moral Considerations in Non-Human Agents 2016 (EM-CAI 2016).

### Some points to note for your essay

- Your essay should be 1500-2000 words (excluding references). Essays longer than 2000 words (excluding references) will not be marked. Please report the number of words on the front cover.
- For readability please typeset the text and use a serified font, e.g. Times rather than Arial.
- Use 1.5 line or double spacing (in MS Word: Format → Paragraph → Line Spacing). Margins should be at least 2cm.
- Use Harvard format for your references. For assistance with this see: <http://guides.is.uwa.edu.au/harvard>
- Don't forget to put your name and student number on the front.
- Please understand that if you refer to someone else's work you will normally paraphrase their remarks and include a citation (in Harvard format). If you use the authors' exact words that is a quotation, so should be surrounded by quotation marks and include a citation. Cutting and pasting without quotation marks, and particularly without a citation, is **plagiarism**. BTW, paraphrasing involves reading something, understanding it (which may involve reading background documents) and then summarising the main points in your own words. Minor syntactic transformations on the starting text is not paraphrasing.

If you are unsure of the difference I suggest you go to:

[http://www.studentservices.uwa.edu.au/ss/learning/online\\_services/plagiarism\\_portal](http://www.studentservices.uwa.edu.au/ss/learning/online_services/plagiarism_portal)

In particular, follow the link to the FAQs and the link to the examples of what not to do.

- Hint 1: Excluding quotes – which must be accurate and within quotation marks – you should be aware that contractions, such as “doesn’t”, “won’t”, etc, are inappropriate in formal writing; spell the phrases out: do not, will not, etc.
- Hint 2: If you are using MS-Word, I strongly suggest you do not ignore the red lining (assumed spelling mistakes) and green lining (assumed grammatical issues). Both can be wrong, e.g. the difference between American and Australian/British spelling or specialist words unknown to the dictionary, but often these suggestions are right. Think about each one before dismissing it.
- Hint 3: A necessary condition for a good mark is at least one reference not given to you in the assignment description.
- A marking rubric can be found linked to the the CITS3200 unit timetable near where you find this essay description.

Examples of excellent essays have been kindly provided by 2013 CITS3200 students Jack Golding and Michael Martis (and made available here with their permission, of course).

[http://undergraduate.csse.uwa.edu.au/units/CITS3200/essay/Jack\\_Golding.pdf](http://undergraduate.csse.uwa.edu.au/units/CITS3200/essay/Jack_Golding.pdf)

[http://undergraduate.csse.uwa.edu.au/units/CITS3200/essay/Michael\\_Martis.pdf](http://undergraduate.csse.uwa.edu.au/units/CITS3200/essay/Michael_Martis.pdf)

Finally, please submit the electronic source document version using cssubmit, i.e. doc, docx, odt or tex. See the Unit Timetable on the Project web page for the date and time that the submissions close.

Michael J. Wise  
August 11, 2017