

CS265- Big Data Systems (In Progress)

NOTE: this module should be labeled as “under development.”

Course Level:

Graduate

Course Description:

“Big data is everywhere. A fundamental goal across numerous modern businesses and sciences is to be able to utilize as many machines as possible, to consume as much information as possible and as fast as possible. The big challenge is how to turn data into useful knowledge. This is a moving target as both the underlying hardware and our ability to collect data evolve. In this class, we discuss how to design data systems, data structures, and algorithms for key data-driven areas, including relational systems, distributed systems, graph systems, noSQL, newSQL, machine learning, and neural networks. We see how they all rely on the same set of very basic concepts and we learn how to synthesize efficient solutions for any problem across these areas using those basic concepts.” [\(Course Description\)](#)

Module Topic:

Electronic Privacy and Big Data Systems

Module Author:

Diana Acosta-Navas

Semesters Taught:

Spring 2019

Tags:

- big data (CS)
- machine learning (CS)
- privacy (phil)
- moral rights (phil)
- consent (phil)
- stakeholder rights and interests (phil)

Module Overview:

In this module, we focus on the challenges that big data systems pose for some common views about what privacy is and why it matters. Through the discussion of three case studies, students are presented with philosophical concepts that capture various moral intuitions about electronic privacy. They are then prompted to consider the applicability of these concepts in the context of big data systems.

The module begins with a case study in which individuals' data was collected and analyzed with the aim of influencing their voting behavior. Students are then asked to consider the moral permissibility of employing user data and artificial intelligence for the benefit of data owners. These two cases serve to introduce the notions of stakeholders' rights and interests as potential grounds for privacy protections. Lastly students are asked to consider these concepts in light of a third case study focusing on the use of facial recognition software. The instructor moderates a discussion of the ways in which the case challenges our common assumptions about privacy and informed consent.

Connection to Course Technical Material:

During the course of the semester, students are exposed to state-of-the-art research on big data systems. They are also trained to produce original research in the area. This module guides students through an ethical analysis of recent case studies in which the collection and analysis of large datasets raises important privacy issues. In this way, it is meant to sensitize students to important ethical issues that may arise in their own work in the field.

Module Goals:

1. Introduce students to philosophical concepts that are useful for thinking about privacy in big data systems (including formal definitions of data privacy and electronic privacy, as well as the notions of stakeholder rights and interests).
2. Discuss how big data systems challenge common assumptions about privacy and consensual data use.
3. Use case studies to train students to identify and analyze ethical issues in the context of complex real-world scenarios.

Key Philosophical Questions:

1. How are individuals wronged when their data is used without their consent?
2. Do we have a moral right to electronic privacy?
3. If there is a moral right to electronic privacy, how does it apply in the context of big data systems?
4. What constitutes informed consent to data use in the context of big data systems?

Key Philosophical Concepts:

- Data Privacy

- Electronic Privacy
- Instrumental/final value
- Stakeholder rights and interests
- Moral rights
- Consent

Class Agenda:

1. Case Study 1: Facebook and Cambridge Analytica.
 - a) Small-group discussion and class-wide debrief: what was wrong with this use of data?
 - b) Overview of different types of privacy.
2. Case Study 2: Facebook's suicide prevention program.
 - a) Small-group discussion and class-wide debrief: is it morally permissible to violate a person's privacy in order to benefit them?
 - b) Key philosophical concepts: instrumental and final value, stakeholder rights and interests.
3. Case Study 3: Facial Recognition Software and Sexual Orientation
 - a) Think-Pair-Share: How does this case challenge our assumptions about privacy and informed consent?
 - b) Class-wide debrief:
 - i. The transparency paradox.
 - ii. The tyranny of the minority.

Sample Class Activity:

Students are asked to consider Facebook's suicide prevention program, which uses AI to analyze users' posts to rank their risk of suicide. Using PollEverywhere, students report anonymously whether they believe it is permissible to collect and analyze a person's data without their consent, provided you do so in order to benefit them.

After the class' responses have been projected on the classroom screen, the Embedded EthiCS fellow leads a class-wide debrief. Students on both sides of the debate are asked to explain their responses. The Embedded EthiCS fellow then introduces the concepts of stakeholder rights and stakeholder interest as a framework to conceptualize different ways in which individuals can be wronged when their privacy is breached.

These concepts are then employed to help students articulate the moral intuitions behind their responses. When privacy is conceived of as a right, we are less prone to think that it can be overridden by other considerations. When we conceive of it as an interest, we are more willing to weigh it against countervailing interests. Based on this distinction, the class is asked to reflect on whether we have a moral right to electronic privacy.