Artificial Intelligence in Society:
New Challenges, Ethical Futures

(Draft of August 20, 2025, for AI Literacy Working Group)

Course description

Emerging artificial intelligence technologies create new challenges for traditional ideas about human ethics. By exploring how artificial intelligence creates specific problems, students learn to think about possible solutions. They do that by learning traditional ways of thinking about ideas like right action, goodness, personal character, rights and responsibilities, and fairness and justice. They observe how these ideas interact with artificial intelligence in specific Cases, and they practice logical reasoning about what to do with AI using those concepts. They explore where our society might need new ethical ideas, and they think about how to build a promising future society that includes AI.

Learning objectives

Content-specific learning objectives

- Students learn how major ethical concepts are defined and supported by investigating how emerging technologies exemplify traditional ethical challenges.
- Students recognize new, possible challenges to what we value arising from current and possible artificial intelligence.
- Students develop familiarity with the terrain of arguments and counter-arguments about the value and threats of artificial intelligence.
- Students develop ideas about ethical, beneficial, and sustainable use of and relationships to artificial intelligence.

Learning objectives for Philosophy courses

- Objective 1a: Students give accurate and relevant answers, complete with supporting details, to specific questions about philosophical ideas relevant to the course.
- Objective 1b: Students give accurate accounts of philosophical ideas relevant to the course in the context of criticizing or assessing those ideas.
- Objective 2a: Students speculate, in well-informed, well-supported, and plausible fashion, about what a given philosopher would say about a novel issue or problem.
- Objective 2b: Students extrapolate creatively and plausibly from their knowledge of philosophers or philosophical positions in developing their own related ideas.
- Objective 3a: Students write paragraphs that exhibit clarity, focus, a good command of the subject matter, and an orderly development of ideas.
- Objective 3b: Students write multi-paragraph pieces that exhibit clarity, focus, a good command of the subject matter and an ability to work with that subject matter creatively, and an orderly development of ideas both within and across paragraphs.

- Objective 4a: Students speak in clear, focused, well-informed, and orderly fashion.
- Objective 5a: Students state arguments accurately and clearly, and identify strengths and weaknesses of different arguments.
- Objective 5b: Students develop and defend their own arguments, taking into account a variety of philosophical positions but adding original insights or emphases.
- Objective 7a: Students explain difficult passages clearly, accurately, and thoroughly.
- Objective 7b: Students use apt quotations and creative, critical, plausible readings of texts in their writing.
- Objective 8a: Students are able to explain the weaknesses of their own present positions, and the strengths of competing positions.
- Objective 8b: Students are able to explain why their pre-theoretical commitments have or have not changed as a result of what they have learned in the course, and if they have changed how they have done so.

Weekly topics

Week 1: What do we mean by "artificial intelligence"?

- A brief history of AI; the main varieties of AI
- What does "intelligence" mean? In what sense are AIs intelligent?

Week 2: Introduction to ethics and value theory

- Major approaches to reasoning about rightness and wrongness, goodness and badness: (a) consequentialist approaches, (b) right action approaches, (c) character approaches, (d) justice and fairness approaches

Week 3: Privacy and big data sets

- Ways of thinking about the value of privacy
- Case: high school student data collection
- Case: automated surveillance

Week 4: Responsibility when AI reasoning is hidden

- Ways of thinking about responsibility
- What aspects of AI activity are unobservable, even to the creators?
- How do we think about responsibility when unobservable processes yield objectionable results?

Week 5: Self-controlled systems

- Case: are self-driving cars safer or a menace?
- Case: do armed robots battles save soldiers' lives?
- Case: AI out of control

Week 6: Automated decision-making, fairness and bias

- Ways of thinking about fairness and avoiding bias
- How can AI systems introduce biases?
- Case: automated student admissions
- Case: bail hearings and automated assessment of recidivism probability

Week 7: AIs manipulating and deceiving people

- Ways of thinking about the problems with manipulation and deception
- Case: AI systems evading our cognitive defenses
- Measures for avoiding AI deception and manipulation

Week 8: Should we create or ban super-intelligences?

- What are the risks and benefits of independent, super-human intelligences?
- Examples of technology and research bans and their histories

Week 9: What would give an AI rights?

- What are the criteria for being morally considerable?
- What would a robot or AI need to do to earn rights in our society?

Week 10: How does AI challenge authorship and intellectual property?

- Approaches to creators' rights in ethics and law
- Case: training large language models on text and images
- Case: original works produced by generative AI

Week 11: How might AI change labor and employment

- Ways of thinking about rights, and possibility of a right to work
- Arguments for automating work as the key to the future
- Case: industries being replaced by automation

Week 12: How AI changes communication: slop, fakes, deepfakes, and trust

- Ways of thinking about the moral significance of trust in communication
- Ways of thinking about freedom of speech and expression
- Case: news slop
- Case: political deepfakes

Week 13: Should we work to preserve our humanity in a world with AI?

- Ways of thinking about justice, agency, and dignity
- What are possible challenges to global justice from AI?
- Does AI offer ways to rethink our human potential?

Week 14: AI and our environment

- Why would we ever care about our environment? — main arguments
- Kinds of environmental risks of AI activity and power consumption
- AI in the service of environmental preservation
- Visions of sustainable AI