

# AI Ethics

A sample syllabus by Tiina C Rosenqvist

[This syllabus outlines an introductory course in AI ethics.]

## 1. Course Description

This course offers an exploration of the ethical issues that arise from the development and deployment of artificial intelligence (AI) across various fields. We begin by covering foundational topics in both philosophy and AI, including how to construct, evaluate, and engage with ethical arguments in a rapidly evolving technological landscape. From there, we will explore key ethical frameworks and see how these frameworks help us critically analyze and address real-world AI challenges.

Central to the course are topics like algorithmic fairness, where we examine how AI systems can introduce or exacerbate bias in areas such as criminal justice and healthcare. We will investigate privacy concerns related to AI-driven surveillance technologies and the ethical dilemmas they create. Additionally, we'll address the growing threat of online manipulation, focusing on how AI can influence behavior and decision-making through tailored content, deepfakes, and other tactics, as well as the ethics of generative AI.

The course ends with a close reading of Shannon Vallor's *The AI Mirror*. We will critically engage with her reflections on how AI technologies mirror and transform human values, ethics, and our understanding of ourselves in a digitally driven world.

## 2. Course Objectives

Through active participation in this course, you will:

- **Gain a comprehensive understanding of AI ethics:** you will explore key issues in AI ethics, including algorithmic fairness, the implications of surveillance technologies, and the ethical challenges posed by online manipulation and generative AI.
- **Apply ethical frameworks to real-world cases:** through readings, class activities, and assignments, you will develop the tools necessary to critically assess the ethical implications of AI technologies. By the end of the course, you will contribute thoughtfully to ongoing discussions in this rapidly evolving field.
- **Enhance your overall critical thinking skills:** you will strengthen your ability to think critically about complex philosophical issues, identify underlying assumptions, and construct effective counterexamples.
- **Refine your argumentation skills:** you will practice constructing and evaluating philosophical arguments, both orally and in writing. You will learn to create well-organized essays that effectively convey your arguments and analyses, articulate your ideas with clarity, and engage thoughtfully with the perspectives of others.

### 3. Course Structure

#### *In-person sessions:*

The standard format for the in-person sessions is lecture and discussion. There will also be occasional group work. You are expected to participate actively in discussions and other in-class activities. At the end of the course, you will present a 5-minute talk on your final paper.

#### *Out-of-class activities:*

Outside of class, you will write two papers, complete a minimum of three module quizzes, and collaborate on a group poster and project. Additionally, you are expected to finish the assigned readings before the relevant class sessions..

### 4. Assignments & Evaluation

#### *Individual assignments (total: 65%)*

- Attendance & Participation—10%
- Module Quizzes (3) —15%
- Paper 1 (3-4 pages) —10%
- Paper 2 (4-5 pages) —25%
- 5-Minute Talk - 5%

#### *Group assignment (35%)*

- Poster —10%
- Project —25%

#### *Attendance and Participation*

Attendance is mandatory, but you are allowed to miss two classes without prior notice, except on the day when you are scheduled to give a group presentation. If you anticipate missing more than two classes, prior notice is required and you will be asked to complete a make-up assignment. Failure to provide notice may result in a deduction from your participation grade. Repeated lateness can also negatively impact your participation grade.

Active participation in discussions and class activities is expected. This requires completing assigned readings beforehand and coming to seminar prepared with ideas and questions.

#### *Module Quizzes*

There will be quizzes at the end of modules I, III, IV, and V. Only the best three quiz scores count towards your final grade (each of the three is worth 5% of your final grade, so together they're worth 15%). This means that you can either not take one of the four quizzes, or you can take all four and have the lowest score dropped (the recommended option).

The quizzes are untimed and open-book, and you have two attempts for each quiz (your **latest** score is the one that is recorded; this means that your recorded score can go down if your second attempt results in a lower score than the first).

### *Papers*

You are required to write two papers. The first paper (3-4 pages) is worth 15% of your final grade and the second paper (4-5 pages) is worth 25%. Prompts will be provided.

### *Group Poster + Project*

Your group will collaborate to create a poster that focuses on a specific issue in AI ethics, which can take the form of an infographic.

Additionally, you will work together on a project that may or may not relate to your poster topic. Possible options include conducting a recorded interview with an expert (using thoughtfully crafted questions), producing a scripted podcast or animation, or writing an op-ed. This project should be well-researched, incorporating additional scholarly sources, and demonstrate critical thinking and a deep engagement with the concepts and tools covered in the course. Your projects will be presented to the class during our group project showcase.

## 5. PHILSKILLS Workshops

Throughout the semester, we will focus on developing and refining the essential skills required for success in this course. These workshops, labeled “**PHILSKILLS**” on the class schedule, will help you practice the following: reconstructing and evaluating arguments, reading and interpreting philosophical texts, writing clear and persuasive philosophy papers, engaging in thoughtful discussions and debates on philosophical topics.

Whether this is your first philosophy course or you have previous experience, these workshops are designed to enhance your skills and deepen your understanding. Don't worry if you are new to philosophy; I do not expect you to already know how to do these things. For those with a background in philosophy, there's always room for growth!

## 6. Course Policies

[omitted; see other syllabi]

## 7. Resources

[omitted; see other syllabi]

## 8. Class & Readings Schedule

### Module I: Preliminaries

.....

In this module, we will explore some foundational topics: What is philosophy, and how do we construct, reconstruct, and evaluate arguments? What exactly is AI, and where is it applied today? Can morality be objective, and what kinds of questions do AI ethicists ask?

.....

(1) INTRODUCTION TO PHILOSOPHY; *PHILSKILLS: What (even) is an argument*

(2) WHAT (EVEN) IS AI?

- **Required:** Heaven (2024): “What is AI?”

(3) AI APPLICATIONS

- **Required:** Xu et al. (2021): “Artificial Intelligence: A powerful paradigm for scientific research”

(4) METAETHICS: MORALITY & OBJECTIVITY

- **Required:** Street (2015): “Does anything really matter or did we just evolve to think so?”

(5) WHAT (EVEN) IS AI ETHICS?; *PHILSKILLS: How to read philosophy*

- **Required:** Stahl (2021): “Concept of ethics and their application to AI”

*Module I Quiz*

### Module II: Ethical Theories

.....

This module introduces us to key ethical theories and invites us to explore how AI brings these theories to life.

.....

(6) CONSEQUENTIALISM

- **Required:** Singer (1972): “Famine, affluence, and morality”

(7) DEONTOLOGY

- **Required:** O’Neill (2003): “A Kantian approach to famine relief”

(8) VIRTUE ETHICS

- **Required:** Hursthouse (1996): “Normative virtue ethics”

(9) RELATIONAL ETHICS

- **Required:** Metz and Clark Miller (2013): “Relational ethics”

(10) AI ETHICS AND VALUES

- **Required:** Gabriel (2020): “Artificial Intelligence, Values, and Alignment”
- **Optional:** Zeynep Tufekci (2016): “Machine intelligence makes human morals more important” (TED Talk)

- (11) REVIEW & DISCUSSION, *PHILSKILLS: How to talk philosophy*
- **Required:** LeGuin (1973): “The ones who walk away from Omelas”

*Module II Quiz*

### Module III: Algorithmic Fairness

.....

We’ll start the module with an overview of machine learning algorithms, followed by an exploration of algorithmic bias in criminal justice and healthcare.

.....

- (12) WHAT ARE MACHINE LEARNING ALGORITHMS
- **Required:** Piano (2020): “Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward”
- (13) ALGORITHMIC BIAS IN CRIMINAL JUSTICE
- **Required:** “Algorithms were supposed to reduce bias in criminal justice - do they?”
- (14) RACIST ALGORITHMS?
- **Required:** Benjamin (2019): “Default Discrimination: Is the Glitch Systemic?”
- (15) ALGORITHMIC FAIRNESS IN MEDICINE
- **Required:** Chen et al. (2023): “Algorithmic fairness in artificial intelligence for medicine and healthcare”
- (16) RELATIONAL ETHICS AND ALGORITHMIC FAIRNESS
- **Required:** Birhane (2021): “Algorithmic injustice: a relational ethics approach”
- (17) CATCH-UP / REVIEW; *PHILSKILLS: How to write a philosophy paper*

*Paper 1*

### Module IV: Surveillance, Manipulation, and Generation

.....

In this module, we will investigate privacy concerns related to AI-driven surveillance technologies and the ethical dilemmas they create, address the growing threat of online manipulation, and explore the ethics of generative AI.

.....

- (18) PRIVACY & SURVEILLANCE
- **Required:** Selinger & Leong (2021): “The ethics of facial recognition technology” (excerpts)
  - **Optional:** “AI & Surveillance – Striking The Balance For Ethical Deployment”
- (19) ONLINE MANIPULATION
- **Required:** Botes (2023): “Autonomy and the social dilemma of online manipulative behavior”

(20) DEEPFAKES

- **Required:** Diakopoulos and Johnson (2019): “Anticipating and Addressing the Ethical Implications of Deepfakes in the Context of Elections” (excerpts)

(21) GENERATIVE AI

- **Required:** Zohny et al. (2023): “Ethics of generative AI”

(22) REVIEW / GROUP POSTER

*Module IV Quiz  
Poster*

## Module V: Humanity in the Age of AI

.....

In this module, we will read Shannon Vallor's *The AI Mirror*, and critically engage with her reflections on how AI technologies mirror and transform human values, ethics, and our understanding of ourselves in a digitally driven world.

.....

(23) AI AND HUMANITY

- **Required:** Vallor (2024): *The AI Mirror*, Chs. 1-2

(24) AI AND HUMANITY

- **Required:** Vallor (2024): *The AI Mirror*, Ch. 3

*Group Project*

(25) AI AND HUMANITY

- **Required:** Vallor (2024): *The AI Mirror*, Chs. 4-5

(26) AI AND HUMANITY

- **Required:** Vallor (2024): *The AI Mirror*, Chs. 6-7

*Module V Quiz*

## Module VI: Wrapping up

.....

The last module is dedicated to showcasing your learning and ideas.

.....

(27) 5-MINUTE TALKS

(28) GROUP PROJECT SHOWCASE

*Paper 2*