

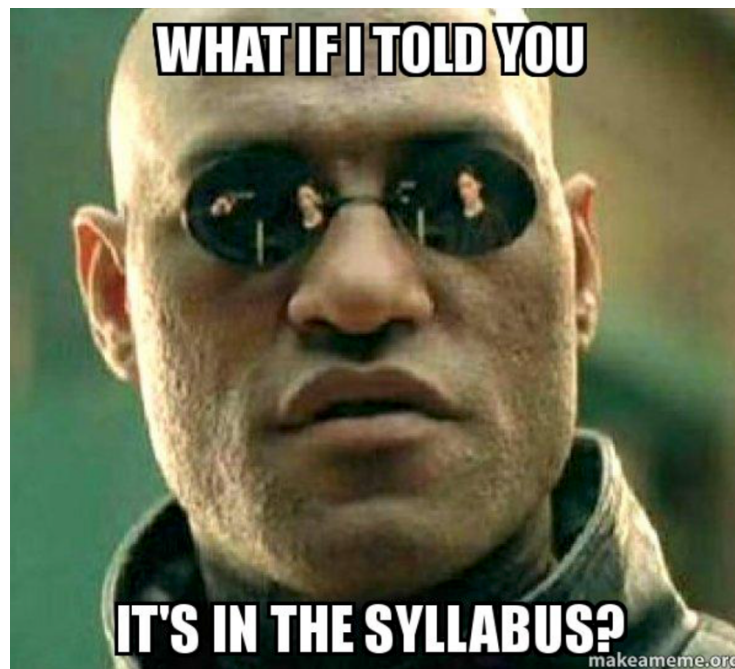
# Ethics, Data, and Technology (PHI 3681)

## Read the syllabus!

---

**Please read this syllabus carefully.** Look, I understand that syllabi are tedious to read. Nobody *wants* to read a syllabus (except maybe the stuff about topics, readings, and assignments). However, you really do need to read this document carefully, as you need to know what it says in order to (a) benefit as much as possible from the course and (b) get a good grade. I promise reading it will be well worth your time!

Besides, if you don't read this syllabus carefully, and you email me to ask a question that is answered explicitly in it, I am going to annoy you by responding with this (admittedly pretty dated) meme:



# General information

---

## Course description

In this course, we will explore questions about how emerging technologies should be designed and regulated. Why do AI systems often behave in unexpected (and sometimes harmful) ways? Should we be concerned about the fact that technology companies such as Facebook gather vast amounts of data about our online activities? What does it mean to say that a machine learning algorithm is a “black box,” and is there something unfair about using such algorithms to decide how to treat people?? As we investigate these and other questions about emerging technologies, we will draw on concepts and readings from a variety of different fields, including philosophy, economics, computer science, data science, and law.

Credit hours: 3

## Meeting times and location

### Lecture

Monday/Wednesday 10:40–11:30 AM in MCCA G186

### Section

**This course has a required discussion section on Fridays (meeting times and locations vary).**

## AI Credit

### Ethical-AI

AI Ethics/Policy: Human-centered considerations (e.g., policy, fairness, accountability, transparency, ethics, safety).

AI course content is over 50%

*This course accomplishes the AI Designation objectives of the subject areas listed above. The course explores a variety of ethical challenges raised by contemporary AI systems, as well as ethical frameworks that can help us think more clearly about those challenges.*

## Teaching team

### Instructor

David Gray Grant  
Assistant Professor of Philosophy  
University of Florida  
[david.grant@ufl.edu](mailto:david.grant@ufl.edu)  
Office hours: Mondays 2:00–4:00 PM  
Office location: Griffin Floyd Hall 301

### Teaching assistants

Name: Dennis Kyrychenko  
Email: [dkyrychenko@ufl.edu](mailto:dkyrychenko@ufl.edu)  
Office hours: Wednesdays 11:45 to 2:45 PM  
Office location: TBD, but almost certainly on the third floor of Griffin Floyd Hall;  
please email Dennis if you would like to meet before the location is announced

Name: Luke McGrath  
Email: [lukemcgrath@ufl.edu](mailto:lukemcgrath@ufl.edu)  
Office hours: Mondays 12:00-3:00 PM  
Office location: Griffin-Floyd 316

## Materials and fees

No required materials or fees. All readings and other course materials will be posted to or linked from Canvas.

# Communication policy

---

## Announcements

**Course announcements will be posted on Canvas. You are responsible for checking Canvas and your UF email at least once a week** (and ideally every day that UF is in session) to make sure that you do not miss important announcements.

## Contacting your TA

**Your teaching assistant is your primary point of contact for questions about assignments, grades, attendance, extensions, and other administrative matters for the course.** Your TA will let you know the best way to contact them.

**Please contact your TA first if you have questions about one of these administrative matters.** If you contact me first about assignments, grades, attendance, or other administrative matters, I will almost certainly have to refer you back to your TA, as they:

1. maintain our attendance records;
2. handle excused absences and assignment extensions;
3. grade your assignments and maintain our grading records; and
4. address clarificatory questions about assignments and course policies.

If you are unable to resolve the issue after speaking with your TA, or if your TA does not respond within two business days, or if the matter is genuinely urgent and you are not sure that your TA can help, please feel welcome to reach out to me directly and I will be happy to assist you.

## Contacting your instructor

**For questions about course content, or for other issues that you are unable to resolve by communicating with your TA, please feel welcome to reach**

**out to me (Dr. Grant) directly by email or by coming to office hours.** I'm here to help!

- Email ([david.grant@ufl.edu](mailto:david.grant@ufl.edu)) and office hours are the most reliable way to get in touch with me outside of class.
- **I make every effort to respond to email from students within two (2) business days.** Please do not contact me again if two business days have not passed unless the matter is genuinely urgent.
- Note that emails do sometimes get lost (due to spam filtering, for instance). **Please feel welcome to send me another email or come up to me after class if you do not hear back within two business days.** I promise I am not ignoring you intentionally!

**When you email me, please (always) specify both (1) the title or course number of the class your question is about and (2) your TA's name.** I regularly teach two entirely different classes with hundreds of students and several TAs, so it will be much easier for me to help you if you provide this information.

## Tech support

**We are unable to provide tech support** for software that you will (or may) be using in this class, such as Canvas, UF email, Perusall, iClicker, Google Drive, etc., as we lack the requisite back-end access to these systems, and are in any case not experts. However, **we do need to know if you are experiencing technical issues** that are interfering with your ability to fully participate in the course.

So, if you are experiencing technical issues, please:

1. Contact the UF Help Desk (<https://it.ufl.edu/helpdesk/>) by phone, chat, or email. (They are available 24/7.)
2. Let your TA know if the issue is interfering with your ability to participate in the class, complete assignments, etc.

# Graded work

---

## Description of graded work

	Description	Points
<b>Perusall</b>	Participation on Perusall	150
<b>Participation</b>	Participation in lectures and discussion sections	150
<b>Case studies</b>	Three small-group activities focused on evaluating a real-world case study	300
<b>Midterm exam</b>	Take-home exam on the material for the first half of the course	200
<b>Final exam</b>	In-class exam on the material for the second half of the course	200

## Description of assignments

### Overview

Your grade in this class will be determined by two things:

1. the quantity and quality of your participation in lecture and section, as well as your annotations of course readings in Perusall; and
2. your performance on several written assignments, which consist largely of short answer and short essay questions.

**Note for non-humanities students:** I recognize that many of you are not humanities students and do not have much (or any) experience with the kind of writing that is expected in philosophy courses. Please rest assured that the course has been designed with you in mind: prior experience with college-level humanities courses is not a prerequisite for success, and hard-working students with no prior exposure to college-level humanities courses routinely earn a full 'A' in the course.

**Note for humanities students:** I also recognize that many of you *are* humanities students and do not have technical backgrounds in computer or data science.

Please rest assured that the course has *a/so* been designed with you in mind: we will presuppose no prior familiarity with the technical topics we will discuss, and hard-working students with no prior exposure to computer or data science routinely earn a full 'A' in the course.

## Perusall

Collaborative annotation of weekly readings on the Perusall online collaborative annotation platform. You will be asked to annotate at most one reading per week (minus three that you will be permitted to skip without penalty). Your grade will be determined by the quantity and quality of your participation in these collaborative discussions over the course of the semester.

Perusall annotations are required and will not be accepted late.

- You can skip three Perusall annotation assignments without penalty.
- Each subsequent Perusall annotation assignment that you fail to submit will result in a 10% deduction from your Perusall grade.

**You are responsible for doing all assigned readings, regardless of whether there is an associated Perusall assignment.**

- Note that while many of the readings for this course are short, they are also very dense—you will need to read them *carefully* and *multiple times* in order to fully absorb the material.

## Participation

A significant portion of your grade will be based on the overall **quantity** and **quality** of your participation in lecture and section. Below is a summary of course standards for different letter grades. (Borderline cases will receive "+" or "-" grades as appropriate.)

- A (100%): Outstanding participation. Highly engaged during lecture and section, with active participation in small group discussions and frequent constructive contributions to full-class discussions. In-class contributions

reflect excellent understanding of and critical engagement with the content of lectures and assigned readings.

- B (86%): Good participation. Moderately engaged during lecture and section, with active participation in small group discussions and occasional constructive contributions to full-class discussions. In-class contributions reflect significant understanding of and critical engagement with the content of lectures and assigned readings.
- C (76%): Minimally adequate participation. Minimally engaged during lecture and section. In-class contributions are infrequent and reflect limited understanding of and critical engagement with the content of lectures and assigned readings.
- D (66%): Poor participation. Attentive during lecture and section, but participating only very infrequently in small-group and full-class discussions. In-class contributions do not reflect understanding of or critical engagement with the content of lectures and assigned readings.
- F (0%): Inadequate participation. Not attentive during lecture and section and/or not participating in either small-group or full-class discussions.

Note that **attendance in lecture and section is required** and having too many unexcused absences will negatively affect your participation grade. See "Attendance requirement" below for details.

## Emerging technology case studies

Three small group exercises focused on evaluating a case study concerning one or more emerging technologies. The exercises will be graded on the basis of a written report. Each member of the group will submit their own version, which will be revised by the group on the basis of in-class discussion.

- Group members will normally receive the same grade, based on the overall quality of the report submitted by the group (see the rubric for written assignments below). However, students that fail to contribute adequately (based on peer evaluations) may be required to re-do the assignment for an individualized grade. Students that repeatedly fail to adequately contribute to group assignments may also lose participation points.
- You will be given time to work on small group exercises during section or lecture. If you are unable to attend the relevant section or lecture, you may be asked to complete the assignment individually.



- Students that do not submit an individual version of the report will receive a zero for the assignment.

### Take-home midterm exam

Take-home exam covering the material from the first half of the course.

### In-class final exam

In-class exam covering material from the second half of the course. The exam will take place in lecture; more details will be provided later in the semester.

- Students with accommodations will be asked to take the final exam at the Disability Resource Center.)

### **Note that the final exam will take place in lecture, not during exam week.**

For each course, UF automatically sets aside a time and location for an in-person final exam, and automatically reports that information to students on one.uf. This course will not have an in-person exam during exam week, so you can disregard the final exam time/location specified for this course on one.uf. Details about the final exam, including the (finalized) date, will be announced on Canvas.

## Attendance requirement

**Attendance in both lecture and discussion section is required, and excessive absences will negatively affect your participation grade.**

- **You are allowed six (6) unexcused absences total without penalty.**  
We do this in part to allow for students to miss class for reasons that, while perfectly legitimate, would not normally qualify for an excused absence (such as travel undertaken for pleasure, rather than to address emergencies or participate in official UF activities).
- **Each unexcused absence beyond the first six will result in a 10% deduction from your total participation grade for the course.**
- Unexcused absences will not be penalized during the first two weeks of class, as we know that enrollment will be in flux.

- If you need to miss lecture or discussion section for any reason, please contact your TA as soon as possible so that you can discuss whether the absence will be excused and what sort of documentation will be necessary (if applicable).
- **If you do not contact your TA about an absence within one week following the absence, it will not be excused except under exceptional circumstances.**
- Poor conduct in class will also result in a 10% deduction. Examples of poor conduct include arriving late, disruptive behavior, falling asleep, failing to participate in small group activities, sending text messages, checking email, etc. See “Other course policies” below for examples of conduct that may result in a penalty.

We will take attendance in two ways:

**1. Lecture attendance will be taken using iClicker.**

- a. You will receive an email asking you to download the iClicker app, which you will use to mark yourself present *before each lecture begins*.
- b. If you are unable to sign in using iClicker for any reason, please let your TA know at the end of class and they will update our attendance records. We understand that iClicker is not perfectly reliable!
- c. You must login to iClicker *using your UF Gatorlink*. When you do, the course should appear automatically in the app’s interface. (If you have just joined the course, it may take a few days for the course to show up—just notify your TA.)

**2. Section attendance will be taken using the Canvas “Roll Call” attendance tool.**

Some notes on attendance records:

- The aforementioned attendance-taking tools will be used to keep track of *all* of your absences, both excused and unexcused.
- Your TA will maintain a separate list of your excused absences, so that we can calculate your total number of unexcused absences from lecture and section at the end of the semester.

- The “Attendance” grade that Canvas automatically generates is not directly related to your participation grade in the class, as discussed below (see “The ‘Attendance’ grade in Canvas”).

## Deadlines and extensions

**All assignment deadlines will be posted on Canvas.**

**If you think you may need an extension, please reach out to your TA to explain the situation and see if you qualify.**

- We are understanding people, so please reach out to your TA if you think you *might* qualify for an extension, even if you are not sure.
- Note that rates of mental illness among college-age adults are at historically high levels (as we will discuss in the course in connection with our discussion of social media algorithms). In general, mental health issues *will* qualify you for an excused absence, even if you do not have an official diagnosis (though we may require you to reach out to the UF Counseling and Wellness Center for assistance and a doctor’s note).

## Verbal preparation rule

**Whenever you submit work for this course, you must be prepared to present it orally, and to answer questions about it.** This will help ensure that you do not submit work that you do not fully understand (either because it was prepared too hastily, or because it was prepared with unauthorized assistance, such as unauthorized generative AI use). It will also help us—and your other instructors with similar policies—demonstrate to employers and others that will be evaluating you in the future that grades at UF mean what they are supposed to mean, in the sense that they are a reflection of students’ actual performance in the course.

**To ensure that you are prepared to orally present each assignment that you submit, we will randomly select several students to orally defend their submissions.** Defenses will take place in either the required discussion sections or your TA’s office hours (as determined by the instructor). Oral defenses will have two parts:

1. *Brief presentation.* The student will give a brief presentation outlining the main ideas contained in their submission.
2. *Brief Q&A.* The student will be asked questions about those ideas by their TA and/or classmates.

During your presentation, you are welcome to have your paper in front of you as a memory aid. You are also welcome to refer to notes as you speak. However, you cannot simply read out a presentation that you prepared beforehand; you should explain your paper informally and largely from memory.

**Oral defenses will be graded pass/fail.** A passing performance simply means demonstrating that you understand what you wrote—that you can explain your argument in your own words and engage thoughtfully with questions. I expect the only failing performances will be cases where the presenter is unable to explain the ideas expressed in their own paper (which should be straightforward if you did the work yourself).

- Since being prepared for a possible oral defense is part of the assignment itself, and since the defense is graded pass/fail, students that pass will not receive extra credit (though they will benefit from the additional feedback that they receive during the defense).
- Students that fail the defense may be required to complete a new, substitute assignment in a controlled environment and without external aids.

**If you suspect that defending your submissions in front of your fellow students would cause you significant anxiety, please let us know at the beginning of the semester,** and we will allow you to conduct your defenses (if any) during office hours.

## Final grades

### Grading scale

The following grade scale will be used to assign final letter grades for the course.

Points	Grade Scale
940 - 1000	94 - < 100 = A
900 - < 940	90 - < 94 = A-
870 - < 900	87 - < 90 = B+
840 - < 870	84 - < 87 = B
800 - < 840	80 - < 84 = B-
770 - < 800	77 - < 80 = C+
740 - < 770	74 - < 77 = C
700 - < 740	70 - < 74 = C-
670 - < 700	67 - < 70 = D+
640 - < 670	64 - < 67 = D
600 - < 640	60 - < 64 = D-
0 - < 600	0 - < 60 = E

## No rounding

**Final grades will not be rounded (so please don't ask).** This means that, to receive a given grade, you must meet or exceed the lower threshold specified; for instance, to get an 'A' you must score at least 940 points. Neither points nor grades will be rounded.

## The Canvas Gradebook

We will report grades and other grade-related information to you on Canvas in two ways:

1. assignment grades will be posted on Canvas and viewable in the Canvas gradebook for our course; and
2. attendance records for the required discussion sections will be maintained using the Canvas Roll Call tool.

Note that Canvas automatically reports some misleading information about grades, as discussed below.

## The “Total” column in Canvas

**Ignore the “Total” grade calculated by Canvas.** During the semester, the “Total” column in Canvas is not a reliable guide to how well you are doing in the course, as it does not take into account your performance on assignments that have not yet been graded. Since many students misinterpret the Total column, I have hidden it on our Canvas gradebook. If you would like to determine how your final grade will depend on your performance on future assignments, you can determine that using a grade calculator or (even better) basic algebra. I encourage you to learn how to do this sort of calculation in an Excel spreadsheet, as you will find that to be a useful skill going forward.

## The “Attendance” grade in Canvas

**The “Attendance” grade calculated by Canvas is not directly related to your grade in the course.** Your TA will use the Canvas Attendance tool to track your attendance in our required discussion sections. Canvas automatically calculates an “Attendance” grade on the basis of the resulting attendance records. However, that grade is not used to calculate your final grade in the course, and is only indirectly related to your participation grade. You should view the Attendance grade reported by Canvas as a rough (and not necessarily reliable) summary of how many total absences you have accrued so far—both excused and unexcused.

- Please see the “Participation” and “Attendance” subsections above for more information about how attendance and participation will affect your final grade in the course.
- Please feel welcome to reach out to your TA if you have any questions about how many unexcused absences you have accrued, their present assessment of the quality and quantity of your participation in the course so far, etc.

# Other course policies

---

## Classroom conduct

Students and faculty each have responsibility for maintaining an appropriate learning environment. Those who fail to adhere to such behavioral standards may be subject to loss of participation points or other penalties (as appropriate).

**In order to avoid unnecessary disruptions to the learning environment, you are expected to conform to the following policies** (in addition to other commonsense and generally accepted norms of classroom conduct that are not explicitly identified below). First-time violations of these policies will receive a verbal warning. Subsequent violations will yield a loss of participation points.

### Side conversations, etc.

**Side conversations (including whispered ones) and other forms of disruptive behavior that make it difficult for others to concentrate are not permitted.**

- While lecture is in progress, you are expected to remain silent unless you have been called on to speak by the lecturer or are participating in a small group activity.
- Note that this is the expectation in most formal settings with large groups (for obvious reasons).

### Timely arrival and departure

You are expected to:

1. **Arrive on time.** Please be seated and ready for lecture to begin by the official start time specified above. This includes checking yourself in with iClicker (if we are using it) and putting your devices away *before* the official start time for the class.
2. **Remain seated and engaged until class officially ends.** Please remain seated and engaged with lecture until the official end time specified above.

Please do feel free (of course) to leave the room to go to the bathroom, address genuine emergencies, etc.

3. **Refrain from packing up early.** Avoiding disruptions includes refraining from packing up your belongings until the official end time specified above or when the instructor announces that the lecture is over (whichever comes first). If you need to leave class early, please let us know (in advance if possible).

We understand that students sometimes need to arrive late or leave early for legitimate reasons. If you need to leave class early, please contact your TA as soon as possible to explain.

## Laptop and Cellphone Policy

**Laptop and cellphone use is prohibited in this class**, in part because it has been shown to negatively affect learning outcomes for *other* students, and not just the student using the device.

- You may use a tablet if it is lying flat and opened to a note-taking app or course reading.
- Please let us know as soon as possible if you have (or think you may need) an accommodation letter from the University permitting you to use a laptop to take notes.

## Accommodations for students with disabilities

Students with disabilities who experience learning barriers and would like to request academic accommodations should connect with the disability Resource Center. [Click here to get started with the Disability Resource Center.](#) **It is important for students with accommodations to share their accommodation letter with their instructor and discuss their access needs as early as possible in the semester.**



# Plagiarism

## Acknowledgement of sources

Per the UF Honor Code and Student Conduct Code, “[s]ubmitting materials from any source without proper attribution” constitutes plagiarism and carries potentially serious penalties. **To avoid plagiarism, all sources and assistance used in preparing your assignments must be precisely and explicitly acknowledged.**

- The web creates special risks here. Cutting and pasting even a few words from a web page or paraphrasing material without a reference constitutes plagiarism. If you are not sure how to refer to something you find on the internet, you can always give the URL.
- You can find a copy of the UF Honor Code and Student Conduct Code here: <https://policy.ufl.edu/regulation/4-040/>.

## Unauthorized generative AI use

**Students are not permitted to use large language models (such as ChatGPT) or other AI-based text generation tools to complete written assignments for this course unless specifically authorized to do so by the relevant assignment prompt.** Submitting work produced using one of these tools without prior written authorization (such as explicit authorization from the assignment prompt) constitutes plagiarism.

I’m sure that none of you would do such a thing intentionally, in part because you want to learn the material we will cover, and in part because you want to develop the (useful and marketable) skills that you will develop by doing the work yourself. However, I have heard that some students believe that unauthorized generative AI use is not detectable, and so will not result in adverse consequences.

While it is of course difficult to prove that a student has used generative AI in an unauthorized way, students that do so are *routinely* found guilty of violating other provisions of the honor code by the Student Conduct and Conflict Resolution Committee. The reason for this is that chatbots have a well-known tendency to both:

1. plagiarize other authors without attribution; and
2. hallucinate sources and citations.

These are much easier to prove, and either is sufficient for a guilty finding for plagiarism as defined by UF. So, please don't do it—for your sake, for our sake, and for the sake of your fellow students! If you have any questions about what constitute unauthorized use of sources, plagiarism, etc., please let us know and we will be happy to discuss it. (You should of course review

### Procedure for suspected cases

**Students that the teaching team suspects have engaged in plagiarism may be required to undergo an oral defense of their submission**, and may be reported to the Student Conduct and Conflict Resolution Committee (SCCR).

**In the event that the SCCR finds that you are guilty of plagiarism or other forms of academic misconduct prohibited by the UF Honor Code and Student Conduct Code, you should expect the minimum penalty to be a grade of E for the class (not just the assignment).**

### Perusall privacy and accessibility policies

You can find Perusall's privacy policy [here](#) and their accessibility statement [here](#).

## UF policies

---

This course complies with all UF academic policies. For information on those policies and for resources for students, please visit [this link](#).

## Course objectives

---

1. Develop a basic vocabulary for discussing the ethical dimensions of data science and its applications.

2. Analyze issues and policies concerning emerging “big data” technologies through the application of ethical concepts.
3. Critique public policies, social practices, and social institutions that shape, and are shaped by, scientific discovery and technological design.
4. Discern the structure of arguments, represent them fairly and clearly, and evaluate them for cogency.
5. Formulate original arguments, anticipate objections, and respond in a conscientious fashion.
6. Read and discuss complex philosophical texts from both historical sources and contemporary works.
7. Speak and write clearly and persuasively about abstract and conceptually elusive matters.

## Tentative schedule and readings

---

**Please note that the following is tentative and subject to change. Please see Canvas for an updated syllabus, as well as for assignment deadlines.**

### Unit 1: The alignment problem

Mon 01/12 · L1 · Introduction

- No readings

Wed 01/14 · L2 · Do artifacts have politics?

- Winner, "Do Artifacts Have Politics?"

Mon 01/19 · No class (holiday)

- No readings

Wed 01/21 · L3 · AI safety

- Amodei et al., "Concrete Problems in AI Safety"
- USC Center for Artificial Intelligence in Society, "Social Networks and Substance Abuse Prevention for Homeless Youth" (webpage with videos)

Mon 01/26 · L4 · Consequentialism I

- Shafer-Landau, "Consequentialism: Its Nature and Attractions"

Wed 01/28 · L5 · TBA

- Readings TBA

Mon 02/02 · L6 · Consequentialism II

- No new readings

Wed 02/04 · L7 · Moral goals and moral constraints

- Nozick, "Moral Constraints and Moral Goals"
- Wells et al., "Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show"
- Lewis, "'Our minds can be hijacked': the tech insiders who fear a smartphone dystopia" (The Guardian)

## Unit 2: Autonomy

Mon 02/09 · L8 · Fake news and free speech I

- Hsu et al., "OpenAI's Sora Makes Disinformation Extremely Easy and Extremely Real" (NYT)
- Pontin, "The Case for Less Speech" (Wired)

Wed 02/11 · L9 · Fake news and free speech II

- Mill, On Liberty, ch. 2 ("Of the Liberty of Thought and Discussion") (Excerpts)

Mon 02/16 · L10 · Digital minimalism

- Aylsworth and Castro, "Is There a Duty to Be a Digital Minimalist?"

Wed 02/18 · L11 · Is the attention economy noxious?

- Castro and Pham, "Is the Attention Economy Noxious?"

## Unit 3: Privacy

Mon 02/23 · L12 · Why does privacy matter? I

- Rachels, "Why Privacy is Important"

- King and Meinhardt, "Rethinking Privacy in the AI Era," ch. 3 ("Provocations and Predictions") (skim this one)

Wed 02/25 · L13 · Why does privacy matter? II

- Marlinspike, "Why 'I Have Nothing to Hide' is the Wrong Way to Think About Surveillance"

Mon 03/02 · L14 · The traditional privacy protection paradigm I

- Barocas and Nissenbaum, "Big Data's End Run Around Anonymity and Consent"

Wed 03/04 · L15 · The traditional privacy protection paradigm II

- No new readings

Mon 03/09 · L16 · Anonymization and differential privacy

- Kearns and Roth, The Ethical Algorithm, "Algorithmic Privacy: From Anonymity to Noise"

Wed 03/11 · L17 · A puzzle about privacy

- No new readings

## Unit 4: Fairness

Mon 03/16 · No class (spring break)

- No readings

Wed 03/18 · No class (spring break)

- No readings

Mon 03/23 · L18 · Algorithmic fairness I

- Barocas et al., Fairness and Machine Learning: Limitations and Opportunities, ch. 1 ("Introduction") (<https://fairmlbook.org/introduction.html>)

Wed 03/25 · L19 · Algorithmic fairness II

- Angwin et al., "Machine Bias" (ProPublica)
- Corbett-Davies et al., "A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear." (Washington Post)

Mon 03/30 · L20 · Algorithmic fairness III

- Mayson, “Bias in, bias out,” pp. 2221–2250

Wed 04/01 · L21 · TBA

- Readings TBA

Mon 04/06 · L22 · Algorithmic fairness IV

- Mayson, “Bias in, bias out,” pp. 2251–2261

## Unit 5: Transparency

Wed 04/08 · L23 · The black box problem

- Christian, The Alignment Problem, ch. 3 ("Transparency")

Mon 04/13 · L24 · The right to explanation

- Vredenburg, “The Right to Explanation”

Wed 04/15 · L25 · The right to explanation

- No new readings

Mon 04/20 · L26 · Final exam

- No readings

Wed 04/22 · L27 · Conclusion

- No readings