Repository Entry Template
Embedded EthiCS @ Harvard Teaching Lab

Overview

**Course:** CS 182: Artificial Intelligence

**Course Level:** Upper-level undergraduate

**Course Description:** “Artificial Intelligence (AI) is already making a powerful impact on modern technology, and is expected to be even more transformative in the near future. The course introduces the ideas and techniques underlying this exciting field, with the goal of teaching students to identify effective representations and approaches for a wide variety of computational tasks. Topics covered in this course are broadly divided into search and planning, optimization and games, and uncertainty and learning. Special attention is given to ethical considerations in AI and to applications that benefit society.”

**Module Topic:** AI, Responsibility, and Impact

**Module Author:** Anni Räty

**Semesters Taught:** Fall 2023

**Tags:** AI for social impact [CS], artificial intelligence [CS], optimization [CS], benefit [phil], decision-making [phil], welfare [phil], respect [phil], justice [phil]

**Module Overview:** In this module students practice ethically evaluating the consequences of computing professionals’ choices. The module introduces students to ACM’s code of ethics, and to philosophical concepts of welfare, respect, and justice. The module focuses on a case study where AI is used to optimize NGO interventions to improve maternal and child health outcomes in rural India (SAHELI).

**Connection to Course Material:** In this class students learn about topics in artificial intelligence. The SAHELI case study connects with two prior class modules on restless multi-armed bandits (RMAB).

This module is for a class where students learn about topics in artificial intelligence. The lectures cover various applications of AI, including the SAHELI system, which was recommended as a topic by the CS course heads. Many other applications covered in class would also make a good case study for a similar module.

Goals

**Module Goals:**

1. Introduce students to the idea of professional responsibility.
2. Provide students with tools for ethically evaluating the impact of AI systems.
3. Practice identifying interventions into an AI system that affect its ethical impact.

---

1 https://seas.harvard.edu/computer-science/courses
Key Philosophical Questions:
1. How can we ethically assess outcomes?
2. How can a computing professional make responsible decisions?
3. How might a system that aims to do good, fall short of norms of justice and respect?

Materials

Key Philosophical Concepts:
- Welfare
- Justice
- Respect
- Responsibility

In this module students are introduced to the concepts of welfare, justice, and respect, as tools for ethically evaluating the impact of AI systems. In class activities, students learn about how welfare and justice/respect can sometimes be at odds in cases where AI is used for social good. The concept of responsibility is discussed in connection with ACM’s code of ethics.

Assigned Readings:

This paper covers the SAHELI case study. It introduces the application, and also briefly discusses the ethical questions that pertain to it.

Implementation

Class Agenda:
1. Introducing Embedded EthiCS, the ACM code of ethics, and professional responsibility.
2. Think-Pair-Share: We often have to ask ourselves what would be the responsible thing to do (voting, consumer decisions, vaccinations). Share the last time you had to do this, and how you made a decision.
3. Three kinds of ethical impact:
   a. Welfare
   b. Respect
   c. Justice
5. Group discussion: Not knowing how SAHELI actually impacted the participants in Arman’s program, how might a system like this...
   a. promote/fail to promote welfare?
   b. promote/fail to promote respect?
   c. promote/fail to promote justice?

6. Discussion of SAHELI’s actual impact. What kinds of changes could be made to the system to improve its impact?

<table>
<thead>
<tr>
<th>Sample Class Activity:</th>
<th>The module assignment can be used as a third class activity.</th>
</tr>
</thead>
<tbody>
<tr>
<td>This module begins with a “Think-Pair-Share” exercise where students identify occasions where they had to decide what would be the responsible thing to do. Later in the module, students discuss the SAHELI case study and its potential ethical impact in small groups (3-4).</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Module Assignment:</th>
<th>For this iteration of the module, students were asked to write a 400 word forum post responding to the prompt. This module was taught at the very end of term, so time was constrained; an earlier module might involve more writing.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consider the applications for AI that you have encountered in this class. Choose one to evaluate its impact in terms of either welfare, justice, or respect. What changes would improve the impact of the application? For example...</td>
<td></td>
</tr>
<tr>
<td>a. Changes in design?</td>
<td></td>
</tr>
<tr>
<td>b. Changes in data?</td>
<td></td>
</tr>
<tr>
<td>c. Changes in the social context?</td>
<td></td>
</tr>
<tr>
<td>d. Changes to policy?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lessons Learned:</th>
<th>Student reception was overall positive. The SAHELI case study worked well for this module, but there are other options in the course syllabus as well. This is a large lecture course where students may not be used to group discussion, so structuring the group discussions and prompting students to share is essential.</th>
</tr>
</thead>
<tbody>
<tr>
<td>This is a large lecture course where students may not be used to group discussion, so structuring the group discussions and prompting students to share is essential.</td>
<td></td>
</tr>
</tbody>
</table>
Additional Research Notes

1. While conducting your research for this module, what materials (articles, blog posts, podcasts, etc.) did you find most helpful? Please include one or two sentences by way of explanation if it won’t be obvious to a future GF what role the material played in your preparation for the module.
   a.

2. While developing this module, did you have any ideas that were left on the cutting room floor (i.e. ideas about potential topics, readings, activities etc. that were not ultimately incorporated in the module or final repository entry)? If so, please record them here and briefly explain why they did not make the final cut (e.g. time constraints, CS instructor preferences, etc.).
   a. Many of the applications discussed in this course could be used as a case study. CS course heads had recommendations available.

3. A more informal take on lessons learned: What else should a future GF know if we have the opportunity to run this module again? For example, based on the actual performance of the module you may have additional insights or speculations to share – If a class activity was successful, do you think the class size was a key factor? Did you perceive any differences between undergraduate vs. graduate students with respect to receptiveness to the module or success on the assignment? Etc.
   a. Class activities went well, but some students required a bit of nudging to join in the conversation.
   b. The module assignment was very short for this iteration, in part because the module was delivered on the last day of class. There could be a more substantial writing assignment if the module is delivered earlier in the semester.

4. After reviewing the student feedback form for the module, were there any comments or general takeaways that you think would be useful for future GFs to take into consideration if they are tasked with repeating this module?
   a.