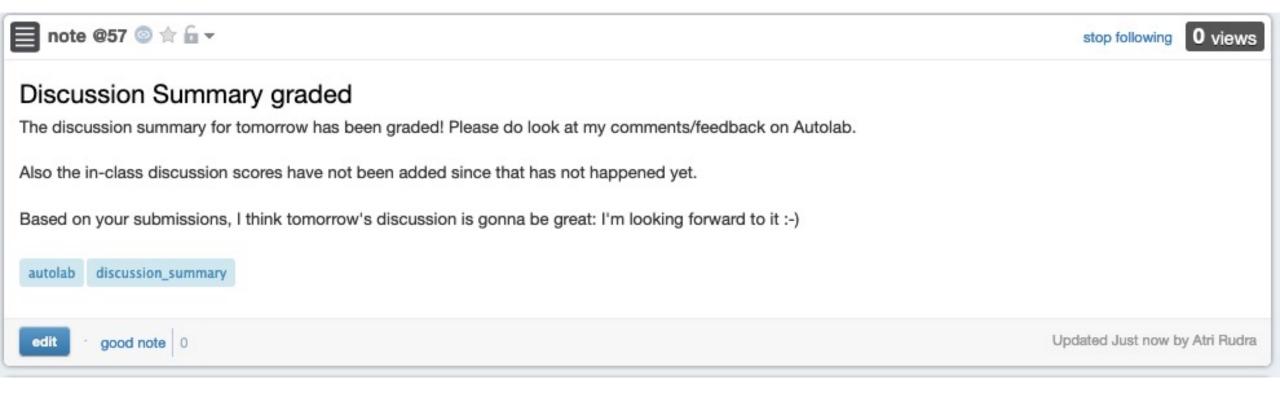# ML and Society

Apr 14, 2022

Discussion summaries graded

In-class discussion

Fairness and Abstraction in Sociotechnical Systems

ANDREW D. SELBST, Data & Society Research Institute
DANAH BOYD, Microsoft Research and
Data & Society Research Institute
SORELLE A. FRIEDLER, Haverford College, PA
SURESH VENKATASUBRAMANIAN, University of Utah
JANET VERTESI, Princeton University

A key goal of the fair-ML community is to develop machine-learning based systems that, once introduced into a social context, can achieve social and legal outcomes such as fairness, justice, and due process. Bedrock concepts in computer science—such as abstraction and modular design—are used to define notions of fairness and discrimination, to produce fairness-aware learning algorithms, and to intervene at different stages of a decision-making pipeline to produce "fair" outcomes. In this paper, however, we contend that these concepts render technical interventions ineffective, inaccurate, and sometimes dangerously misguided when they enter the societal context that surrounds decision-making systems. We outline this mismatch with five "traps" that fair-ML work can fall into even as it attempts to be more context-aware in comparison to traditional data science. We draw on studies of sociotechnical systems in Science and Technology Studies to explain why such traps occur and how to avoid them. Finally, we suggest ways in which technical designers can mitigate the traps through a refocusing of design in terms of process rather than solutions, and by drawing abstraction boundaries to include social actors rather than purely technical ones.

Thing to keep in mind

You are expected to participate ☺

Discussion Participation

During the in-class discussion, y'all will form groups of size three (3) with perhaps one of two exceptions (to form groups of size two).

What happens in the group discussion

The goal of the group discussion is to come up with two top group responses for each part of the discussion summary: `Thoughts`, `Questions` and `Epiphanies`. *Ideally*, these responses should come from one of the group members discussion summary submission. However, it is OK to come up with a new response if e.g. if the group felt it would be better to synthesize the individual group member's responses.

After the group discussion is done, each group member will present two group responses. (It is up to the group on how to divide among the `Thoughts`, `Questions` and `Epiphanies`.) I will be keeping track of individual participation and you will be graded as follows.

Discussion participation grading rubric

- `Level 0`: No participation.
- `Level 1`: Exactly one non-trivial question asked or one non-trivial answer given.
- `Level 2`: At least two non-trivial questions asked or one non-trivial answers given.

What is a non-trivial question/answer?

I do not want to formally define what questions/answers are non-trivial since this is somewhat subjective. But just to give an idea: If the question was "What did you think about the paper assigned for today's in-class discussion?". An answer "Great!" will be considered trivial whereas a non-trivial answer would be one that goes into the specifics of what part(s) of the paper you though were great. Perhaps a better phrase for non-trivial would be *thoughtful*.

Discuss!

Fairness and Abstraction in Sociotechnical Systems

ANDREW D. SELBST, Data & Society Research Institute
DANAH BOYD, Microsoft Research and
Data & Society Research Institute
SORELLE A. FRIEDLER, Haverford College, PA
SURESH VENKATASUBRAMANIAN, University of Utah
JANET VERTESI, Princeton University

A key goal of the fair-ML community is to develop machine-learning based systems that, once introduced into a social context, can achieve social and legal outcomes such as fairness, justice, and due process. Bedrock concepts in computer science—such as abstraction and modular design—are used to define notions of fairness and discrimination, to produce fairness-aware learning algorithms, and to intervene at different stages of a decision-making pipeline to produce "fair" outcomes. In this paper, however, we contend that these concepts render technical interventions ineffective, inaccurate, and sometimes dangerously misguided when they enter the societal context that surrounds decision-making systems. We outline this mismatch with five "traps" that fair-ML work can fall into even as it attempts to be more context-aware in comparison to traditional data science. We draw on studies of sociotechnical systems in Science and Technology Studies to explain why such traps occur and how to avoid them. Finally, we suggest ways in which technical designers can mitigate the traps through a refocusing of design in terms of process rather than solutions, and by drawing abstraction boundaries to include social actors rather than purely technical ones.

Thoughts

Fairness and Abstraction in Sociotechnical Systems

ANDREW D. SELBST, Data & Society Research Institute
DANAH BOYD, Microsoft Research and
Data & Society Research Institute
SORELLE A. FRIEDLER, Haverford College, PA
SURESH VENKATASUBRAMANIAN, University of Utah
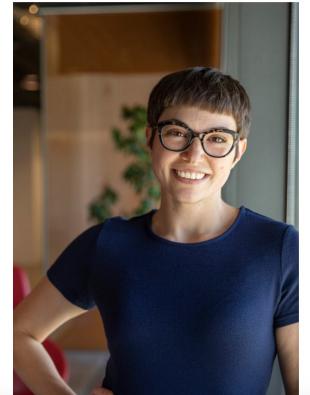JANET VERTESI, Princeton University

A key goal of the fair-ML community is to develop machine-learning based systems that, once introduced into a social context, can achieve social and legal outcomes such as fairness, justice, and due process. Bedrock concepts in computer science—such as abstraction and modular design—are used to define notions of fairness and discrimination, to produce fairness-aware learning algorithms, and to intervene at different stages of a decision-making pipeline to produce "fair" outcomes. In this paper, however, we contend that these concepts render technical interventions ineffective, inaccurate, and sometimes dangerously misguided when they enter the societal context that surrounds decision-making systems. We outline this mismatch with five "traps" that fair-ML work can fall into even as it attempts to be more context-aware in comparison to traditional data science. We draw on studies of sociotechnical systems in Science and Technology Studies to explain why such traps occur and how to avoid them. Finally, we suggest ways in which technical designers can mitigate the traps through a refocusing of design in terms of process rather than solutions, and by drawing abstraction boundaries to include social actors rather than purely technical ones.

Passphrase for today: Karen Levy

Karen Levy

Publications   Essays + Press   Teaching   Students   CV   Contact

I'm an assistant professor in the Department of Information Science at Cornell University, associated faculty at Cornell Law School, and field faculty in Cornell's Department of Science and Technology Studies and Department of Sociology.

I research the **legal, organizational, social, and ethical aspects of data-intensive technologies**. I am interested in what happens when we use digital technologies to enforce rules and make decisions about people, particularly in contexts marked by conditions of inequality. A good deal of my research considers the impact of data-intensive technologies on work and workers; I am currently writing a book analyzing the emergence of electronic monitoring in the long-haul trucking industry. I also study the role of data collection technologies in intimate relationships, and how they contribute to practices of both care and control. (Vox made a really nice video about my trucking research, and Cornell Alumni magazine profiled my work on data collection and privacy.)

I have a PhD in Sociology from Princeton University, where my dissertation examined the development of legal and organizational surveillance in the United States trucking industry. I have a JD from Indiana University's Maurer School of Law. I've been a New America National Fellow, a postdoctoral fellow at New York University School of Law's

Questions + Epiphanies

Fairness and Abstraction in Sociotechnical Systems

ANDREW D. SELBST, Data & Society Research Institute
DANAH BOYD, Microsoft Research and
Data & Society Research Institute
SORELLE A. FRIEDLER, Haverford College, PA
SURESH VENKATASUBRAMANIAN, University of Utah
JANET VERTESI, Princeton University

A key goal of the fair-ML community is to develop machine-learning based systems that, once introduced into a social context, can achieve social and legal outcomes such as fairness, justice, and due process. Bedrock concepts in computer science—such as abstraction and modular design—are used to define notions of fairness and discrimination, to produce fairness-aware learning algorithms, and to intervene at different stages of a decision-making pipeline to produce "fair" outcomes. In this paper, however, we contend that these concepts render technical interventions ineffective, inaccurate, and sometimes dangerously misguided when they enter the societal context that surrounds decision-making systems. We outline this mismatch with five "traps" that fair-ML work can fall into even as it attempts to be more context-aware in comparison to traditional data science. We draw on studies of sociotechnical systems in Science and Technology Studies to explain why such traps occur and how to avoid them. Finally, we suggest ways in which technical designers can mitigate the traps through a refocusing of design in terms of process rather than solutions, and by drawing abstraction boundaries to include social actors rather than purely technical ones.