# The Professor Consultation Room: A Thought Experiment on Understanding, Expertise, and the Cognitive Limits of Large Language Models

(Position Paper, Preprint)

**Jingde Cheng**

(Professor Emeritus)

Department of Information and Computer Sciences, Saitama University

Saitama, Japan

cheng@aise.ics.saitama-u.ac.jp, jingde.cheng@gmail.com

***Abstract***: This position paper proposes a new thought experiment, the **Professor Consultation Room**, which bears on epistemology, philosophy of mind, ontology, philosophy of logic, intelligence science, and artificial intelligence.

### *Background and Motivation*

In 1980, the American philosopher John Rogers Searle (1932–2025) proposed the **Chinese Room** thought experiment to argue that a symbol-manipulation system does not possess genuine semantic understanding [1–3].

With the development of expert systems and, more recently, large language models (LLMs), the central controversy raised by the Chinese Room—the distinction between intelligence in performance and intelligence as an intrinsic cognitive capacity—has become increasingly significant. However, the Chinese Room scenario relies on the artificial assumption that a system merely manipulates syntactic symbols without understanding their semantics, an assumption that does not fully correspond to the mechanisms of real-world AI systems.

Particularly in the context of expert systems, people are often less concerned with whether an AI system *truly understands* than with whether its outputs can serve as reliable and trustworthy evidence of expertise. For example, in an actual educational setting, if an expert system can provide correct answers to all questions in a course, students may naturally regard it as equivalent to a "real professor." Yet a crucial question remains: can even the most advanced expert system genuinely answer all students' questions and provide guidance in the same way that a real professor can?

To address this more realistic yet insufficiently analyzed phenomenon, and to explore the ongoing debates concerning the understanding and creativity of LLMs, the author proposes a new thought experiment: the **Professor Consultation Room**. The central question of this thought experiment is no longer merely whether a system possesses understanding. Rather, it asks:

> Can an external questioner determine, solely on the basis of question-and-answer interactions, whether the entity inside the room possesses cognitive abilities that have not yet been demonstrated by current LLMs?

### *The Professor Consultation Room Scenario*

Professor P, a professor at a world-class university, teaches a course in logic to a group of students S1, S2, ..., Sn. Professor P possesses a high level of intelligence and academic expertise, and the students likewise possess considerable intelligence and knowledge.

At any time, students may submit, via a computer terminal located outside Professor P's office, any question related to logic in textual form. The questions are not restricted to material covered in lectures, handouts, or assigned textbooks; the only restriction is that they must concern logic.

Professor P may answer the questions personally. Alternatively, when absent, Professor P may employ a logic-oriented LLM expert system that he has developed. This system possesses all of Professor P's logical knowledge and can answer questions on the basis of that knowledge. However, it does not possess cognitive abilities that exceed the generally recognized capability boundaries of existing LLMs.

All responses are displayed in textual form on the computer terminal outside the room.

### *Questions of the Professor Consultation Room*

**Question 1.**
Can any student, solely by asking a finite number of questions from outside the room and receiving answers, determine with certainty—and provide a clear and explainable justification—whether the respondent inside the room is Professor P himself or the logic-oriented LLM expert system?

**Question 2.**
If the answer to Question 1 is affirmative, what is the student's clear and explainable justification?

**Question 3.**
If the answer to Question 1 is negative, what minimum necessary conditions must Professor P's logic-oriented LLM expert system satisfy in order to remain indistinguishable from Professor P under such questioning?

### *References*

[1] J. Searle, “Minds, Brains, and Programs,” *Behavioral and Brain Sciences*, Vol. 3, No. 3, pp. 417–424, 1980.

[2] D. Cole, “The Chinese Room Argument,” *The Stanford Encyclopedia of Philosophy*, Center for the Study of Language and Information (CSLI), Stanford University, 2004–2024.

[3] L. Hauser, “Chinese Room Argument,” *Internet Encyclopedia of Philosophy (IEP)*.