Embedded EthiCS: Origins, Evolutions & Lessons

Barbara J. Grosz

Harvard University

Embedded EthiCS @ Harvard

Roots of Embedded EthiCS at Harvard

- Roots of Embedded EthiCS at HarvardCS108
  - Intelligent Systems: Design and Ethical Challenges
    - Classroom activity surprise
    - Demand
  - Graduate Teaching Fellows, Co-Instructor
  - “More, please …” and an introduction to Alison Simmons
- Some lessons and looking forward

Embedded EthiCS @ Harvard

An Enlightening Class Activity

- Class activity in "Intelligent Systems: Design and Ethical Challenges" (CS 108)

  Imagine that you are a member of the group at some social media company that has as its main responsibility making revenue from advertisements. This group, which we'll dub the "money team", has just signed a major client who is entering the market with an innovative and empirically-validated fitness solution. Based on your knowledge as a user of features of Facebook or some other social networking platform, please do the following:

1. List 3-4 comparative advantages/disadvantages of ….
2. List 3-4 advantages/disadvantages of user-user vs. item-item algorithms
3. List 5 features of a user's profile/data you want your team to consider using in the algorithm that decides to which users to post this ad and where to post it, for content- or collaborative-based recommendations (your choice)

Embedded EthiCS @ Harvard

An Unmet Challenge: Dialogue Systems From Gaffes to Ethical Problems

And now there's ChatGPT and its siblings.

B: ''What's something nice that your sister does for you?''

T: ''She does nothing nice to me.''

B: ''Well, what is the last nice thing your sister did?''

T: ''She helped me with my project — and then she destroyed it.''

B: ''Oh, yeah, tell me more!''

…

B: ''Have you told your sister lately how cool she is?''

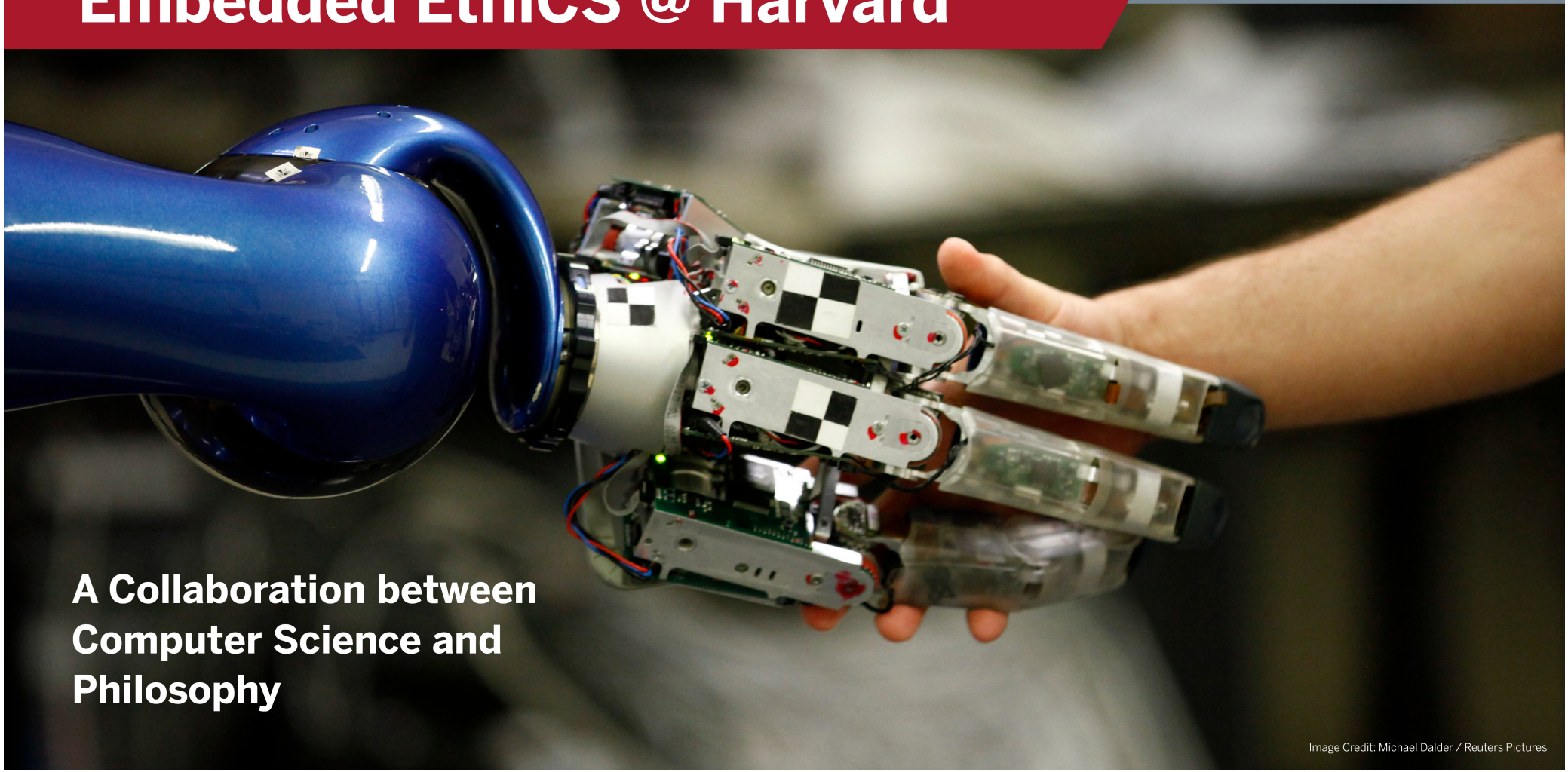
T: ''No. She is not cool.''

B: ''You never know, she might appreciate hearing it.''

NYTM September 2015

Embedded EthiCS @ Harvard

Embedded EthiCS @ Harvard



A Collaboration between Computer Science and Philosophy

Image Credit: Michael Dalder / Reuters Pictures

Evolution: From 4 courses and 1 Grad Fellow to ….

- Spring 2017:
  - 4 courses (“great ideas” intro), data science, networks (theory), HCI
  - 1 graduate fellow

- Spring 2023:
  - 11-12 courses each semester
  - 47 distinct courses
  - 9500 students, and more as extension school students
  - 125 modules delivered
  - **Module repository**: open source, augmented with teaching advice materials
  - **Teaching lab**: philosophers and computer scientists meet weekly to collaborate on module development
  - Modules delivered by **postdocs** and **co-piloted** with faculty.

Embedded EthiCS @ Harvard

Some Lessons from the Launching Years

- Start small
- Embedding is key and philosophical expertise matters
- Philosophy graduate and CS & Philosophy postdoctoral fellows contribute essentially.
- The Teaching Lab is crucial for cross-disciplinary teamwork
- Collaboration is essential

Embedded EthiCS @ Harvard

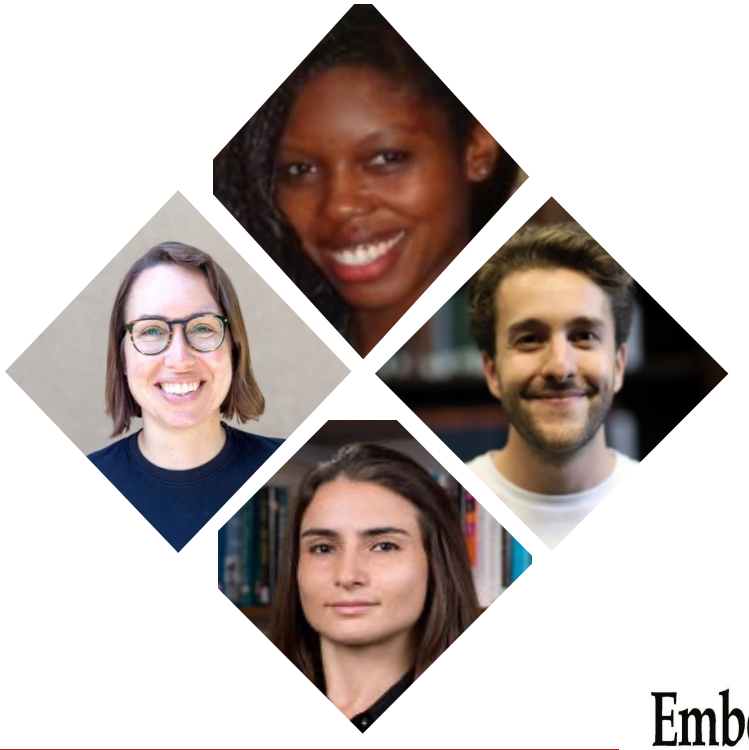
Benefits for Philosophy Graduate Students and Postdocs

• …working with a group of people who truly believe they can make a positive difference in the world - even though they clearly understand just how challenging it is. Working with that ambition and that hope in mind is extremely energizing. It's definitely one of the most rewarding intellectual experiences I've had. (D)

• I realized how much overlap there is with research questions that are central in our department. (E)

• We learned to design a lecture from the ground up. We're also learning to do research in a field of practical ethics, which is a different skill set from those one develops as a regular TF…and the conditions for doing that research are ideal: you have access to some expert in CS and can ask them as many questions as you want while you're working things out. (S)

• Topics for two of my dissertation papers; an AOC [area of competence] that is very useful on the job market; and fun times with [TL colleagues] that are non-replicable. (K)



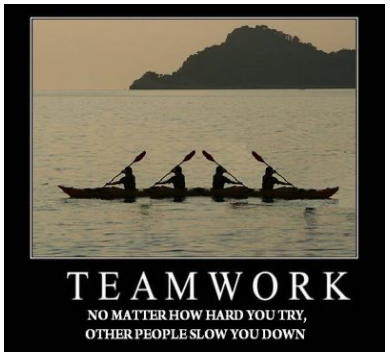
Embedded EthiCS @ Harvard

Embedded EthiCS @ …

Embedded EthiCS @ Harvard

Embedded EthiCS Collaborations Current and Future

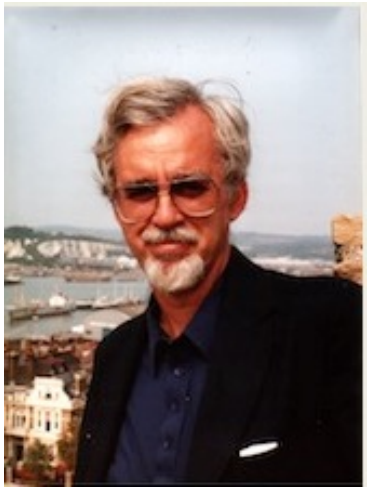
- Hub website(s): Linking websites with modules, teaching tips and more: https://embeddedethics.seas.harvard.edu/
- Joint ventures:
  - Development of methods for assessment
  - Workshops for interested colleges/universities
  - Support of industry practitioners
  - K-12
  - Beyond CS
- CS, philosophy, social sciences: see National Academies report: Fostering Responsible Computing Research: Foundations and Practices



Embedded EthiCS @ Harvard

Earlier Roots:
CSLI's Situated Language Project

David Israel, SRI



John Perry,
Stanford University

Embedded EthiCS @ Harvard

Many Teams, Evolving Approaches, Bountiful Thanks

Embedded EthiCS @ Harvard