

## COMM 4940: Investigating Alleged AI Failures

J. Nathan Matias <nathan.matias AT cornell.edu> , Department of Communication, Field Member, Information Science

- Spring 2026 (see course listing)
- Time: DAY & DAY TIME - TIME
- Location: TBD

AI systems that monitor and interact with people are everywhere—directing the behavior of law enforcement, giving us relationship advice, managing the world’s financial systems, shaping our cultures, and flipping a coin on the success or failure of movements for change. How can we know if they are any good, and if accusations of their harms are true?

In 2023, this challenge became even clearer when Attorney General Letitia James and 31 other state Attorneys General filed a federal lawsuit against Meta for allegedly “harming young people’s mental health and contributing to the youth mental health crisis.” One of the lawsuit’s claims is that Meta, through Facebook and Instagram, encouraged compulsive use of the platform through its algorithms and design features—contributing to a widespread mental health crisis. In the next two years as generative AI was deployed to millions of people, people started to raise the alarm that chatbots were also contributing to deaths by suicide. In response, company CEOs sometimes argued that deeper social problems, not their products were to blame.

In this 3 credit course for (15) upper-level undergraduates and up to (5) PhD students, you will learn about the design of interactive algorithms and the feedback patterns they create with human behavior. You will learn about the challenge they represent for social policy, ways to research their behavior, debates over how to identify AI failures, and emerging policy ideas for investigating and governing these complex patterns. Along the way, you will hear from pioneers in policy, advocacy, and scholarship.

This course is an excellent stepping stone for anyone interested in a career in policy, advocacy, academia, or industry research.

This course requires students to have taken introductory statistics and be able to conduct multiple regression. This could include ILRST2110, AEM 2100, CRP 1200, ENGRD 2700, HADM 2010, MATH 1710, PUBPOL 2100, PUBPOL 2101, PSYCH 2500, SOC 3010, STSCI 2100, STSCI 2150, BTRY 3020, STSCI 2110, STSCI 3200. If you have any questions about this requirement, please contact the instructor.

### How to Register for this class

**Upper Level Undergraduates** To join this class, register for COMM 4940 like any class.

**PhD Students** To join this class, you have two options:

1. Enroll in an independent study (COMM 7970) with Professor Matias and then attend the class (preferred). Before enrolling, please write a brief note to Professor Matias with “Enrolling in COMM [4940]” in the subject. The email should:  
Introduce yourself, your program, and where you are at (1-2 sentences)  
Explain why you want to take the course. For example, is it related to your research? An interest in policy? A project you want to develop in the course? (2-3 sentences)  
Describe the perspectives and skills you bring to the class (a few bullet points) (for example, an understanding of certain theory, or capabilities at qualitative research, data analysis, policy writing, advocacy, media-making, or software development)
2. Enroll in COMM 4940 as a piece of elective coursework

## What You Learn

In this seminar class, you will engage with the science and policy challenges of regulating human-algorithm behavior. Along the way, you will work with scientific issues of methods, theory, and ethics, policy questions about how to govern these situations, and how to bridge between science and policy in a democracy. For a final project, student teams (3-4) will produce a novel project that makes a contribution at the intersection of science and policy.

By the end of the semester, students will be able to:

- Identify, analyze, and evaluate claims about how human and AI behavior interact
- Summarize and criticize major approaches to making claims about AI failures, from the perspective of computer science and the social sciences
- Describe and evaluate major governance approaches
- Design and analyze research methodologies for transparency, accountability, and product improvement
- Understand the uses of social science and computer science research in the policy process
- Design an intervention into scholarly and policy conversations on alleged AI failures

PhD students are encouraged to connect the course to an existing research question that they wish to connect with policy, or to develop a project that could become part of their wider research. In addition to the above learning outcomes, students will be able to:

- Identify ways to link their research interests to policy
- Design studies that can inform policy processes beyond their scholarly field

## Activities

*Weekly Activities:* Throughout the semester, students will read a selection of articles and discuss that reading in class and on Canvas. Once teams have been formed, students will also submit regular progress reports on their final project.

*Algorithm Incident Report:* The midterm is an analysis of an algorithm-involved event in the news, based on publicly-available information.

*Project Proposal:* Your project proposal will include a description of the project, a bibliography, a list of the roles that team members will play, and a timeline.

*Final Project:* An intervention into academic and/or policy conversations on human-algorithm research. In 2024, the class will support projects in two areas:

- An in-depth case study and/or forensic analysis of some alleged AI system failure (see the AI Incident Database for ideas)
- A proposal for a “black box” or data schema of what to record to inform future analysis of incidents
- A research project of your own choosing, if your team includes a PhD student

*Grading:* Participation in class & online: 30%. Midterm: 30%. Final project: 40%.

## About the Instructor

Dr. J. Nathan Matias (@natematias)

organizes community behavioral science for a safer, fairer, more understanding internet. Nathan is an assistant professor in the Cornell University Department of Communication. He is also a field member in Information Science.

Nathan is the founder of the Citizens and Technology Lab, a public-interest project at Cornell that supports community-led behavioral science—conducting independent, public-interest audits and evaluations of social technologies. CAT Lab achieves this through software systems that coordinate communities to conduct their own research on social issues. Nathan has worked with communities of millions of people to test ideas to prevent online harassment, broaden gender diversity on social media, manage human/algorithmic misinformation, and audit algorithms.

## Office Hours

I am available for office hours in person and online during the following times:

- DAY (TIME-TIME)
- DAY (TIME-TIME)

Please feel free to show up or schedule in advance. The most productive office hours involve coming to my office with a sense of your question and all of the

needed information ready to go. I also enjoy just having interesting conversations with students about things you're interested in as well!

## Schedule

### 1.1 What is an AI Failure and Why Does it Matter? (Intro class)

This class session will introduce the course and pose initial questions.

- Sophie Mellor (2022) After a 14-year-old Instagram user's suicide, Meta apologizes for (some of) the self-harm and suicide content she saw. Yahoo! News.

---

**1.2 Tech and Methods: How Do AI Systems Work? What Does It Mean to Fail?** In this class, we will discuss the code and mathematics behind adaptive algorithms, as well as what it means for an AI system to fail.

- Paresh Dave (2023) The 5 Instagram Features That US States Say Ruin Teens' Mental Health. WIRED
- Narayanan, A. (2023). Understanding social media recommendation algorithms . Knight First Amendment Institute.

Graduate student reading:

- Angel, M. P., & boyd, d. (2024, March). Techno-legal solutionism: Regulating children's online safety in the United States. In Proceedings of the 2024 Symposium on Computer Science and Law (pp. 86-97).

Questions to consider:

- What are the specific claims being made about Meta and mental health?
- How does the description of the systems underlying the "algorithm" change how you think about how harms might occur?
- What harms are so important that they deserve investigation and redress by designers, courts, or law? What harms are not?
- How do we make sense of claims that the attempt to attribute harm to technology is an unhelpful form of techno-determinism?

---

### 2.1 Feedback Loops: Simple Models to Explain Complex Behaviors

- Clegg, N. (2021, March 31). You and the Algorithm: It Takes Two to Tango. Medium.
- Kimmerer, R. (2013) "Windigo Footprints" from Braiding Sweetgrass. Milkweed Editions

Graduate student reading:

- Matias, J.N. (2023) Humans and algorithms work together — so study them together. *Nature*.

Questions to think about from Windigo Footprints:

- Think about an event related to algorithms and AI systems that matters to you. It might be an incident related to mental health, or discrimination, or something else
- Imagine the forces akin to “hunger” (the drive for more) and “gratitude” (a drive for less of something) that direct individuals and institutions toward **more** and which ones direct them toward **less**
- Reflect on how those forces might have contributed to that event.
- How persuasive do you find Clegg’s argument that companies should be held less responsible for harms since their systems react to human behavior?

---

**2.2 Incidents and Disasters** In this class, we’re going to ask the question “What is an incident” and “What is an accident/disaster/catastrophe?” while also discussing how understanding them might help society: Readings:

- Marsh, Allison (2021) The Inventor of the Black Box Was Told to Drop the Idea and “Get On With Blowing Up Fuel Tanks”. *IEEE Spectrum*.
- Pinch, T. J. (1991). How do we treat technical uncertainty in systems failure? The case of the space shuttle Challenger. In *Social responses to large technical systems: control or anticipation* (pp. 143-158). Dordrecht: Springer Netherlands.

Graduate student reading:

- Knowles, S. (2014). Engineering risk and disaster: Disaster-STS and the American history of technology. *Engineering Studies*, 6(3), 227-248.

Questions to think about — choose a specific risk, harm, or incident involving AI, and consider:

- Based on the article by Pinch, list out at least six of the “causes” of the Challenger disaster, and label them (were they technical, political social, etc). Could you name the most important one?
- List out things that might possibly go wrong
- What might you want to record in order to go back and figure out what went wrong?

---

**3.1 The Lifecycle of a Population Harm in Democracy** In this class, we will consider the lifecycle of a conversation about harms over the long arc. Choose one of the following stories to learn more about, and link up with other students looking into the same case study.

- **Air pollution**, environmental toxicology, and the clean air act:
  - Jacobs, E. T., Burgess, J. L., & Abbott, M. B. (2018). The Donora smog revisited: 70 years after the event that inspired the clean air act. *American journal of public health*, 108(S2), S85-S88.
  - Costa, D., & Gordon, T. (2000). Mary O. Amdur. *Toxicological Sciences*, 56(1), 5-7.
- The **automotive safety** and crash injury studies:
  - Gangloff, A. (2013). Safety in accidents: hugh dehaven and the development of crash injury studies. *Technology and Culture*, 54(1), 40-61.
  - Mars, R. (2017) “The Nut Behind the Wheel.” 99% Invisible
- **Nuclear testing**, radiation exposure, and Lukemia
  - Documentary: Downwinders and the Radioactive West. PBS Utah
  - Machado, S. G., Land, C. E., & McKay, F. W. (1987). Cancer mortality and radioactive fallout in southwestern Utah. *American journal of epidemiology*, 125(1), 44-61.
  - Beck, H. L., & Krey, P. W. (1983). Radiation exposures in Utah from Nevada nuclear tests. *Science*, 220(4592), 18-24.
- **Smoking** and lung cancer:
  - Parascandola, M. (2004). Two approaches to etiology: the debate over smoking and lung cancer in the 1950s. *Endeavour*
  - Documentary: (2015) Merchants of Doubt. Sony Pictures Classics.
- **Per- and polyfluoroalkyl substances (PFAS)**:
  - Lerner, S. (2024) Toxic Gaslighting: How 3M Executives Convinced a Scientist the Forever Chemicals She Found in Human Blood Were Safe. ProPublica.
  - Film: (2019) Dark Waters. Focus Features.

For whichever of these stories you choose, the group of students studying the same case should create a bullet point timeline of:

- When did people first start worrying about the problem?
- When did people start collecting data about the problem?
- Who stood to gain or lose from the results of the decision?
- What institutions faced important decisions (in companies, governments, courts, etc)
- How were scientists brought into the issue?
- How did scientists handle disagreements?
- If any consensus was reached on the issue, how long did it take?
- How long did it take between the first expression of concern and a systemic solution to the problem?

---

**3.2 Misattribution and Mis-estimation of Harm** While unaddressed harm can lead to compounding problems over time, misattribution and mis-estimation of harms can also create serious problems. It can lead to injustice

toward accused product creators and distract attention from the real problem. The consequences of a false positive for institutions and people's lives can cause irreversible harm and can for generations. In this class, we will discuss what happens when scientists and policymakers are wrong about the causes of an evident harm.

- Pierson, E., Li, S. (2015) A better way to gauge how common sexual assault is on college campuses. *The Washington Post*
- Gantman, A. P., & Paluck, E. L. (2025). What is the psychological appeal of the serial rapist model? *Worldviews predicting endorsement. Behavioural Public Policy*, 9(2), 461-476.

Graduate student reading (choose one):

- Rosenberg, M., Townes, A., Taylor, S., Luetke, M., & Herbenick, D. (2019). Quantifying the magnitude and potential influence of missing data in campus sexual assault surveys: A systematic review of surveys, 2010–2016. *Journal of American college health*, 67(1), 42-50.
- Porat, R., Gantman, A., Green, S. A., Pezzuto, J. H., & Paluck, E. L. (2024). Preventing sexual violence: A behavioral problem without a behaviorally informed solution. *Psychological Science in the Public Interest*, 25(1), 4-29.

---

**4.1 Making Empirical Claims About Technology Harms** In this two-part series, we will discuss a case study in an alleged algorithm failure. Students will choose among one of the following datasets for project teams to analyze. Teams will be assigned to one of two sides: the side defending technology makers, and the side making an accusation about a technology failure.

Before the class, students will read a paper related to the topic and come prepared to discuss the broader issue and receive information on how to analyze the dataset.

- Primary plan: Data from the European Union on the number of child safety reports for different online platforms
- Backup (if above data collection proves impossible):
  - Abdel-Aty, M., & Ding, S. (2024). A matched case-control analysis of autonomous vs human-driven vehicle accidents. *Nature communications*, 15(1), 4931.

---

**4.2 Analyzing Archives of AI Failures** In this class, teams will present competing analyses to try to convince the class of their point of view.

We will then discuss what we learned from the experience.

---

**5.1 What is an AI Incident?** In order for algorithms to be governed, we need to be able to describe what happens when one goes wrong. So what is an incident anyway, and how would we establish that harm had occurred?

- Matias et al (under review). Accelerating the Science of Digital Harms through Data Donations to a Digital Harms Research Center.
- McGregor, S. (2021, May). Preventing repeated real world AI failures by cataloging incidents: The AI incident database. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 17, pp. 15458-15463).

Graduate Student Reading:

- Singh, G., Patel, R. H., Vaqar, S., & Boster, J. (2024). Root cause analysis and medical error prevention. In StatPearls. StatPearls Publishing.

To prepare for class, think about the story by Sophie Mellor that we read at the beginning of class. Imagine that you are an employee at Instagram who is asked to investigate a mental-health related incident experienced by a young person who was an Instagram user.

- create a definition for the incident type— what is it and what isn't it? How would you tell what counts as an incident or not?
- make a list of 6 things that you would want to record from what might be known by Instagram, friends/family, and institutions

In class, we will work to collaborate on an incident report template for several kinds of mental health related incidents.

---

**5.2 Discussing Final Project Ideas** In this class, we will discuss possible final project ideas for the course.

In preparation for the class, meet with a possible project team to develop ideas for collective feedback and discussion.

As a reminder, final project ideas include one of three options:

- An in-depth case study and/or forensic analysis of some failure of an adaptive algorithm (see the AI Incident Database for ideas)
- A proposal for a “black box” or data schema of what to record to inform future analysis of incidents
- An independent project involving a graduate student team member

---

FEBRUARY BREAK

---

**6.2 Methods: Auditing and Red Teaming an AI System** How do researchers identify systemic failures in the behavior of an AI system? Audits and Red Teaming are two popular methods.



- Metaxa, D., Park, J. S., Robertson, R. E., Karahalios, K., Wilson, C., Hancock, J., & Sandvig, C. (2021). “Introduction to Auditing” Auditing algorithms: Understanding algorithmic systems from the outside in. *Foundations and Trends® in Human–Computer Interaction*, 14(4), 272-344.  
– If “Auditing AI” is available from MIT Press, we may use chapters from that book instead
- Bullwinkel, B., Minnich, A., Chawla, S., Lopez, G., Pouliot, M., Maxwell, W., ... & Russinovich, M. (2025). Lessons from red teaming 100 generative ai products. arXiv preprint arXiv:2501.07238.

Graduate student reading:

- Gerchick, M., Jegede, T., Shah, T., Gutierrez, A., Beiers, S., Shemtov, N., ... & Horowitz, A. (2023, June). The devil is in the details: Interrogating values embedded in the allegheny family screening tool. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1292-1310).

Before class, write how you would test the AI system you want to study for an outcome that matters to you.

---

**7.1 Methods: Investigating and Describing an Incident** So far this semester, we have looked at simple incident reports such as the systems used by the U.S. National Transportation Safety Board to report a crash— introduced in the 1960s. Since then, scholars have developed more comprehensive ideas about how to map and describe failures. How far should we go toward offering more complex models? And how much should we understand/engage with the politics at play?

In this class, we’re going to learn more about those systems and work through the case of the Richmond CA Refinery fire of 2012. We’re also going to consider how incident reporting fits into the wider politics of error reporting and the challenges of academic access to sensitive information.

- Animation of Fire at Chevron’s Richmond Refinery, August 6, 2012. US Chemical Safety and Hazard Investigation Board.
- Sadisivam, N. (2021) A California law gave the people power to cut pollution. Why isn’t it working?. *Grist*.
- Yousefi, A., Rodriguez Hernandez, M., & Lopez Peña, V. (2019). Systemic accident analysis models: A comparison study between AcciMap, FRAM, and STAMP. *Process Safety Progress*, 38(2), e12002. (in Canvas)

Graduate student reading:

- Larsen, R. (2022). ‘Information Pressures’ and the Facebook Files: Navigating questions around leaked platform data. *Digital Journalism*, 10(9), 1591-1603.

Before the class, reflect on what might have been different if someone had done a systematic analysis of the type of alleged AI failure you are hoping to study.

---

**7.2 Finalizing Student Projects** In your final-project teams, students will present a 3 minute presentation to summarize and analyze an issue in human-algorithm feedback that you would like to work on for your final. Fellow students will submit feedback and suggestions.

---

### 8.1 Methods: Causality

- Odgers, C. L. (2024). The great rewiring: is social media really behind an epidemic of teenage mental illness? *Nature*, 628(8006), 29-30.
- Matias, N., & Penney, J. (2025). Science and Causality in Technology Litigation. *Journal of Online Trust and Safety*, 2(5).

Graduate student reading:

- Office of the Surgeon General. (2023). Social media and youth mental health: The US Surgeon general's advisory. U.S. Department of Health and Human Services.

In this class, we will discuss the question of causality and methods for establishing the different kinds of causality that might be important to answer practical questions about technology benefits, harms, and interventions to transform them. For your project topic:

- Identify the kinds of causality that matter
  - Identify hurdles to obtaining those causal inferences
  - Guess how many scientists, studies, and years you think it might take to arrive at confident causal claims about the harm you are studying
- 

**8.2 How Can Evidence Inform Governance?** What is evidence-based governance and how might research contribute (or not) to policy? In this session, we'll discuss a model where researchers are advisors, versus a model where researchers are producing evidence that is used in courts and other contestational settings.

- Spruijt, P., Knol, A. B., Vasileiadou, E., Devilee, J., Lebrecht, E., & Petersen, A. C. (2014). Roles of scientists as policy advisers on complex issues: A literature review. *Environmental Science & Policy*, 40, 16-25.
- Systemic Justice (2023) Strategic Litigation: A Guide for Legal Action. Systemic Justice.

Graduate Student Reading:

- Weiss, C. H. (1979). The many meanings of research utilization. Public administration review, 39(5), 426-431.

Questions for discussion in Canvas and in class: \* For your team's final project area:

- What kind of evidence might be needed for policymaking
- Is that kind of evidence possible yet?
- Does the evidence already exist?
- How does evidence already influence policy, if at all?
- What forces are preventing evidence from being available?->

**9.1 Team Check-In** In this class, we will:

- Support group collaboration on student projects
  - Discuss common questions about projects and final proposals
- 

**9.2 Guest Panel: Processing Posthumous Digital Remains** In this class, we will hear from a group of scientists developing a center that can receive posthumous digital remains.

- Orben, A., & Matias, J. N. (2025). Fixing the science of digital technology harms. Science, 388(6743), 152-155.
- 

**SPRING BREAK (Syllabus Adjustment Check-In)**

---

**10.1 Policy Topic: How Does Evidence Actually Inform Policy?** In previous classes, we discussed theories for how evidence could inform policy. How does it really happen?

- **Each Team** share on Canvas an example of a piece of research, journalism, or activism that genuinely impacted policy (ideally related to your project), and comment on how it did so. A especially high quality submission will be able to cite an example where someone made a claim about the relationship between the corporate/government policy change and the project. This will require some research beyond the single article or project page. Comment on other team's examples, come prepared to discuss your team's example, and be prepared to ask questions and discuss other team's examples.

Graduate Student Reading (choose one):

- Specter, M., Sellars, A. (2025 - Forthcoming) Software Screws Around, Reverse Engineering Finds Out: How Independent, Adversarial Research Informs Government Regulation.
  - Contandriopoulos, D., Lemire, M., Denis, J. L., & Tremblay, É. (2010). Knowledge exchange processes in organizations and policy arenas: a narrative systematic review of the literature. *The Milbank Quarterly*, 88(4), 444-483.
- 

**10.2 Policy Topic: Algorithmic Literacy** In order for democratic societies to govern a problem, people need to know about it and care enough to do something. How does the public understand and care about human-algorithm feedback?

- Druga, S., Christoph, F., & Ko, A. J. (2022). Family as a Third Space for AI Literacies: How do children and parents learn about AI together?.
- Rainie, L., Anderson, J. (2017) “The need grows for algorithmic literacy, transparency and oversight” from *Code-Dependent: Pros and Cons of the Algorithm Age*. Pew Research Center

Graduate student reading:

- Menon, A., & Zhang, B. (2025). Future Shock or Future Shrug? Public Responses to Varied Artificial Intelligence Development Timelines. *The Journal of Politics*.

Discussion questions:

- If the design of algorithms is controlled by engineers and most people aren’t engineers, how would algorithm literacy actually make a difference in people’s lives if at all?
  - What kinds of knowledge about algorithms does a democratic public need?
  - How could your project area be transformed by algorithmic literacy — and how not?
- 

**11.1 Policy Topic: Public Engagement with Incident Reporting** If incident reporting is to work, the public need to understand incidents. In this class, we will look at the history of the Dark Patterns Tip Line.

- Nguyen, Stephanie (2021) Key learnings from the Dark Patterns Tip Line. Rita Allen Foundation
- King, J., & Stephan, A. (2021). Regulating Privacy Dark Patterns in Practice-Drawing Inspiration from the California Privacy Rights Act. *Georgetown Law Technology Review*, 5(2), 250-276.

Graduate Student Reading:

- Blake, C., Rhanor, A., & Pajic, C. (2020). The demographics of citizen science participation and its implications for data quality and environmental justice. *Citizen Science: Theory and Practice*, 5(1).

Questions for discussion on Canvas and in class:

- For your team's topic, how might it be possible to ask the public to report incidents?
- Would the public be able to notice and distinguish possible incidents?
- What kind of public awareness will be necessary?
- How might inequalities in incident report influence who gets served by policy?

---

**11.2 Policy Topic: Creating Change With Courts** What influence can courts have on the behavior of companies, populations, and algorithms?

- Charles, S. (2020, January 27). CPD decommissions 'Strategic Subject List.' *Chicago Sun-Times*.
- Kaplan, J. (2017). *Predictive Policing and the Long Road to Transparency*

Graduat Student Reading:

- Jillson, Elisa. (2021) *Aiming for truth, fairness, and equity in your company's use of AI*. Federal Trade Commission.

Questions to consider:

- What role did courts play in the policy debates about the Strategic Subjects List?
- How can laws and agencies influence what courts are able to do?
- How might courts be important to the project you're working on?

---

**12.1 Final Project Check-in** In the class, teams will be supported to meet in person, workshop your ideas with peers, and get feedback on your project progress.

---

**12.2 Case Study: TBD** In this class, we will review a consequential case study that made it into the courts and hear from guests who were involved on different sides of the case.

Readings: TBD

---

**13.1 Final Presentations Part I** In this session, teams will give a final presentation of their project, with an opportunity to receive feedback from peers.

In the discussion, please post in advance a question for feedback you would like your peers to consider during your session in the class.

---

**13.2 Final Presentations Part II** In this session, teams will give a final presentation of their project, with an opportunity to receive feedback from peers.

In the discussion, please post in advance a question for feedback you would like your peers to consider during your session in the class.

---

#### **14.1 Reflections on Learning & the Future of Algorithms and Society**

In this session, we will reflect on the topics in the class. Please submit a discussion post, which will inform our group conversation:

- Name something you believed at the beginning of the course
  - Reflect on how your thinking about that question has changed
  - Name one way this topic might continue to be relevant to your future life, whether in your role as citizen or your work
- 

**Final Project Deadline: TBD**

---

## **Course Practices and Policies**

### **Weekly Workload**

Before each class, I will assign **two readings** I expect you to read before class. As part of class participation, you will submit a reaction comment on one of the readings to the relevant discussion on Canvas and respond to at least one other student's comment by 9pm Eastern the night before class. Please come to class with:

- The ability to summarize:
  - the goal/question of the paper
  - the field or ecosystem that the author is in
  - what constitutes an advancement in that field
  - how the paper advances the conversation
  - a question or observation that links the paper to the theme of the day or the theme of the course
  - you can choose to respond to one of the prompt questions, but are not obligated to

Each week, students are expected to post at least one news article to Canvas add at least one comment in response to one other student's posted link.

Graduate students will have additional readings. Graduate students will also rotate responsibility for summarizing their reading to the class. Summarizing the reading will involve:

- Creating 2-3 slides for the reading
  - Slide one: introducing the core question of the paper, and the authors
  - Slide two: summarize the methods and findings of the paper
  - Slide three: one to two discussion questions that link the paper to the theme of the class
- Alt: if the author of the paper is a guest speaker, the graduate student will interview the guest

Since this is a discussion course, attendance is expected.

### **Participation in Discussions**

Please post discussions about readings to Canvas.

Whenever the course has assigned reading for a session, you are expected to post at least one top-level comment and one reply to someone else's comment by the end of the day before class. Participation will be 30% of your course grade.

### **Team Progress Reports**

During the project period of the class, teams will submit on Canvas a weekly progress report no more than one page long, as group homework. This progress report will be graded 0/1 based on whether it was submitted. Reports should include the following details (a sample template is available here):

- What your team made progress on
- What your team is doing next
- Who contributed what
- Where your team is stuck
- Any updates to your timeline

### **Formats**

Written assignments should be uploaded to Canvas in one of the following formats:

- PDF
- Text file
- Word-compatible document

Slide decks should be submitted in one of the following formats:

- PowerPoint

- Keynote
- Google Slides
- A Markdown slide presentation system (such as Marp)
- PDF

In cases where students choose to submit code or analysis as part of a project, I can accept assignments in the following languages:

- R (including R Markdown or Sweave)
  - For more guidance on R, see this excellent Cornell course, which has full public materials.
- Python
- Ruby

### Group Work and Academic Integrity

I expect students to follow the Cornell University Code of Academic Integrity. You should submit your work as your own, cite sources and outside assistance, and credit people for their contributions.

This class includes group work and individual assignments. On group projects, you are encouraged to work together on the activities for the class, but you are only only able to put your name on projects to which you made an intellectual contribution. If you have any questions about what is appropriate, ask me.

Because my main goal is for you to develop the ability to read, think, develop ideas, and analyze data, this course is an “Approved Tools Only” class for Generative AI. If you are thinking about using a generative tool, please consult the professor in advance, consult any team-mates, and acknowledge its use in any submissions. At the beginning of the class, I will issue a survey for students to share the tools you find helpful and will pre-approve ones that are compatible with the learning goals of the course.

As a sign of respect for others, please verify your own work before submitting it as an assignment or sharing it with a team-mate.

**Approved Generative AI Tools:** The list of approved AI tools will be updated in this table.

Purpose	Tool category	Use Suggestion
Writing and editing	Grammar checking tools	Use to review text after writing
Information retrieval	Search engines Information discovery apps	Use for discovery and inspiration, but always verify by reading anything you cite
—	—	



## **Recordings and Classroom Privacy**

Since this is a small course focused on highly sensitive issues, we need to take special care to encourage and protect everyone's capacity to speak freely. Each person in this class is expected to respect the principles of academic freedom for instructors and classmates and will maintain the privacy of the classroom environment, as outlined in Cornell's Code of Academic Integrity.

This commitment to building respect and trust in the classroom means members of this class will not: record, photograph, or share online any interactions that involve classmates or any member of the teaching team. Students will also respect the intellectual property rights of the instructor, and will not share or otherwise make accessible any course materials to anyone not enrolled in the course, without the instructor's written permission.

This policy is not meant to restrict students' ability to use classroom recordings in ways beneficial to their learning. Students who may benefit from recorded lectures and lecture playback, including students who use English as an additional language or who have accommodations from SDS, should speak to the course instructor to maintain transparency and trust in the classroom. Students approved to record lectures are expected to maintain the respect and privacy of the learning environment, as stated above.

## **Grading**

I use the following grading scale, derived from Matt Salganik's grading practices.

All grades are final. There will not be any make-up or extra credit assignments offered.

Per university procedures, I will only give a grade of A+ in exceptional circumstances.

Letter Grade Numeric Grade

- A 93 - 100
- A- 90 - 92.99
- B+ 87 - 89.99
- B 83 - 86.99
- B- 80 - 82.99
- C+ 77 - 79.99
- C 73 - 76.99
- C- 70 - 72.99
- D 60 - 66.99
- F 0 - 59.99

## **Grace Period**

Late Submissions: All assignments have an automatic one-day grace period. On time and early papers are always encouraged and will be graded the same week.

Students can turn in the assignment up to a day late, no questions asked, with the expectation that it could take substantially longer to receive a grade and feedback. After that, an automatic one grade (A to B, B to C, etc) is dropped on the assignment.

If you turned in an assignment on time and haven't received a grade within the expected period, please contact the instructor in case of a technical glitch.

### **Accommodations**

I am committed to working with students with recognized SDSes to ensure you have the best possible class experience. Please follow these steps to get started.

### **Health Accommodation for Professor Matias**

As part of a medical condition of my own, I may on rare occasion need accommodations myself. I will introduce the details of this accommodation and related processes in the first class.

### **Counseling Resources**

This class is covering some very difficult topics that relate to severe harms. If you are someone who wrestles with mental health issues, I encourage you to take a moment to ensure that you have the support you need to take this class effectively. If you're unsure what kind of support would be helpful, we can discuss ideas in office hours.

It is common for students to experience stressful events at some point during graduate school. Students sometimes experience depression, anxiety, family stress, the loss of loved ones, financial strain, and other stressors. It is perfectly normal for students to seek the service of mental health professionals to provide them with support and skills to cope with these experiences. Below I have provided the contact information for some of the mental health services available to Cornell University students so that you will know where you can go if you or a friend would like to take advantage of these resources.

Cornell Health: 110 Ho Plaza, Ithaca NY. Phone: 607-255-5155.

### **Acknowledgments**

I am grateful to Lucas Wright for his contributions to a review article that this course is based on, and to the Human-Algorithm-Behavior Research Collective for input. Several sections of these policies have been inspired by syllabi from Matthew Salganik, Adrienne Keene, and Neil Lewis Jr.

### **Attention to Detail**

Reading instructions carefully, paying attention to detail, and planning your time wisely are all essential for thriving in this course. All of these will be rewarded

with learning, and in the case of this syllabus—this link to a cute baby owl getting a massage.