

Course Introduction

Manoel Horta Ribeiro
manoel@cs.princeton.edu

COS 424 / Fall 2026



MENU

APPETIZER (10 minutes)
Manoel's intro

1st COURSE (15 minutes)
Why this course?

2nd COURSE (15 minutes)
What are we going to learn?

DESSERT (5 minutes)
Course Logistics



MENU

APPETIZER (10 minutes) **Manoel's Intro**

1st COURSE (15 minutes)
Why this course?

2nd COURSE (15 minutes)
What are we going to learn?

DESSERT (5 minutes)
Course Logistics



BSc and MSc @ UFMG, Brazil 🇧🇷

PhD @ EPFL, Switzerland, 🇨🇭

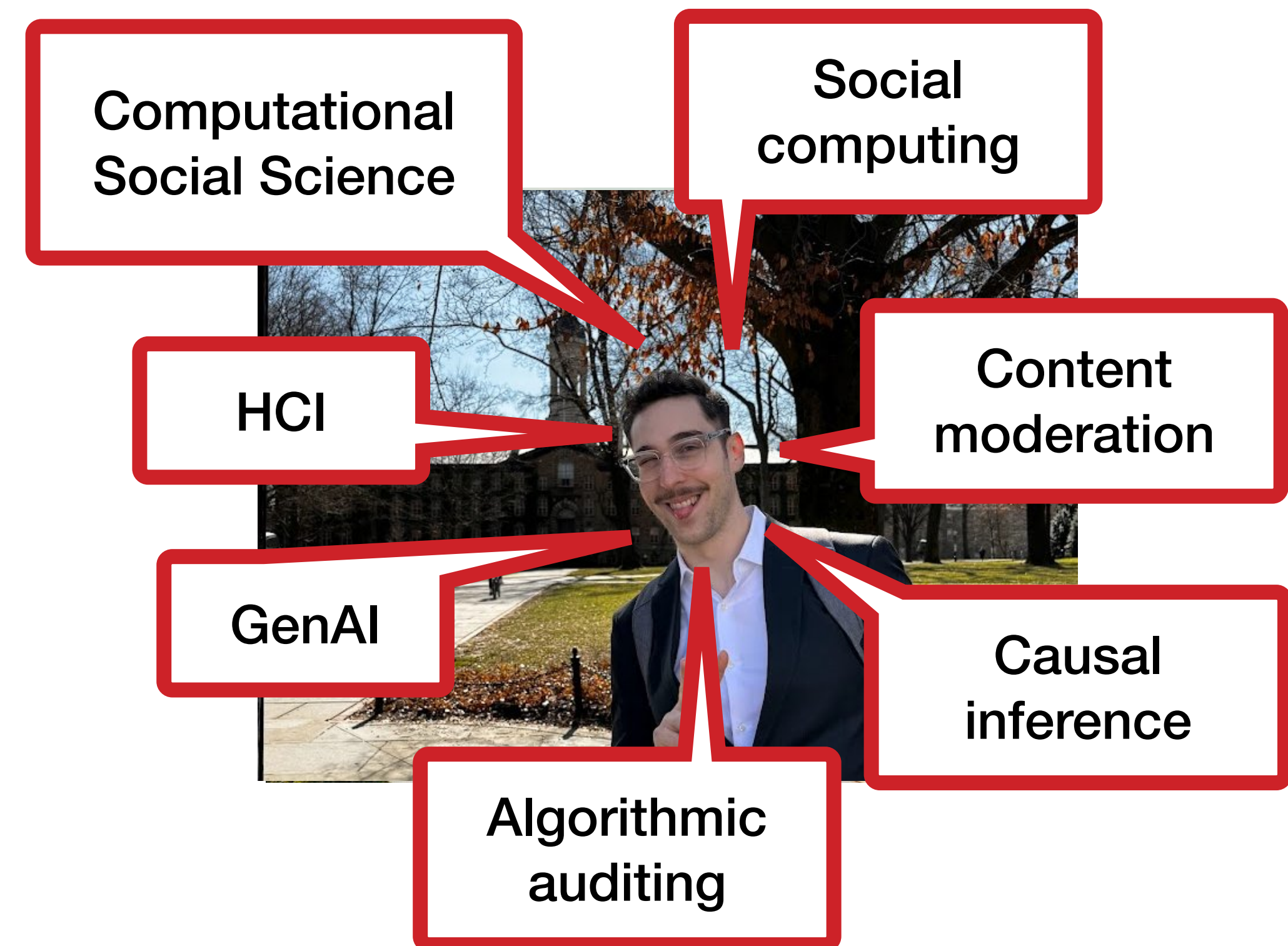
Since 2025 @ Princeton, 🇺🇸



 Meta



Microsoft Research



Course Staff

Undergrad Course Assistants



Xiaoyan Li

Senior Lecturer
Department of Computer Science
Princeton University
xiaoyan@cs.princeton.edu



cocogong@princeton.edu



allenliu@princeton.edu



MENU

APPETIZER (5 minutes)
Manoel's Intro

1st COURSE (15 minutes)
Why this course?

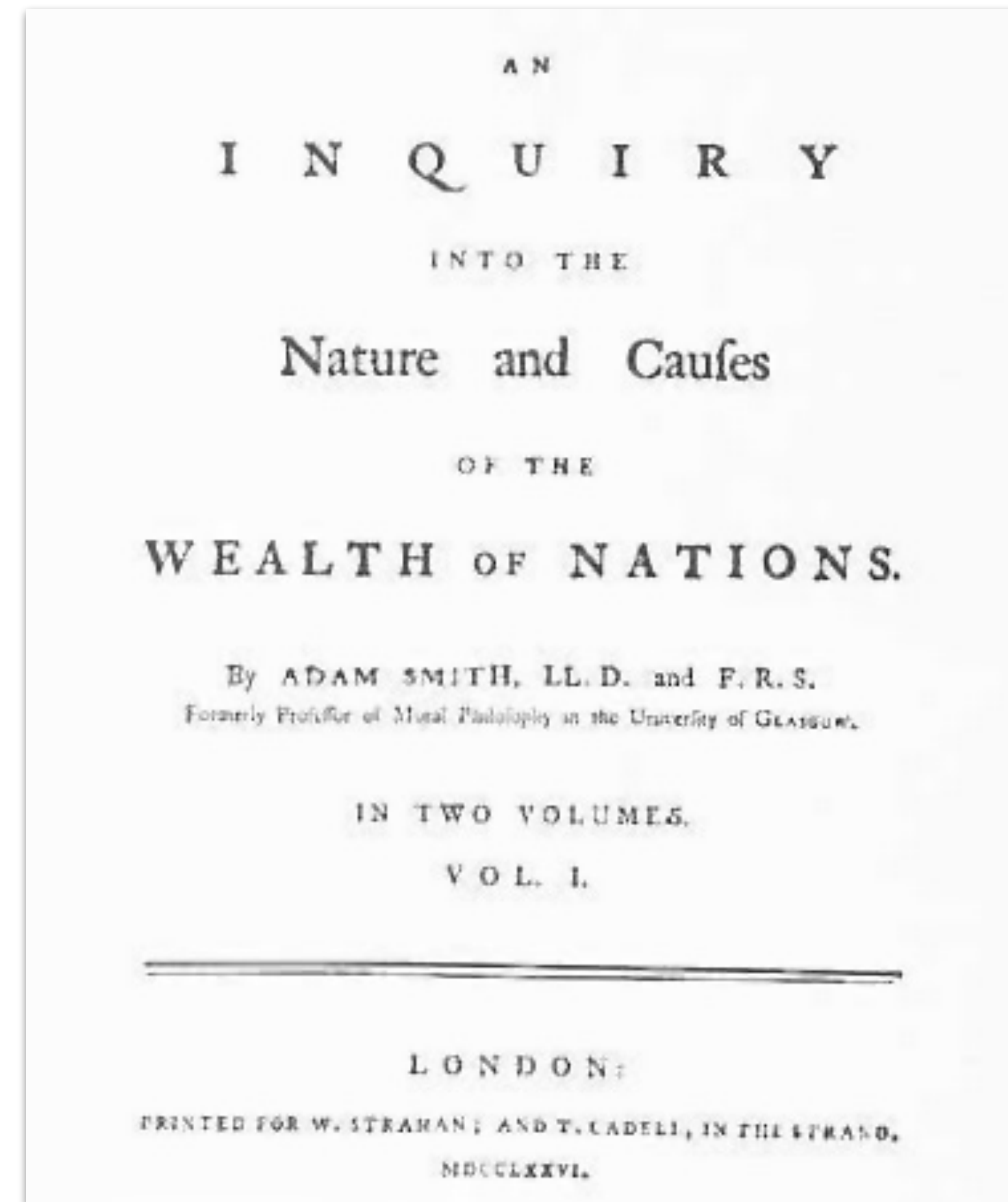
2nd COURSE (15 minutes)
What are we going to learn?

DESSERT (5 minutes)
Course Logistics

Founding Myths

Scientific disciplines have
“founding myths” that shape:

- How they see themselves;
- What they value;
- What they believe their role ought to be in society;



Computer Science's Myths

CS's founding myths:

- theory;
- system building;

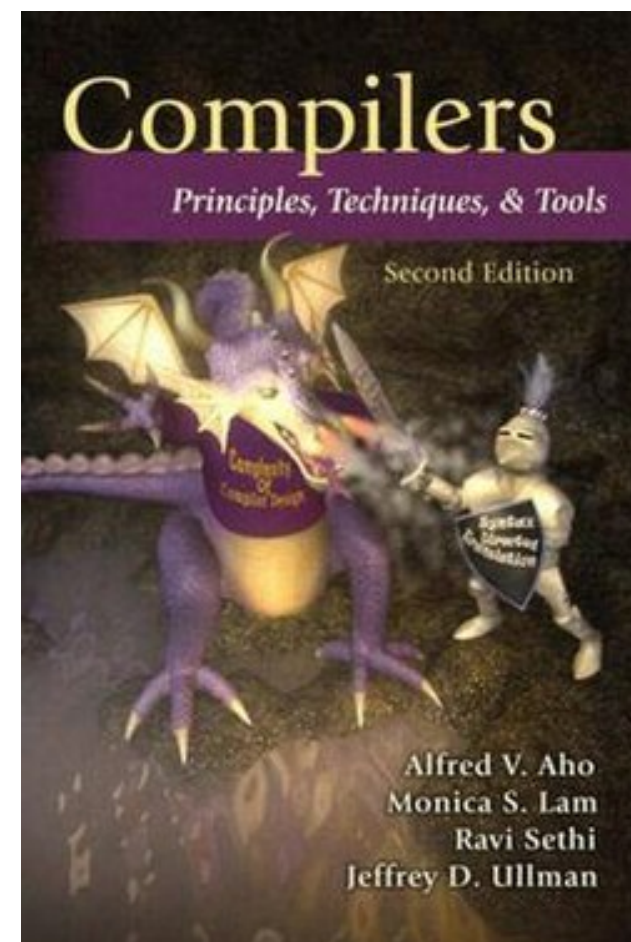
COS 583: Great Moments in Computing

Foundations of Digital Logic
[Boole, 1854]
[Shannon, 1938]

Operating Systems
[Ritchie & Thompson, 1974]
[Engler et al. 1995]

Crypto and Encryption
[Diffie & Hellman, 1976]
[Rivest et al., 1978]

Compilers
[Hopper, 1952]
[Backus et al. 1957]



Yet, Modern CS is Empirical

Modern CS is driven by observation and experiment!



Technical Staff
OpenYay



SWE
FriendBook



PhD Student
Printstone

Yet, Modern CS is Empirical

Modern CS is driven by observation and experiment!

Characterizing and Detecting Propaganda-Spreading Accounts on Telegram

Klim Kireev^{§†}, Yevhen Mykhno, Carmela Troncoso^{§†}, Rebekah Overdorf^{‡*}

[§] EPFL, [†] MPI-SP Max Plank Institute for Security and Privacy

[‡] Ruhr University Bochum (RUB), Research Center Trustworthy Data Science and Security in University Alliance Ruhr

^{*} University of Lausanne

USENIX 2025 (Security & Privacy)

“I’ve talked to ChatGPT about my issues last night.”: Examining Mental Health Conversations with Large Language Models through Reddit Analysis

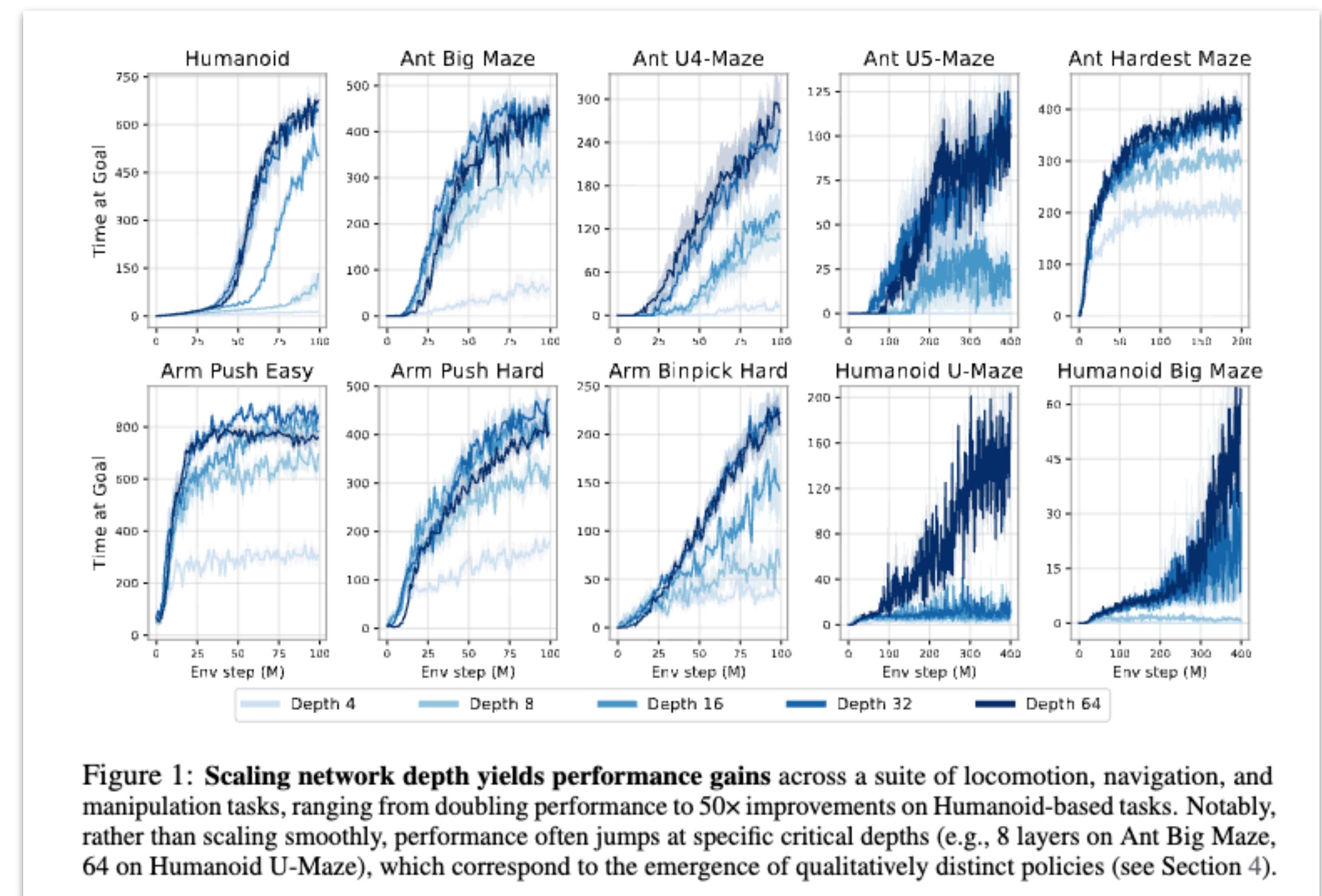
KYUHA JUNG, University of California, Irvine, USA

GYUHO LEE, Seoul National University, Republic of Korea

YUANHUI HUANG, University of California, Irvine, USA

YUNAN CHEN, University of California, Irvine, USA

CSCW 2025 (Human Comp. Inter.)



Wang et al.; NeurIPS 2025
(Machine Learning)

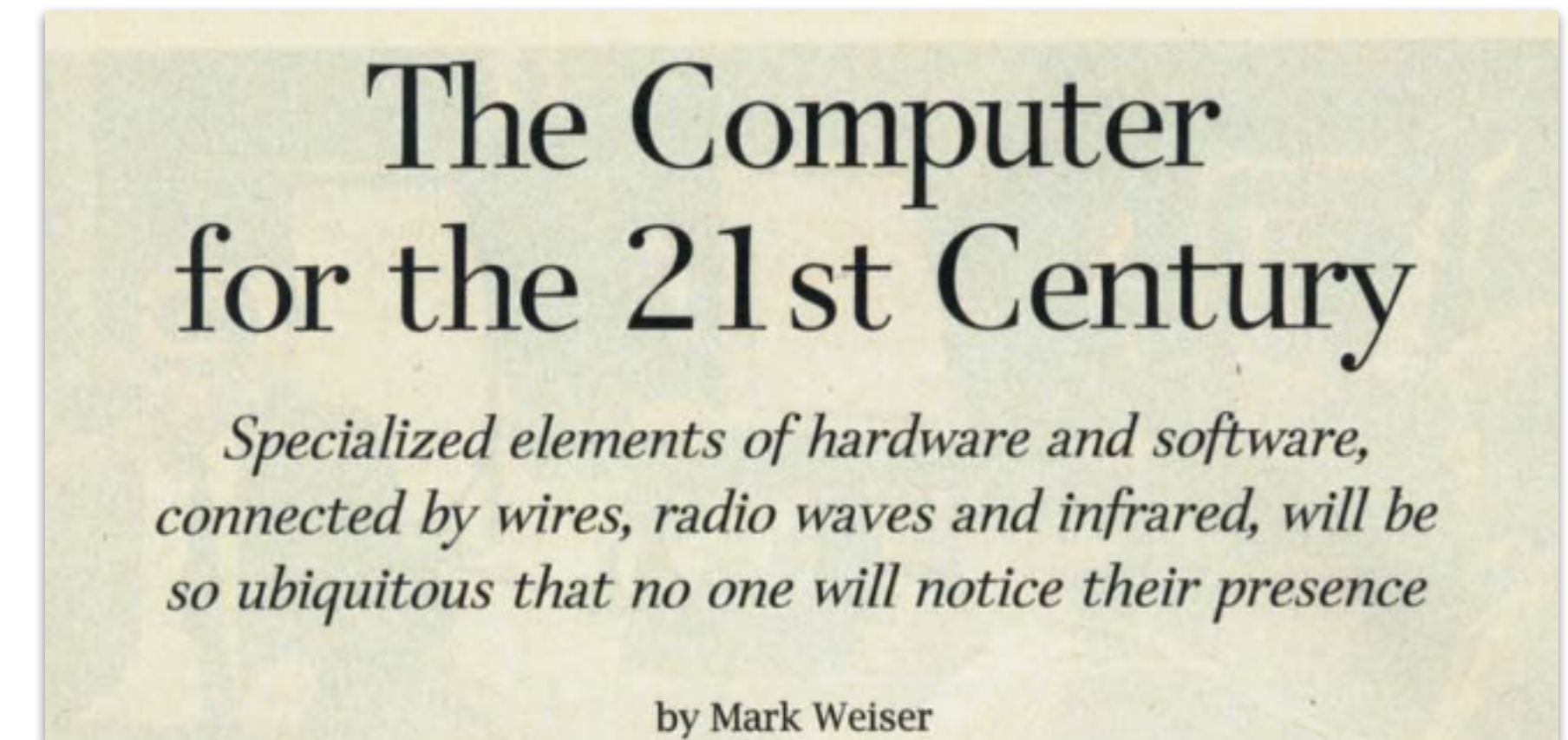
Why did this happen?

**Reason #1: Socio-technical
integration of computing**

“Scope Creep”

Computer Science Broadened its scope. Why?

- Computing became ubiquitous.
- Problems started involving humans using computers.



(Link)

“Only when things disappear in this way are we freed to use them without thinking and so to focus on beyond them on new goals”

Example: Security and Privacy

Before: Privacy is about cryptography;
it is a “technical property.”

After: Privacy is also about people:
social norms, expectations, context.

Copyright © 2004 by Washington Law Review Association

PRIVACY AS CONTEXTUAL INTEGRITY

Helen Nissenbaum*

Abstract: The practices of public surveillance, which include the monitoring of individuals in public through a variety of media (e.g., video, data, online), are among the least understood and controversial challenges to privacy in an age of information technologies. The fragmentary nature of privacy policy in the United States reflects not only the oppositional pulls of diverse vested interests, but also the ambivalence of unsettled intuitions on mundane phenomena such as shopper cards, closed-circuit television, and biometrics. This Article, which extends earlier work on the problem of privacy in public, explains why some of the prominent theoretical approaches to privacy, which were developed over time to meet traditional privacy challenges, yield unsatisfactory conclusions in the case of public surveillance. It posits a new construct, “contextual integrity,” as an alternative benchmark for privacy, to capture the nature of challenges posed by information technologies. Contextual integrity ties adequate protection for privacy to norms of specific contexts, demanding that information gathering and dissemination be appropriate to that context and obey the governing norms of distribution within it. Building on the idea of “spheres of justice,” developed by political philosopher Michael Walzer, this Article argues that public surveillance violates a right to privacy because it violates contextual integrity; as such, it constitutes injustice and even tyranny.

I. INTRODUCTION

Privacy is one of the most enduring social issues associated with information technologies. It has been a fixture in public discourse through radical transformations of technology from stand-alone computers, housing massive databases of government and other large institutions, to the current distributed network of computers with linked information systems, such as the World Wide Web, networked mobile devices, video and radio-frequency surveillance systems, and computer-enabled biometric identification. Among many privacy controversies that have stirred public concern, a particular set of cases, to which I have applied the label “public surveillance,” remains vexing not only because these cases drive opponents into seemingly irreconcilable stances, but because traditional theoretical insights fail to clarify the sources of their controversial nature.¹ This Article seeks to shed light on the problem of

* Associate Professor, Department of Culture & Communication, New York University, East Building 7th Floor, 239 Greene Street, New York, New York 10003. E-mail address: helen.nissenbaum@nyu.edu.

Many people and institutions have inspired and helped me in this endeavor, beginning with the Institute for Advanced Study, School of Social Sciences, where I wrote and presented early drafts.

(Link)

But also...

**Human Computer
Interaction**

studying interaction made user studies central.

Systems Research

the live Internet made measurement essential.

**Computational Social
Science**

digital traces blurred CS and social science.

Reason #2: The Triumph of Empiricism in Machine Learning

How is Knowledge Accumulated?

Before: through theory

Vapnik (1999): Use SVMs because they are based on the developed theory.

Intuition ➡ Theory ➡ Model ➡ Experiment

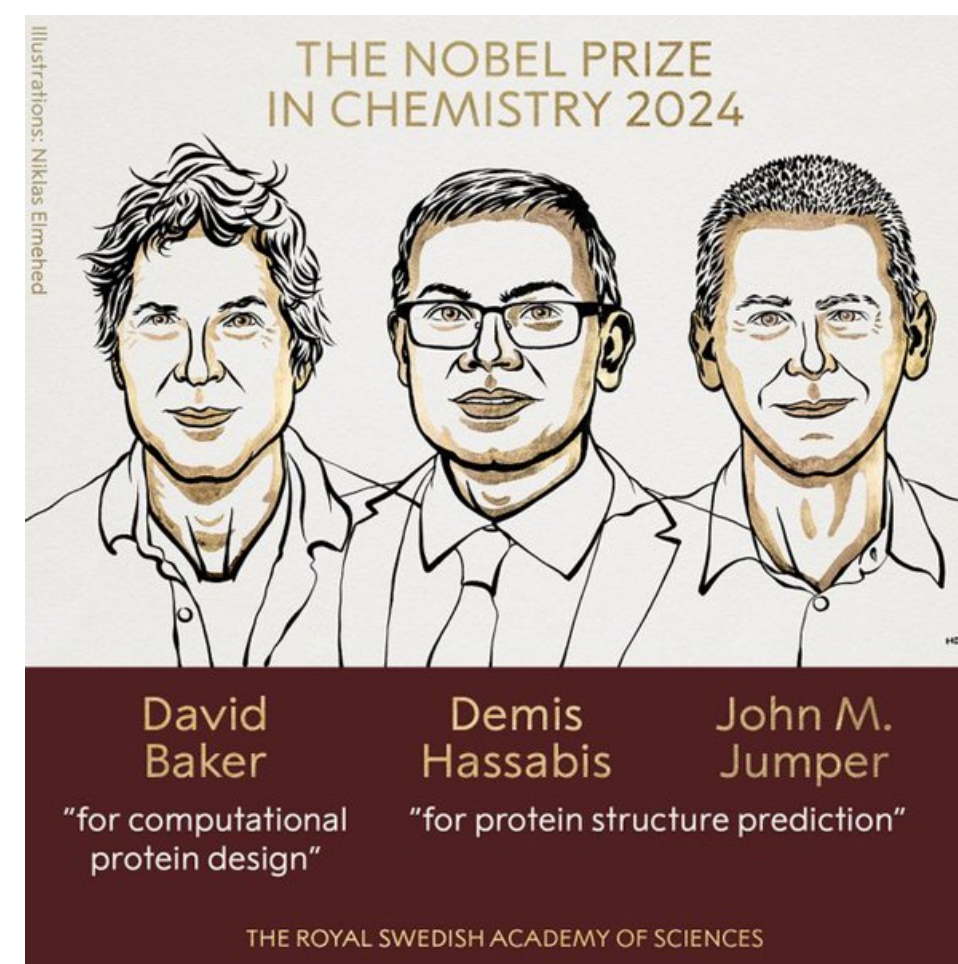
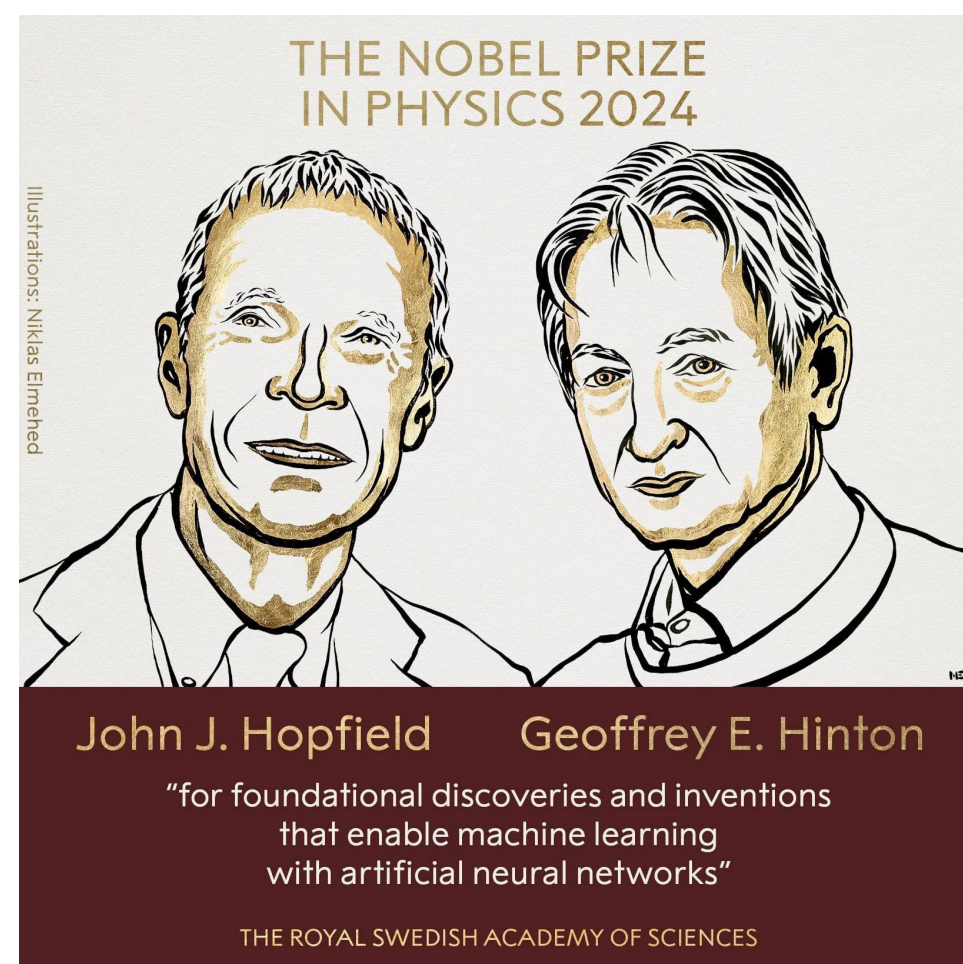
After: through experimentation

Srivastava (2014): “A motivation for dropout comes from a theory of the role of sex in evolution (...). A closely related motivation for dropout comes from thinking about successful conspiracies.”

Intuition ➡ Model ➡ Experiment ➡ Theory

Machine Learning is Eating CS Science

- Already true in the Deep Learning era — two Nobel prizes in 2024!
- Generative AI may become a “method layer” that percolates science, and even mathematics (see Terrence Tao’s account of Erdős #1026)



The story of Erdős problem #1026

8 December, 2025 in [expository](#), [math.CA](#), [math.CO](#) | Tags: [AI](#), [AlphaEvolve](#), [Erdos](#) | by [Terence Tao](#)

[Problem 1026 on the Erdős problem web site](#) recently got solved through an interesting combination of existing literature, online collaboration, and AI tools. The purpose of this blog post is to try to tell the story of this collaboration, and also to supply a complete proof.

The original problem of Erdős, [posed in 1975](#), is rather ambiguous. Erdős starts by [recalling his famous theorem with Szekeres](#) that says that given a sequence of $k^2 + 1$ distinct real numbers, one can find a subsequence of length $k + 1$ which is either increasing or decreasing; and that one cannot improve the $k^2 + 1$ to k^2 , by considering for instance a sequence of k blocks of length k , with the numbers in each block decreasing, but the blocks themselves increasing. He also noted a [result of Hanani](#) that every sequence of length $k(k + 3)/2$ can be decomposed into the union of k monotone sequences. He then wrote “As far as I know the following question is not yet settled. Let $x_1, \dots, x_n$ be a sequence of distinct numbers, determine

$$S(x_1, \dots, x_n) = \max_r \sum x_{i_r}$$



MENU

APPETIZER (5 minutes)
Manoel's Intro

1st COURSE (20 minutes)
Why this course?

2nd COURSE (15 minutes)
What are we going to learn?

DESSERT (5 minutes)
Course Logistics

So what is this course?

- This course is about *breadth*, not depth.
- What methods and tools are out there?
- How can we use them in a principled fashion?



Syllabus

Basics

**Describing &
Visualizing Data**

Network & Text Data

Applied ML

Causal Inference

Causal Inference

Regression Analysis

Experiments

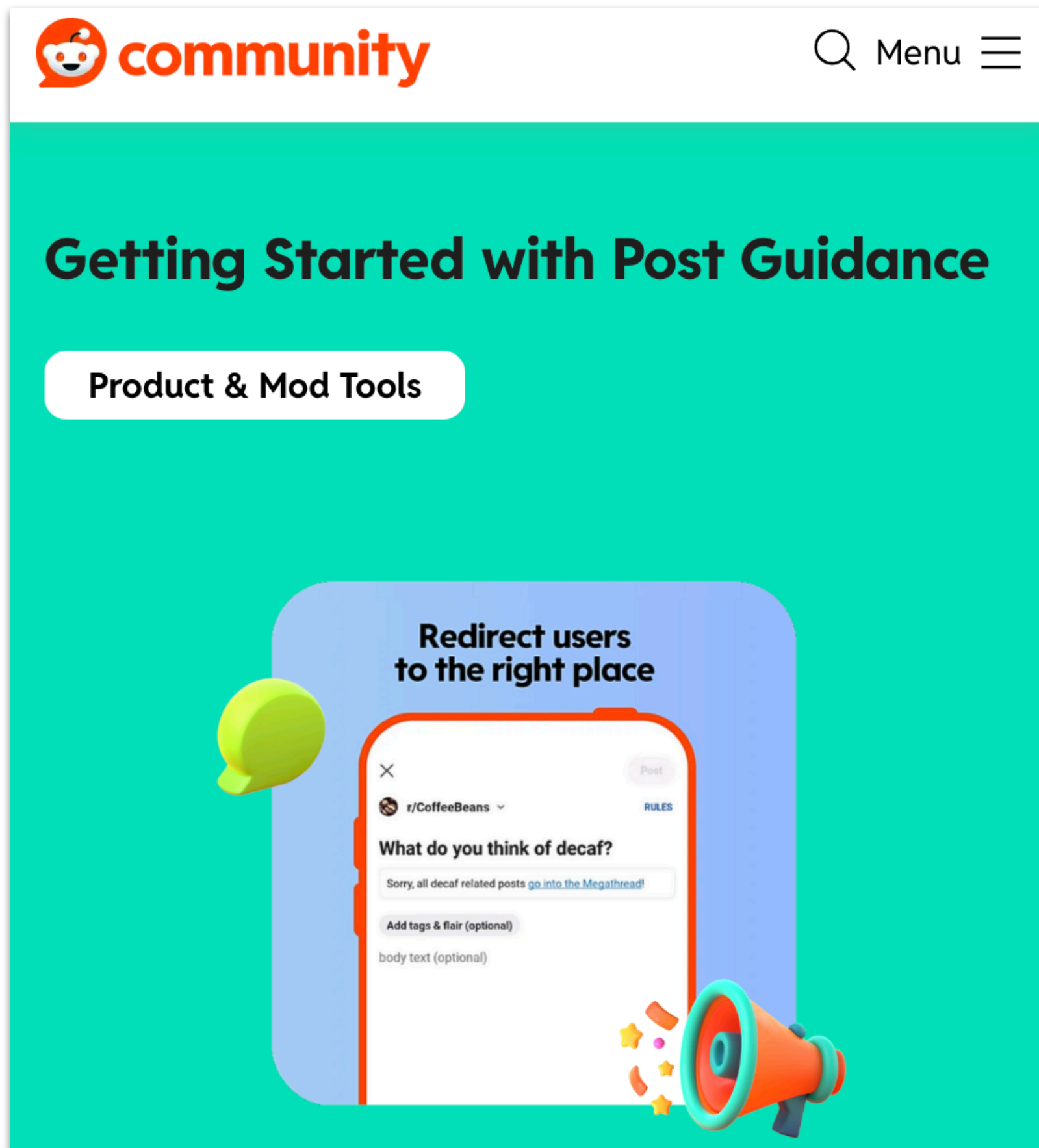
Advanced Topics

Quasi-experiments

LLMs in practice

Causal ML

Example: Post Guidance



(Link)

Post Guidance for Online Communities

MANOEL HORTA RIBEIRO, EPFL, Lausanne, Switzerland

ROBERT WEST, EPFL, Lausanne, Switzerland

RYAN LEWIS, Reddit, San Francisco, US

SANJAY KAIRAM*, Reddit, San Francisco, US

Effective content moderation in online communities is often a delicate balance between maintaining content quality and fostering user participation. In this paper, we introduce *post guidance*, a novel approach to community moderation that proactively guides users' contributions using rules that trigger interventions as users draft a post to be submitted. For instance, rules can surface messages to users, prevent post submissions, or flag posted content for review. This uniquely *community-specific*, *proactive*, and *user-centric* approach can increase adherence to rules without imposing additional burdens on moderators. We evaluate a version of Post Guidance implemented on Reddit, which enables the creation of rules based on both post content and account characteristics. Specifically, we conduct a large randomized experiment, capturing activity from 97,616 posters in 33 subreddits over 63 days. We find that Post Guidance (1) increased the number of "successful posts" (posts not removed after 72 hours), (2) decreased moderators' workload in terms of manually-reviewed reports, (3) increased contribution quality, as measured by community engagement, and (4) had no impact on posters' own subsequent activity, within communities adopting the feature. Post Guidance on Reddit was similarly effective for community veterans and newcomers, with greater benefits in communities that used the feature more extensively. Our findings indicate that *post guidance* represents a transformative approach to content moderation, embodying a paradigm that can be easily adapted to other platforms to improve online communities across the Web.

CCS Concepts: • **Human-centered computing** → **Social networking sites**; **Empirical studies in collaborative and social computing**; **Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: content moderation, online experiments, social media

(Link)



“Digital Kingdom of Fiefs”



General rules



Specific rules



Sanjay Kairam
u/PeerRevue

Roam around
communities



Reddit Rules

Rules

Rule 1

Remember the human. Reddit is a place for creating community and belonging, not for attacking marginalized or vulnerable groups of people. Everyone has a right to use Reddit free of **harassment, bullying**, and threats of **violence**. Communities and users that incite violence or that promote hate based on **identity or vulnerability** will be banned.

Rule 2

Abide by community rules. Post authentic content into communities where you have a personal interest, and do not cheat or engage in content manipulation (including **spamming, vote manipulation**, ban evasion, or subscriber fraud) or otherwise interfere with or disrupt Reddit communities.

Rule 3

Respect the privacy of others. Instigating harassment, for example by revealing someone's **personal or confidential information**, is not allowed. Never post or threaten to post **intimate or sexually-explicit media** of someone without their consent.

Rule 5

You don't have to use your real name to use Reddit, but don't **impersonate** an individual or an entity in a misleading or deceptive manner.

Rule 6

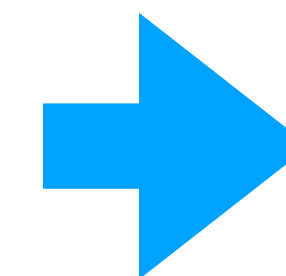
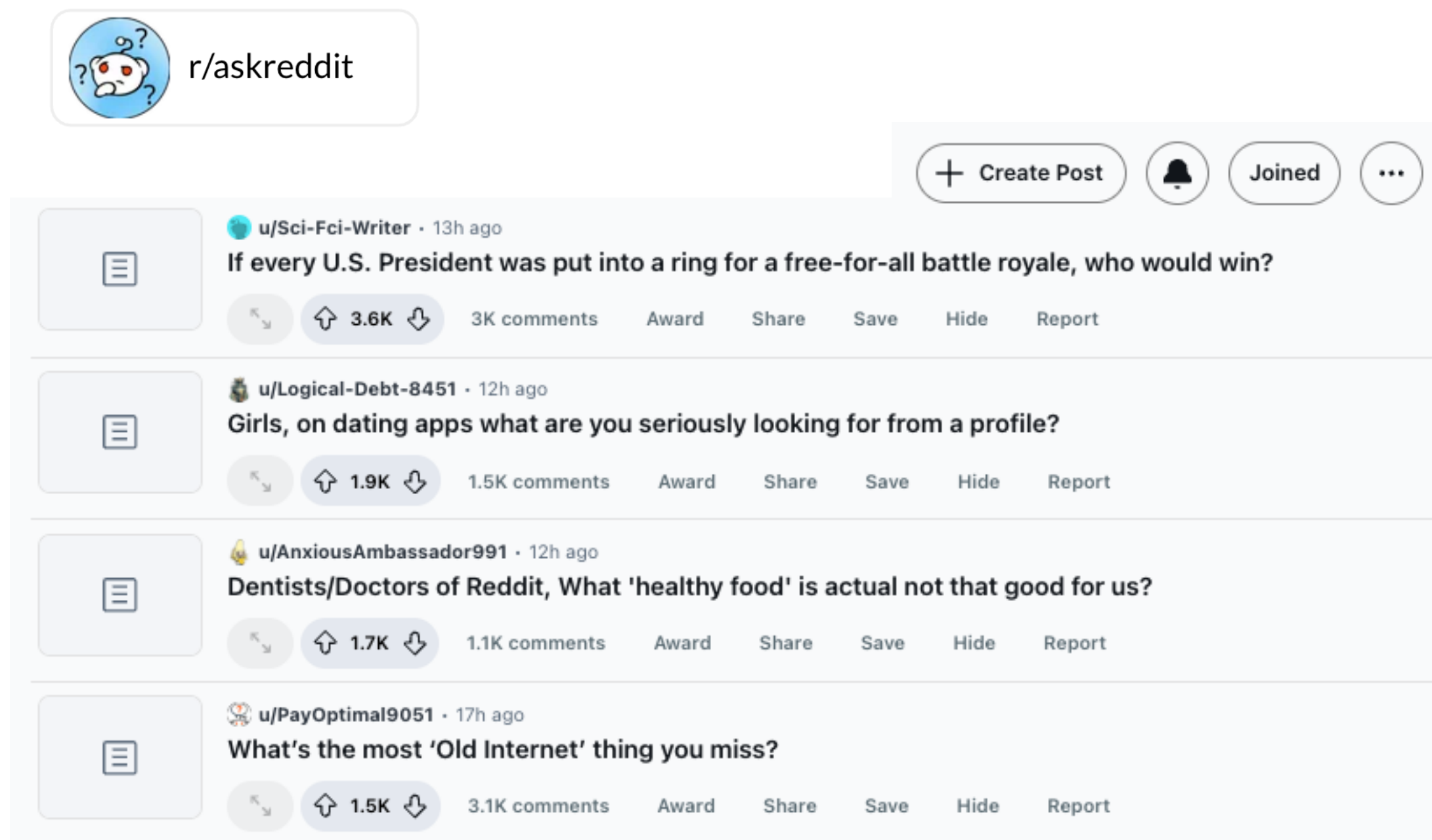
Ensure people have predictable experiences on Reddit by properly labeling content and communities, particularly content that is graphic, sexually-explicit, or offensive.

Rule 7

Keep it legal, and avoid posting illegal content or soliciting or facilitating **illegal or prohibited transactions**.

Rule 8

Don't **break the site** or do anything that interferes with normal use of Reddit.



RULES		
1	Rule 1 - Questions must be clear and direct and may not use the body textbox	▼
2	Rule 2 - No personal or professional advice requests	▼
3	Rule 3 - Open ended questions only	▼
4	Rule 4 - No personal info	▼
5	Rule 5 - No loaded questions	▼
6	Rule 6 - [Serious] tagged posts are off-limits to jokes or irrelevant replies	▼
7	Rule 7 - No begging for goods or services, or Reddit rewards	▼
8	Rule 8 - Be respectful to other users at all times and conduct your behaviour in a civil manner	▼
9	Rule 9 - No image only replies or large ascii art	▼
10	Rule 10 - Harmful Misinformation	▼
11	Rule 11 - No spam, machine generated content or karma farming	▼

Manual Review +



Enter Post Guidance

Create a post

DRAFTS 0

r/CoffeeBeans

Remember not to attack others if they have different tastes.

Post

Images & Video

Link

Poll

Title

0/300

Markdown ⓘ

Switch to Fancy Pants Editor

Text (optional)

×

Post

r/CoffeeBeans

RULES

Title

Add tags & flair (optional)

body text (optional)

1 2 3 4 5 6 7 8 9 0

Typical Reddit Experience...

If you were analyzing Reddit logs like I was... You'd find something like...

User posts something ...

... only to get their post deleted instantly...

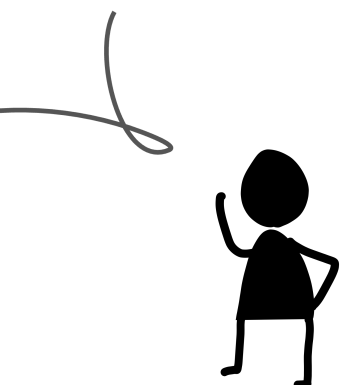
... for an absolutely avoidable reason.

User gets mad...

and sends a bunch of messages to moderators...

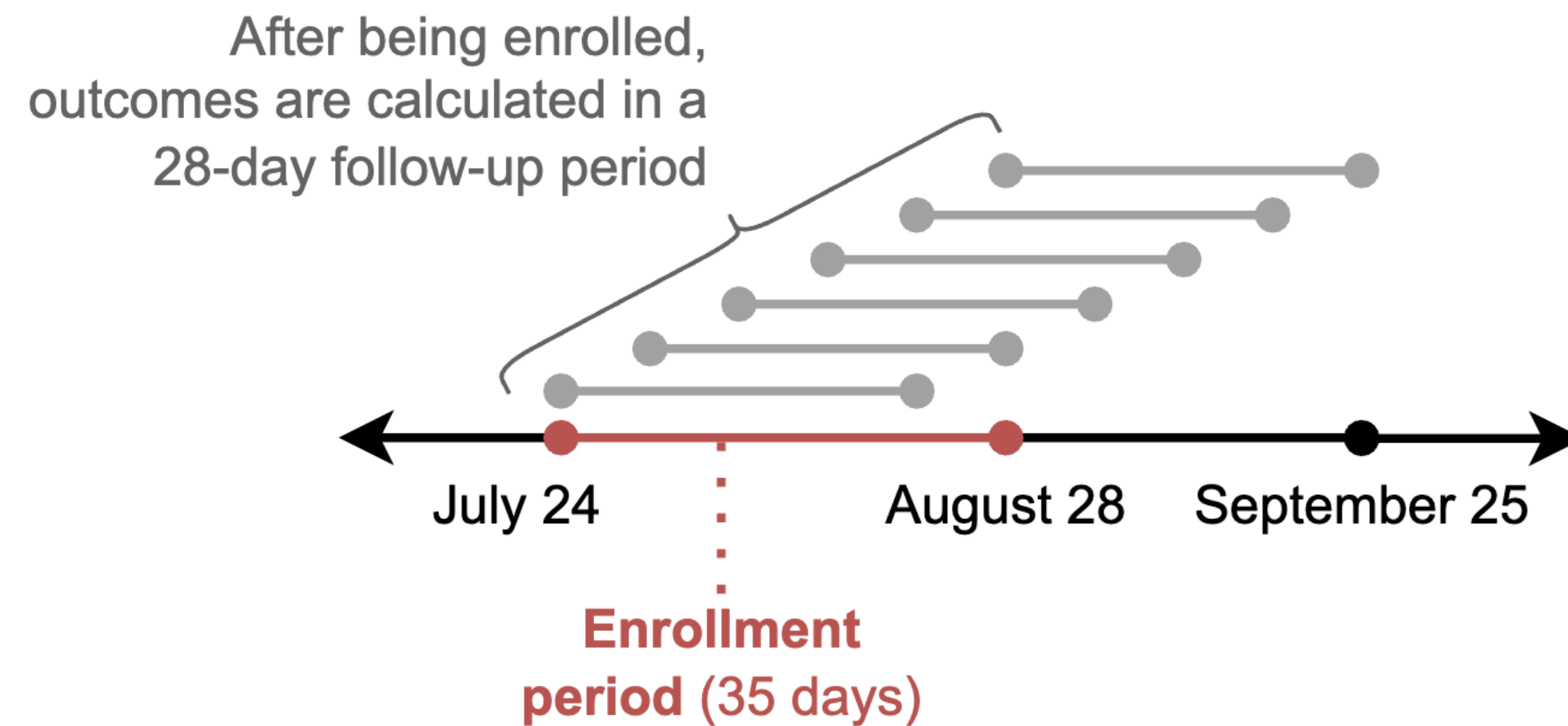
**Describing &
Visualizing Data**

Network & Text Data



But does it work?

Testing Post Guidance! An RCT



What sample size do we need?

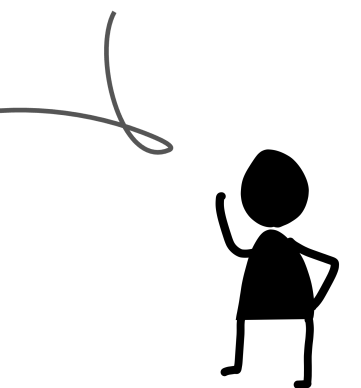
What is the best unit of analysis here?

What are threats to the validity of the experiment?

Causal Inference

Regression Analysis

Experiments

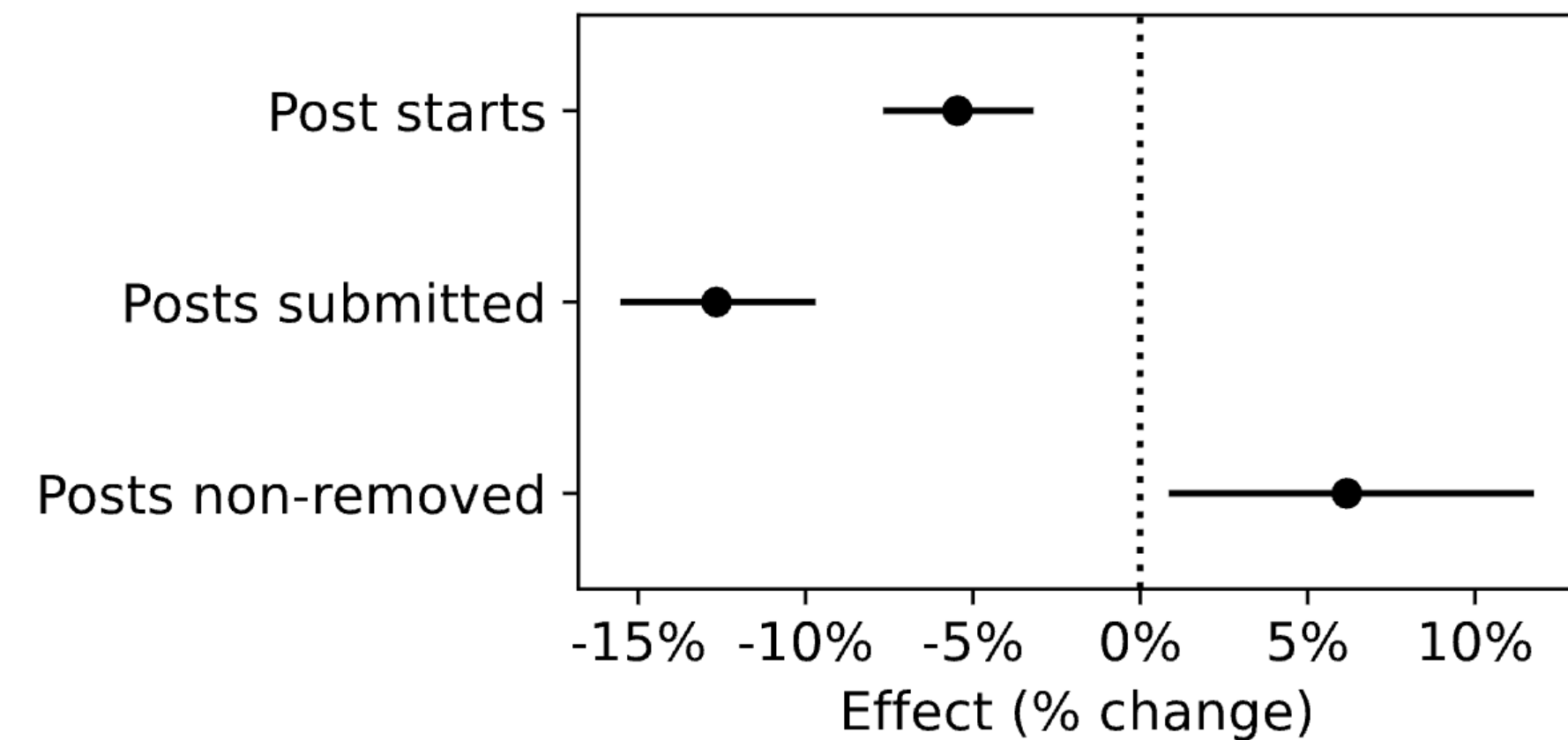


Post guidance...

... increases the number of non-removed contributions to communities adopting it.

... decreases the workload for moderators in communities adopting it.

... increases the quality of contributions to communities adopting it.



-9.4%*** Number of reports

-34.9%*** Posts removed by Automod

+28.6%** Comments

+36.1%** Upvotes

What's next?

Applied ML

Training a classifier to detect posts likely to be removed

Quasi-experiments

How can we answer causal questions without experiments?

Causal ML

Determining which communities will post guidance work

LLMs in practice

Allowing mods to customize prompts instead of regexes



MENU

APPETIZER (5 minutes)
Manoel's Intro

1st COURSE (15 minutes)
Why this course?

2nd COURSE (15 minutes)
What are we going to learn?

DESSERT (5 minutes)
Course Logistics

Meeting Logistics

- **Lectures:** Tuesdays and Thursdays (2:55 pm – 4:15 pm)
- **Quizzes:** “live” during the first 15 minutes of (most) Thursday lectures
- **Precepts:** Fridays! We will give you exercises that will prepare you for homework and the project.
 - We have 25-seat slots either 10:40 am to 11:30 am; or 12:15 am to 1:05 pm.

Course Components

- **Project (50% of grade):** End-to-end data science project in groups of ~5! (AI use encouraged!). Three milestones
 - **M1 (2%):** Propose project idea
 - **M2 (8%):** Initial data wrangling and result.
 - **M3 (40%):** Final report and presentation.
- **Quizzes (25% of grade):** Multiple-choice questions administered at the beginning of most Thursday lectures.
 - The two worst grades are not considered!

Course Components

- **Homework (25% of grade):** Short exercises assigned at the end of each precept. (AI use forbidden!)
 - Graded on a fail/pass basis. **Pass** = At least 60% correct!
 - **Released:** Fridays; **Deadline:** the next Friday.
 - The worst grade will be dropped; it will not be considered!
- **Precepts:** Not graded but essential to the class! You...
 - Can ask questions about homework & projects.
 - Will go through tutorials that will teach practical skills!

Final Project Submission

Following the university policy, you must submit your final project report during the 3-hour window on December 17th from 8:30 am to 11:30 am ET set by the university registrar!

Project

Precepts

Hwork

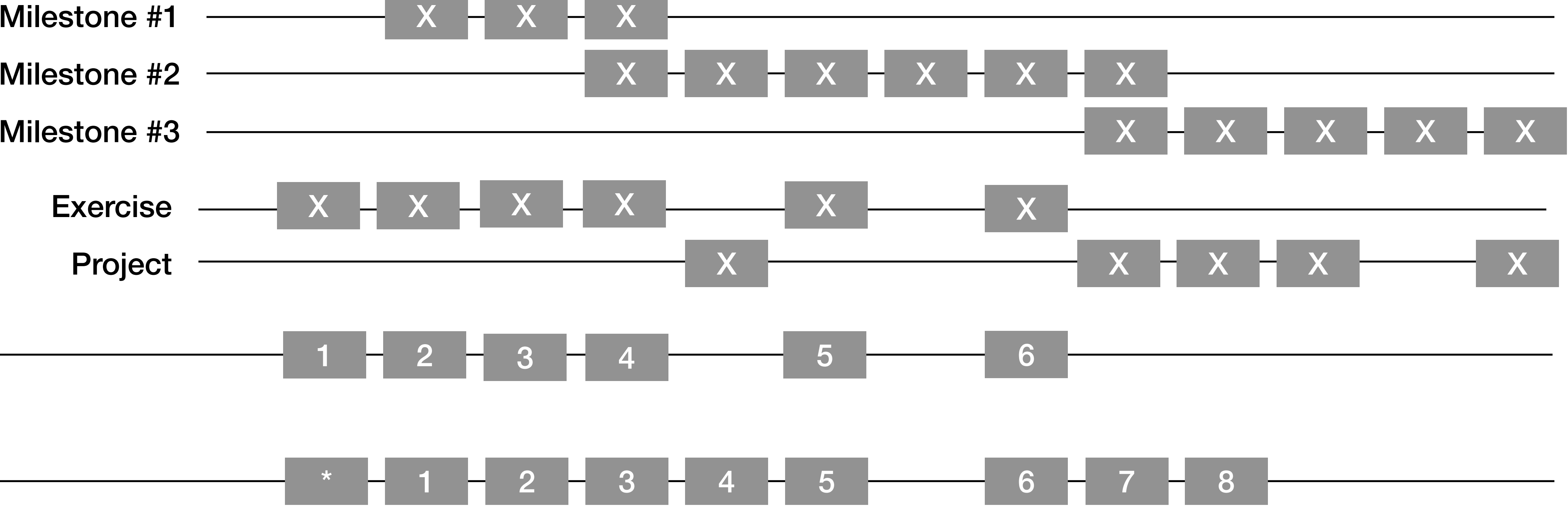
Quizzes

Basics

Causal Inference

Advanced

W1* W2 W3 W4 W5 W6 W7 W8 W9 W10 W11 W12 W13^ W14





Questions?

