From Ethical Challenges of Intelligent Systems to Embedding Ethics in Computer Science Education

Barbara J. Grosz
School of Engineering and Applied Sciences
Harvard University

AI Limitations and Smart Toys
The Need to Address Ethical Challenges

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!''

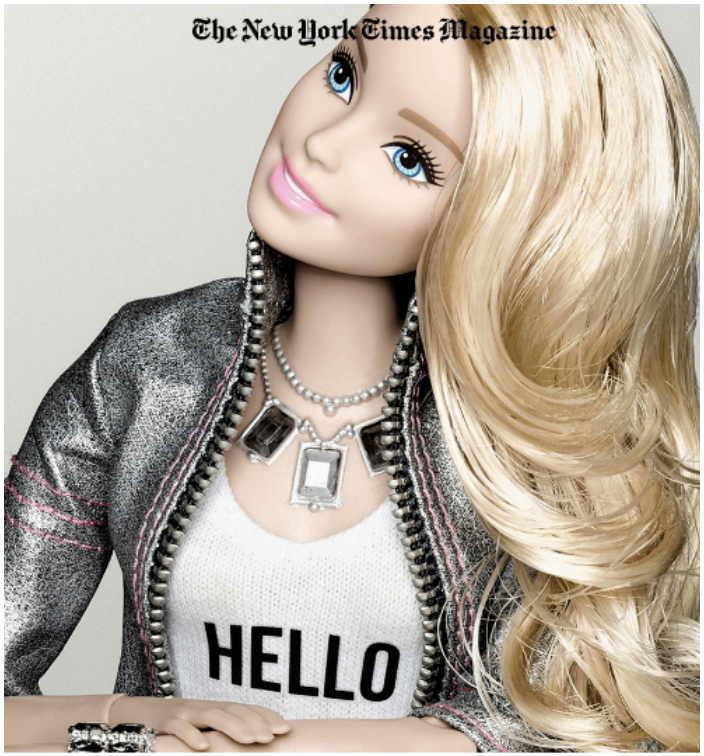
…

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.''

It's hard to keep people from "straying" from the topic (that the bot thinks is the topic).



Everyday Interactions and Failures to Meet Cognitively Reasonable Expectations

- Where is the nearest ER?
- Where can I get a flu shot?
  - Sample answers:

    What kind of businesses are you looking for?

    Launches iTunes
- Where can I go to get a sprained ankle treated?"
  - Sample answer:

    "Here's what I got"

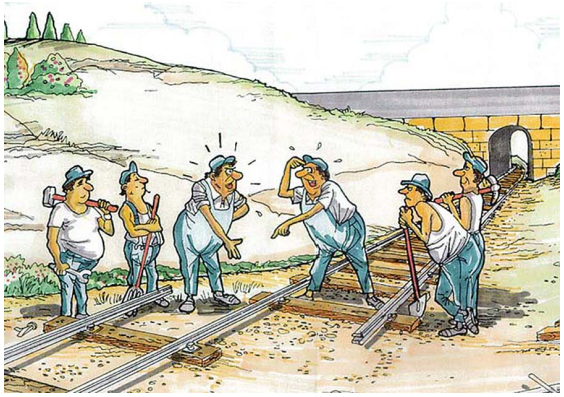
    <list of webpages that describe how to treat a sprained ankle>

Not simply annoying or slightly amusing, but raising ethical issues.

credit to students in CS108: Intelligent Systems: Design and Ethical Challenges

Set the Right Goal: AI Complementing People Rather Than AI Replacing People

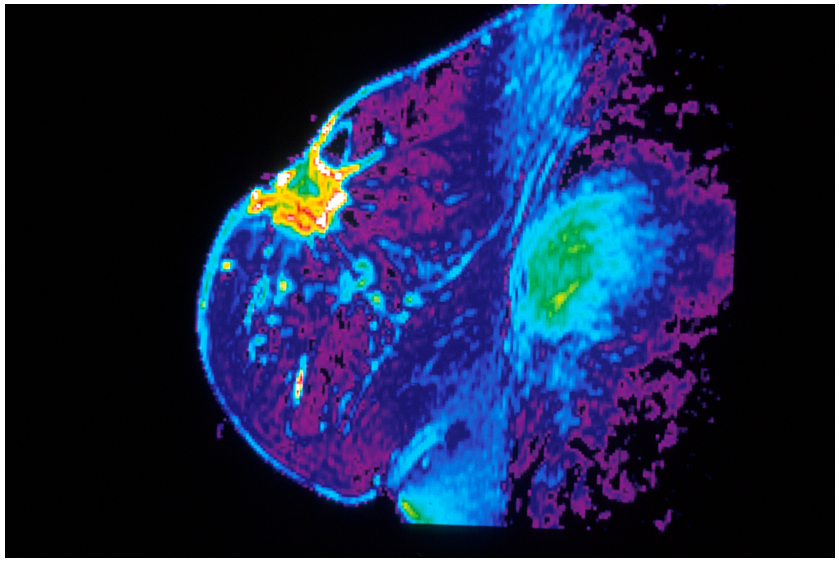
Design of Teamwork-Capable Systems

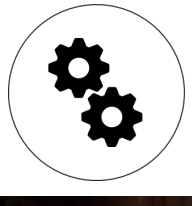


Teamwork is not simply the sum of individual plans.

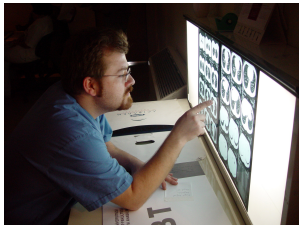
"..the capabilities needed for collaboration cannot be patched on but must be designed in from the start. " (Grosz,1994)

Complementary Approach for Data-dependent approaches: Radiology Example

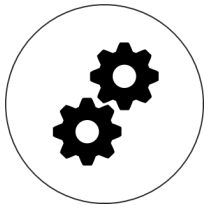




6% error



3.4% error



+

0.52% error

Example credit: E. Kamar, Microsoft Research, "Hybrid Intelligence".

Cognitive Influences on Ethical Choices?
CS 108 recommender-system exercise experience

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 content-based vs. collaborative filtering.
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

Ethics from the start, continuously, ubiquitously

- Challenges of Teamwork: "*..the capabilities needed for collaboration cannot be patched on but must be designed in from the start.* " (Grosz, 1994)
- Ethics must be taken into account from the start...
  - Design with ethics principles in mind
  - Design ethical reasoning systems
- For systems to better in the future:
  - Educate computer scientists about the challenges: Harvard launch of Embedded EthiCS.
  - Including ethics in project/product design.
  - Educate the public to set expectations correctly

Embedded EthiCS: What and Why?

- Goals:
  - Integrate ethics into the computer science curriculum as a whole, providing expertise in ethics as deep and as broad as that in computer science itself.
  - Students learn to identify and reason clearly about the ethical implications of technology *while* they are learning ways to develop and implement algorithms, design interaction systems, and code, and not as a separate endeavor.
- Rationale*:* compounding of modules throughout the curriculum will
  - demonstrate breadth of applicability across areas of CS,
  - reinforce the importance of bringing ethical reasoning to bear, and
  - *habituate* students to thinking ethically,
  - rather than just interrupting their technically-focused course of study for one moment.

Embedded Ethics: How?

- A practical approach: embed philosophers with expertise in ethics directly into the teaching of CS courses.
- Embedded EthiCS TFs: Ph.D. students and postdoctoral fellows in Philosophy with strong backgrounds in ethics and proven teaching abilities.
- Ethics "mini-modules":
  - class sessions with "active learning" engagement with an ethical challenge raised by some topic covered in the class
  - assignment related to the class sessions

Embedded EthiCS Pilot Courses and Assessments

AREA	TITLE	CHALLENGES
Introductory Courses	Great Ideas in Computer Science	The Ethics of Electronic Privacy
	Introduction to Computer Science II	Morally Responsible Software Engineering
	Advanced Topics in Data Science	Moral Considerations for Data Science Decisions
Theory	Fairness, Privacy, and Validity in Data Analysis	Diversity and Equality of Opportunity in Automated Hiring Systems
CS and Economics	Topics at the Interface of Economics and Computing	Interpretability and Fairness
	Networks	Facebook, Fake News, and the Ethics of Censorship
	Economics and Computing	Matching Mechanisms and Fairness
Programming Languages and Computer Systems	Programming Languages	Verifiably Ethical Software Systems
	Data Systems	Data and Privacy
	Big Data Systems	Privacy and Statistical Inference from Data
Human-Computer Interaction	***Design of Useful and Usable Interactive Systems***	Inclusive Design and Equality of Opportunity
Artificial Intelligence	***Machine Learning***	Machine Learning and Discrimination
	Introduction to AI	Machines and Moral Decision-Making
	Autonomous Robot Systems	Robots and Work

Toward Sustainability

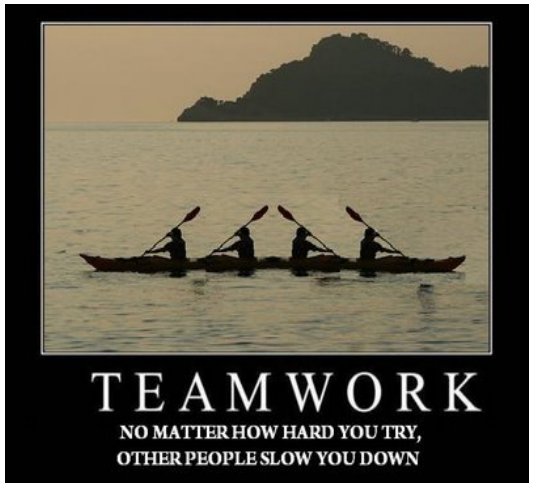
- Faculty roles
  - participate in module design
  - are at the class session
  - commit to assignment
- TFs roles
  - answer questions with philosophical depth and expertise
  - can embed in 3-4 courses; many benefits
- Key features of successful modules
  - relevant topic
  - active learning: participatory activities
  - assignment
  - links back in subsequent class meetings

Challenges to Implementation

- Measuring long-term effectiveness.
- Cross-disciplinary "ignorance concerns": do I know enough to participate in this kind of teaching?
  - Philosophers worry about lack of technical expertise.
  - Computer science faculty and TFs worry about their lack of knowledge of ethical reasoning and philosophical theories.
- Building a corpus of successful Embedded EthiCS modules.
- Building corps of Philosophy Ph.D. students and postdoctoral fellows for the program and integrating into their education and training programs.
- University support for mounting of the program: design of course content, TF support, administrative support.

Teamwork Needed
Ethics is Everyone's Responsibility

To be truly "smart" a system must (be designed to) work well *with* people.



YES!