

APHI 213 – Ethics and Philosophy of Artificial Intelligence

“...The ultimate engineering task is the fragile, endless, and sublime human project of using the culture we produce to make ourselves into the beings we wish to become.” (S. Vallor)

Instructor: Alessandra Buccella (Dr. B) – abuccella2@albany.edu

Teaching Assistant: Yulin Wu - ywu28@albany.edu

Class time and place: Tuesdays & Thursdays, 10:30 – 11.50 AM, HU 124.

Office hours:

DR. B: Mondays, 10am – 12 pm, HU 253 & by appointment

YULIN (TA): Wednesdays, 3 – 4 pm, HU 255

Course description

Have you heard that artificial intelligence will solve climate change, cure all illnesses, and make poverty a thing of the past? Or have you heard that AI is going to cause the extinction of humanity and destroy the planet? The truth is that AI is transforming our world in ways very few people expected and fully understand. Where there is little understanding, extreme reactions are natural, and the ‘right’ or ‘ethical’ thing to do becomes harder to figure out. So, gaining an understanding of AI and its effects on human existence will be our primary focus in this course. We will start from an overview of the theoretical foundations of AI, and then we will explore its ethical-social implications. We will think about what kind of abilities current AI systems have and about whether they can exercise these abilities in ways that benefit humanity and the world more generally. We will also discuss whether AI’s achievements are being accurately portrayed by the media and what can be done to make the general public more aware of both AI’s benefits and its risks.

This course is a Humanities General Education course.

After taking this course, you will be able to:

- Understand how philosophy works when thinking about AI.
- Spot and think deeply about ethical and philosophical problems in AI.
- See how AI reflects the culture of its creators.
- Know why ethics still matters in AI today.
- Understand and question the ideas behind AI ethics.
- Communicate clearly and think critically.

Readings

All readings and resources (links to videos, podcasts etc.) are available on the Brightspace page of the course, divided by week.

Some readings in this class are optional. Check the schedule regularly to make sure you don’t skip any of the non-optional ones!

What to do to succeed in this course:

- **COME TO CLASS**, be on time, don't leave early unless it is an emergency. If you must miss a class or part of a class, please let me know in advance.
- **ACTIVELY PARTICIPATE AND ENGAGE**. This can be done in many ways; you don't have to raise your hand and speak up in class if you don't feel comfortable doing it.
- **TRY TO BE KIND AND RESPECTFUL**. Why is respect important? And what does "respect" mean to you? Let's talk about it!
- **COMMUNICATE AND BE RESPONSIVE**. If something I say in class isn't clear, ask. If you are falling behind, do not be afraid to reach out. If I email you, reply promptly (I will do the same!). If you have doubts about how to complete an assignment or have questions about what's expected of you, share those with me right away.
- **TAKE CARE OF YOURSELF**. Semesters are long and taxing, and you will naturally have ups and downs. If you notice your mental health isn't great at any point in the semester, try your best not to isolate yourself: keep coming to class, and communicate with me so I can help you stay on track with your coursework.

Assignment 'Bundles':

Completing each 'bundle' will result in a certain grade at the end.

TO GET A 'D' (NO FINAL EXAM):

- Attend 50% or more of the classes
- Get a D or higher on the midterm exam
- Complete the Debate questionnaire
- Choose one between a written personal reflection OR a non-written final project.

TO GET A 'C' (NO FINAL EXAM):

- Attend 50% or more of the classes
- Get a C or higher on the midterm exam
- Complete the Debate questionnaire
- Choose one between a written personal reflection OR a non-written final project.

TO GET A 'B':

- Attend 75% or more of the classes
- Get a B- or higher on the midterm exam
- Complete the debate questionnaire
- Choose one between the written personal reflection OR a non-written final project
- Get a C+ or higher on the final exam

TO GET AN 'A':

- Attend 75% or more of the classes
- Get a A- or higher on the midterm exam
- Complete the debate questionnaire
- Complete both the written personal reflection AND a non-written final project
- Get a B+ or higher on the final exam

Deadlines

The midterm and final exams will be on particular days, but the 'deadline' for the written reflection and non-written project simply aligns with the 'natural' end of the semester (12/16, last day of Finals Week). The debate questionnaire is due on 11/25 (see schedule below).

Religious Commitments

If you think your religious commitments may impact your course work at any point during the semester, please let me know as soon as possible so that we can discuss the best strategy for accommodating your needs.

Academic Integrity

All forms of academic dishonesty, including but not limited to cheating and plagiarism (i.e. using the words of other people or AI bots as your own), will not be tolerated and will result in failing the course. As UAlbany students, you make a commitment to respecting the Standards of Academic Integrity Policy. You should familiarize yourselves with the policy and its subsections at <https://www.albany.edu/dean-students/community-standards/standards-academic-integrity> and take advantage of the resources offered by the University Libraries regarding cheating, plagiarism, and how to properly cite your sources.

A note about Generative AI use (ChatGPT etc.)

Generative AI bots like ChatGPT, Gemini, Claude, etc. *do not* replace your own ability to think. Indeed, extensive use of ChatGPT has been recently linked to declining academic performance, increased procrastination, and memory loss (<https://tinyurl.com/2byaafxh>).

In this course, using the words of AI bots like ChatGPT as your own on *any* assignment is considered cheating. As such, it will result in being failed for the course and reported to the appropriate UAlbany authorities.

AI bots can however be used to *assist* you in completing assignments by brainstorming ideas, summarizing the readings, composing 'cheat sheets' to study for the exams, etc. We will spend some time discussing appropriate and inappropriate uses of chatbots and we'll agree on some ground rules as a group.

(Dis)ability and Learning

If you have a disability and would like to request accommodations, you should get in touch with Disability Access and Inclusion Student Services (DAISS) at <https://www.albany.edu/dean-students/disability>. All information and documentation you provide is confidential.

Even if you do not require specific accommodations, please let me know if there is anything I can do (within reason) to facilitate your learning.

Electronics in the classroom

You are allowed to use computers or tablets in the classroom to take notes if you so choose, but be aware that they can be sources of distraction and interfere with your (and others') ability to follow

what's going in class. You are not as good at multitasking as you might think:
<https://time.com/4737286/multitasking-mental-health-stress-texting-depression/>

You are encouraged to take notes by hand, if you are able (especially in preparation for the midterm and final exams).

Communications

I will communicate with you via email and through Brightspace, so make sure you check your university email (username@albany.edu) regularly. I will try to answer your emails within 24 hours (excluding weekends): if you don't receive an answer from me by then, feel free to reach out again!

Assignments Descriptions/Instructions

Midterm exam & Final exam:

In-class, hand-written exams. During the exams, you are allowed to consult your own hand-written notes *only*. Exceptions will be made for students requiring relevant accommodations.

Written personal reflection:

This is a 2-3 pages (double-spaced, 11 or 12 font size) free-form personal reflection on the following prompt:

Choose one example of how AI is impacting your own life. Did this course change any of your beliefs or attitudes towards it? If you were in charge of that particular AI application, what changes would you make?

Submission through Brightspace.

Assignment-specific AI policy: For this assignment, any use of AI chatbots should be *disclosed* through an additional document submitted alongside the paper. In the additional document, you should explain which chatbot you used, when, and for what (e.g. generating initial ideas, improving clarity, assistance with summarizing and explaining concepts, etc.). Screenshots of your actual exchanges with the chatbot should also be included when appropriate.

Non-written final project:

This can be almost anything other than a written essay: a poem, a musical composition, an artwork, a video, a podcast episode, an infographic or poster, and more.

There is only one 'rule': the project must express your views on AI and its impacts on your personal life.

Submission through Brightspace.

Assignment-specific AI policy: For this assignment, any use of AI chatbots should be *disclosed* through an additional document submitted alongside the project in which you explain which chatbot you used, when, and for what (e.g. generating initial ideas, improving clarity, assistance with summarizing and explaining concepts, etc.). Screenshots of your actual exchanges with the chatbot should also be included when appropriate.

Debate questionnaire:

Answer the questions before class on 11/25. The link to access the questionnaire is on Brightspace (Content > Week 13).

Grading systems and conversions

→ I do not grade on a curve. I make this choice for several reasons that align with what I believe teaching and learning should be about. If you are curious about what my specific reasons for not curving grades are, feel free to ask!

The midterm and the final exams will be graded on a 60-point scale, with each question being worth 10 points. The letter grade will be based on the percentage (see below).

The debate questionnaire, the personal statement, and the non-written project will be pass/fail.

Conversion table for letter grades:

A: 93-100
A-: 90-92
B+: 87-89
B: 83-86
B-: 80-82
C+: 77-79
C: 73-76
C-: 70-72
D+: 67-69
D: 63-66
D-: 60-62
E: < 60

Tentative schedule and readings

(Optional readings are clearly marked)

Week 1: Introduction to the course

8/26 Introductions, course ground rules, initial discussion

8/28 Initial discussion continued

Week 2: What is AI, *really*?

9/2 Functionalism, Computationalism, and the origins of AI

9/4 Artificial Neural Networks and Machine Learning

Week 3: Artificial ‘dumbness’?

9/9 Deep Learning and Adversarial examples

Optional reading: Buckner, “Deep learning: A philosophical introduction”

9/11 Large Language Models and Meaning

Optional readings: Milliére & Buckner, “A philosophical introduction to Large Language Models, part 1”

Bender et al., “On the dangers of stochastic parrots: Can language models be too big?”

Week 4: What's AI got to do with ethics?

9/16 The different meanings of "AI ethics"

9/18 Discussion of practical AI ethics cases

Week 5: AI, empathy, and emotions

9/23 Read: Inzlicht et al., "In praise of empathic AI"

9/25 *Choose one of the following to either read or listen to:*

Read: Turkle, "Who do we become when we talk to machines?"

<https://mit-genai.pubpub.org/pub/uawlth3j/release/2>

Bloom, "AI is about to solve loneliness: that's a problem"

<https://www.newyorker.com/magazine/2025/07/21/ai-is-about-to-solve-loneliness-thats-a-problem>

Listen: "Love in the time of Replika" (Warning: it contains non-explicit sexual content)

<https://hiphination.org/season-6-episodes/s6-episode-3-love-in-time-of-replika/>

Week 6: AI responsibility

9/30 Read: *AI Ethics*, ch. 8

Optional reading: Bryson, "Robots should be slaves"; Listen: Bryson on Why Robots Should Be

Slaves: [https://philosophicaldisquisitions.blogspot.com/2017/06/episode-24-bryson-on-why-](https://philosophicaldisquisitions.blogspot.com/2017/06/episode-24-bryson-on-why-robots-should.html)

[robots-should.html](https://philosophicaldisquisitions.blogspot.com/2017/06/episode-24-bryson-on-why-robots-should.html)

10/2 Discussion continued.

Week 7: Midterm week

10/7 *Midterm exam review*

10/9 *Midterm exam*

Week 8: AI bias

10/14 *No class (Fall break)*

10/16 Read: *AI Ethics*, ch. 9

Optional readings: Johnson, "Algorithmic bias: on the implicit biases of social technology"

Birhane, "Algorithmic injustice: a relational ethics approach"

Week 9: AI fairness

10/21 What is "fairness"?

Optional reading: Binns, "Fairness in Machine Learning: Lessons from Political Philosophy"

10/23 What is "equality"?

Week 10: It's all in the data!

10/28 Introducing the idea of 'Big data'

Optional reading: Cukier & Mayer-Schoenberger, "The Rise of Big Data: How It's Changing the Way We

Think About the World"

10/30 The 'quantified self'

Week 11: Trust and transparency

11/4 What does it mean to 'trust' AI?

Optional reading: Ryan, "In AI we Trust: Ethics, Artificial Intelligence, and Reliability"

11/6 The 'cult' of transparency

Week 12: AI and social justice

11/11 Read: Green, “Data Science as Political Action: Grounding Data Science in a Politics of Justice”

11/13 Read: Buccella, “AI for all is a matter of social justice”

Week 13: Examining our priorities: What do we want AI for?

11/18 Read: *AI Ethics*, ch. 12.

11/20 Humans don’t know how to desire

Week 14: The ‘Existential risks’ of AI

11/25 Discussion on debate (Watch debate and complete questionnaire *before class!*)

<https://www.youtube.com/watch?v=2xyQktWIRNg&t=18s>

11/27 *No class (Thanksgiving break)*

Week 15:

12/2 *Guest speaker (TBD)*

12/4 *Final exam review*

12/11 *Final exam (10:30 am – 12:30 pm)*

12/16 *Last day to submit written reflections and non-written final projects.*