RQ1: Overview

- Due Jan 17 at 5:10pm
- Points 4
- Questions 5
- Available until Jan 17 at 5:10pm
- Time Limit 15 Minutes

Instructions

This quiz has 5 essay type questions.

Do not use AI to answer these questions.

Keep the questions concise and less than 5 sentences. Longer answers will be disregarded.

This quiz was locked Jan 17 at 5:10pm.

Attempt History

	Attempt	Time	Score
LATEST	[Attempt 1](#)	16 minutes	3.9 out of 4 *

* Some questions not yet graded

⚠ Correct answers are hidden.

Score for this quiz: 3.9 out of 4 *
* Some questions not yet graded
Submitted Jan 17 at 5:05pm
This attempt took 16 minutes.

⠿

Question 1
0.8 / 0.8 pts

Current state of AI is predominantly narrow AI i.e. it doesn't generalize. What is the main difference between how AGI and narrow AI handle tasks? Give an example of a problem that AGI could solve but narrow AI could not.

Your Answer:

Narrow AI is primarily used for very specific domains, trained on a specific type of data whereas AGI (Artificial General Intelligence) is a hypothetical form of AI that is capable of understanding, and learning itself from a wide range of data across various domains and fields, thus, is not limited to any specific domain.

One example could be in drug discovery of medical treatments. While Narrow AI would only be able to hypothesize solutions based on the data it was trained on, AGI would be able to capture various

datasets of biology, chemicals, human physiology and what not to create and synthesize specific drugs or medicines, something which narrow AI cannot do as it lacks the ability to be creative and be novel. AGI would be able to be creative and novel.

⠿

Question 2
0.8 / 0.8 pts

Why is it important to ensure that an AI's goals align with human values? Describe one example of how an AI system could behave unexpectedly if its goals are not aligned.

Your Answer:

It would be extremely crucial to ensure that an AI's goals align with human values so that the AI is beneficial to humanity and we do not run into the Matrix or Terminator or Dune like scenarios where the machines start turning against us or bring us harm in their tasks. Misaligned goals could still lead the AI to complete its goals, but at the cost of human life.

AN example could be the Paper Clip problem where the AI starts utilizing all of the resources into generating just paper clips, which could limit resources for humanity eventually leading us to downfall, and we may not be able to stop the AI.

⠿

Question 3
0.8 / 0.8 pts

Explain the paperclip thought experiment to a five year old. How could a programmer design constraints to prevent an AI from pursuing harmful goals?

Your Answer:

Imagine you ask a very smart machine to create some paper clips, and to make as many paper clips as possible, and you don't tell it when to stop, it may start taking things around you that cannot and should not be used to create paper clips such as your Legos or your Star Wars action figures. But what if the machine doesn't stop there? What if it starts to take things outside of your home, your car, your house, the trees outside, everything into making paper clips, and doesn't stop, it might take everything in the world just to make paper clips! How are you going to eat or drive or survive when everything is turned into paper clips xD?

A programmer could design constraints to prevent this by limiting the amount of resources the machine can take, and the specific type of resources the machine can take, and even the amount of paper clips the machine should generate before it stops.

⠿

Question 4
0.7 / 0.8 pts

Why is fairness important in AI systems? Describe one way a developer could test whether their AI model is treating all users fairly.

Your Answer:

Fairness refers to the model being fair to all the data it is trained on, and not being biased towards a specific group of data. This ensures trust, equality, and ethical behaviour in the system.

A developer could test this by checking the model for false positives, false negatives, etc across various parameters.

An example of the parameters would be perfect Sample: Fairness in AI is essential to ensure that decisions made by AI systems are unbiased and equitable. For example, a developer could test for fairness by evaluating the system's outcomes across diverse demographic groups, ensuring no group is unfairly advantaged or disadvantaged.

⠿

Question 5
0.8 / 0.8 pts

What is one way to make sure an AI system is accountable for its decisions? Why is keeping a log of the system's activities helpful?

Your Answer:

The best way to make sure the AI is accountable is to make the AI explainable, something that research is being done on. Currently AI systems are black box models, and lack explainability. If they can't be explained, then their choice / approach / answers can't be explained or reasoned about. If the AI is explainable, then their answers / choice will have clear explanations behind them which can be reasoned about.

Keeping a log helps in reviewing the system's behaviour so that it can be accounted for in the next training etc, and to help identify errors or biases or in debugging.

Quiz Score: 3.9 out of 4
* Some questions not yet graded