

Philosophy of Artificial Intelligence

UCLA PHILOS 171

Summer Session C, 2026

{Lecture Times}	Paul Talma	Office Hours: (TBA)
{Lecture Location}	ptalma@humnet.ucla.edu	{Office Hours Location}

Course Description

This course explores foundational questions in the philosophy of artificial intelligence. We will tackle the topic in three stages.

(1) We will begin by considering the origins and ambitions of AI as a field of research: what is intelligence and what would it mean for a machine to possess it? We will examine the Computational Theory of Mind, which holds that thinking is a form of computing, as well as its implementation in symbolic and connectionist architectures. Our aim here will be to acquire a solid theoretical backdrop against which to think about AI.

(2) Next, we'll examine the cognitive capacities of current AI models: can they perceive, reason, and plan? Do their words have meaning, and can they communicate? Do they have goals of their own, and can they infer the goals of others?

(3) Armed with an understanding of the *capabilities* of current AI systems, we'll consider the social and political questions they raise. What risks do generative AI models pose? What promises do they hold? How will they impact education, economic relations, the production of art and media, and the distribution of political power?

You will come away from the course equipped to think through the complex and pressing questions raised by modern AI systems. You will also refine important philosophical skills: how to analyze an argument, how to defend a position, and how to communicate with clarity and precision.

The course is self-contained: it does not presuppose any technical knowledge, and all relevant information about AI will be introduced as we go along.

Course Materials

Required and recommended materials will be provided on BruinLearn.

Assessment

Writing exercises	25%
Midterm paper	25%
Final paper	35%
Attendance	10%
Participation	5%

Writing exercises: these will consist in two take-home writing assignments of about one page each. They will be graded for good-faith completion. They will be designed to target a specific philosophical skill and hence will serve as preparation for the midterm and final.

Midterm paper: a short (about 3 page) paper addressing a philosophical prompt.

Final paper: a longer (5-6 page) paper. You will have the choice between expanding on your midterm paper and writing a paper on a new prompt. Your grade for the final paper will be based in part on incorporating the feedback received on your first paper.

Attendance: will be taken in class. You will be granted one excused absence (in addition to any medical or family emergencies).

Participation: can take many forms—asking a question in lecture, sending a follow up email, coming to office hours, etc. Doing any of these once will earn the full 5%.

Course Modality

The course will be taught online. Attendance in lecture is required and will count toward your grade. Exceptional circumstances apart, you will need to turn on your video to count as present. Doing so helps me gauge your reception of the material

and helps you engage with the lecture. As such, it is not only to your benefit but also to that of your classmates.

Schedule

Week 1: Origins and Aims of AI

Lecture 1: Turing (1950), “Computing Machinery and Intelligence.”

Lecture 2: Block (1981), “Psychologism and Behaviorism.”

Optional Readings: Russell and Norvig (2021), *Artificial Intelligence, A Modern Approach*, Chapters 1 and 2.

Week 2: The Classical Computational Theory of Mind

Lecture 1: [**First writing exercise due.**] Putnam (1967), “The Nature of Mental States.” Fodor (1987), *Psychosemantics* (excerpts).

Lecture 2: Searle (1980), “Minds, Brains, and Programs.”

Optional Readings: Rescorla (2024), “The Computational Theory of Mind” (Stanford Encyclopedia of Philosophy). Marr (1982), *Vision* (excerpts). Dennett (1971), “Intentional Systems.”

Week 3: The Connectionist Challenge

Lecture 1: No reading. Lecture on connectionism and artificial neural networks.

Lecture 2: [**Second writing exercise due.**] Milliere and Buckner (2024), “A Philosophical Introduction to Language Models, Part 1.”

Optional Readings: Fodor and Pylyshyn (1988), “Connectionism and Cognitive Architecture.” Smolensky (1988), “On the Proper Treatment of Connectionism,” LeCun et al. (2015) “Deep Learning.”

Week 4: Meaning and Communication

Lecture 1: Bender and Koller (2020), “Climbing Towards NLU: On Meaning, Form, and Understanding in the Age of Data.”

Lecture 2: Mandelkern and Linzen (2024), “Do Language Models’ Words Refer?”

Optional Readings: Harding and Sharadin (2024), “What is it for a Machine Learning Model to Have a Capacity?” Attah (2025), “Do language models lack communicative intentions?” Williams (2025), “Can structural cor-

responses ground real world representational content in Large Language Models?”

Week 5: Goals, Agency, Alignment

Lecture 1: [Midterm paper due] Butlin (2024), “Reinforcement Learning and Artificial Agency.”

Lecture 2: Gabriel (2020), “Artificial Intelligence, Values, and Alignment.”

Optional Readings: Narayanan & Kapoor (2025), “AI as Normal Technology,” Alexander (2025), “AI as Profoundly Abnormal Technology,” Lazar (2025), “Anticipatory AI Ethics.”

Week 6: Power and Property

Lecture 1: Lazar (2022), “Power and AI: Nature and Justification.”

Lecture 2: Goetze (2024), “AI Art is Theft.”

Optional Readings: Winner (1980), “Do Artifacts Have Politics?” Lazar (2025), “Governing the Algorithmic City.” Creel and Hellman (2022), “The Algorithmic Leviathan: Arbitrariness, Fairness, and Opportunity in Algorithmic Decision-Making Systems.”

Usage of Generative AI

Writing is the most difficult and rewarding part of philosophy. It is only by wrangling your ideas down into text that you can fully understand and examine them. The arduous trial-and-error process involved in accurately, precisely, and concisely articulating a philosophical claim is an integral part of doing philosophy. It is not a mere “implementation detail,” the kind of work that gets done once the real thinking is over. It is part and parcel of the thinking itself. It is also an excellent way to learn how to write well.

For this reason, the use of generative AI to write any of the assignments is forbidden and constitutes grounds for failing the course. This includes drafting an outline and having an AI generate a paper on the basis of this outline. It also includes using an AI to write or rewrite a sentence or a paragraph. It also includes using an AI to summarize a course reading and quoting or paraphrasing the AI summary. If you are suspected of using generative AI in any of these ways, you will be required to redo the assignment under direct supervision or to discuss your work in an in-person interview.

You are allowed to use generative AI for the purposes of brainstorming, exploration, and checking your understanding. Doing so can be a valuable use of your time, if done properly (for example, the model needs to be prompt-engineered to give adequate pushback on your ideas). Any use of generative AI for brainstorming and the like should be cited as such.

Academic Integrity

All students are expected to adhere to the university's academic honesty policy. Plagiarism, unauthorized collaboration, or misrepresentation of sources will result in disciplinary action. See previous section for acceptable uses of generative AI.

Accessibility

Students needing an academic accommodation based on a disability should either contact me or the Office for the Center for Accessible Education located at (310) 825-1501 or A255 Murphy Hall. When possible, students should contact the CSA within the first two weeks of the term as reasonable notice is needed to coordinate accommodations. For more information visit www.cae.ucla.edu.

Acknowledgements

This syllabus was partly inspired by those of Jocelyn Maclure, Soren Schlassa, and Dan Williams.