Repository Entry - CS 181
Embedded EthiCS @ Harvard Teaching Lab

Overview		
Course:	CS 181: Machine Learning	
Course Level:	Upper-level undergraduate	
Course Description:	"Introduction to machine learning, providing a probabilistic view on artificial intelligence and reasoning under uncertainty. Topics include: supervised learning, ensemble methods and boosting, neural networks, support vector machines, kernel methods, clustering and unsupervised learning, maximum likelihood, graphical models, hidden Markov models, inference methods, and computational learning theory. Students should feel comfortable with multivariate calculus, linear algebra, probability theory, and complexity theory. Students will be required to produce non-trivial programs in Python."	
Module Topic:	Moral Responsibility in Development	
Module Author:	Ellie Lasater-Guttmann	
Semesters Taught:	Spring 2022	
Tags:	Machine learning [CS], causal chain [phil], moral responsibility [phil], backward-looking responsibility [phil], forward-looking responsibility [phil]	
Module Overview:	The module uses a case of racial bias in healthcare to model forward-looking and backward-looking moral responsibility. Students build causal chains to pinpoint what went wrong in the healthcare case and how agents could have acted differently to prevent bad outcomes.	
Connection to Course Material:	Students had spent the previous month on prediction problems in machine learning. The module's centerpiece case is a ML algorithm that predicts healthcare expenditure.	The healthcare case was an example of the exact type of models the students had been building. It also presented interesting challenges philosophically because the causal chains were complex.

Goals		
Module Goals:	- Work through choice points in designing an algorithm to improve a healthcare system, given a desired outcome - Reevaluate whether that outcome is in fact the proper outcome - Draw a causal chain from each design decision to a bad outcome - Determine where different aspects of responsibility lie for the bad outcome, including backward-looking and forward-looking responsibilities	
Key Philosophical Questions:	1. When is a developer morally required to mitigate bad outcomes?	These questions build over the course of the module, as students

	2. Can we be morally responsible, even if our design choices are not the sole cause of an outcome? 3. What design choices contribute to bad outcomes? 4. How do forward-looking and backward-looking responsibilities differ?	perform different steps in the in-class activity.

Materials		
Key Philosophical Concepts:	● Causal chain / choice points ● Causal responsibility ○ Sufficient cause ● Moral responsibility ○ Backward-looking ○ Forward-looking	Students are in a position to respond to question #1 once they have understood the nature of different types of moral responsibilities and how they can relate to causal responsibility. Causal chains/choice points illuminate the concepts of causal and ethical responsibility by showing how decisions can cause certain outcomes, which then have corresponding ethical responsibility.
Assigned Readings:	● Not applicable (students were not in a position to be able to complete a reading assignment ahead of the module).	Had students been able to complete a reading, I would have had them read the healthcare case on which the module activity is based: https://www.science.org/doi/10.1126/science.aax2342 .

Implementation		
Class Agenda:	1. Lecture on causal chains and moral responsibility (20 minutes) 2. Scenario Part 1 (10 minutes) 3. Classwide regroup (5 minutes) 4. Scenario Part 2 (5 minutes) 5. Classwide regroup (5 minutes) 6. Scenario Part 3 (5 minutes) 7. Classwide regroup (5 minutes) 8. Scenario Part 4 (15 minutes) 9. Classwide regroup (5 minutes)	I strongly recommend the classwide regroups after each section of the activity, to ensure we're keeping the philosophical learning at the forefront.
Sample Class Activity:	*Part 1 - How do you approximate healthcare need?* Students were given several options for how a program can approximate future healthcare need. They must choose one, and then anticipate the ethical consequences. Similarly, for future parts, the students go through decision-points in the algorithm	This module centered on this interactive activity. The module would have been substantially less effective if it had been removed. Students enjoyed being participants in the healthcare case, rather than having an

	design and implementation that lead to ethical implications. Part 2 - What does your company do with the predications it calculates? Part 3 - What would a successful outcome look like? What about a failure? Part 4 - The actual outcome was a failure. Draw a causal chain that led to this outcome. Identify forward-looking and backward-looking responsibilities of the agents.	observational third-party perspective on it.
Module Assignment:	Select a real-life outcome in Artificial Intelligence or Machine Learning that you believe is morally wrong. You can select your own outcome from the news or select one of the outcomes in the two options below: ● COMPAS, a case management tool predicting recidivism that flagged "blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend" (Angwin 2016). ● An NLP algorithm filled in the inference "Man is to ____ as woman is to _____" with "Man is to computer programmer as woman is to homemaker" (Bolukbasi et al, 2016). Draw a causal chain that resulted in this outcome and circle the choice points that were the largest contributors to the outcome. At each morally relevant choice point, write two alternative decisions that could have prevented the outcome.	Students were specifically tasked with investigating another case that we did not cover, as a small research assignment. If there is time for an assignment of this kind, I would strongly recommend it as it prompted the students to take their discoveries from the module into the wild.
Lessons Learned:	1. The activity was successful and integral to learning our philosophical concepts. 2. I would recommend solidifying for the students why the company would have used healthcare cost as a proxy for healthcare need. This is the most contested aspect of the case, and the real-life reason is a disappointing one that would have fueled additional discussion.	