PH3035—Philosophy of Artificial Intelligence

Essay Topics

2nd Semester, 2023-24

Topics:

Write an essay of approximately 2500 words on **one** of the following:

1. Conduct a Turing Test (or, rather, play Turing's 'imitation game') with ChatGPT. Given what Turing says in his 1950 paper[1] do you think that ChatGPT passes or fails? How, and why? What kinds—or topics—of conversation would make it more likely to pass, or more likely to fail? What does this show about the Turing Test? What does this show about ChatGPT?

2. Ask ChatGPT whether it has a mind, or whether it can think. It will say 'no', but it will also give some reasons as to why. Use its answers to ask further questions, picking up on aspects of what it says that you find significant, whilst also providing your own critique of what it says (using the material we have discussed in this module). In the light of this conversation, and your own reasoning, what do you think would have to be true of ChatGPT before it would be correct to say that it has a mind, or that it is capable of thinking?

3. Design-your-own-Essay. If you wish to do so, you may devise a topic of your own on which to write. It should find a way to explore some connection between ChatGPT and the topics of the module (based on specific, documented, conversations that you have with it). If you choose this option, your topic must be cleared with me by **19th March** (although I encourage you to talk to me about it sooner); I would like to see an essay question (analogous to those above), a thesis statement, and a brief outline before granting permission, and I reserve the right to refuse it.

Notes on ChatGPT:

- You can find ChatGPT, create a free account, and interact with it here: https://chat.openai.com/chat. Occasionally, the server is too busy, but refreshing your page a couple of times (or re-trying in a minute or two) is usually enough to get in; be patient and don't leave things to the last minute. If you run into problems and you're really stuck, send me an email.

- Your essay **must** include screenshots of your conversations with ChatGPT (these are not part of the word-count). Such images should be sensibly formatted and included in the text itself (analogous to a block quotation). You may also include other screenshots and supplementary material, if necessary, in an appendix to the paper. I will be looking not only at the responses that ChatGPT gives, but also (more importantly) at the prompts that you use,[2] and the reasons that you give for doing so

---

[1] Turing, A. (1950) "Computing Machinery and Intelligence" *Mind*, 59, 433–460

[2] Notice that I *deliberatley* put this in terms of 'prompts' and 'responses' and definitely *not* 'questions' and 'answers'; it's not an oracle or a search engine!

- Treat ChatGPT as a source which you can cite and evaluate critically, just like any other text; keep in mind how it works, how it has been trained, and its consequent limitations. It is, in effect, auto-complete on steroids (and hallucinogens); it is up to you to make sensible judgements (*on which you should comment*) about the reliability, trustworthiness, and accuracy of its outputs.

- This Panopto video from UCC's Derek Bridge is a really good overview of ChatGPT and other related "Large Language Models" and well worth your time: https://bit.ly/3myuFZ4

Notes on Practicalities:

- This essay is due before midnight on **Monday 8th April, 2024**, and is worth 50% of your overall grade for the course. Essays should be submitted via Canvas, either in WS-Word (.doc or .docx) or PDF format, by uploading it on the relevant 'assignments' section.

  – Iff you ask for it (in writing, by email), you will automatically be granted an extension (no questions asked) until Friday 12th April (*you do have to ask for it, though.*).

- If you feel you would like some further advice about general essay-writing skills, don't hesitate to contact me. The department webpage http://www.ucc.ie/philosophy has some useful information under the "Undergraduate/Other Documents" section, and the following two links are very useful:

  – http://www.jimpryor.net/teaching/guidelines/writing.html
  – http://prezi.com/z4h1_fwilbxj/a-sample-philosophy-paper/

- Uploaded essays will be automatically scanned by Canvas's plagiarism detection software. As always, the penalties for plagiarism are very severe. Plagiarism includes (but is not limited to): buying an essay or having someone else (human or AI) write it for you; submitting an essay that you wrote for another course; copying passages from a book or website without proper acknowledgement etc. Anything that prevents you from demonstrating your original attempt to grapple with an issue is a **B**ad **T**hing, both pragmatically and intellectually. Don't do it.

—◇◇—