Introduction to Artificial Intelligence
Introduction/Agents

-

Exercise 1

Read Turing's original paper on AI (Turing, 1950). In the paper, he discusses several objections to his proposed enterprise and his test for intelligence. Which objections still carry weight? Are his refutations valid? Can you think of new objections arising from developments since he wrote the paper? In the paper, he predicts that, by the year 2000, a computer will have a 30% chance of passing a five-minute Turing Test with an unskilled interrogator. What chance do you think a computer would have today? In another 25 years?

Link: here

Exercise 2

Study the 2023 EU "Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)" (or its final version, if available). Which provisions appear to have the most direct and tangible impact on what kinds of AI systems can be deployed? (Link: here)

Exercise 3

Many researchers have pointed to the possibility that machine learning algorithms will produce classifiers that display racial, gender, or other forms of bias. How does this bias arise? Is it possible to constrain machine learning algorithms to produce rigorously fair predictions?

Exercise 4

Summarize the pros and cons of allowing the development, deployment, and use of lethal autonomous weapons. (Hint: A good place to start is the Congressional Research Service report "International Discus- sions Concerning Lethal Autonomous Weapon Systems," dated October 15, 2020.)

Exercise 5

Investigate the state of the art for domestic robots: what can be done (with what assump- tions and restrictions on the environment) and what problems remain unsolved? Where is research most needed?

Exercise 6

The vast majority of academic publications and media articles report on successes of AI. Examine recent articles describing failures of AI. (IBM Watson for Oncology, Tesla Autopilot Crashes, Microsoft Tay, Amazon AI Recruiting Tool are good examples, but feel free to find your own.) How serious are the problems identified? Are they examples of overclaiming of or fundamental problems with the technical approach?

Exercise 7

Various subfields of AI have held contests by defining a standard task and inviting re- searchers to do their best. Examples include the ImageNet competition for computer vision, the DARPA Grand Challenge for robotic cars, the International Planning Competition, the Robocup robotic soccer league, the TREC information retrieval event, and contests in ma- chine translation, speech recognition, and other fields. Investigate one of these contests, and describe the progress made over the years. To what degree have the contests advanced the state of the art in AI? Do what degree do they hurt the field by drawing energy away from new ideas?

Exercise 8

Survey a few language transformation models throughout the history of AI, from early rule-based systems to modern deep learning approaches, e.g. ELIZA, ALICE, IBM Watson, and GPT-3/ChatGPT. Analyze how each model processes and generates human language, the underlying techniques (e.g., pattern matching, statistical methods, neural networks), their strengths and weaknesses, and their impact on the development of AI.

Exercise 9

"Hallucinations" in language models refer to instances where the AI generates information that is factually incorrect, logically inconsistent, or entirely fabricated, despite being presented in a confident and coherent manner. Why do they occur? In which areas is it critical to address them? How are these addressed?

Each group will be given one exercise. Prepare to present your answers and discuss in about 15 minutes in class (using blackboard or slides).