PHIL351 – PHILOSOPHICAL PERSPECTIVES ON COGNITIVE SYSTEMS RESEARCH (2023)

Winter 2023, Term 1
Department of Philosophy, University of British Columbia

1. CONTACT

INSTRUCTOR

Name: Dr. Aaron Henry
Lecture Location: LASR 104 Or you can attend online - without penalty – via Panopto.
Lecture Times: Tues & Thurs 12:30-2:00

Office Hours: Wed & Fri 1:30-2:30
Office: Over Zoom: https://ubc.zoom.us/j/61631323776?pwd=amg3emRuamVkUmpiejVaSFVrVzFpQT09

Email: aaron.henry@ubc.ca

E-mail policy: E-mails must be sent from your UBC e-mail address and must include the course code (PHIL351) in the subject line. E-mails are for administrative purposes only – questions about course material will be addressed during office hours. I aim to reply to e-mails within one or two business days.

TEACHING ASSISTANTS

Tyeson Daviesbarton: tdaviesb@mail.ubc.ca
Spencer Knafelc: knafelcs@mail.ubc.ca

2. COURSE

OVERVIEW

With the release of large language models like ChatGPT, there has been renewed public debate about the prospects of Artificial General Intelligence (AGI) and its existential risks. While not always recognized as such, many of these debates are philosophical in nature. Among other things, they require us to ask:

- *What is it* for a system to be intelligent and to engage in intelligent activities like thinking, reasoning, and understanding?
- When are we *justified* in concluding that a system is intelligent? For example, do large language models like ChatGPT merely imitate intelligent language usage or do they manifest genuine understanding?
- What are the *ethical consequences* of the answers we give to the above questions for the ways we use large language models and related computing technologies?

We will approach these metaphysical, epistemological, and ethical questions by considering a suite of interlocking debates within the philosophy of cognitive science regarding the mechanistic underpinnings of human cognition. Our focus will be on the extremely prevalent thesis within cognitive science that human cognition is a species of *computation* and on why this thesis licenses the theoretical project of building AGI. We will examine various ways one might understand the claim that cognition is computation and various conceptual and empirical challenges that have been posed to this thesis, including from proponents of 'embodied' cognition (or '4E cognitive science'). We will also ask how the outcome of these debates about the metaphysics of intelligence and cognition have for the project of designing 'ethical AI.'

LEARNING OBJECTIVES

A primary objective of any philosophy course is skill development. Some of the skills you will be developing in this course include:

- ability to read, analyze, and critically assess a philosophical text;

- ability to defend your views, both in writing and in conversation.

In addition, you will acquire:

- grasp of some of the central problems and controversies in philosophy of mind and philosophy of cognitive science/artificial intelligence.

EXPECTATIONS

What I expect from you:

- to complete assignments on time and according to the instructions;
- to treat your peers with respect;
- to ask questions and seek help when you don't understand something;
- to take responsibility for your own learning.

What you can expect from me and your TAs:

- to promote a positive and stimulating learning environment;
- to provide support throughout the term;
- to offer constructive feedback on your work;
- to treat you with respect;
- to think carefully about your questions and make a serious effort to answer them.

TEXTS

All readings will be available through the course website (see §4 for details). However, we will draw frequently on the following books (all of which are available electronically through the UBC library system):

Cain, M. J. (2016). *The philosophy of cognitive science*. Polity.

Dietrich, E., Fields, C., Sullins, J. P., Van Heuveln, B., & Zebrowski, R. (2021). *Great Philosophical Objections to Artificial Intelligence: The History and Legacy of the AI Wars*. Bloomsbury Academic.

Smith, B. C. (2019). *The Promise of Artificial Intelligence: Reckoning and Judgment*. MIT Press.

3. ASSESSMENT

Reflection paper	(15% of final)	Oct 6th, 11:59 PM
Midterm	(25% of final)	Oct 19
Final paper	(25% of final)	Dec 8th, 11:59 PM
Final exam	(35% of final)	TBD

4. POLICIES

COURSE WEBSITE

All announcements and course documents will be posted on Canvas. To access this site, go to https://canvas.ubc.ca/and login with your CWLid and password. PHL3510H will appear under the "courses" portion of the welcome page, on the left hand side. Click on the link to access our site. You should check this site regularly for updates.

LATENESS

Assignments will be penalized 1/3 a letter grade for each day that they are late. Extensions may be granted if extraordinary circumstances are documented, but students should contact me to request an extension before the due date. Any assignment that is more than 5 days late will not be accepted.

Academic Integrity

Academic integrity is essential to the pursuit of learning and scholarship in a university, and to ensuring that a degree from the University of British Columbia is a strong signal of each student's individual academic achievement. Accordingly, the University treats cases of cheating and plagiarism very seriously. Plagiarism, which is intellectual theft, occurs where an individual submits or presents the oral or written work of another person as his or her own. Scholarship quite properly rests upon examining and referring to the thoughts and writings of others. However, when another person's words (i.e. phrases, sentences, or paragraphs), ideas, or entire works are used, the author must be acknowledged in the text, in footnotes, in endnotes, or in another accepted form of academic citation. Where direct quotations are made, they must be clearly delineated (for example, within quotation marks or separately indented). Failure to provide proper attribution is plagiarism because it represents someone else's work as one's own. Plagiarism should not occur in submitted drafts or final works. A student who seeks assistance from a tutor or other scholastic aids must ensure that the work submitted is the student's own. Students are responsible for ensuring that any work submitted does not constitute plagiarism. Students who are in any doubt as to what constitutes plagiarism should consult their instructor before handing in any assignments. A link about Academic Misconduct:
http://www.calendar.ubc.ca/Vancouver/index.cfm?tree=3,54,111,959

Use of Generative AI (Artificial Intelligence) Tools

If you make use of generative artificial intelligence tools (e.g., ChatGPT) to complete any course-related work, the generated material must be clearly and correctly indicated, and cited/referenced using an accepted referencing style. Failure to clearly indicate and reference AI-generated material will be reported as academic misconduct. You should consult your instructor if you have any questions about the use of generative AI tools.

Accessibility Needs

The University of British Columbia is committed to accessibility. If you have a disability that may interfere with your ability to successfully take this course, then please email me in the first few weeks. You must also register with Access and Diversity, so that they can help provide support (grant extra time on exam, note taker, etc.):
https://students.ubc.ca/about-student-services/access-diversity

5. Support

Advice

The term goes *very* quickly! It's therefore important to keep up with the lectures and readings. If you feel you need additional help with any of the course material, please don't wait to contact me.

Internet Resources

Here is a link to information about the University of British Columbia's writing resources:
https://writing.library.ubc.ca/
Here are links to information about how to write a good philosophy paper:
http://www.jimpryor.net/teaching/guidelines/writing.html
http://www.public.asu.edu/~dportmor/tips.pdf
http://catpages.nwmissouri.edu/m/rfield/guide.html
Here is a link to information about the 'Cornell' note taking system:
http://lifehacker.com/202418/geek-to-live--take-study+worthy-lecture-notes

6. Schedule

Please note that this schedule may change at the instructor's discretion to suit the pace of the course and the interests of the class.

DATE	TOPIC	READING
Sept 7	Introduction to the course and a philosophical pre-history of cognitive science	Cain, Chapter 1, §§1-3 Cantwell-Smith, 'Introduction' **Optional:** Dietrich et al., Introduction (pp. 1-5) Haugeland, 'The Sage of the Modern Mind' (Chapter 1 of *Artificial Intelligence: The Very Idea*)
Sept 12	Recognizing intelligence when we see it: the Turing Test and ChatGPT	Dietrich et al. (pp. 17-42) **Optional:** Turing, 'Computing machinery and Intelligence' Dennett, 'Intentional systems theory'
Sept 14	Modern information and computation theory and some distinctions in the metaphysics of mind (I)	Cain, Chapter 1 §§4 to end Dietrich et al. (pp. 45-55) Cantwell-Smith, Chapter 1 ('Background') **Optional:** Kind, pp. 76-86
Sept 19	Modern information and computation theory and some distinctions in the metaphysics of mind (II)	No new readings
Sept 21	Classical 'symbolic' cognitive science and Good Old Fashioned Artificial Intelligence (GOFAI)	Cantwell-Smith, Chapter 2 ('History') **Optional:** Bermudez, Chapters 1-2 Newell and Simon, "Computer Science as Empirical Enquiry: Symbols and Search"; Marr, "Vision"
Sept 26	Classical cognitive science continued: The Language of Thought Hypothesis (LOT) and peripheral vs. central cognition	Cain, Chapter 2, §§1-3 **Optional:** Fodor, The persistence of the attitudes' Bermudez, Ch. 8 (pp. 203-210)
Sept 28	From GOFAI to artificial neural networks (ANNs)	Dietrich et al. (pp. 65-73) Cantwell-Smith, Chapter 3 ('Failure')
Oct 3	ANNs cont'd	Cantwell-Smith, Chapter 5 Cain, Chapter 2, §§4-5
Oct 5	ANNs cont'd	Cantwell-Smith, Chapter 6
Oct 10	The problem of meaning (I)	Dietrich et al., Ch. 5 (pp. 87-134) **Optional:** Searle, "Minds, brains and programs" David Cole "The Chinese room" (focus on §4, but you are welcome to read the entire entry)
Oct 12	No class	No class
Oct 17	The problem of meaning (II): the interpretation problem for ANNs	Cappelen & Dever, *Making AI intelligible*, Introduction and Ch. 1
Oct 19	**Midterm**	No readings
Oct 24	The problem of relevance (the 'frame problem')	Dietrich et al., Ch. 6 (pp. 137-169)
Oct 26	The problem of consciousness	Dietrich et al., Ch. 7 (pp. 171-196)
Oct 31	The problem of consciousness (II)	TBD

Nov 2	Bringing it all together: Judgment and reckoning	Smith, Chapters 7-9
Nov 7	Bringing it all together: Judgment and reckoning	Smith, Chapters 10-13
Nov 9	AI Ethics (I): Alignment problems	Dietrich et al. Ch. 8 (197-217)
Nov 14	No class	
Nov 16	AI Ethics (II): Algorithmic bias	Whittaker et al., 'Disability, Bias, and AI' **Optional:** Johnson, 'Algorithmic bias, on the implicit biases of social technology'
Nov 21	4E cognitive science (I): Introducing the 4 E's	Corris and Chemero, 'Embodiment and enactivism'; Shapiro, 'Embodied Cognition'
Nov 23	4E cognitive science (II): conservatives, reformists, and radicals	
Nov 28	4E cognitive science (II): Ecological psychology **Guest Lecture**	Anderson, Chapter 5 of *After Phrenology*
Nov 30	4E cognitive science (III): Language and higher cognition	Anderson, Chapters 6-7 Dietrich et al., pp. 244-55
Dec 5	Back to AI Ethics: Is embodiment the key to ethical AI?	Dietrich et al., the rest of Ch. 9
Dec 7	Wrap up and review	No new readings