AI5 - Humans and Intelligent Machines: Coursework 1

Testing for Intelligence

There is a whole industry of annual AI testing that is based on the Turing Test, or more generally Turing Test-like approaches, in which mathematics or programming teams attempt to benchmark their algorithms against one another through standardised scenarios, rules and parameters. Some competitions continue to attempt to be faithful to what they perceive to be the original Turing Test (though there are many interpretations), while other tests diverge from the original and establish new forms of tests purporting to be more relevant to modern AI, or to a particular technology or methodology.

At the more traditional end of the spectrum of tests is the Loebner Prize, which is an example of a test so faithfully aligned to the original that it is easy to game the system to appear human-like using chatbot techniques, and therefore the results are not considered relevant by serious AI researchers.

While the original Turing Test still has many adherents and defenders[1], more recently, researchers have updated the test in various ways such as the Winograd Schema test which aim to undermine ‘cheating’ chatbot approaches and produce more substantial, practical results[2].

At the more modern end of the AI testing spectrum are challenges such as the DARPA Grand Challenge in autonomous vehicles[3] or the recent Alexa Prize from Amazon [https://developer.amazon.com/alexaprize] which define new challenges for competitive teams over potentially years of development. In the latter case, the competition is highly technically challenging and potentially lucrative with funding of a quarter of a million dollars for each selected team to develop their solution and an additional half a million dollars prize for the winning entrants.

Funding new algorithms and approaches is one side of modern competitive AI, while competitions also exist to protect against, detect or evaluate AI. A good example is Facebook’s recent Deepfake Detection Challenge[4], which aims to develop robust counter-measures and identification techniques against the use of AI in producing fake video representations. Some authors consider these approaches to be ‘destructive Turing tests’ i.e., they are designed to deceive rather than encourage advances in general AI, but many would argue the processes and results are similar to those originally imagined by Turing, even if the foundational ethics and technologies have changed substantially.

---

[1] https://link.springer.com/article/10.1007/s00146-020-00946-8, https://www.forbes.com/sites/tomtaulli/2020/11/27/turing-test-at-70-still-relevant-for-ai-artificial-intelligence/

[2] http://www.cs.virginia.edu/~robins/Am_I_Human.pdf , https://www.nextplatform.com/2019/03/18/modernizing-the-turing-test-for- 21st-century-ai/

[3] https://en.wikipedia.org/wiki/DARPA_Grand_Challenge

[4] https://ai.facebook.com/datasets/dfdc/

Your task

This is a **group assessment**. With your group, write an essay of approximately 1000 words describing the design of a test for intelligence in a competition format. The essay can include diagrams, tables and figures.

Your test should attempt to differentiate human intelligence from an algorithm, but you may test any human characteristic using any methodology and technology you like, although your test must be able to meet the criteria below well.

Assessment

In total, this work is worth up to **30 marks**, which matches the percentage contribution to the overall mark of the module.

After giving your test a memorable name, you should describe:

- **Define** human intelligence and **which aspect of human intelligence** your test focuses on, why, and how [10];
- The **conditions and process** of the test (e.g. as a flow chart or rules-based sequence), including how entrants to the competition will be **graded or ranked** against the test and the **end conditions** by which entrants will be able to say that they have succeeded/won or failed [10];
- A paragraph describing how **feasible** the test will initially be, why the test will be **challenging** for teams in a modern context and – referring to current advances in AI technologies – how well you expect entrants to perform over a few years of your competition [10].

Why a group assessment

You will find that group assignments are present in multiple units. This is for a few reasons:

1. You can get far more out of some experiences if you share them with others. In particular, it is much more enriching to **discuss** challenging topics like intelligence and ethics **with other people**, some of which may have different views.
2. Remote Flexi work is increasingly popular and this is a good opportunity for working in a **distributed team**.
3. With an online course, it can be hard for students to interact and communicate with each other. This assignment is a way for you to simply **talk to** other students.

Word count

1000 words (+/- 10%). This limit does not include references, diagrams, tables and figures. Keep your diagrams, tables and figure short and clear. They should be there only to support your argument not to replace content in the essay.

Remember to not put your name or other personally identifying information on the coursework so anonymous grading can be undertaken.