

AIDE Summer Syllabus - Summer 2025

Program Faculty and Staff

Faculty Leads

Staff Leads

Northeastern Faculty Participants

Visiting Faculty Participants

Program Content

Seminar Reading List

Workshops

Speaker Series

CS Bootcamp

Mentoring

Graduate Participant Contributions

Seminar Participation

Short Argument Summaries

CS Lab Assignments

Final Projects

Comprehensive Schedule

Week 1

Week 2

Week 3

Week 4

Week 5

Week 6

Week 7

Week 8

Week 9

Program Faculty and Staff

Faculty Leads

John Basl (j.basl@northeastern.edu)

Kathleen Creel (k.creel@northeastern.edu)

Staff

Keanna Smigliani (k.smigliani@northeastern.edu)

Amod Lele

Deryni Mahoney

Northeastern Faculty Participants

[Liz Bucar](#)

[Sina Fazelpour](#)

[Chad Lee-Stonch](#)

[Vance Ricks](#)

[Christo Wilson](#)

Visiting Faculty Participants

[Clinton Castro](#) (University of Wisconsin-Madison)

[Alice Huang](#) (Western University)

[Renée Jørgensen](#) (University of Michigan)

[Rima Basu](#) (Claremont McKenna College)

[Gabby Johnson](#) (Claremont McKenna College)

[Solon Barocas](#) (Microsoft Research)

[Milo Phillips-Brown](#) (The University of Edinburgh)

Program Content

Seminar Reading List

Week No.	Readings
1	Monday: • Winner (1980) “Do Artefacts Have Politics” • Pacey (1985) “The Culture of Technology” • Liao & Huebner (2020), “Oppressive Things” • (Highly Recommended) Daston (2022) Rules: A Short History of What We Live By, pages 1-20, 82-147 (aka Chapters 1, 4, and 5) Wednesday: • “Classification”, FairML Book • Fazelpour & Danks (2021) “Algorithmic Bias: Senses, Sources, Solutions” • Grant (2023) “Equalized Odds is a Requirement of Algorithmic Fairness” • Recommended (and assigned next week): Hedden (2021) “On Statistical Criteria of Algorithmic Fairness” Background reading: • Jonas (1973) “Technology & Responsibility: Reflections on the New Task of Ethics” • Haraway (1985) “A Cyborg Manifesto” • Borgmann (1987) “Focal Things and Practices” • Rogaway (1995) “The Moral Character of Cryptographic Work” • Friedman & Nissenbaum (1996) “Bias in Computer Systems” • Hutchinson & Mitchell (2018) “50 Years of Test Unfairness”
2	Monday: • Hu (2025) “Does calibration mean what they say it means; or, the reference class problem rises again” • Johnson (2021) “Algorithmic Bias: On the Implicit Biases of Social Technology” Thursday: • Background: Chouldechova (2017) “Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments” • Hedden (2021) “On Statistical Criteria of Algorithmic Fairness”

	● Fazelpour & Lipton (2020) “Algorithmic Fairness from a Non-Ideal Perspective” ● Fazelpour (2025) “Disciplining Deliberation: A Socio-technical Perspective on Machine Learning Trade-Offs”
3	Monday: ● Creel & Hellman (2022) “The Algorithmic Leviathan: Arbitrariness, Fairness, and Opportunity in Algorithmic Decision-Making Systems” ● Lazar (2025) “Governing the Algorithmic City” Thursday: ● Zerilli, Knott, Maclaurin, & Gavaghan (2018) “Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?” ● Basl, Behrends, Grant (2023) "What We Owe to Decision Subjects: Beyond Transparency and Explanation in Automated Decision Making"
4	Monday ● Thomson (1975) “The Right to Privacy” ● Reiman, (1976) “Privacy, Intimacy, and Personhood” ● Marmor (2015) “What is the Right to Privacy” Wednesday: ● Nissenbaum (2009) <i>Privacy in Context</i>, “Part III The Framework of Contextual Integrity,” Chapters 7, 8, 9 Thursday ● Lee-Stronach, book chapter on privacy
5	Monday: ● Johnson & Dupre (forthcoming) “Uncanny Performance, Divergent Competence” ● Kasirzadeh & Gabriel (2023) “In Conversation with Artificial Intelligence:Aligning Language Models with Human Values” Tuesday ● Jorgensen (2021) “Algorithms and the Individual in Criminal Law” Thursday: ● Weidinger et al. (2021) “Ethical and social risks of harm from Language Models” ● Huang et al. (2024), “Collective Constitutional AI” ● Fazelpour & Magnani (ms), “Aspirational Affordances”

6	Monday: ● Creel (2022) “Transparency in Complex Computational Systems” ● Nguyen (2021) “Transparency is Surveillance” Thursday ● Johnson (2025), “The hard proxy problem: proxies aren’t intentional; they're intentional” ● (OPTIONAL) Hu (2024), “What is “Race” in Algorithmic Discrimination on the Basis of Race?” ● (OPTIONAL) Cossette-Lefebvre, Lippert-Rasmussen (2024), “Neither Direct, Nor Indirect: Understanding Proxy-Based Algorithmic Discrimination”
7	Monday: ● Vredenburg (2022), “Right to Explanation” ● Karlan & Kugelberg (2025), “No right to an explanation” Wednesday: ● Gallow (2024), “Instrumental Divergence” ● Thorstad (2024), “Against the Singularity Hypothesis” ● Harding & Kirk-Gianni (2025), “What is AI safety? What do we want it to be?”
8	Friday (OPTIONAL) ● Kopec et. al. (20203), “The effectiveness of embedded values analysis modules in Computer Science education: An empirical study” ● Davis (2006), “Integrating Ethics into Technical Courses: Micro-Insertion”

Professional Development Workshops

Week No.	Workshop Name
1	● Public Scholarship with Liz Bucar
2	—
3	—
4	● Tech Ethics Syllabus Design Workshop ● Professional Development Workshop: Finding Your Place in the Conversation
5	—
6	—
7	● The Job Market
8	● (Optional) Embedding Ethics in CS/Engineering Courses with Katie Creel
9	● Journal Selection

Speaker Series

Week No.	Guest Speaker
1	● Clinton Castro (Simon Fraser University)
2	● Sina Fazelpour (Northeastern)
3	● Katie Creel (Northeastern) ● Alice Huang (Western University)
4	● Chad Lee-Stronach (Northeastern)
5	● Renée Jørgensen (University of Michigan)
6	● Rima Basu (Claremont McKenna) ● Gabby Johnson (Claremont McKenna)

7	—
8	—
9	• Solon Barocas (Microsoft Research) • Milo Phillips-Brown (The University of Edinburgh)

CS Bootcamp

Week No.	Topic
1	• Introduction to Programming and Python (Refresher) • Data, Decision Trees, and Linear Regression
2	• Fairness, Prediction, and COMPAS
3	• Comparative Fairness Analyses with Decision Trees
4	• Differential Privacy
5	• Neural Networks: Perceptron & Transformer Math by Hand
6	• Capstone Project: A Philosophical LLM
7	• Capstone Project: A Philosophical LLM

Mentoring

All graduate student participants will be assigned faculty mentors. Mentors will, among other things, provide guidance on your final projects (see below), help identify relevant literature or

faculty that can provide additional guidance, and help you network with scholars working in the area of AI ethics.

Graduate Participant Contributions

Seminar Participation

Seminar meetings will be discussion based. Faculty leads will often start by providing useful background information and context for assigned readings and fill in gaps during discussion. However, the expectation is that all participants will have come having read the assigned readings and the submitted short argument summaries (see below) and be ready to discuss them.

Seminar participants have varying expertise, background training, and familiarity with concepts, distinctions, arguments, theories, etc.. As participants introduce technical terminology or materials from their area of focus, they should be prepared to explain that material to other participants.

Short Argument Summaries

Seminar participants will be divided into groups. At each seminar meeting, one group will submit a Short Argument Summary (SAS) which includes both an argument reconstruction and an analysis.

Your short argument summary should be a maximum of 750 words.

If you have a summary due for a given meeting, you should submit your short argument summary to the shared Dropbox folder by 9am on the day that it will be discussed.

All cohort members should read all short argument summaries prior to the day's meetings.

For each short argument summary you should reconstruct a single argument from any of the assigned readings that you find interesting, insightful, confusing, puzzling, or worth talking about for some other reason.

Your argument reconstruction should consist of no more than two premises. It should be an obvious instance of either *modus ponens*, *modus tollens*, or hypothetical syllogism. Please come talk to John or Katie if you have questions about how to do this or would like to talk through your first draft.

You should not try to reconstruct the primary argument for the whole paper. Instead, direct our attention towards a sub-argument so that we can focus our discussion on the details.

After reconstructing an argument, you should use the remaining available space to offer some commentary on the argument. You might choose to elaborate on the justification of a premise, raise an objection to a premise, explain some confusion about how a premise is justified, etc. The goal of the analysis is to focus our discussion in a productive and useful way in the context of your chosen argument.

CS Lab Assignments

Participants without prior experience with programming will begin CS Lab with an introduction to programming in Python and basic data concepts. All seminar participants will take part in the second half of CS Lab, which focuses on machine learning and large language models.

Final Papers

Seminar participants will work on a research paper over the course of the program. They will present a full draft of their paper during a week-long works in progress workshop during the last week of the program.

Comprehensive Schedule

Week 1

Topic: Philosophy of Technology

Monday, Jun 2, 2025

Readings

- Winner (1980) “Do Artefacts Have Politics”
- Pacey (1985) “The Culture of Technology”
- Liao & Huebner (2020), “Oppressive Things”
- (Highly Recommended) Daston (2022) Rules: A Short History of What WeLive By, pages 1-20, 82-147 (aka Chapters 1, 4, and 5)

Schedule

Time	Topic
10-11:30	Seminar
11:30-1:30	Lunch
12:30-2	Seminar
2:30-4	CS Lab 1
5:30-7:30	Opening Event

Tuesday, Jun 3, 2025

Speaker(s)

- [Clinton Castro](#) (University of Wisconsin-Madison)

Schedule

Time	Topic
12-1:30	Catered Lunch w/ Christo Wilson
1:30-3:30	Clinton Castro Speaker Session
3:30-5	Discussion w/ Clinton Castro
5:45-8	Dinner w/ speaker

Wednesday, Jun 4, 2025

Readings

- “Classification”, FairML Book
- Fazelpour & Danks (2021) “Algorithmic Bias: Senses, Sources, Solutions”
- Grant (2023) “Equalized Odds is a Requirement of Algorithmic Fairness”
- Recommended (and assigned next week): Hedden (2021) “On Statistical Criteria of Algorithmic Fairness”

Schedule

Time	Topic
12-1:30	Lunch Block
1:30-4:30	Seminar (SAS Group A)

Thursday, Jun 5, 2025

Speaker(s)

- Liz Bucar (Northeastern University)

Schedule

Time	Topic
9-11:30	Public Scholarship w/ Liz Bucar
12-1:30	Catered Lunch with Liz Bucar
1:30-5	Public Scholarship w/ Liz Bucar
5-8	Optional: Dinner on your own + visit to the Museum of Fine Arts

Week 2

Topic: Fairness I

Monday, Jun 9, 2025

Readings

- Hu (2025) Does calibration mean what they say it means; or, the reference class problem rises again
- Johnson (2021) Algorithmic Bias: On the Implicit Biases of Social Technology

Schedule

Time	Topic
10-12	CS Lab 2
12-1:30	Lunch Block
1:30-4:30	Seminar (SAS Group B)

Wednesday, Jun 11, 2025

Speaker(s)

- [Sina Fazelpour](#) (Northeastern)

Schedule

Time	Topic
10:30-12	CS Lab 3

Thursday, Jun 12, 2025

Speaker(s)

- [Sina Fazelpour](#) (Northeastern)

Readings

- Background: Chouldechova (2017) “Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments”
- Hedden (2021) “On Statistical Criteria of Algorithmic Fairness”
- Fazelpour & Lipton (2020) “Algorithmic Fairness from a Non-Ideal Perspective”
- Fazelpour (2025) “Disciplining Deliberation: A Socio-technical Perspective on Machine Learning Trade-Offs

Schedule

Time	Topic
10-12	Seminar (SAS Group C)
12-1:30	Catered Lunch
1:30-3	Sina Fazelpour Speaker Session
5-8	Drinks w/ Sina

Week 3

Topic: Fairness II

Monday, Jun 16, 2025

Speaker(s)

- [Katie Creel](#) (Northeastern)

Reading and Preparation

- Creel & Hellman (2022) “The Algorithmic Leviathan: Arbitrariness, Fairness, and Opportunity in Algorithmic Decision-Making Systems”
- Lazar (2025) “Governing the Algorithmic City”

Time	Topic
10-11:30	Seminar (SAS Group A)
11:30-12:30	Lunch Block
1:30-2	Seminar (SAS Group A)
2:30-4	Katie Creel Speaker Session

Tuesday, Jun 17, 2025

Speaker(s)

- Alice Huang

Time	Topic
10-12	CS Lab 4
12-1:30	Catered Lunch - Alice Huang Talk

Wednesday, Jun 18, 2025

Readings

- Kasirzadeh (2022) “Algorithmic Fairness and Structural Injustice: Insights from Feminist Political Philosophy”
- Lin & Chen, (2022) “Artificial Intelligence in a Structurally Unjust Society”

Time	Topic
3:30-6:30	Seminar (SAS Group B)

Thursday, Jun 19, 2025

Readings

- Zerilli, Knott, Maclaurin, & Gavaghan (2018) “Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?”
- Basl, Behrends, Grant (2023) "What We Owe to Decision Subjects: Beyond Transparency and Explanation in Automated Decision Making"

Week 4

Topic: Privacy

Monday, Jun 23, 2025

Readings

- Thomson (1975) “The Right to Privacy”
- Reiman, (1976) “Privacy, Intimacy, and Personhood”
- Marmor (2015) “What is the Right to Privacy”

Schedule

Time	Topic
12-1:30	Catered Lunch: Syllabus Design Workshop
1:30-4:30	Seminar (SAS Group C)

Tuesday, Jun 24, 2025

Wednesday, Jun 25, 2025

Readings

- Nissenbaum (2010) “Privacy in Context: Technology, Policy, and the Integrity of Social Life” (chapters 7, 8, 9)

Schedule

Time	Topic
10-12	CS Lab 5
12-1:30	Lunch Block
1:30-4:30	Seminar (SAS Group A)

Thursday, Jun 26, 2025

Speaker(s)

- [Chad Lee-Stronach](#) (Northeastern)

Readings

- Lee-Stronach, book chapter on privacy

Schedule

Time	Topic
10:30-11:30	Professional Development Workshop: Finding Your Place in the Conversation
12-1:30	Catered Lunch w/ Chad Lee-Stronach
1:30-3	Chad Speaker Session
5-8	Drinks w/ Speaker

Week 5

Topic: id

Monday, Jun 30, 2025

Readings

- Johnson & Dupre(forthcoming)“Uncanny Performance, Divergent Competence”
- Kasirzadeh & Gabriel (2023) “In Conversation with Artificial Intelligence:Aligning Language Models with Human Values”

Schedule

Time	Topic
10-12	CS Lab 6: Neural Networks: Perceptron & Transformer Math by Hand
12-1:30	Lunch Block
1:30-4:30	Seminar (SAS Group B)

Tuesday, Jul 1, 2025

Speaker(s)

- [Renée Jørgensen](#) (University of Michigan)

Readings

- Jørgensen, (2021) “Algorithms and the Individual in Criminal Law”

Time	Topic
12-1:30	Catered Lunch w/ Renée Jørgensen
12-1:30	Catered Lunch with Speaker
3-5	Meetings with Speaker
5:30-7:30	Dinner with Speaker

Wednesday, Jul 2, 2025

Thursday, Jul 3, 2025

Readings

- Weidinger et.al (2021) “Ethical and social risks of harm from LanguageModels”
- Huang et al. (2024), “Collective Constitutional AI”
- Fazelpour & Magnani(ms), “Aspirational Affordances”

Schedule

Time	Topic
10-12	CS Lab 7
12-1:30	Lunch Block
1:30-4:30	Seminar (Group C)

Week 6

Topic: Transparency + LLMs

Monday, Jul 7, 2025

Speaker

- Rima Basu (Claremont McKenna)

Readings

- Creel (2022) “Transparency in Complex Computational Systems”
- Nguyen (2021) “Transparency is Surveillance”

Schedule

Time	Topic
10:30-12	Speaker Session: Rima Basu
12-1:30	Catered Lunch w/ Rima Basu
1:30-4:30	Seminar (SAS Group C)

Tuesday, Jul 8, 2025

Wednesday, Jul 9, 2025

Schedule

Time	Topic
10:30-12	Interpretability Lab: LogitLens
12-1:30	Catered Lunch for LogitLens
1:30-4:30	Interpretability Lab: LogitLens

Thursday, Jul 10, 2025

Speaker(s)

- Gabby Johnson (Claremont McKenna)

Readings

- Johnson(2025), "The hard proxy problem: proxies aren't intentional; they're intentional"
- (OPTIONAL) Hu (2024), "What is "Race" in Algorithmic Discrimination on the Basis of Race?"
- (OPTIONAL)Cossette-Lefebvre, Lippert-Rasmussen (2024), Neither Direct,Nor Indirect: Understanding Proxy-Based Algorithmic Discrimination

Schedule

Time	Topic
10:30-12	CS Lab 8
12-1:30	Catered Lunch w/ Gabby Johnson
1:30-3	Speaker Session: Gabby Johnson
3-4:30	Meetings with Speaker
5:30-7:30	Dinner with Speakers

Week 7

Trends in Phil AI: X-Risk and Social Media

Monday, Jul 14, 2025

Readings

- Vredenburg (2022) “Right to Explanation”
- Karlan & Kugelberg (2025), “No right to an explanation”

Schedule

Time	Topic
10-12	CS Lab 9
12-1:30	Catered Lunch: Job Market Workshop
1:30-4:30	Seminar (SAS Group B)

Tuesday, Jul 15, 2025

Wednesday, Jul 16, 2025

Readings

- Gallow(2024),“Instrumental Divergence”
- Thorstad(2024),“Against the Singularity Hypothesis”
- Harding & Kirk-Gianni (2025),“What is AI safety? What do we want it to be?”

Schedule

Time	Topic
10-12	CS Lab 10
12-1:30	Lunch Block
1:30-4:30	Seminar (SAD Group C)

Thursday, Jul 18, 2024

Time	Topic
All day	Silent Work Day in RP 909 with Katie

Week 8

Topic: Paper Writing Week

Monday, Jul 21, 2025

Paper Writing

Time	Topic
All day	Paper Writing

Tuesday, Jul 22, 2025

Schedule

Time	Topic
All Day	Paper Writing

Wednesday, Jul 24, 2024

Schedule

Time	Topic
All Day	Paper Writing

Thursday, Jul 24, 2025

Schedule

Time	Topic
All Day	Paper Writing

Friday, Jul 25, 2025

Time	Topic
All Day	Paper Writing

Week 9

Topic: Works in Progress Workshop

Monday, Jul 28, 2025

Speaker(s)

- [Solon Barocas](#)
- [Milo Phillips-Brown](#) (The University of Edinburgh)

Schedule

Time	Topic
10-11:30	Solon Barocas Speaker Sessions
11:30-1	Catered speaker lunch with Kate Vredenburgh
3:30-5	Milo Philips-Brown Speaker Session
5:30-7:30	Dinner with Speaker: Kate Vredenburgh

Tuesday, Jul 29, 2025

Schedule

Time	Topic
10-11	Student Presentation Slot 1
11-12	Student Presentation Slot 2
12-1:30	Catered Lunch: Journal Selection Workshop
1:30-2:30	Student Presentation Slot 3
3-4	Student Presentation Slot 4

Wednesday, Jul 30, 2025

Schedule

Time	Topic
10-11	Student Presentation Slot 5
11-12	Student Presentation Slot 6
12-1:30	Break
1:30-2:30	Student Presentation Slot 7
3-4	Student Presentation Slot 8

Thursday, Jul 31, 2025

Schedule

Time	Topic
10-11	Student Presentation Slot 9
11-12	Student Presentation Slot 10
12-1	Catered Lunch
1-2	Student Presentation Slot 11
2-4	Wrap-up and Evaluations
5:30-7:30	Closing Dinner + Bowling