

## Embedded EthiCS@Harvard: S20 Repository Entries

Diana Acosta-Navas

1. **Course.** CS 265: Big Data Systems
2. **Course Level.** Graduate
3. **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)
4. **Module Topic.** Electronic Privacy and Big Data Systems
5. **Module Author.** Diana Acosta-Navas
6. **Semesters Taught.** Spring 2017-2018, Spring 2018-2019, Spring 2019-2020
7. **Tags.** big data [CS], machine learning [CS], privacy [phil], moral rights [phil], consent [phil]; stakeholder [phil]; rights [phil]; interests [phil]
8. **Module Overview.** In this module, we focus on the challenges that big data systems pose to our understanding of data privacy and the value we accord to it. Through the discussion of three case studies, students are presented with philosophical concepts that capture various moral intuitions about electronic privacy. Each case study is meant to test and challenge students’ intuitions, posing increasingly complex scenarios. The module begins with a case of unlawful data collection for the purpose of influencing voting behavior. Next, students are asked to consider the moral permissibility of employing user data and artificial intelligence for the benefit of data owners. Lastly, students are asked to reflect on the use of location tracking on smartphones for the purpose of contact-tracing. The notions of *informed consent*, *stakeholders’ rights* and *stakeholders’ interests* are introduced to inform these discussions. Through these different cases, the module seeks to enhance students’ understanding of the ethical nuances involved in the use of big data.
9. **Connection to Course 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.
10. **Module Goals.**

1. Understanding the formal definitions of philosophical concepts like data privacy, electronic privacy, and privacy violations, as well as the notions of stakeholder rights and interests.
2. Identifying potential privacy violations in big data systems.
3. Analyzing complex case studies to diagnose ethical issues in the context of complex real-world scenarios.

### 11. Key Philosophical Questions.

1. Why and how are individuals wronged when their personal data is used without their consent?
2. Is it ethically permissible to use personal data, when it is done for the benefit of its owners?
3. What benefits to individuals justify the use of their personal data?
4. Are there extenuating circumstances in which our privacy rights are overridden?
5. How to balance privacy protections and the protection of the common good in these circumstances?
6. What constitutes informed consent in the context of big data systems?

### 12. Key Philosophical Concepts.

- Data Privacy
- Electronic Privacy
- Privacy violation
- Interests
- Moral rights
- Consent

### 13. Assigned Readings.

<ul style="list-style-type: none"> <li>● Raskar, <i>et al.</i> (2020). "Apps Gone Rogue: Maintaining Personal Privacy in an Epidemic." <i>ArXiv.org</i>, ArXiv.org, Mar 19, 2020.</li> </ul>	<p>This paper describes a privacy-focused contact-tracing app: <i>Private Kit: Safe Paths</i>. This app was developed at the MIT Media Lab with the goal of curbing the spread of coronavirus. The paper identifies some of the most salient privacy risks posed by the development of contact-tracing apps and discusses how <i>Private Kit: Safe Paths</i> addresses such concerns.</p>
--	---

### 14. Class Agenda.

1. Case Study 1: Facebook and Cambridge Analytica  
Privacy as a descriptive and normative concept
2. Case Study 2: Facebook's suicide prevention program  
Stakeholder rights and interests.
3. Case Study 3: Contact-Tracing and Privacy  
Privacy rights in extenuating circumstances

### **15. Sample Class Activity.**

Students are asked to consider Facebook's suicide prevention program, which analyzes users' posts in order to assess and rank suicide risk. Using *Poll Everywhere*, students report anonymously whether they believe it is permissible to collect and analyze a person's data without their consent, provided that it is done for their benefit. After the class' responses have been projected on the classroom screen, the Embedded EthiCS TA leads a class-wide debrief. Students on both sides of the debate are asked to explain their responses. The TA then introduces the concepts of stakeholder rights and interests 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 that motivate 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. Afterwards, students are prompted to reflect on the notion of informed consent and its application in the context of big data systems.

### **16. Module Assignment. N/A**

### **17. Lessons Learned.**

Student response to this module was overall positive both times it was taught (Spring 2019 and 2020). A few lessons stand out.

- The module was taught on Zoom in Spring 2020. The active learning exercises and activities took longer than they did during the in-person version of the module. For this reason, the final discussion had to be cut short and many of the important points were not developed in depth. For modules taught on Zoom, instructors might consider reducing content by 10-15%.
- Though asking students to contribute at various points is important to keep the class active and engaged, these spaces should be clearly marked, as student participation takes longer on Zoom.
- Students were quite interested in the contact-tracing discussion. More time should be allocated for this case-study.
- Someone should be assigned to monitor the Zoom chat and students' raised hands.