CS 4530
Fundamentals of Software Engineering

Module 16A: Engineering Ethical Software

Adeel Bhutta, Joydeep Mitra and Mitch Wand
Khoury College of Computer Sciences

Learning Goals

- By the end of this lesson, you should be able to…
  - Illustrate how software can cause inadvertent harm or amplify inequities
  - Explain the role of human values in designing software systems
  - Explain some techniques that software engineers can use in producing software systems that are more congruent with human values.

Ethically and morally problematic technology is **everywhere**!

- Algorithms that gate access to loans, insurance, employment, government services…
- Algorithms that perpetuate or exacerbate existing discrimination
- Bad medical software can kill people (Therac-25)
- UIs that discriminate against differently-abled people
- Third-party data collection for hyper-targeted advertising
- LLM's that harvest copyright or personal data
- And on… and on… and on…

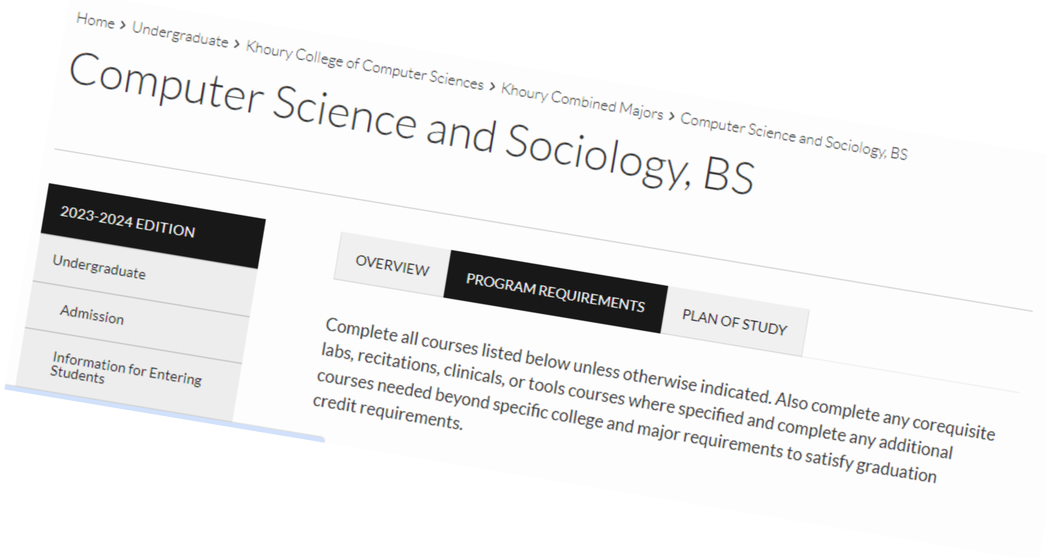
And this is only the tip of the iceberg

- Other Challenges:
  - interfaces and systems designed to be addictive;
  - corporate ownership of personal data;
  - weak cyber security and personally identifiable information (PII) protection;
  - and many more …

SOCL 4528. Computers and Society. (4 Hours)

Focuses on the social and political context of technological change and development. Through readings, course assignments, and class discussions, offers students an opportunity to learn to analyze the ways that the internet, artificial intelligence, and other technological advances have required a reworking of every human institution—both to facilitate the development of these technologies and in response to their adoption.

Attribute(s): NUpath Difference/Diversity, NUpath Societies/Institutions



Home › Undergraduate › Khoury College of Computer Sciences › Khoury Combined Majors › Computer Science and Sociology, BS

Computer Science and Sociology, BS

2023-2024 EDITION

Undergraduate

Admission

Information for Entering Students

OVERVIEW | PROGRAM REQUIREMENTS | PLAN OF STUDY

Complete all courses listed below unless otherwise indicated. Also complete any corequisite labs, recitations, clinicals, or tools courses where specified and complete any additional courses needed beyond specific college and major requirements to satisfy graduation credit requirements.

Equity and Software

As new as software engineering is, we're newer still at understanding its impact on underrepresented people and diverse societies.

We must recognize imbalance of power between those who make development decisions that impact the world.

and those who simply must accept and live with those decisions that sometimes disadvantage already marginalized communities globally.

Quote: "Software Engineering at Google: Lessons Learned from Programming Over Time," Wright et al, 2020 (O'Reilly)

Recognize inequities in your software

One mark of an exceptional engineer is the ability to understand how products can advantage and disadvantage different groups of human beings

Engineers are expected to have technical aptitude, but they should also have the discernment to know when to build something and when not to

Demma Rodriguez
Head of Equity Engineering, Google 2018-2020
Meta 2020-2022
AirBnB 2022-present

Quote: “Software Engineering at Google: Lessons Learned from Programming Over Time,” Wright et al, 2020 (O’Reilly)

More than *don't be evil*

- Engineering equitable software requires conscious effort
- How do we weigh competing interests and values?
  - How do we determine what "the right thing" is?
  - How do we convince our investors/managers to take this action?

Identify Unintended Consequences

- Technology *will* be adopted in unanticipated ways. Being intellectually rigorous means considering and mitigating risks in designs ahead of time.
- What if:
  - Our recommendation system promotes misinformation or hate speech?
  - Our database is breached and publicly released?
  - Our facial recognition AI is used to identify and harass peaceful protestors?
  - Our child safety app is used to stalk women?
  - Our chatbot is sexist or racist?

Identify the human and social contexts in which our software will run

- Social Context
- Business Context
- Legal and Regulatory Context
- Others?

What is the social context?

- what categories of people will benefit from our software?
- what categories of people will be harmed by the use of our software?

What is the business context?

- Who is going to pay for this software?
  - users?
  - advertisers?
  - sponsors or sponsoring agencies?
- What are their incentives?
  - how do their incentives affect (or distort) our priorities in designing or developing this software?
    - if our sponsors are selling advertising, then we may be pressed to prioritize "engagement" (maybe undesirable)
    - if our sponsors want to help disadvantaged people to connect with services, we may be pressed to prioritize accessibility.

Who is selling what to who?

- "If you're not paying for the product, then you are the product"

--

Generally credited to Richard Serra and Carlota Fay Schoolman (1974, about TV advertising)

What is the legal and regulatory context?

- Americans with Disabilities Act (ADA)
- Litigation-averse sponsors may insist on elaborate Terms & Conditions
- Software for use in the EU may need to comply with the GPDR.
- What about financial software?
- What about personal data?
- What about leaving cookies, etc., on our machines?

Is the activity of the software transparent?

- What data does it collect about the individual user?
- Does it store things on our computers?
- Does it touch our files?
  - Does it violate the "CIA" of computer security?

Special considerations for AI

- Incentives for AI system may be particularly mysterious
- What objective function is your LLM optimizing?
- AI "agents" that operate in the real world present particularly complicated risks

A short course in AI safety

aisafety.dance

A short plug for Nicky Case

ncase.me

What can we do?

- There is **not** a single standard that everyone follows for developing ethical or equitable software.
- There have been a number of attempts:
  - Value Sensitive Design
  - Microsoft's Responsible AI Standard
  - Gender Inclusive Design
  - and many more

Value **S**ensitive **D**esign: an approach to consider human values in the design process.

Technology is the result of human imagination

All technology involves design

All design involves choices among possible options

All choices reflects values

Therefore, all technologies reflect and affect human values

Ignoring values in the design process is irresponsible

- Identify Direct and Indirect Stakeholders
- Identify Human Values
- Study Impact of Values on Stakeholders
- Resolve Value Tensions and Identify unintended consequences

https://vsd.ccs.neu.edu/introduction/what-is-vsd/

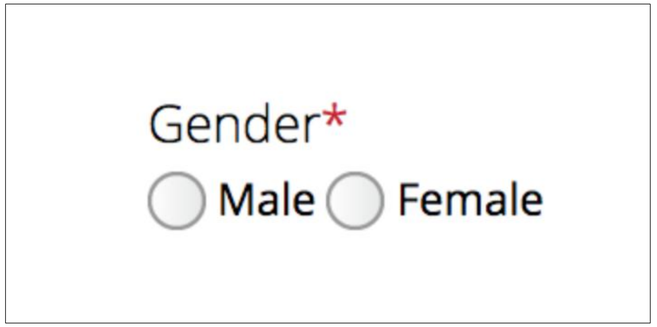
Microsoft’s **Trust Code** includes a standards that is based on 6 values

- Microsoft’s company-wide "Trust Code," covers specific policies like the: **Responsible AI Standard**, and codes of conduct for open-source communities
- The Responsible AI Standard focuses on ***fairness, privacy, reliability, inclusiveness, transparency, and accountability*** for AI systems.

https://www.microsoft.com/en-us/legal/compliance/sbc

Gender Inclusive Design

- Gender Stereotypes are fairly common in real world and many apps
- GenderMag is a structured approach for evaluating – and fixing – technology products and services
  - A set of cognitive styles for using technology that statistically cluster by gender. These are the **5 facets**.
  - A set of a fictional people (called **personas**)
  - **Evaluation walkthrough**: a prescribed series of steps to uncover gender-inclusion issues



https://gendermag.org/

GenderMag

Facets

- Information processing style
- Learning style for new technology
- Computer self-efficacy
- Attitude to risk
- Motivations

Personas

- Abi, Tim and Pat

Walkthrough

- Answer a series of questions for each persona
- Generate Report on usage differs

Subgoal report form

Use case (What is to be achieved overall):

(e.g., Abi wants to find a science fiction book.)

Subgoal

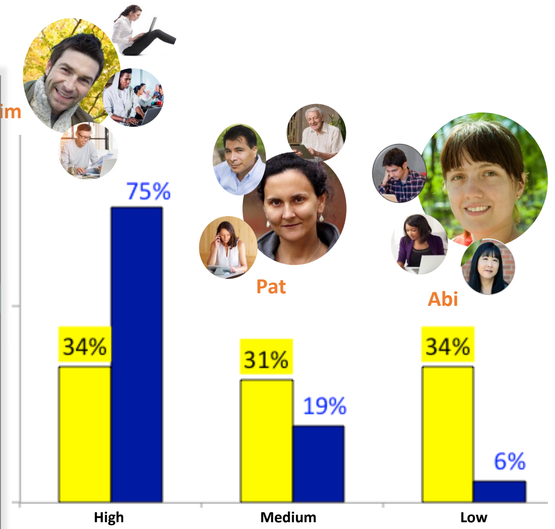
(e.g., See bookstore map.)

Will have thought of this as a step toward achieving the overall use case?

(fill in persona name)

☐ **Yes**	Facets Considered? ☐ Motivations ☐ Information Processing Style ☐ Computer Self-Efficacy ☐ Attitude Towards Risk ☐ Learning: by Process vs. by Tinkering ☐ None of the above	**Why?**
☐ **Maybe**	☐ Motivations ☐ Information Processing Style ☐ Computer Self-Efficacy ☐ Attitude Towards Risk ☐ Learning: by Process vs. by Tinkering ☐ None of the above	
☐ **No**	☐ Motivations ☐ Information Processing Style ☐ Computer Self-Efficacy ☐ Attitude Towards Risk ☐ Learning: by Process vs. by Tinkering ☐ None of the above	

Go to Action report form next.



Don’t ignore the challenges

- **Have a core set of Principles of Ethical SE/AI**
  - **Harvard**: Fairness, Transparency, Accountability, Privacy, Security
  - **AWS**: Fairness, Explainability, Privacy and security, Safety, Controllability, Veracity and robustness
  - **Google AI**: Design for safety, Be transparent, Secure AI systems
- **Follow guidelines like - *Balance technical and moral responsibility***
  - Be Proactive
  - Be honest
  - Be accountable
  - Be a responsible citizen

There are SE-level mitigations for some of these risks

- Form a diverse team
  - People from diverse backgrounds bring different experiences and different perspectives
- Consider human values throughout the project
- Rely on standards when possible
  - ADA
  - ARIA
  - etc.
- Monitor actual usage & misusage, user feedback
  - who? what? when? how?

Systemic mitigations

- Work for systemic change?
  - political, social, etc.
- Work to convince developers to consider human values?

Where does this leave us?

- **So that we can sleep at night**
  - Consider the different ways that our software may impact others
  - Consider the ways in which our software interacts with the political, social, and economic systems in which we and our users live
  - Follow best practices, and actively push to improve them
  - Encourage diversity in our development teams
  - Engage in honest conversations with our co-workers and supervisors to explore possible ethical issues and their implications.