INLS 690
PHILOSOPHY AND ETHICS OF AI

Instructor: Fox
Email: foxm@unc.edu
Room: Manning 303
Schedule: M/W 1:25-2:40pm
Office: Zoom, Greenlaw 505
Office Hours: 3:30-5:30pm on Wednesdays or by appointment

DESCRIPTION AND OBJECTIVES

This course covers such topics as the Turing test and the frame problem in the philosophy of AI, superintelligence and the so-called singularity—when AI becomes uncontrollable and independent, when it governs and improves itself—the existential risk of Artificial General Intelligence (AGI), the moral status of AI, the design of ethical AI, and the possible implications of merging AI with humans. We will engage with leading philosophers and ethicists in the field, reflect on stories by Isaac Asimov and Harlan Ellison, and critique the films *2001: A Space Odyssey* and *Her*. We will also explore what some consider to be myths about AI. Overall, this course will teach you how to think critically about AI from philosophical, ethical, and imaginative perspectives.

TEXTS

All readings for this course are available under Course Reserves on Canvas.

REQUIREMENTS (AND GRADE DISTRIBUTION)

Participate (20%)

This course will be conducted as a seminar, so your participation in class discussion will be integral to its success. Please come to class well-prepared to discuss the readings. Before each class period, you'll submit a brief response to that class period's readings to a Canvas discussion board. You can address part of a reading or the reading as a whole. The purpose of these responses is to help you grasp what you've read and to help prepare the class for in-class discussion.

Your participation grade will be based mostly on the quality of your commentary on Canvas and in-class. Try to have something to say each class period.

I'll provide participation grades at the end of the semester.

Reflect (30%)

You'll submit 2 1000-word pieces that will each critically or philosophically engage with one or more of the texts or films included in a particular week's list of materials.

Write (40%)

For your term paper, which should be 2500-3000 words, you have several options:

1) An extended critical or philosophical engagement of a topic from the course or a related topic
2) A well-researched magazine article in the style of an article from Nautilus Magazine that addresses some aspect of the current or future state of AI and its philosophical, ethical, or cultural implications
3) A critical essay on the representation of AI in literature or film that addresses philosophical, ethical, or other cultural concerns

Present (10%)

At the end of the semester, you'll present a lightning talk, about 5 minutes long, on your term paper.

GRADING SCALE

Undergraduate

A	> 92
A-	90-92
B+	88-89
B	83-87
B-	80-82
C+	78-79
C	73-77
C-	70-72
D+	68-69
D	63-67
D-	60-62
F	< 60

Graduate

H – High Pass
L – Low Pass
P – Pass
F – Fail
I – Incomplete

HONOR CODE

Your full participation and observance of the University's Honor Code are expected. And you are not permitted to upload any content from this course, including recordings you may be

permitted to make through Accessibility Resources & Service, to the web in any form, including but not limited to Chegg, Course Hero, Coursera, Google Drive, etc. If you post course content, you may be violating my intellectual property rights. If you post your own work from this course, you are allowing sites to profit from your intellectual property. In utilizing web sources to upload or download course content, you risk violating the University's Honor Code.

ACCESSIBILITY RESOURCES AND SERVICE

The University of North Carolina at Chapel Hill facilitates the implementation of reasonable accommodations, including resources and services, for students with disabilities, chronic medical conditions, a temporary disability or pregnancy complications resulting in barriers to fully accessing University courses, programs and activities. Accommodations are determined through the Office of Accessibility Resources and Service (ARS) for individuals with documented qualifying disabilities in accordance with applicable state and federal laws. See the ARS Website for contact information (https://ars.unc.edu) or email ars@unc.edu.

MISCELLANEOUS

I may make changes to the syllabus as necessary over the course of the semester.

SCHEDULE

A * indicates a due date.

Jan. 9	**Overview and Introductions**
Jan. 11	**The Turing Test** Turing. "Computing Machinery and Intelligence." *PEAI*
Jan. 16	**No class (MLK Day)**
Jan. 18	**The Chinese Room Thought Experiment 1** Searle. "Minds, Brains, and Programs." *PEAI*
Jan. 23	**The Chinese Room Thought Experiment 2** Boden. "Escaping from the Chinese Room" *PEAI*
Jan. 25	**Mind as Computer** Searle. "Is The Brain a Digital Computer?" *Proceedings and Addresses of the American Philosophical Association*
Jan. 30	**The Frame Problem 1** Dennett, "Cognitive Wheels: The Frame Problem of AI." *PEAI* Optional reading: Hanahan, "The Frame Problem." *The Stanford Encyclopedia of Philosophy*
Feb. 1	**The Frame Problem 2** Searle, "The Background of Meaning, *Speech Act Theory and Pragmatics*
Feb. 6	**Superintelligence and the Singularity 1** Chalmers, "The Singularity: A Philosophical Analysis," *Journal of Consciousness Studies*

Feb. 8	**Superintelligence and the Singularity 2** Goertzel. "Should Humanity Build a Global AI Nanny to Delay the Singularity Until Its Better Understood?" *Journal of Consciousness Studies*
Feb. 13	**No class (Well-being Day)**
***Feb. 15**	**The Singularity and the Computer Simulation Hypothesis** Prinz, "Singularity and Inevitable Doom." *Journal of Consciousness Studies* ***Response Paper 1 due*** Optional reading: Bostrom, "Are We Living in a Computer Simulation?" *Philosophical Quarterly*
Feb. 20	**Artificial General Intelligence (AGI) and Existential Risk 1** Bostrom. "The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents. Minds and Machines." *Mind and Machines* Optional reading: Omohundro. "The Basic AI Drives." *Proceedings of the AGI08 Workshop.* Armstrong. "General Purpose Intelligence: Arguing the Orthogonality Thesis." *Analysis and Metaphysics*
Feb. 22	**Artificial General Intelligence (AGI) and Existential Risk 2** Häggström. "Challenges to the Omohundro–Bostrom Framework for AI Motivations." *Foresight* Optional reading: Goertzel. "Superintelligence: Fears, Promises and Potentials." *Journal of Evolution and Technology*
Feb. 27	**Moral Status of AI 1** Asimov. "The Bicentennial Man." *The Bicentennial Man and Other Stories.*
Mar. 1	**Moral Status of AI 2** Anderson. "The Unacceptability of Asimov's Three Laws of Robotics as a Basis for Machine Ethics." *Machine Ethics*. Optional reading: Michael LaBossiere (2017). Testing the Moral Status of Artificial Beings; or I'm Going to Ask You Some Questions." *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*
Mar. 6	**Ethical AI 1** Allen et al. "Prolegomena to Any Future Artificial Moral Agent." *Journal of Experimental & Theoretical Artificial Intelligence* Optional reading: Awad et al. "The Moral Machine Experiment." *Nature*
Mar. 8	**Ethical AI 2**

	Yampolskiy et al. "Safety Engineering for Artificial General Intelligence. *Topoi*
Mar. 13	**No class (Spring Break)**
Mar. 15	**No class (Spring Break)**
***Mar. 20**	**Artificial You 1** Schneider. Introduction-Chapter 2, *Artificial You* ***Response Paper 2 due***
Mar. 22	**Artificial You 2** Schneider. Chapters 3-4, *Artificial You*
Mar. 27	**Artificial You 3** Schneider. Chapters 5-7, *Artificial You*
Mar. 29	**Artificial You 4** Schneider. Chapter 8 and Conclusions, *Artificial You*
Apr. 3	**Workshop** Come prepared to discuss ideas for your term paper.
Apr. 5	**Literature** Ellison, "I Have No Mouth and I Must Scream."
Apr. 10	**Film 1** *2001: A Space Odyssey*
Apr. 12	**Film 2** *2001: A Space Odyssey*
Apr. 17	**Film 3** *Her*
Apr. 19	**Film 4 (cont'd)** *Her*
Apr. 24	**Presentations 1**
Apr. 26	**Presentations 2**
***May 1**	***Term Paper due***