

1. Environments

- a. What does it mean for an environment to be stochastic? The self-driving car is an example of a stochastic environment; why is this?**

An environment is stochastic when an action could have multiple outcomes instead of being guaranteed to have a single effect, like in a deterministic environment. Self-driving cars operate in a stochastic environment because their actions may not have perfectly predictable results due to unforeseen circumstances such as tire slippage, weather conditions, or unexpected obstacles.

- b. What does it mean for an environment to be partially observable? The self-driving car is also partially observable. Why is this, and how does it differ from a stochastic environment?**

An environment is partially observable when an agent can't completely perceive information about all parts of an environment and must make guesses instead. Self-driving cars operate in partially observable environments because they can't see behind obstacles and don't know other drivers' intentions. Observability relates to the input/perception of an agent, while stochasticity relates to its output/action.

- c. What does it mean for an environment to be sequential? Chess is an example of a sequential environment. Why is this?**

An environment is sequential when an agent must take a series of connected actions to complete a task, where each action affects future possibilities. Chess is sequential because each move affects possible future moves, so players must think many moves ahead.

- d. What does it mean for an environment to be dynamic? A video game is an example of an environment that is usually dynamic. Why is this?**

An environment is dynamic when it continues to change while an agent decides on an action.

Many video games are dynamic because the game world moves and changes while a player makes decisions. For example, an enemy may continue attacking while a player plans their next move.

- e. What does it mean for an environment to be multiagent? Card games such as poker can be treated as multiagent. How does this approach change the way we think about the problem?**

An environment is multi-agent when multiple actors exist and interact. The actors can be human or computational. This approach requires each agent to consider others' intentions when making decisions.

- 8. Large Language Models, or LLMs, have had a massive impact both on the field of AI and on culture more broadly since their introduction a few years ago.**

For this question, we'll consider three of the most well-known LLMS:

ChatGPT from OpenAI.

Gemini from Google

Claude from Anthropic.

You will need to create accounts for each of these, but should not need to sign up for any paid services.

Choose one of the environment questions from question 1, and paste it into one of these LLMs, and provide a copy of their response. Are they correct? Can you distinguish between your answer and an LLM?

Now pick one of the LLMs and craft a prompt in which you give it the original question, your answer, and an answer from one of the other LLMs, and ask it to determine which answer was written by a human and why. Does it get the answer right?

Include a copy of your prompt and the LLM's reply.

First, I asked Claude question 1e: *"What does it mean for an environment to be multi-agent? Card games such as poker can be treated as multiagent. How does this approach change the way we think about the problem?"*

Claude's response:

"A multiagent environment contains multiple intentional actors (human or computational) whose decisions affect each other, requiring agents to reason about others' intentions and potential actions. In poker, this means players must consider not just their own cards and optimal moves, but also anticipate opponents' strategies, bluffs, and reactions - making the problem more complex than simply calculating probabilities or following a fixed strategy."

Claude's response is correct in explaining a multiagent environment. However, I can tell this is an AI-generated response because its diction and syntax are unnecessarily complicated. It also integrates the poker example into its explanation, even though the question only mentions poker

as a helpful example for students to understand the concept. A student response would normally be more straightforward and minimal, like my answer.

Next, I provided Claude with the two answers (mine and Claude's) and asked it to determine which was which. Here was my prompt:

"I will show you two answers to this question:

Answer 1: A multiagent environment contains multiple intentional actors (human or computational) whose decisions affect each other, requiring agents to reason about others' intentions and potential actions. In poker, this means players must consider not just their own cards and optimal moves, but also anticipate opponents' strategies, bluffs, and reactions - making the problem more complex than simply calculating probabilities or following a fixed strategy.

Answer 2: An environment is multi-agent when multiple actors exist and interact. The actors can be human or computational. This approach requires each agent to consider others' intentions when making decisions.

One of these answers was written by a human and one was written by an AI language model. Can you determine which is which and explain your reasoning?"

Interestingly, Claude incorrectly identified Answer 1 as human-written when it was actually AI-generated, and Answer 2 as AI-generated when it was human-written in its response:

"Let me analyze these answers carefully:

Answer 1 is likely written by a human, while Answer 2 appears to be AI-generated. Here's my reasoning:

1. Structure and Flow:

- Answer 1 has a more natural flow with interconnected ideas. It builds its explanation organically, connecting concepts with fluid transitions and using the poker example to illustrate the broader point.*
- Answer 2 is more formulaic and compartmentalized, with shorter, more distinct statements that don't flow as naturally into each other.*

2. Depth and Nuance:

- Answer 1 provides richer detail and specific examples, like mentioning "bluffs" and "reactions" in poker. It shows understanding of how the concept applies in practice.*
- Answer 2 remains at a more surface level, stating the basic definition without diving into meaningful implications or examples.*

3. Language Patterns:

- Answer 1 uses more varied and natural language construction, with complex sentences that combine multiple related ideas.*
- Answer 2 uses simpler, more repetitive sentence structures ("... when ..." followed by "... can be ...") which is characteristic of AI-generated text.*

4. Integration of Examples:

- Answer 1 weaves the poker example naturally into the explanation, using it to illustrate specific aspects of multiagent environments.*
- Answer 2 doesn't meaningfully integrate or expand upon examples, merely mentioning that actors can be human or computational without explaining why this matters.*

The differences suggest that Answer 1 reflects human thought patterns and writing style, while Answer 2 shows the more systematic but somewhat shallow and formulaic approach typical of AI language models. ”