Artificial Intelligence: Brains, Minds, and Machines

PHIL209M
Summer 2020
(last updated: June 25, 2020)

INSTRUCTOR:	Shen Pan	MEET TIMES:	M–F 11:00 AM – 1:00 PM
EMAIL:	shenpan@umd.edu	OFFICE HOURS:	By appointment

Course Description: This course provides an introduction to the philosophy of artificial intelligence. We are on the verge of constructing AI systems that will rival or surpass human minds in many ways. Indeed, the current world champions in chess, go, and Jeopardy! are all AI-powered computers. Despite the promise and prominence of artificial intelligence, however, there remain controversial conceptual and foundational issues concerning both the nature and the creation of artificial intelligence. Just what it is for machines—or, for that matter, any systems—to be *intelligent* is not at all obvious. Is intelligence simply a capacity for problem solving of some sort? Or is something else required? When AI-powered computers look for solutions to complex problems, do they engage in *thinking*? If they do, might they engage in a different *form* of thinking from that of human beings? If not, might it be *better* if they do? *Can* they? Indeed, how do we even *tell* whether AI-powered computers engage in thinking or not? Is that an *important* question?

The ever-increasing ubiquity of AI technology also raises new and, in some cases, urgent ethical questions. How will artificial intelligence impact the world, and what can we do to make sure that the impact will be beneficial? More specifically, how can we make sure that AI systems will respect our ethical principles when they make decisions at speeds mere human beings cannot achieve, or on the basis of processes that mere human beings cannot comprehend? Concerning AI systems themselves, should we think of them merely as sophisticated machines? Or are they a new form of life in some sense? Could they ever possess consciousness or free will? What legal and moral rights, if any, should we grant them?

These are only some of the fascinating but also challenging questions in the philosophy of artificial intelligence. The ultimate goal of this course, however, is *not* to arrive at definitive, settled answers to these questions. Rather, it is to introduce to students relevant philosophical frameworks for thinking about them as clearly as possible, while trying to achieve a deeper understanding of artificial intelligence. In other words, this course will teach students how to think philosophically about artificial intelligence, at a high level of abstraction and from a certain distance from practice.

Prerequisites: This course has no formal prerequisites; it only requires a willingness to engage with abstract thinking and critical evaluation of argumentation (sometimes of one's own). Since this is a *philosophy* of artificial intelligence course, no knowledge in computer science or neuroscience will be presupposed, though prior exposure to either can be an asset.

Learning Objectives: I have two broad, interconnected goals for all my students: to become better thinkers, and to become better writers. More specifically for this course, if you work hard, at the end of the summer you can expect to be able to:

1. gain a foundation of knowledge in critical concepts of as well as theoretical approaches to artificial intelligence;

2. understand and critically review recent history of thought about artificial intelligence;
3. identify and analyze ethical, social, and political implications of both artificial intelligence and our interaction with it.

This course also has a meta-goal, and that is to teach you how to integrate insights from philosophy and the cognitive sciences, while being critical of both.

About the Instructor: Shen Pan is from the Department of Philosophy at UMD. His research interests lie at the intersection of metaphysics and philosophy of mind, with a specific focus on time. Two overarching questions driving his research are "What is time?" and "How is time represented in cognitive systems?". As an instructor, Shen is an enthusiastic practitioner of the Socratic Method, and believes that philosophy teaches transferable skills such as critical thinking and effective communication that students can use both in their wider studies and in their future careers. During his free time, Shen enjoys cooking, nature, photography, and anything related to cats.

Required Tech Resources: Since this is an online course, a webcam- and microphone-equipped **computer** along with reliable, high-speed **internet** access will be required. At its core, the course comprises 15 **live class sessions via Zoom**. These sessions will be held largely in a seminar format, which requires active preparation and participation by all members to be a success. For that reason, students will also be utilizing two external services, both free of charge, in order to fully engage with course content and the learning community at large. These services are **Discord** (for asynchronous course communication) and **Google Drive** (for certain collaborative assignments).

- Course ELMS site: https://umd.instructure.com/courses/1285069;
  View assignment deadlines and instructions, upload written work, review grades, etc.
- Course Discord Server: *to be disclosed to students before class begins*;
  Receive announcements, schedule office hours, engage in after-class discussion, etc.
- Shared Google Drive Folder: *to be disclosed to students before class begins*;
  Access readings and other course materials, complete Collaborative Reading assignments (see below), etc.

You are responsible for making sure that these softwares are properly installed and configured (instructions will be provided for accessing user-specific content). While I do not anticipate you running into technical trouble, in the rare event that you do, UMD's Division of Information Technology (https://it.umd.edu) has a wide range of relevant resources and their staff members are ready to help you during regular work hours. You may also contact me for technical support directly on the #trouble-shooting channel on Discord.

Learning Materials: With the exception of the first day of class (July 13), there will be required readings assigned for each class session. Corresponding to these, there will be (a total of 14) Collaborative Reading assignments to be completed collaboratively by you and your assigned teammates *before class*. Each class session will then be devoted to building upon and expanding—sometimes significantly—contents from the readings (i.e., *not* explaining them). As you will soon find out, philosophy is a collective, communal enterprise, and the best way to learn philosophy is to actually *do* philosophy.

All learning materials will be posted online (see the last page of the syllabus for a tentative list). This means that there will be no required purchases for this course. Nevertheless, you may wish to consult as background reading Margaret A. Boden's excellent introductory book, *AI: Its Nature and Future* (Oxford University Press, 2016).

> I am always happy to provide you with further learning materials if you want to go more into depth with a topic. Two resources, both free and online, that you may find particularly helpful are the Stanford Encyclopedia of Philosophy (https://plato.stanford.edu) and the Internet Encyclopedia of Philosophy (https://plato.stanford.edu).

Course Assessment: Assessment is based on five categories of graded items, which are worth 750 points in total. The breakdown is as follows:

Collaborative Reading (14)	210 points
Short Argument Reconstruction Essays (2)	160 points
Critical Response Paper (1)	100 points
In-class Presentation (1)	100 points
Unit Exams (3)	180 points

(Instructions and deadlines are individually posted for each graded item on the course ELMS site.)

Your grade is determined by your performance and your performance alone (i.e., not curved or rounded up). To be fair to everyone, I have to establish clear standards and apply them consistently, so please understand that being close to a cutoff is not the same thing as making the cut. For the purpose of calculating final grades, the following conversion rule will be used:

A+: 96%, A: 92%, A-: 89%
B+: 86%, B: 82%, B-: 79%
C+: 76%, C: 72%, C-: 69%
D+: 66%, D: 62%, D-: 59%
F: below 59%

If earning a particular grade is important to you, please speak with me at the beginning of the course so that I can offer some helpful suggestions for achieving your goal. In general, in addition to attending class sessions, you should expect to spend, on average, 4 hours every day on readings and written assignments. *This is in line with standard workload for college courses, which typically require 2 hours outside of class for every hour in class.*

Communication: I use Discord to send out important announcements, and it is your responsibility to check the #announcements channel regularly. The best way to contact me regarding content-related questions is to use Discord also—either through private messaging or posting threads on appropriate channels—as I check it multiple times a day. As indicated above, the course ELMS site will be used for grading and bookkeeping purposes for the most part. If, however, you have grade-related questions, please contact me via the ELMS messaging service only.

Expectations: Be patient with the readings, don't cheat, and be respectful to your peers and instructor. Below I elaborate on each in turn.

Reading philosophy is not easy. Philosophical texts require close attention, careful thought, and active engagement. For this reason, you are not going to get much out of the texts if you just skim them. Moreover, the online nature of this course requires that you take an active role in the learning process. This includes engaging and collaborating with your peers through prior collaborative work as well as in class sessions.

Anyone who engages in academic dishonesty will be immediately reported to the relevant disciplinary authorities for further action. If you have questions about what constitutes cheating or academic dishonesty, please do not hesitate to ask me. But a general rule of thumb is that for any ideas that are not originally yours you should indicate clearly and explicitly where they come from. For details, you may consult the Office of Student Conduct (https://www.studentconduct.umd.edu).

> If you are having trouble completing an assignment or are not satisfied with your performance, please get in touch me as soon as possible so that we can discuss solutions. Do not resort to cheating in any shape or form, which can jeopardize your future at UMD and beyond.

We will be discussing controversial issues in this course, and challenging our beliefs throughout. To do that effectively as a learning community, we must be polite, respectful, and generous in our conversations with others.

Attendance: By default, you are expected to attend every class session (though I will not take attendance). This is for two reasons. First, in general, consistent attendance offers students the most effective opportunity to gain command of course concepts and materials. Second, as noted above, philosophy is inherently a dialectical practice, and the required back-and-forth questioning and answering is not something that can be made up outside of class. In the rare event that a live class session cannot be held, instruction is *not* thereby canceled. Rather, the instructor will announce backup plans, in which you are expected to participate.

Late Work Policy: Because they are an integral part of each class session's learning experience, for Collaborative Reading assignments, late work will *not* be accepted. Likewise, your In-class Presentation cannot be made up. Late work for all other assignments and exams will be accepted, for up to two days, with 10% of the grade deducted per day of being late.

> I do realize that even the most diligent students may have to miss a deadline on occasion due to illness or some other emergency. Should such unfortunate situations occur, please get in touch with me as soon as possible, with relevant documentation, so that we can discuss plans for you promptly.

Accessibility and Inclusion: *Your academic achievement is important to me.* I am committed to the principle that no qualified individual with a disability shall, on the basis of disability, be excluded from participation in or be denied the benefits of the services, programs, or activities of the University, or be subjected to discrimination. If you have a disability that may prevent you from fully demonstrating your abilities, please contact me as soon as possible, so that we can discuss how accommodations can be implemented, to ensure full participation and to make your learning experience comfortable and effective.

The University of Maryland values the diversity of its student body. Along with the University, I am committed to providing a classroom atmosphere that encourages the equitable participation of all students regardless of age, disability, ethnicity, gender, national origin, race, religion, or sexual orientation. Potential devaluation of students in the classroom that can occur by reference to demeaning stereotypes of any group and/or overlooking the contributions of a particular group to the topic under discussion is inappropriate.

Other Course Related Policies: It is our shared responsibility to know and abide by the University of Maryland's policies that relate to all courses. Please visit https://go.umd.edu/ug-policy for the Office of Undergraduate Studies' full list of campus-wide policies, read them carefully, and follow up with me if you have any questions.

Reading List

Almost all readings for the course will be original research articles. The following is a tentative list of readings categorized into the three modules by which the course will be organized. The list is by design an overly inclusive one, and will receive continuous updating to provide students with a bird's-eye view of the breadth and depth of the subject matter. Not every reading will be discussed at length; nor will students be expected to study every item on the list comprehensively. The amount of instruction time allocated to each module is by default one week, which is subject to change, depending on student interest.

Module 1: Philosophical Foundations of Artificial Intelligence

- Ned Block (1978): Troubles with functionalism.
- Ned Block (1995): The mind as the software of the brain.
- Nick Bostrom (1997): How long before superintelligence?
- Robert Epstein (2016): The empty brain.
- John Haugeland (1997): What is mind design?
- Kevin Lande (2019): Do you compute?
- David Marr (1977): Artificial intelligence—a personal view.
- Alan Newell & Herbert Simon (1981): Computer science as empirical inquiry: Symbols and search.
- Hilary Putnam (1967): The nature of mental states.
- John R. Searle (2014): What your computer can't know.
- Alan M. Turing (1950): Computing machinery and intelligence.

Module 2: Philosophical Challenges to Artificial Intelligence

- David Anderson & Jack B. Copeland (2002): Artificial life and the Chinese room argument.
- Margaret A. Boden (1988): Escaping from the Chinese room.
- Jack B. Copeland (1993): The curious case of the Chinese gym.
- Daniel C. Dennett (1987): Cognitive wheels: The frame problem of AI.
- Jerry A. Fodor (1987): Modules, frames, fridgeons, sleeping dogs, and the music of the spheres.
- Terence Horgan (2013): Original intentionality is phenomenal intentionality.
- Frank Jackson (1982): Epiphenomenal qualia.
- Clark Glymour (1987): Android epistemology and the frame problem.
- Michael Graziano (2015): Build a brain.
- Drew McDermott (1987): We've been framed: Or, why AI is innocent of the frame problem.
- George Musser (2016): Consciousness creep.

- John R. Searle (1980): Minds, brains, and programs.
- John R. Searle (1980): The background of meaning.

Module 3: Ethics of Artificial Intelligence

- Colin Allen, Gary Varner & Jason Zinser (2000): Prolegomena to any future artificial moral agent.
- Nick Bostrom & Eliezer Yudkowsky (2014): The ethics of artificial intelligence.
- Awad Edmond, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, & Iyad Rahwan (2018): The moral machine experiment.
- Michael LaBossiere (2017): Testing the moral status of artificial beings, or "I'm going to ask you some questions".
- Deborah G. Johnson (2006): Computer systems: Moral entities but not moral agents.
- Deborah G. Johnson & Mario Verdicchio (2018): Why robots should not be treated like animals.
- Eric Schwitzgebel & Mara Garza (2015): A defense of the rights of artificial intelligences.
- John P. Sullins (2006): When is a robot a moral agent?
- Mary A. Warren (1997): Moral status: Obligations to persons and other living things.