Do LLMs Understand Language?

Prompt A: The Chinese Room Shows That AI Does *Not* Understand Language

John Searle argues that manipulating symbols according to rules is not the same thing as understanding their meaning. In the Chinese Room thought experiment, the person in the room produces correct answers in Chinese without knowing what any of the symbols mean.

Discussion questions:

1. In what sense, if any, is the system in the Chinese Room *missing* something essential to understanding?
2. If a human inside the room does not understand Chinese, why should we say the system as a whole does?
3. How does this thought experiment challenge the idea that correct behavior is sufficient evidence of understanding?
4. Applying this argument to modern AI systems: even if a model produces fluent, coherent language, what reasons might we have to say it still lacks genuine understanding?

Goal:
Defend the claim that **syntactic symbol manipulation alone cannot produce semantic understanding**.

Do LLMs Understand Language?

Prompt B: The Chinese Room Fails to Show That AI Lacks Understanding

Critics of the Chinese Room argue that the thought experiment sets an unrealistically narrow standard for understanding. They claim that understanding can emerge at the level of a *system*, even if no single part of the system understands on its own.

Discussion questions:

1. Why might it be a mistake to focus on the person in the room rather than the system as a whole?
2. Do humans themselves "understand" language in any way other than following learned rules and patterns?
3. If an AI system can use language appropriately across many contexts, explain ideas, and respond flexibly, what more should we require for understanding?
4. Should understanding be judged by *internal processes* or by *observable abilities*? Why?

Goal:
Defend the claim that **functional competence and appropriate use may be enough to count as understanding**.

Do LLMs Understand Language?

Take Home Follow-Up Writing Prompt

In class, you discussed whether large language models (LLMs) can be said to *understand* language, drawing on arguments inspired by John Searle's Chinese Room thought experiment and its modern critics.

Now that you have heard arguments on **both sides**, reflect on how (if at all) your view has changed.

Your task

Write a short response (approximately **300–400 words**) that addresses the following:

1. **Initial position:**
   Briefly state which position you found more convincing *at the start* of the discussion (LLMs do not understand language / LLMs can be said to understand language).
2. **Challenge from the opposing side:**
   Describe **one argument from the opposing position** that you found serious, compelling, or difficult to dismiss. Explain *why* it gave you pause.
3. **Revised view:**
   After considering both sides, explain your current position.
   - Did you change your mind?
   - Refine your original view?
   - Or remain unconvinced—but for clearer reasons?

Expectations

- Engage seriously with **both sides** of the debate.
- Focus on reasoning, not whether LLMs are "good" or "bad."
- Clarity and thoughtfulness matter more than reaching a particular conclusion.