

# AI Ethics

# Big Players in Deep Learning

DeepSeek (R1)	IBM (Watson)	Qualcomm
fast.ai (Edu/Library)	Perplexity	Mistral
US Gov.	Nvidia	Cydoni
OpenAI ( <del>ChatGPT</del> ) (Microsoft)	Oracle	AMD
Anthropic ( <del>Claude</del> )	Tesla*	Amazon
Google ( <del>Gemini</del> ; Gemini)	Boston Dynamics*	Apple
Meta ( <del>Llama</del> )	Xai	Baidu
	Intel	Alibaba
		Qwen

## Chinese AI startup DeepSeek



WSJ WSJ

### [How China's DeepSeek Outsmarted America](#)

6 hours ago

The New York Times

### [DeepSeek's Rise: How a Chinese Start-Up Went From Stock Trader to A.I....](#)



4 hours ago

Forbes

### [Who Is Behind DeepSeek? Here's What To Know About Founder Liang Wenfeng.](#)



3 hours ago

The Atlantic

### [China's DeepSeek Surprise](#)



22 hours ago

## Google News (2025-01-28)

Reuters

### [US tech shares recover some losses from steep DeepSeek selloff](#)



2 hours ago

BBC

### [Nvidia and Microsoft shares steady after DeepSeek AI app shock](#)



1 hour ago

CNN

### [DeepSeek chaos suggests 'America First' may not always win](#)



9 hours ago

AP News

### [What is DeepSeek, the Chinese AI company upending the stock market?](#)



20 hours ago

Yahoo Finance

### [Why market panic over China's DeepSeek is 'overblown,' analysts say](#)



22 hours ago

[View full coverage](#) →

WSJ WSJ

## [How China's DeepSeek Outsmarted America](#)

6 hours ago

 The New York Times

## [DeepSeek's Rise: How a Chinese Start-Up Went From Stock Trader to A.I....](#)

4 hours ago

 Forbes

## [Who Is Behind DeepSeek? Here's What To Know About Founder Liang Wenfeng.](#)



## [Nvidia and Microsoft shares steady after DeepSeek AI app shock](#)

1 hour ago

 CNN

## [DeepSeek chaos suggests 'America First' may not always win](#)

9 hours ago

 AP News

## [What is DeepSeek, the Chinese AI company upending the stock market?](#)





# Chinese AI startup DeepSeek


Google News (2025-01-28)



WSJ WSJ

## How China's DeepSeek Outsmarted America

6 hours ago

 The New York Times

DeepSeek's Rise: How a



 Reuters

## US tech shares recover some losses from steep DeepSeek selloff

2 hours ago




 BBC

## Nvidia and Microsoft shares steady after DeepSeek AI app shock

1 hour ago




 CNN

DeepSeek chaos suggests



6 hours ago

 The New York Times

## DeepSeek's Rise: How a Chinese Start-Up Went From Stock Trader to A.I....



4 hours ago

 Forbes

## Who Is Behind DeepSeek? Here's What To Know About Founder Liang Wenfeng.



3 hours ago

 The Atlantic

## China's DeepSeek Surprise



1 hour ago

 CNN

## DeepSeek chaos suggests 'America First' may not always win




9 hours ago

 AP News

## What is DeepSeek, the Chinese AI company upending the stock market?



20 hours ago

 Yahoo Finance

## Why market panic over





From Stock Trader to A.I....

4 hours ago

**F** Forbes

Who Is Behind DeepSeek?  
Here's What To Know About  
Founder Liang Wenfeng.



3 hours ago

**A** The Atlantic

China's DeepSeek Surprise



22 hours ago

always win

9 hours ago

**AP** AP News

What is DeepSeek, the  
Chinese AI company  
upending the stock market?



20 hours ago

**y!** Yahoo Finance

Why market panic over  
China's DeepSeek is  
'overblown,' analysts say



22 hours ago

[View full coverage](#) →

From Stock Trader to A.I....

4 hours ago

**F** Forbes

Who Is Behind DeepSeek?  
Here's What To Know About  
Founder Liang Wenfeng.



3 hours ago

**A** The Atlantic

China's DeepSeek Surprise



22 hours ago

always win

9 hours ago

**AP** AP News

What is DeepSeek, the  
Chinese AI company  
upending the stock market?



20 hours ago

**y!** Yahoo Finance

Why market panic over  
China's DeepSeek is  
'overblown,' analysts say



22 hours ago

[View full coverage](#) →



# Chinese AI startup DeepSeek



WSJ WSJ

[How China's DeepSeek Outsmarted America](#)

6 hours ago

The New York Times

[DeepSeek's Rise: How a Chinese Start-Up Went From Stock Trader to A.I....](#)

4 hours ago

Forbes

[Who Is Behind DeepSeek? Here's What To Know About Founder Liang Wenfeng.](#)

3 hours ago

The Atlantic

[China's DeepSeek Surprise](#)

22 hours ago

## Google News (2025-01-28)

Reuters

[US tech shares recover some losses from steep DeepSeek selloff](#)

2 hours ago

BBC

[Nvidia and Microsoft steady after DeepSeek app shock](#)

1 hour ago

CNN

[DeepSeek chaos suggests 'America First' may not always win](#)

9 hours ago

AP News

[What is DeepSeek, the Chinese AI company upending the stock market?](#)

20 hours ago

Yahoo Finance

[Why market panic over China's DeepSeek is 'overblown,' analysts say](#)

22 hours ago

**DeepSeek** (Chinese: 深度求索; pinyin: *Shēndù Qiúsuǒ*) is a Chinese artificial intelligence company that develops open-source large language models (LLM). Based in Hangzhou, Zhejiang, it is owned and solely funded by Chinese hedge fund High-Flyer, whose co-founder, Liang Wenfeng, established the company in 2023 and serves as its CEO.

Wikipedia (2025-01-28)

DeepSeek's AI assistant became the No. 1 downloaded free app on Apple's iPhone store Monday, propelled by curiosity about the ChatGPT competitor. Part of what's worrying some U.S. tech industry observers is the idea that the Chinese startup has caught up with the American companies at the forefront of generative AI at a fraction of the cost.

AP News (2025-01-28)

[View full coverage](#) →

# Outline

- How ethics relate to this class
- Computing societies codes of ethics
- How things can go wrong
- Accountability and enforcement
- Strategies (what should you do?)

# 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?

Who are the authors; who is the audience

- Are these discussed at your internships?

- Why is it useful?

- How can it be less useful?

- <https://www.acm.org/code-of-ethics>

# Ethics Practices

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.

# Hugging Face (Private, For-Profit)

## Ethics & Society at Hugging Face



- Rigorous: asking “does it work?”
- Consentful: support the self-determination of those affected
- Socially Conscious: support a stronger society
- Sustainable: ecologically sustainable
- Inclusive: broaden who build and benefits
- Inquisitive: rethink relationship with technology

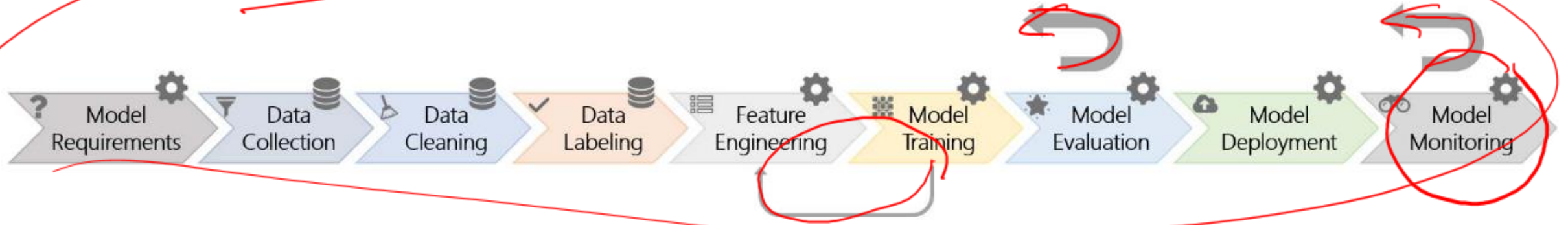
# Awful AI

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

- Look at the list and pick an article (about ~~10~~ minutes) S 2:21
- Summarize in one to three sentences (you'll only have time to skim read)
- Discuss with your neighbors (about ~~10~~ minutes) S 2:26



# Research at Microsoft



## 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

# Enforcement

Who oversees 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 infants
- 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 consider 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

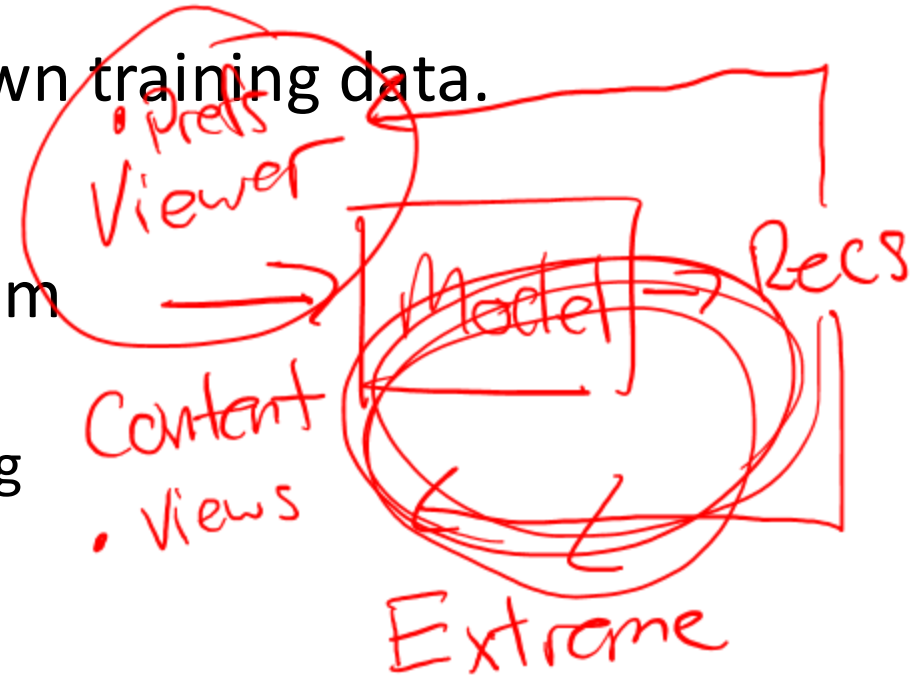
1. AI feedback loops
2. Bias
3. Disinformation

# AI Feedback Loops Monitor

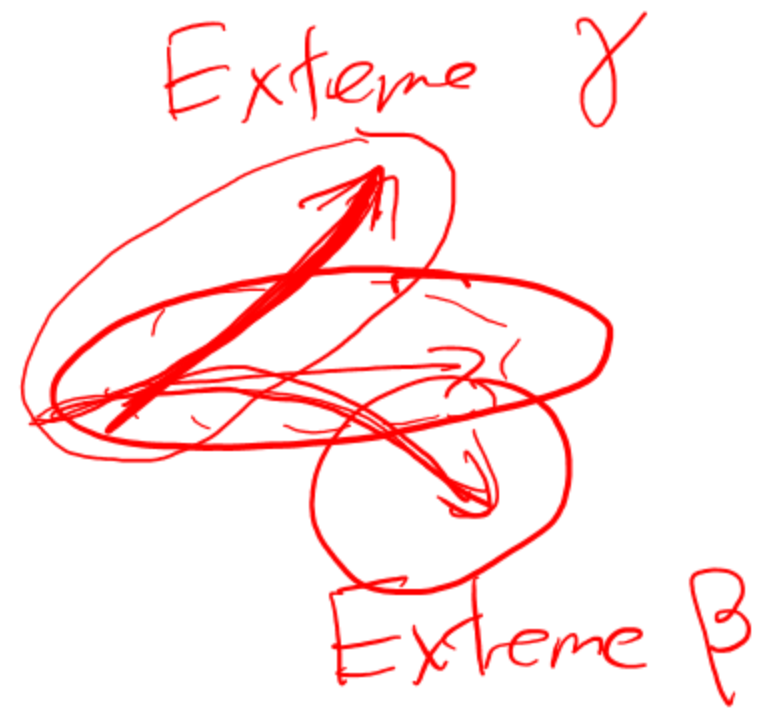
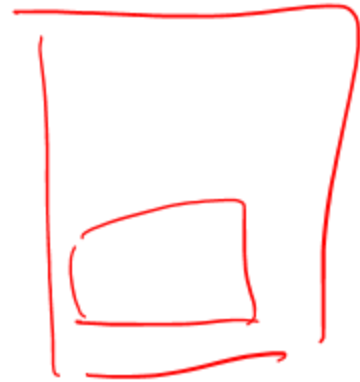
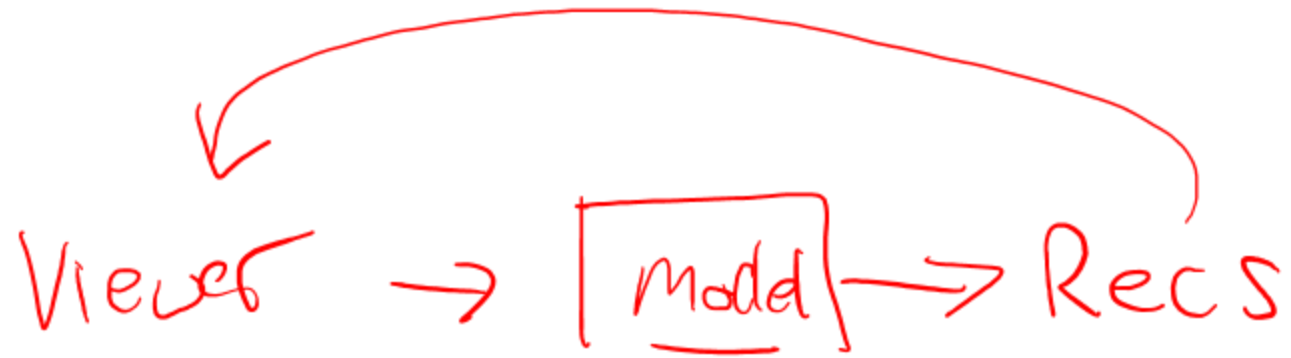
When a **model** has an impact on its own training data.

- 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



# AI Feedback Loops

Metrics lead to (the more you rely on it)

- manipulation
- gaming
- myopic focus on short-term
- unexpected consequence (not looking in the right place)

Cast study: **reduce ER wait times**

- cancel operations
- patients wait in ambulances
- turn stretchers into beds

Case study: **grading essays**

- cannot evaluate creativity/novelty
- gibberish with sophisticated words
- students with regional vernaculars receive lower grades
- Chinese students receive higher scores



# 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 (group individuals that shouldn't be grouped)

## Representation

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

# What would you label this?



Dairy cow, Britannica

# Disinformation



<https://www.youtube.com/watch?v=oxXpB9pSETo>

# Large Language Models (LLMs, 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?

- Ask questions (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?

65%

63%





# Strategies: OK, so what should you do?

- Ask questions (more in the assignments)
- Implement ethics processes ([Markkula Center for Applied Ethics](#)):
  - Regular ethical risk sweeping
  - Expanding the ethical circle
  - Think about the terrible people
  - Closing the loop (feedback and iteration)

# Strategies: OK, so what should you do?

- Ask questions (more in the assignments)
- Implement ethics processes ([Markkula Center for Applied Ethics](#)):
- Push our governments to enact healthy policies and regulations