

INF 385T.12 Ethics of AI: Theorizing Good Systems

28770

Fall 2023

Instructor: Ken Fleischmann

Pronouns: he/him/his

Email: kfleisch@ischool.utexas.edu

Office hours: By appointment

Academic Assistant (AA): Ayodele Oduba

Pronouns: he/him/his

Email: aoduba@utexas.edu

Office hours: Tuesdays & Thursdays 1 - 1:30pm

Course Description

COURSE DESCRIPTION

Artificial intelligence is becoming an increasingly important fixture in our everyday lives. This course will help you to understand the ethical challenges related to AI and how to address them. To do this, the course will provide you with a broad range of ethical theories, including non-Western and feminist theories, and teach you to apply these theories to contemporary ethical challenges resulting from AI.

LEARNING OUTCOMES

1. You will learn about the history of AI and the ethical challenges that arise from recent advances in AI
2. You will gain an awareness of a wide range of ethical theories and learn to apply them to the ethics of AI
3. You will learn about efforts to develop principles for the design of ethical AI and develop your own view

How Will You Learn?

STATEMENT OF LEARNING SUCCESS

The ethics of AI is a timely, important, and expansive topic that defies the boundaries of space and time. Although ethics of AI was considered a fairly narrow, minor, and esoteric topic when I began pursuing research on this topic over 20 years ago, today it is understood as a critical issue across academia, industry, government, and nonprofits. As a result, demand for this course has expanded, leading me to offer this course as an online, asynchronous course that can scale up to meet the increased demand. I hope that this course will provide a launchpad for everyone to begin to explore the ethics of AI, and I will be happy to support your learning during and beyond the semester in any ways that I can. I hope the flexibility of the videos and online discussions will give you space and time to grapple with these important issues and help fit the course into your busy schedules.

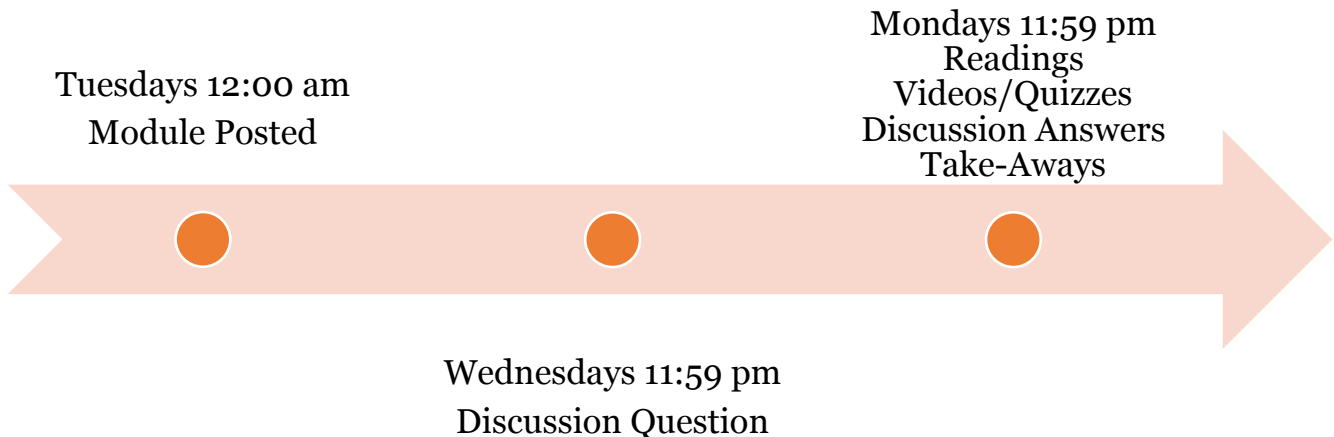
TEACHING MODALITY INFORMATION

This course is fully asynchronous, and work can be done at any point during the week, although it is important to note that topics are covered on a week-by-week basis, and there are multiple consistent deadlines during each week. The deadlines are Wednesdays for discussion questions (which can be submitted before or after completing the readings and watching the videos) and Mondays for quizzes, discussion answers, and take-aways. During each week, you should first post your discussion question, then complete the readings, then watch the videos (including completing the embedded quizzes), then post your answers, then post your take-aways. The goal is to ensure that you are actively engaged in learning each week, as this engagement will assist you in retaining this knowledge in the future. The quizzes and other elements of the course are intentionally designed to be as straightforward as possible – the goal of the course is to assist and support you in your learning journey, and to build a learning community that can facilitate everyone’s learning experience.

WEEKLY ACTIVITIES



WEEKLY SCHEDULE



COMMUNICATION

The course Canvas site can be found at utexas.instructure.com. Any time you have any questions or need to provide any updates, please email both the instructor and the AA via Canvas. You are responsible for ensuring that the primary email address you have recorded with the university is the one you will check regularly for course communications because that is the email address that Canvas uses. Due to the asynchronous, online nature of the course, it is critical that you regularly check Canvas and your Canvas-linked email account.

ASKING FOR HELP

Whenever you have any questions or need help, please email both the instructor and the AA via Canvas. Please list your question or the topic in your email. Most simple inquiries can be figured out via email, so in those cases, please provide enough information for us to be able to reach a prompt resolution. If your question is complex or otherwise likely requires setting up a meeting, please list several available meeting times, to facilitate the process of finding a mutually convenient time to meet. So, please do not just send generic emails asking for help or to meet, please make sure to list the topic of your inquiry as well as enough detail to be able to reach a prompt resolution of the inquiry (simple and straightforward questions or updates) or several available meeting times to facilitate setting up a mutually convenient time (for more complex topics that require synchronous discussion).

DIVERSITY, EQUITY, AND INCLUSION

I am deeply committed to doing everything that I can to ensure that all students, with a broad range of intersectional identities and lived experiences, will be well served by this course, that the course will meet students' learning needs, and that the diversity that students bring to this class can be comfortably expressed and viewed as a resource, strength and benefit to all students. Please reach out at any time with any concerns.

DISABILITY & ACCESS (D&A)

The university is committed to creating an accessible and inclusive learning environment consistent with university policy and federal and state laws. Please let me know if you experience any barriers to learning so I can work with you to ensure you have equal opportunity to participate fully in this course. If you are a student with a disability, or think you may have a disability, and need accommodations please contact Disability & Access (D&A). Please refer to the D&A website for more information: <http://diversity.utexas.edu/disability/>. If you are already registered with D&A, please deliver your Accommodation Letter to me as early as possible in the semester so we can discuss your approved accommodations and needs in this course.

Course Requirements and Grading

REQUIRED MATERIALS AND DEVICES

All required materials are available via the course Canvas site. To be successful in this course, you will need a device capable of accessing the course Canvas site, as well as reliable internet access to use Canvas and email.

CLASS EXPECTATIONS

Class participation Since this is an online, asynchronous course, class participation for this course will involve completing all weekly readings, watching all weekly videos and completing the embedded quizzes, and participating in all weekly discussion boards (discussion questions and answers board and take-aways board). Timely completion of all assignments is essential for you to stay on track and be successful in the class.

Professional standards You are individually responsible for completing all assignments on time and independently – everyone will be graded on their own individual work. We are collectively responsible for maintaining a respectful environment in the discussion boards at all times, and treating peers with respect. Discrimination, personal attacks, or hate speech of any kind will not be tolerated, and will have consequences.

ASSIGNMENTS

The following table represents how you will demonstrate your learning and how we will assess the degree to which you have done so.

Assignments	Percent of Total Grade
Videos	25%
Quizzes	25%
Discussion Questions	10%
Discussion Answers	20%
Take-Aways	20%

Bonus Assignments	Points Added to Total
Meet the Instructor (synchronous Zoom, multiple options)	+1
Introductions (asynchronous video, multiple formats permitted)	+1
Mid-Semester Check-In (synchronous Zoom, multiple options)	+1
End-of-Semester Reflections (synchronous Zoom, multiple options)	+1
Final Take-Aways (asynchronous video, multiple formats permitted)	+1

ASSIGNMENT DEADLINES AND LATE WORK PENALTIES

The Canvas gradebook is set up to automatically deduct 10% of the total assignment grade per day, on top of any deductions based on the assignment rubric. Please make sure to turn in all assignments on time. You are responsible for your own time management and setting aside the time to complete class assignments – think of the videos as a way to more flexibly adjust the learning that you would usually do during a set weekly three-hour block to potentially multiple times throughout the week, and just as the readings and assignments for most classes would be in addition to this class contact time, similarly these activities will be in addition to the videos. Please manage your time accordingly, and make sure that you will be able to set aside enough time each week to complete all of the required activities and assignments. You have some flexibility during the week, and you can always choose to turn in assignments earlier during the week, but setting aside time during each week to devote to completing the weekly class activities and assignments is essential. As such, we will not accept excuses such as having a busy week or not getting around to the work in time – in those cases, the late penalty will stand as is. For emergency situations, please work with Student Emergency Services so that they can notify us, allowing us to extend deadlines as needed to accommodate the circumstances.

EXTRA CREDIT

As noted in the table above, there are opportunities to earn extra credit, as points added to your final grade.

VIDEOS (25%)

Videos are critical learning material for this course. You must watch at least 90% of each video to get credit for watching that video. Please make sure to watch videos through Canvas.

QUIZZES (25%)

To ensure that you are completing, paying attention to, and understanding the readings and the videos, the videos will include embedded quizzes. These quizzes will cover content from the readings and/or videos.

DISCUSSION QUESTIONS (10%)

At the start of the week, post a discussion question about any aspect of any topic covered during the week. Asking the question does not require prior preparation; your question does not need to demonstrate a full comprehension of the week's topic(s), just that you are intellectually curious to learn more about the topic(s).

DISCUSSION ANSWERS (20%)

Choose any two classmates' questions and write one-paragraph answers to each. Please make sure to support your answer with evidence from the readings and videos, citing at least two sources per answer (can be informal, and can include the video – for example, you can just say: “as mentioned in the lecture about Deep Blue”).

TAKE-AWAYS (20%)

In a paragraph, describe one thing that you learned this week about the topic (in weeks with multiple topics, please feel free to choose any of the topics). Start with a topic sentence that summarizes your key take-away, then provide at least two supporting examples from different informally cited sources, and end with a conclusion sentence that provides an example of how you might apply this knowledge in the future (to work, life, etc.).

MEET THE INSTRUCTOR (+1)

To receive one bonus point, sign up for and attend a brief informal small group Zoom chat with the instructor. Multiple timing options will be provided – if you have conflicts with all times, please let me know in advance and I will try to add more time options.

INTRODUCTIONS (+1)

To receive one bonus point, record a brief (2-3 minute) video introducing yourself to the class, including your name, your degree program, your prior academic background, your planned career path, and one thing that interests you about the ethics of AI.

MID-SEMESTER CHECK-IN (+1)

To receive one bonus point, sign up for and attend a brief informal small group Zoom chat with the instructor.

END-OF-SEMESTER REFLECTIONS (+1)

To receive one bonus point, sign up for and attend a brief informal small group Zoom chat with the instructor.

FINAL TAKE-AWAYS (+1)

To receive one bonus point, record a brief (2-3 minute) video describing one thing that you learned from the class about the ethics of AI, and describe how you might apply this insight to your studies, career, and/or life.

GRADE CUTOFFS

Required to count toward a graduate degree



Course Outline

All instructions, assignments, readings, rubrics and essential information will be on the Canvas website at utexas.instructure.com. Check Canvas regularly. **Changes** to the schedule may be made at my discretion if circumstances require. I will announce any such changes in class and will also communicate them via a Canvas announcement. It is your responsibility to note these changes when announced, and I will do my best to ensure that you are notified of changes with as much advance notice as possible.

Week	Date	Class Topic(s)	Readings	Assignments
1	8/21-8/28	Introduction	Class Syllabus	Videos/Quizzes 2 Bonuses
Unit 1: History of Ethics and History of AI				
2	8/29-9/4	Indian Ethics	Bilimoria1993.pdf	Videos/Quizzes Discussion Qs/As Take-Aways
		Classical Chinese Ethics	Hansen1993.pdf	
		Babbage's Engines	Schaffer1994.pdf	
3	9/5-9/11	Buddhist Ethics	DaSilva1993.pdf	Videos/Quizzes Discussion Qs/As Take-Aways
		Islamic Ethics	Nanji1993.pdf	
		Dartmouth Conference on AI	Kline2011.pdf	
4	9/12-9/18	Kantian Ethics	ONeill1993.pdf	Videos/Quizzes Discussion Qs/As Take-Aways
		Consequentialism	Hooker2010.pdf	
		Deep Blue	Hankey2021.pdf	
5	9/19-9/25	Distributive Justice	Murray2014.pdf	Videos/Quizzes Discussion Qs/As Take-Aways
		Virtue Ethics	Slote2010.pdf	
		Watson	Chandrasekar2014.pdf	

6	9/26-10/2	Ethics of Care	Held2005.pdf	Videos/Quizzes Discussion Qs/As Take-Aways
		Ubuntu	Ujomudike2016.pdf	
		Autonomous Cars	OliveiraEtAl2021.pdf	
Unit 2: Ethics and Values of AI: Societal Dimensions				
7	10/3-10/9	Human Values and Value-Sensitive Design	CieciuchEtAl2015.pdf FriedmanEtAl2020.pdf ChengFleischmann2010.pdf	Videos/Quizzes Discussion Qs/As Take-Aways
8	10/10-10/16	Codes of Ethics	ACMCodeOfEthics2018.pdf ALACodeOfEthics2021.pdf ALACoreValues2019.pdf Fleischmann2011.pdf	Videos/Quizzes Discussion Qs/As Take-Aways 1 Bonus
9	10/17-10/23	AI Ethics Guidelines	JobinEtAl2019.pdf Hagendorff2020.pdf Hickok2021.pdf	Videos/Quizzes Discussion Qs/As Take-Aways
Unit 3: Ethics and Values of AI: Technical Dimensions				
10	10/24-10/30	Fairness	BuolamwiniGebbru2018.pdf Lee2018.pdf GuoEtAl2019.pdf	Videos/Quizzes Discussion Qs/As Take-Aways
11	10/31-11/6	Accountability	Shah2018.pdf RajiEtAl2020.pdf SlotaEtAl2021.pdf	Videos/Quizzes Discussion Qs/As Take-Aways
12	11/7-11/13	Transparency	FleischmannWallace2005.pdf Miller2019.pdf LarssonHeintz2020.pdf	Videos/Quizzes Discussion Qs/As Take-Aways
13	11/14-11/27	Power	Kalluri2020.pdf CampoloCrawford2020.pdf BirhaneEtAl2022.pdf	Videos/Quizzes Discussion Qs/As Take-Aways
14	11/28-12/4	Agency	Fleischmann2009.pdf JohnsonVerdicchio2019.pdf NeffNagy2018.pdf	Videos/Quizzes Discussion Qs/As Take-Aways 2 Bonuses

Course Policies and Disclosures

HONOR CODE

The University of Texas at Austin strives to create a dynamic and engaging community of teaching and learning where students feel intellectually challenged; build knowledge and skills; and develop critical thinking, creativity, and intellectual curiosity. As a part of this community, it is important to engage in assignments, exams, and other work for your classes with openness, integrity, and a willingness to make mistakes and learn from them. The UT Austin honor code champions these principles:

I pledge, as a member of the University of Texas community, to do my work honestly, respectfully, and through the intentional pursuit of learning and scholarship.

The honor code affirmation includes three additional principles that elaborate on the core theme:

- I pledge to be honest about what I create and to acknowledge what I use that belongs to others.
- I pledge to value the process of learning in addition to the outcome, while celebrating and learning from mistakes.
- This code encompasses all of the academic and scholarly endeavors of the university community.

The honor code is more than a set of rules, it reflects the values that are foundational to your academic community. By affirming and embracing the honor code, you are both upholding the integrity of your work and contributing to a campus culture of trust and respect.

ACADEMIC INTEGRITY EXPECTATIONS

Students who violate University rules on academic dishonesty are subject to disciplinary penalties, including the possibility of failure in the course and/or dismissal from the University. Since such dishonesty harms the individual, all students, and the integrity of the University, policies on academic dishonesty will be strictly enforced. For further information, please visit the Student Conduct and Academic Integrity website at:

<http://deanofstudents.utexas.edu/conduct>

CONFIDENTIALITY OF CLASS RECORDINGS

All work submitted in this class must be your own original, independent work. Plagiarism will not be tolerated. Class recordings are reserved only for students in this class for educational purposes and are protected under FERPA. The recordings should not be shared outside the class in any form. Violation of this restriction by a student could lead to Student Misconduct proceedings.

GETTING HELP WITH TECHNOLOGY

Students needing help with technology in this course should contact the [ITS Service Desk](#) or the Information Commons team by sending an email to help@ischool.utexas.edu.

CONTENT WARNING

Our classroom provides an open space for the critical and civil exchange of ideas. Some readings and other content in this course will include topics that some students may find offensive and/or traumatizing. I'll aim to forewarn students about potentially disturbing content and I ask all students to help to create an atmosphere of mutual respect and sensitivity.

SHARING OF COURSE MATERIALS IS PROHIBITED

No materials used in this class, including, but not limited to, lecture hand-outs, videos, assessments (quizzes, exams, papers, projects, homework assignments), in-class materials, review sheets, and additional problem sets, may be shared online or with anyone outside of the class without explicit, written permission of the instructor. Unauthorized sharing of materials promotes cheating. The University is well aware of the sites used for sharing materials, and any materials found online that are associated with you, or any suspected unauthorized sharing of materials, will be reported to [Student Conduct and Academic Integrity](#) in the Office of the Dean of Students. These reports can result in sanctions, including failure of the course.

ARTIFICIAL INTELLIGENCE

It is a violation of university policy to misrepresent work that you submit or exchange with your instructor by characterizing it as your own, such as submitting responses to assignments that do not acknowledge the use of generative AI tools. Please feel free to reach out to me with any questions you may have about the use of generative AI tools before submitting any content that has been substantially informed by these tools.

RELIGIOUS HOLY DAYS

By [UT Austin policy](#), you must notify me of your pending absence as far in advance as possible of the date of observance of a religious holy day. If you must miss a class, an examination, a work assignment, or a project in order to observe a religious holy day, you will be given an opportunity to complete the missed work within a reasonable time after the absence.

NAMES AND PRONOUNS

Class rosters are provided to the instructor with the student's legal name, unless they have added a chosen name with the registrar's office. If you have not yet done so, I will gladly honor your request to address you with the name and pronouns that you prefer for me to use for you. It is helpful to advise me of any changes or needs regarding your name and pronouns early in the semester so that I may make appropriate updates to my records and be informed about how to support you in this class.

- For instructions on how to add your pronouns to Canvas, visit [this site](#).
- If you would like to update your chosen name with the registrar's office, you can do so [here](#), and reference [this guide](#).
- For additional guidelines prepared by the Gender and Sexuality Center for changing your name on various campus systems, see the Resources page under UT Resources [here](#).

LAND ACKNOWLEDGMENT

I would like to acknowledge that we are meeting on the Indigenous lands of Turtle Island, the ancestral name for what now is called North America. Moreover, I would like to acknowledge the Alabama-Coushatta, Caddo, Carrizo/Comecrudo, Coahuiltecan, Comanche, Kickapoo, Lipan Apache, Tonkawa and Ysleta Del Sur Pueblo, and all the American Indian and Indigenous Peoples and communities who have been or have become a part of these lands and territories in Texas.

Important Safety Information

CARRYING OF HANDGUNS ON CAMPUS

Students in this class should be aware of the following university policies related to Texas' Open Carry Law:

- Students in this class who hold a license to carry are asked to [review the university policy regarding campus carry](#).
- Individuals who hold a license to carry are eligible to carry a concealed handgun on campus, including in most outdoor areas, buildings and spaces that are accessible to the public, and in classrooms.
- It is the responsibility of concealed-carry license holders to carry their handguns on or about their person at all times while on campus. Open carry is NOT permitted, meaning that a license holder may not carry a partially or wholly visible handgun on campus premises or on any university driveway, street, sidewalk or walkway, parking lot, parking garage, or other parking area.
- Per my right, I prohibit carrying of handguns in my personal office. Note that this information will also be conveyed to all students verbally during the first week of class. This written notice is intended to reinforce the verbal notification, and is not a “legally effective” means of notification in its own right.

TITLE IX DISCLOSURE

Beginning January 1, 2020, Texas Education Code, Section 51.252 (formerly known as Senate Bill 212) requires all employees of Texas universities, including faculty, to report to the [Title IX Office](#) any information regarding incidents of sexual harassment, sexual assault, dating violence, or stalking that is disclosed to them. Texas law requires that all employees who witness or receive information about incidents of this type (including, but not limited to, written forms, applications, one-on-one conversations, class assignments, class discussions, or third-party reports) must report it to the Title IX Coordinator. Before talking with me, or with any faculty or staff member about a Title IX-related incident, please remember that I will be required to report this information.

Faculty members and Academic Assistants are [mandatory reporters](#) under federal Title IX regulations and are required to report [a wide range of behaviors we refer to as sexual misconduct](#), including the types of misconduct covered under Texas Education Code, Section 51.252. Title IX of the Education Amendments of 1972 is a federal civil rights law that prohibits discrimination on the basis of sex – including pregnancy and parental status – in educational programs and activities. The Title IX Office has developed supportive ways and compiled campus resources to support all impacted by a Title IX matter.

If you would like to speak with a case manager, who can provide support, resources, or academic accommodations, in the Title IX Office, please email: supportandresources@austin.utexas.edu. Case managers can also provide support, resources, and accommodations for pregnant, nursing, and parenting students.

For more information about reporting options and resources, please visit: <https://titleix.utexas.edu>, contact the Title IX Office via email at: titleix@austin.utexas.edu, or call 512-471-0419.

CAMPUS SAFETY

The following are recommendations regarding emergency evacuation from the [Office of Emergency Management](#), 512-232-2114:

- Students should sign up for Campus Emergency Text Alerts at the page linked above.
- Occupants of buildings on The University of Texas at Austin campus must evacuate buildings when a fire alarm is activated. Alarm activation or announcement requires exiting and assembling outside.
- Familiarize yourself with all exit doors of each classroom and building you may occupy. Remember that the nearest exit door may not be the one you used when entering the building.
- Students requiring assistance in evacuation shall inform their instructor in writing during the first week of class.
- In the event of an evacuation, follow the instruction of faculty or class instructors. Do not re-enter a building unless given instructions by the following: Austin Fire Department, The University of Texas at Austin Police Department, or Fire Prevention Services office.
- For more information, please visit the [Office of Emergency Management](#).

University Resources for Students

For a list of university resources that may be helpful to you as you engage with and navigate your courses and the university, see the [University Resources Students Canvas page](#).