CS 288 Repository Entry
Embedded EthiCS @ Harvard Teaching Lab

Overview

Course:	CS 288: AI for Social Impact
Course Level:	Graduate
Course Description:	"Recent years have seen AI successfully applied to societal challenge problems; indeed, it has a great potential to provide tremendous social good in the future. In this course, we will discuss the successful deployments and the potential use of AI in various topics that are essential for social good, including but not limited to health, environmental sustainability, public safety and public welfare. We will focus on challenges in "AI for Social Impact" (AI4SI), what makes projects successful, and why projects fail. A key part of this course will be to start AI4SI projects with local area non-profits."[1]
Module Topic:	Value Sensitive Design for AI4SI
Module Author:	Michael Pope
Semesters Taught:	Fall 2022
Tags:	AI for Social Impact [CS], Design [CS], Stakeholders [CS], Responsibility [phil], Fairness [phil] Direct and indirect stakeholders [phil], Value [phil]
Module Overview:	This module focuses on questions of value and their effect on the design and deployment of AI for social impact (AI4SI). The module begins with reflections about what AI4SI systems are for and how ethical values relate to technical trade-offs. The module then introduces a framework for evaluating the values that influence and are influenced by AI4SI, Value Sensitive Design. Students learn to identify direct and indirect stakeholders, their values, and potential impacts of AI interventions over time and at different levels of pervasiveness. Having applied the system to a case study together, students independently examine how Value Sensitive Design helps to identify potential ethical features of COMPAS, an algorithm used in criminal sentencing decisions. To close, we discuss how preemptive consideration of values in AI4SI design and deployment can improve systems like COMPAS.
Connection to Course Material:	Students in this course design their own AI4SI projects. As part of these projects, students are asked to identify some ethical challenges and pitfalls of their project, paying special attention to broader impacts, and explain how their project would address these challenges. The framework and case study in this module help students to identify impacts on direct and indirect stakeholders as well as consider how those impacts change with scale.

Value Sensitive Design provides a framework for assessing the ethical impacts of student projects in the course. A case study in criminal justice was selected in consultation with Professor Tambe. The module provides students with a framework for identifying and evaluating ethical dimensions of AI interventions in real-world contexts. We focused on a case study in criminal justice

[1] AI for Social Impact Harvard course catalog listing: here.

		for two reasons. First, since the system we discussed was already in use, it supplied bountiful evidence of how such systems could be designed, deployed, and reviewed. Second, by situating a system's actual impacts within a design framework that is sensitive to the values of those impacted by the system, students can more proactively consider how their projects *could* impact relevant stakeholders.

Goals		
Module Goals:	1. Introduce students to the Value Sensitive Design framework 2. Identify the role of values in the design and deployment of AI for social impact 3. Engagement with strategies for proactively designing AI that is sensitive to stakeholder perspectives	
Key Philosophical Questions:	1. What goals and criteria for success should computer scientists adopt when designing AI systems focused on social impact? 2. How can sensitivity to stakeholder values produce more responsible AI design and deployment? 3. How could a system's application over time and across multiple contexts alter a system's ethical impact?	Q1: In answering this question, students reflect on what designers aim to achieve through their interventions and how we might assess the success or failure of those systems. The question orients the course toward sensitivity to values and the social impacts of a system. Q2: This question invites students to pivot from identifying stakeholders and their values to evaluating AI interventions according to those values. In the module, we discuss considerations of fairness and impacts on cooperation that arise from racial disparities in COMPAS' use. Q3: This question asks students to reflect on the ways an AI system's use over time and beyond its original design context can impact stakeholders. In the module, we examine some impacts of COMPAS' application to new contexts (e.g., determining appropriate prison security levels).

Materials

Key Philosophical Concepts:

- Value-ladenness of AI
- Direct and indirect stakeholders
- Responsibility
- Fairness

The course begins with a discussion of the ways AI systems are *value-laden*. That is, students reflect on the ways that AI is shaped by values and, in turn, can impact what people value. They identify those impacted by AI systems, namely *direct* and *indirect stakeholders*. Having identified stakeholders, students consider how responsible design is sensitive to and can shape stakeholder values. Through a discussion of racial bias in the case study, we discuss how a system's deployment over time can impact the fairness of the system.

Assigned Readings:

- Satell, Greg and Yassmin Abdel-Magied. 2020. "AI Fairness Isn't Just an Ethical Issue." *Harvard Business Review*.
- Hao, Karen and Jonathan Stray. 2019. "Can you make AI fairer than a judge? Play our courtroom algorithm game?" *MIT Technology Review*.

Satell and Abdel-Magied (2020) discuss the ubiquity of AI systems and introduce ethical issues that arise through their use. They discuss sources of bias (e.g., biased datasets) and present strategies for mitigating bias. These strategies underscore the importance of including ethical thinking at the earliest stages of AI design.

Hao and Stray (2019) allows students to explore ethical dimensions of COMPAS through an interactive game. In the game, students manipulate the threshold for calibrating the algorithm to see if an outcome is more or less fair.

Implementation

Class Agenda:

1. Welcome and introductions
2. Small-group discussion 1 (Reflecting on the criteria for successful impacts)
3. Debrief
4. Introduction to Value Sensitive Design as a means for identifying and evaluating the impacts of AI systems
5. Case Study: COMPAS

	6. Small-group discussion 2 (Applying Value Sensitive Design to COMPAS)	
Sample Class Activity:	Students were asked to discuss the following in small groups, before reporting back to the class for further discussion: 1. Who are the direct and indirect stakeholders in the deployment of COMPAS? 2. What values might direct and indirect stakeholders have in criminal justice contexts? 3. What are short-term and long-term impacts of COMPAS? 4. How are the impacts of COMPAS over time sensitive to its pervasiveness? Following the small-group discussion, students reported their findings to the class.	To facilitate this discussion, students received a handout with questions and additional information for framing their discussions. Students sat at round tables near one another, allowing for easier collaboration within groups. During the large-group debrief, students engaged in back-and-forth over potential impacts of the system, especially across time and pervasiveness dimensions.
Module Assignment:	There was a pre-meeting online discussion of the reading, in which students identified (1) those impacted by COMPAS and (2) how they could be impacted. The post-class assignment for this module was integrated into student reports for final projects. Alongside answers to technical questions about the AI systems they designed, students were asked to write a statement of the ethical impacts of their projects within the Value Sensitive Design framework.	The pre-meeting discussion invited students to identify and reflect on the impacts of the central case study for the class meeting. The post-class assignment required students to identify and assess the ethical challenges and pitfalls of their projects through the lens of Value Sensitive Design.
Lessons Learned:	A sentence or two summarizing student reactions to the module, followed by an enumerated list of pedagogical insights drawn from developing and teaching the module. Engagement with this module was overwhelmingly positive. In particular, students appreciated the ways that the Value Sensitive Design framework allowed them to identify and evaluate the impacts of a system on stakeholders. 1. Graduate students can be counted on to complete the reading and come to class ready to discuss ethical dimensions of their work in depth. 2. Small-group discussions were productive and facilitated a back-and-forth in large-group discussions, organically emphasizing	The first activity in this module is designed to help students brainstorm possible ethical impacts of an AI system. For graduate students in a course on AI for social impact, students come prepared for the more focused discussion activity.

	the ways trade-offs present technical and ethical problems in students' work. 3. While the COMPAS case study garnered high levels of student engagement, the amount of public attention on the case can impede the imaginative work that is crucial for applying Value Sensitive Design.	