AI Ethics for Technology Leaders

EGN 6933/AITL?29530
Class Periods: Wednesdays, Periods 9,10,11, 4.05-7.05PM
Location: remote/zoom
Academic Term: Spring 2025

Instructor:
Sonja Schmer-Galunder
s.schmergalunder@ufl.edu
415.604.6293
Office Hours: Thursdays, 4.30-5.30PM ET, MALA, room 4021 and zoom: https://ufl.zoom.us/j/98717368862

Teaching Assistant/Peer Mentor/Supervised Teaching Student:
Please contact through the Canvas website

- Divyansh Singh, Divyansh.singh@ufl.edu, Mondays, 3.30-5.30PM ET, MALA student village or zoom: https://ufl.zoom.us/my/divyanshmeetingroom

Course Description
This elective 3-credit graduate course for MS students in Data Science (MSADS) and AI systems (MSAIS), as well as graduate CS students. It will equips future technology leaders with a broad understanding of ethical considerations when developing and deploying Al systems and models. Students learn about critical core concepts of Al ethics and learn to apply them to real-world scenarios where leaders may face ethical challenges. The course also looks at the impact of Al within a global context, and how to navigate complex issues while respecting diverse social and cultural values. Further, a big focus of the course is on Al safety. Educating students about beneficial and safe Al may be one of the most important topics of today, and this course aims to provide up-to-date knowledge on current social issues.

Course Pre-Requisites / Co-Requisites
Open to all graduate students. Priority will be given to MS students enrolled in MSADS and MSAIS.

Course Objectives

We live in what is perhaps the most exciting but also challenging period in human history. AI is reshaping the world, bringing both opportunities and challenges for technology leaders. You learn to navigate the ethical decisions you may have to make by introducing you to important domains related to harms and risks, AI safety, accountability, fairness, transparency, explainability and social impact. You learn to:

- Analyze AI systems' technical foundations and evaluate their societal implications through multiple theoretical frameworks. Critique AI's effects on social structures, economic systems, and human behavior.
- Evaluate AI's differential impacts across global markets, diverse societies, and local communities. Synthesize ethical frameworks to formulate and defend complex technological decisions.
- Apply philosophical and ethical frameworks to align AI systems with human values. Analyze historical case studies of technological challenges, examining their impacts on individual and societal wellbeing. Evaluate ambiguous real-world scenarios through varied socio-cultural lenses.
- Analyze value tradeoffs in AI development, particularly regarding safety, responsibility, and transparency. Create ethical frameworks for AI development and deployment. Design strategies to influence organizational and public policy toward beneficial AI outcomes.

Materials and Supply Fees
None

Required Textbooks and Software

- Ethics, Technology, and Engineering: An Introduction
- Ibo van de Poel, Lamber Royakkers
- March, 2011, 1st or 2nd edition, 1st edition is available online: https://cdn.prexams.com/6229/BOOK.pdf
- ISBN: 978-1-444-39571-6 (if course notes derived from various published sources are used, provide information above for each source)

Recommended Materials

- BBC Newsnight: "The trolley problem and ethics of driverless cars" – https://www.youtube.com/watch?v=FypPSJfCRFk (5 minutes)
- Awad, Edmond, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. "The moral machine experiment." *Nature* 563, no. 7729 (2018): 59-64, https://core.ac.uk/download/pdf/231922494.pdf & https://www.moralmachine.net/
- Wiener, Norbert. "Some Moral and Technical Consequences of Automation (1960)." (2021), https://www.cs.umd.edu/users/gasarch/BLOGPAPERS/moral.pdf
- Yong, Ed. "A Popular Algorithm Is No Better at Predicting Crimes Than Random People," *The Atlantic*, January 17, 2018. (1200 words, 5 min) https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/.
- Larson, Jeff, Mattu, Surya, Kirchner, Lauren and Angwin, Julia "How we Analyzed the COMPAS Recidivism Algorithm", ProPublica, 2016 https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92. https://dl.acm.org/doi/pdf/10.1145/3458723
- Raghavan, B. and Schneier, B. "Seeing Like a Data Structure": https://www.belfercenter.org/publication/seeing-data-structure
- Satariano, Adam, and Paul Mozur. "The People Onscreen Are Fake. The Disinformation Is Real." The New York Times, February 7, 2023. https://www.nytimes.com/2023/02/07/technology/artificial-intelligence-training-deepfake.html.
- Crawford, Kate, and Vladan Joler. "Anatomy of an AI System." Anatomy of an AI System (2018). https://anatomyof.ai/ and NEW! https://calculatingempires.net/ (Use the audio-guide to explore the map!)
- Winner, Langdon. "Do artifacts have politics?." In Computer Ethics, pp. 177-192. Routledge, 2017. https://faculty.cc.gatech.edu/~beki/cs4001/Winner.pdf
- Atari, Mohammad, Mona J. Xue, Peter S. Park, Damián Blasi, and Joseph Henrich. "Which humans?." (2023), https://hmpa.hms.harvard.edu/sites/projects.iq.harvard.edu/files/culture_cognition_coevol_lab/files/which_humans_09222023.pdf
- Hoffman, Mia, "The EU AI Act: A Primer," Center for Security and Emerging Technology, September 26, 2023. https://cset.georgetown.edu/article/the-eu-ai-act-a-primer/
- Roose, Kevin. "Bing's A.I. Chat: 'I Want to Be Alive. 😈,'" February 16, 2023. https://www.nytimes.com/2023/02/16/technology/bing-chatbot-transcript.html
- Chiang, Ted, "ChatGPT is a blurry JPEG of the Web," The New Yorker, Feb 9, 2023. https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web (3300 words, 15 minutes)
- Robinson-Early, Nick, "AI's 'Oppenheimer moment': autonomous weapons enter the battlefield", The Guardian, July 2024, https://www.theguardian.com/technology/article/2024/jul/14/ais-oppenheimer-moment-autonomous-weapons-enter-the-battlefield
- Russell, Stuart. "If We Succeed." Daedalus: AI and Society, Spring 2022. https://www.amacad.org/publication/if-we-succeed
- Kasirzadeh, A. (2024). Two Types of AI Existential Risk: Decisive and Accumulative. arXiv preprint, https://arxiv.org/pdf/2401.07836
- World Economic Forum, "The Future of Jobs Report", https://www.weforum.org/publications/the-future-of-jobs-report-2023/

- Graeber, D. “On the Phenomenon of Bullshit Jobs: A Work Rant. https://web.archive.org/web/20180807024932/http://strikemag.org/bullshit-jobs/
- Behrend, Tara S., Daniel M. Ravid, and Cort W. Rudolph. "Technology and the changing nature of work." Journal of Vocational Behavior (2024): 104028, https://www.sciencedirect.com/science/article/pii/S0001879124000691?casa_token=gUXe962I8ZwAAA

Required Computer

UF student computing requirement: *https://news.it.ufl.edu/education/student-computing-requirements-for-uf/*

Course Schedule

Week	Date	Topic	Description
1	1/15/2025	Introductions & Goal Setting	In this session we will introduce the course, study materials, course structure, grading, motivation and goals. Students will learn about the course objectives, meet each other, instructors and learn basic definitions, concepts and terminology.
2	1/22/2025	AI Ethics and Leadership	We talk about AI Ethics as a field and why it is important. We discuss various theories relevant to the ethical development and use of technology and AI. We discuss and define what ethical leadership means and think about important tools for future AI leaders to make ethical decisions for the deployment of technologies.
3	1/29/2025	Artificial Intelligence and the Problem of Unintended Consequences	In this class we talk about unintended consequences, harms and risks from AI systems and what technology leaders should think about. We talk about various types of problems that can arise when optimizing for objective functions.
4	2/5/2025	Bias: Human vs. Algorithmic	We learn about various types of biases. Why are AI systems biased and what can we do about it? How can we measure bias? How can we mitigate biases and improve fairness?
5	2/12/2025	Data: What is data?	In this session students will learn about the role of data in AI with a focus on ethical considerations, e.g. representation, inclusion, bias, consent, processing, etc. We will discuss data annotation processes and concerns related to public information, as well as trade-offs between data quality and quantity.
6	2/19/2025	Values and Norms	We take a deeper dive into norms, values and virtues. We learn about deductive and non-deductive arguments, fallacies and risks in ethical discussions, and the value alignment problem.
7	2/26/2025	Privacy and Transparency	We learn about the ethical questions related to privacy and privacy protection.
8	3/5/2025	Truth and Misinformation	We will talk about mis- and disinformation, and the importance of truth and trust. We learn about the use of algorithms that are optimized to influence people, and technology to make information more trustworthy and safe.
9	3/12/2025	Society: The Politics of Algorithms	In this class we will learn about the global context for the use and deployment of AI, including, geopolitical impacts and national security risks.
SPRING BREAK - NO CLASS			

10	3/26/2025	Culture: Human diversity in AI systems	In this session we talk about the importance of understanding the global market, value pluralism, why human diversity is important and how AI is a form of collective intelligence.
11	4/2/2025	Generative Models I: Large Language Models	We will talk about what generative models are, how they work and what type of ethical and social issues they pose. We will also discuss the risks and potential of LLMs and tradeoffs between open-source and closed models.
12	4/9/2025	Generative Models II: AI Safety and Catastrophic Risk	In this session we will discuss potential catastrophic risk from AI. We talk about Artificial General Intelligence (AGI) and different schools of thought related to AI Safety.
13	4/16/2025	Regulation and Governance of AI	We learn about the history and current state of AI regulation, why it is important to follow and understand AI regulation in the US and elsewhere.
14	4/23/2025	The Status and Future of AI in the workplace	In the last class we will discuss the future of the job market and how AI will transform future societies.
15	4/30/2025	Finals: Project Work	Students will present projects they have been working on throughout the semester. The goal of the project is to show how ethical consideration will make AI development more fair, beneficial and safe.

Important Dates
3/12/2025 Final Paper Teaming, Topic, Abstract and Outline due
4/30/2025 Final Paper Presentation

Attendance Policy, Class Expectations, and Make-Up Policy

Class Participation (30%): All students need to attend class each week. If you cannot attend, please contact the TA (cc the instructor) immediately by email. Everyone is expected to participate verbally every class session, coming prepared, having done all assigned readings/media, and contributing thoughtful ideas to all class discussions. Students are expected to be prepared for cold calls during class.
Class Quizzes (15%): We will administer a short quiz at the beginning of class each week related to the reading material and media you had to prepare before class.
In-Class Presentation (25%): Each student will present "AI News" once during the semester. This is a 5-10 min news update meant to look at current developments in AI through an ethical lens.
Break-Out Assignments: You will have the chance to work on use cases in small teams during breakout rooms. It is important that each student has their camera and microphone on and participates throughout the breakout session.
Final Class Presentation (30%): During the last week of the semester students will present their final papers on one of the topics of the syllabus. The presentation should not be longer than 10 min, with 5 min questions.

Scheduling Conflicts:
Please notify us by email (divyansh.singh@ufl.edu) by the second week of the term about any known or potential extracurricular conflicts (such as religious observances, interviews, or other activities). We will try our best to help you with making accommodations, but cannot promise them in all cases. In the event there is no mutually-workable solution, you may be dropped from the class.

Attendance:

Attendance is **mandatory. One absence per semester** does not affect your grade. This includes doctor appointments, family emergencies, recruiting sessions, etc. If you have a second absence for a reason beyond your control, please email Divyansh Singh (divyansh.singh@ufl.edu), and cc the instructor (s.schmergalunder@ufl.edu) and submit a note from the doctor, etc., specifying the date of class and that you can't attend. **Showing up at the beginning of class and leaving in the middle will result in zero attendance points** unless you explain why in advance via email. If you explain in advance, you can get a half-point if you stay for more than half the class.
Excused absences must be consistent with university policies in the Graduate Catalog (https://catalog.ufl.edu/graduate/regulations) and require appropriate documentation. Additional information can be found here: https://gradcatalog.ufl.edu/graduate/regulations/

Students **must** be prepared to have their **CAMERAS ON** during Zoom sessions, with appropriate attire and backgrounds. Let the instructor know ahead of time if you have a technical issue.

Cheating:
Anyone caught cheating will receive a failing grade and will also be reported to the University Office of Student Conduct.

Plagiarism/Self-plagiarism/Use of AI tools:
You must be original in composing your work in this class. To copy text or ideas from another source (including your own previously, or concurrently, submitted coursework) without appropriate reference is plagiarism and will result in a failing grade for your assignment and usually further disciplinary action. For additional information on plagiarism, self-plagiarism, and how to avoid it, see, for example: https://gradadvance.graduateschool.ufl.edu/media/gradadvancegraduateschoolufledu/OGPD_Plagiarism_Workshop_20221019.pdf

Note that cheating on exams and plagiarism are examples of violations in the realm of ethics and integrity. Honesty, integrity, and ethical behavior are of great importance in all facets of life.

Use of AI tools (based on Mollick): During this course, we may ask you to use AI (ChatGPT, or a similar tool provided by UF etc), and you may decide to use these tools on your own. **When you use AI tools, you must acknowledge them, and not include work created by a tool (even as bullets in your presentations, for example) without mentioning it.** We ask that you add a final slide to each of your presentations explaining what you used AI for and what prompts you used to get the results. You will still be responsible for errors in the work. **We may use Turnitin or other similar tools to compare your writing to existing work.** Be aware that information provided by AI tools can be false. You are responsible to provide correct information and original source citations of your work.

State whether attendance is required and if so, how will it be monitored? What are the penalties for absence, tardiness, cell phone policy, laptop policy, etc. What are the arrangements for missed homework, missed quizzes, and missed exams?
This statement is required:
Excused absences must be consistent with university policies in the Graduate Catalog (https://catalog.ufl.edu/graduate/regulations) and require appropriate documentation. Additional information can be found here: https://gradcatalog.ufl.edu/graduate/regulations/

Evaluation of Grades

Assignment	Total Points	Percentage of Final Grade
Attendance and Participation (Assignments)	100	20%
Quizzes (2 question beginning of class)	100	20%

In-class presentations (AI news)	100	15%
Final Paper	100	20%
Final Presentation	100	25%
		100%

Grading Policy

The following is given as an example only.

Percent	Grade	Grade Points
93.4 - 100	A	4.00
90.0 - 93.3	A-	3.67
86.7 - 89.9	B+	3.33
83.4 - 86.6	B	3.00
80.0 - 83.3	B-	2.67
76.7 - 79.9	C+	2.33
73.4 - 76.6	C	2.00
70.0 - 73.3	C-	1.67
66.7 - 69.9	D+	1.33
63.4 - 66.6	D	1.00
60.0 - 63.3	D-	0.67
0 - 59.9	E	0.00

More information on UF grading policy may be found at:
UF Graduate Catalog
Grades and Grading Policies

Students Requiring Accommodations

Students with disabilities who experience learning barriers and would like to request academic accommodations should connect with the disability Resource Center by visiting https://disability.ufl.edu/students/get-started/. It is important for students to share their accommodation letter with their instructor and discuss their access needs, as early as possible in the semester.

Course Evaluation

Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available at https://gatorevals.aa.ufl.edu/students/. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via https://ufl.bluera.com/ufl/. Summaries of course evaluation results are available to students at https://gatorevals.aa.ufl.edu/public-results/.

In-Class Recording

Students are allowed to record video or audio of class lectures. However, the purposes for which these recordings may be used are strictly controlled. The only allowable purposes are (1) for personal educational use, (2) in connection with a complaint to the university, or (3) as evidence in, or in preparation for, a criminal or civil proceeding. All other purposes are prohibited. Specifically, students may not publish recorded lectures without the written consent of the instructor.

A "class lecture" is an educational presentation intended to inform or teach enrolled students about a particular subject, including any instructor-led discussions that form part of the presentation, and delivered by any instructor hired or appointed by the University, or by a guest instructor, as part of a University of Florida course. A class

lecture does not include lab sessions, student presentations, clinical presentations such as patient history, academic exercises involving solely student participation, assessments (quizzes, tests, exams), field trips, private conversations between students in the class or between a student and the faculty or lecturer during a class session.

Publication without permission of the instructor is prohibited. To "publish" means to share, transmit, circulate, distribute, or provide access to a recording, regardless of format or medium, to another person (or persons), including but not limited to another student within the same class section. Additionally, a recording, or transcript of a recording, is considered published if it is posted on or uploaded to, in whole or in part, any media platform, including but not limited to social media, book, magazine, newspaper, leaflet, or third party note/tutoring services. A student who publishes a recording without written consent may be subject to a civil cause of action instituted by a person injured by the publication and/or discipline under UF Regulation 4.040 Student Honor Code and Student Conduct Code.

University Honesty Policy

UF students are bound by The Honor Pledge which states, "We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honor and integrity by abiding by the Honor Code. On all work submitted for credit by students at the University of Florida, the following pledge is either required or implied: "On my honor, I have neither given nor received unauthorized aid in doing this assignment." The Honor Code (https://sccr.dso.ufl.edu/process/student-conduct-code/) specifies a number of behaviors that are in violation of this code and the possible sanctions. Furthermore, you are obligated to report any condition that facilitates academic misconduct to appropriate personnel. If you have any questions or concerns, please consult with the instructor or TAs in this class.

Commitment to a Safe and Inclusive Learning Environment

The Herbert Wertheim College of Engineering values varied perspectives and lived experiences within our community and is committed to supporting the University's core values, including the elimination of discrimination. It is expected that every person in this class will treat one another with dignity and respect regardless of race, creed, color, religion, age, disability, sex, sexual orientation, gender identity and expression, marital status, national origin, political opinions or affiliations, genetic information, and veteran status.

If you feel like your performance in class is being impacted by discrimination or harassment of any kind, please contact your instructor or any of the following:

- Your academic advisor or Graduate Coordinator
- HWCOE Human Resources, 352-392-0904, student-support-hr@eng.ufl.edu
- Pam Dickrell, Associate Dean of Student Affairs, 352-392-2177, pld@ufl.edu
- Toshikazu Nishida, Associate Dean of Academic Affairs, 352-392-0943, nishida@eng.ufl.edu

Software Use

All faculty, staff, and students of the University are required and expected to obey the laws and legal agreements governing software use. Failure to do so can lead to monetary damages and/or criminal penalties for the individual violator. Because such violations are also against University policies and rules, disciplinary action will be taken as appropriate. We, the members of the University of Florida community, pledge to uphold ourselves and our peers to the highest standards of honesty and integrity.

Student Privacy

There are federal laws protecting your privacy with regards to grades earned in courses and on individual assignments. For more information, please see: https://registrar.ufl.edu/ferpa.html

Campus Resources:

Health and Wellness

U Matter, We Care:

Your well-being is important to the University of Florida. The U Matter, We Care initiative is committed to creating a culture of care on our campus by encouraging members of our community to look out for one another and to reach out for help if a member of our community is in need. If you or a friend is in distress, please contact umatter@ufl.edu so that the U Matter, We Care Team can reach out to the student in distress. A nighttime and weekend crisis counselor is available by phone at 352-392-1575. The U Matter, We Care Team can help connect students to the many other helping resources available including, but not limited to, Victim Advocates, Housing staff, and the Counseling and Wellness Center. Please remember that asking for help is a sign of strength. In case of emergency, call 9-1-1.

Counseling and Wellness Center: https://counseling.ufl.edu, and 392-1575; and the University Police Department: 392-1111 or 9-1-1 for emergencies.

Sexual Discrimination, Harassment, Assault, or Violence
If you or a friend has been subjected to sexual discrimination, sexual harassment, sexual assault, or violence contact the **Office of Title IX Compliance**, located at Yon Hall Room 427, 1908 Stadium Road, (352) 273-1094, title-ix@ufl.edu

Sexual Assault Recovery Services (SARS)
Student Health Care Center, 392-1161.

University Police Department at 392-1111 (or 9-1-1 for emergencies), or http://www.police.ufl.edu/.

Academic Resources

E-learning technical suppor*t*, 352-392-4357 (select option 2) or e-mail to Learning-support@ufl.edu. https://elearning.ufl.edu/.

Career Connections Center, Reitz Union, 392-1601. Career assistance and counseling; https://career.ufl.edu.

Library Support, http://cms.uflib.ufl.edu/ask. Various ways to receive assistance with respect to using the libraries or finding resources.

Teaching Center, Broward Hall, 392-2010 or 392-6420. General study skills and tutoring. https://teachingcenter.ufl.edu/.

Writing Studio, 302 Tigert Hall, 846-1138. Help brainstorming, formatting, and writing papers. https://writing.ufl.edu/writing-studio/.

Student Complaints Campus*:* https://sccr.dso.ufl.edu/policies/student-honor-code-student-conduct-code/;https://care.dso.ufl.edu.

On-Line Students Complaints*: https://distance.ufl.edu/getting-help/; https://distance.ufl.edu/state-authorization-status/#student-complaint.*