AI Ethics
SpeakUp

Join by URL: https://web.speakup.info/room/join/80740

Join by key: web.speakup.info → use key 80740

Join by saved session: web.speakup.info

Join by QR:
Projects

• All projects need an ethics discussion
• Some project milestones require an *ethical sweep*

• Some projects will not have a *natural* ethical component
• In these cases, your project group can write about anything you’d like
Ethics and AI

How is ethics important to AI?

(one possible answer) It helps us answer the questions:

• What should we not build?

• What should we build?

• What do we need to considering during a build?
ACM Code of Ethics

• What did you find surprising?

• Are these discussed at your internships?

• Why is it useful?

• How can it be less useful?

• [https://www.acm.org/code-of-ethics](https://www.acm.org/code-of-ethics)
We’re still trying to figure this stuff out.

Browse Standards | Get Program | IEEE Xplore (from 2020 and 2021)

- IEEE Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children
- IEEE Standard Model Process for Addressing Ethical Concerns during System Design
- IEEE Standard for Transparency of Autonomous Systems
- IEEE Standard for Data Privacy Process
- IEEE Standard for Transparent Employer Data Governance
- IEEE Ontological Standard for Ethically Driven Robotics and Automation Systems
- IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being
IEEE Ethically Aligned Design

General Principles as Imperatives

1. **Human Rights**: A/IS shall be created and operated to respect, promote, and protect internationally recognized human rights.

2. **Well-being**: A/IS creators shall adopt increased human well-being as a primary success criterion for development.

3. **Data Agency**: A/IS creators shall empower individuals with the ability to access and securely share their data, to maintain people’s capacity to have control over their identity.

4. **Effectiveness**: A/IS creators and operators shall provide evidence of the effectiveness and fitness for purpose of A/IS.

5. **Transparency**: The basis of a particular A/IS decision should always be discoverable.

6. **Accountability**: A/IS shall be created and operated to provide an unambiguous rationale for all decisions made.

7. **Awareness of Misuse**: A/IS creators shall guard against all potential misuses and risks of A/IS in operation.

8. **Competence**: A/IS creators shall specify and operators shall adhere to the knowledge and skill required for safe and effective operation.

A/IS → autonomous and intelligent system
Awful AI

https://github.com/daviddao/awful-ai

• Look at the list and pick an article (about 10 minutes)

• Summarize in one to three sentences (you’ll only have time to skim read)

• Discuss with your neighbors (about 10 minutes)
Enforcement

Who is in charge of ethical enforcement?

• No precise answer, but there is no special ethics force out there.

• Everyone and nobody.

• Part of regular meetings.
On Accountability

• IBM developed systems to help Nazis track members of the Jewish community
• Volkswagen cheated on emissions testing
• A California database of **suspected gang members** included 42 babies

• Who is responsible?

• How would you feel as the developer?
• How would you feel if someone was injured?
• What can you do as the developer?
• Who is ultimately responsible?
When?

When should you bring up ethical considerations?

From the very beginning.
• easier to avoid pitfalls
• easier to analyze results
• prevent wasting of time
• the system will be ethical

• Ethics is hard.
A Few Key Issues

Just a few selected topics

1. AI feedback loops
2. Bias
3. Disinformation
AI Feedback Loops

• When a model has an impact on its own training data.

• For example, the YouTube recommendation system
  • How do you think performance (success) measured?
  • The algorithm pushes data that keeps people watching
  • What kind of videos keep people watching?

• (Probably) not directly intended, but the model interacts with the real world
• Our economy values profit over people
• The real problem: metrics are just a proxy for what we really care about
• What is the solution?
Software Engineering for Machine Learning: A Case Study

Saleema Amershi  
*Microsoft Research*  
Redmond, WA USA  
samershi@microsoft.com

Andrew Begel  
*Microsoft Research*  
Redmond, WA USA  
andrew.begel@microsoft.com

Christian Bird  
*Microsoft Research*  
Redmond, WA USA  
cbird@microsoft.com

Robert DeLine  
*Microsoft Research*  
Redmond, WA USA  
rdeline@microsoft.com

Harald Gall  
*University of Zurich*  
Zurich, Switzerland  
gall@ifi.uzh.ch

Ece Kamar  
*Microsoft Research*  
Redmond, WA USA  
eckamar@microsoft.com

Nachiappan Nagappan  
*Microsoft Research*  
Redmond, WA USA  
nachin@microsoft.com

Besmira Nushi  
*Microsoft Research*  
Redmond, WA USA  
besmira.nushi@microsoft.com

Thomas Zimmermann  
*Microsoft Research*  
Redmond, WA USA  
tzimmer@microsoft.com
All Data is Biased

Bias comes in many forms

• Historical
  • Cultural context is important
  • Data is not used thoughtfully (train a model to predict crime)
  • Not thinking about algorithmic colonialism (look at the origins of ImageNet)

• Measurement
  • Not measuring what you think you’re measuring
  • “Having a colonoscopy is a predictor of having a stroke”

• Aggregation
  • Inappropriately combining several factors

• Representation
  • Inadvertently amplifying imbalances
  • What is the easiest way to get a passing percentage when asked to predict the gender of a software developer?
Disinformation

https://www.youtube.com/watch?v=oxxBp9pSETo
Large Language Models (e.g., ChatGPT)

• What are your thoughts?
• Is a calculator a fair comparison?
• Can we know without decades of context?
• Ban?
• Change assignments?
• Do you agree with an embrace-and-disclose approach for students?

“ChatGPT is highly skilled at ‘pastiche’ and is essentially a glorified and very fast algorithm for cut/paste from many, MANY, online resources to create a stylistically believable 'autocomplete’.”

— Gary Marcus, an AI specialist, scientist and NYU professor emeritus
Strategies: OK, so what should you do?

• Questions to ask (more in the assignments)
  • Should we even be doing this?
  • What bias is in the data? (All data contains bias.)
  • Can the code and data be audited?
  • What are errors rates for different sub-groups?
  • What is the accuracy of a simple rule-based alternative?
  • What processes are in place to handle appeals/mistakes?
  • How diverse is the team?

• Practices you can implement ([Markkula Center for Applied Ethics](https://markkula.org/ethics)):  
  • Regular ethical risk sweeping
  • Expanding the ethical circle
  • Think about the terrible people
  • Closing the loop (feedback and iteration)

• Push our governments to enact healthy policies and regulations