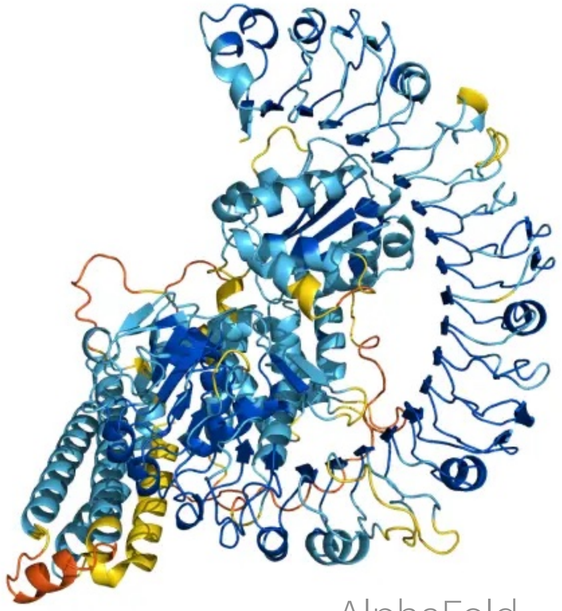
AI and algorithmic decision making for social good

Lily Xu | ORCS E8100 | Spring 2026 | Columbia University

AI can save the world

Bureau of Labor Statistics

International Medical Corps



AlphaFold

AI can destroy the world

Radiolab

GEP

ProPublica

Is technology neutral?

What levers are at our control? How much willful ignorance can we claim?

How do we expect the unexpected?

What does it mean to be a responsible technologist?

Are research innovation and social good at odds?

What can I do with my degree, anyways?

In this course...

What problems to work on?

Whom to involve?

How to build the right model?

- Potential
- Pitfalls
- Using foundation models

What are algorithmic harms, and how to mitigate them?

How to be “fair”?

How to involve people and their preferences in decision making?

How to design for people? Complement, not replace human knowledge?

When and how can we generalize?


Sustainability
THE OCEAN CLEANUP


EUROPEAN INTERBALKAN MEDICAL CENTER
COVID-19 TESTING POINT
Molecular method, through rRT-PCR
Healthcare
YEARS OF EXCELLENCE
ATHENS MEDICAL GROUP
one step ahead!
#safe_ahead
Medical Excellence & Healthcare Innovation
EUROPEAN INTERBALKAN MEDICAL CENTER
#safe_ahead
www.athensmedicalgroup.com


Government


SCHOOL BUS
STOP
135
CHAMPAIGN COMMUNITY UNIT SCHOOL DISTRICT NO. 4
Education

Our goal

The goal will be to learn **principled underpinnings** of ethical, intentional, and impactful research alongside **rigorous tools** drawing from machine learning, statistics, and operations research.

We'll complement these with a breadth of examples across sustainability, healthcare, government/nonprofit, and education — each that contribute research advances while also positively impacting society.

These principles and techniques will help not just when working on "social good" type research, but anytime you're developing or implementing technology that affects people — which is always!

A few logistics

Course expectations

- In-class paper presentations (30%)
- Project (45%)
- Participation (20%)
- Paper responses (5%)

Paper presentations (30%)

- Two types of papers: Application-focused and Methodological / conceptual
- **Application domains**
  - Each person sign up to ***solo present one paper***
- **Methodological/conceptual**
  - Each person sign up to ***(jointly) present three papers*** (two in pairs, one solo)
- At least one presentation before spring break; at least one after spring break

Note, this is 51 papers covered but only 44 slots.

There is some intentional redundancy here!

Paper presentations (30%)

- Class: 75 min long
- Two papers per class
  - Aim for a **15 minute presentation** and be prepared to lead 15 min discussion per paper
- 15 minutes additional for background, collective discussion, and administrivia
- Upload your slides after class (on CourseWorks)

Presentation format

Structure

- **Motivate:** (Intro) What is the problem? Why is it important?
- **Contextualize:** (Related work) How have others tried to address this or similar problems?
- **Explain:** (Methodology) What does this paper propose? What's novel about it? What are the core technical insights?
- **Critique:** (Reviewer) Where may failure modes be? How could this paper be improved?
- **Reflect:** (Discussion) What are implications? Was this paper successful?

Post-presentation discussion

- Prepare ~5 or more discussion questions to help guide the conversation
- These could be questions you are genuinely curious about

Paper responses (5%)

>3 sentences ***per paper*** with your reactions to the paper.

Consider taking on one or more of the following roles:

- **Advocate:** Highlight your favorite aspects of the paper
- **Critic:** Critique of the paper
- **Dreamer:** Idea for an extension of this paper, or a separate project idea inspired by the paper
- **Connector:** Draw links with other topics/papers from the course
- **Academic:** Connect with your own research

Paper responses (5%)

- Must be submitted by 11am the day of class (**no credit after this**)
- You do not need to submit for papers that you're presenting
- You get two excused non-submissions

- Do NOT use LLMs to do this work for you. Doing so will be considered an academic honesty violation and will be treated as such.
  - And remember, paper responses are just 5% of your grade, but in-class participation is 20%

Participation (20%)

- We'll take attendance!
- We expect all enrolled students to meaningfully engage in discussions
- I may call on you in reaction to your written paper responses

Project timeline (45%)

- Group declaration (February 18)
- Project proposal (March 9)
- Project pitch presentation (March 11)
- Check-in
- Final presentations (April 29 and May 4)
- Project writeup (during finals period)

On CourseWorks

- Paper presentation signups
- Form that must be completed each lecture (no later than **11am**!)
- PDFs of papers

Schedule FYI

- Class will be remote (on zoom) the following days:
  - Monday, February 23
  - Wednesday, February 25
  - Wednesday, April 8 (**guest lecture**)
- No class
  - Monday, March 9

Wicked problems

Churchman. *Management Science* 1967. Wicked Problems

Some background

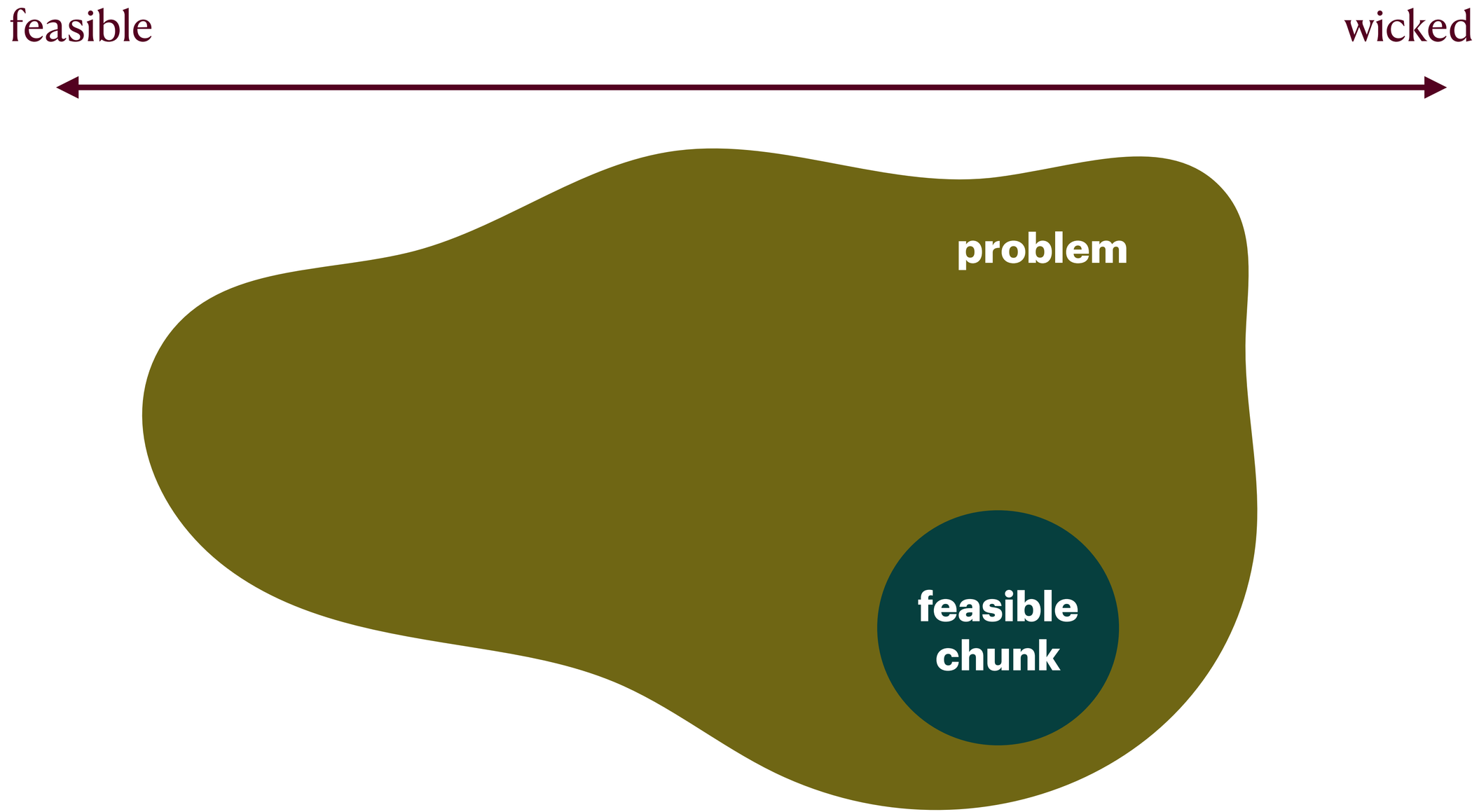
Horst Rittel

UC Berkeley, Science of Design

C. West Churchman

UC Berkeley, Business Administration & Peace and Conflict Studies

Just how extensive are the wicked problems, he did not tell us, but one was led to conclude from the discussion that the membership in the class of non-wicked problems is restricted to the arena of play: nursery school, academia and the like.


feasible
wicked
problem
feasible
chunk

Characteristics of wicked problems

1. There is no definitive formulation of a wicked problem.
2. Wicked problems have no stopping rule.
3. Solutions to wicked problems are not true-or-false, but better or worse.
4. There is no immediate and no ultimate test of a solution to a wicked problem.
5. Every solution to a wicked problem is a "one-shot operation"; because there is no opportunity to learn by trial and error, every attempt counts significantly.
6. Wicked problems do not have an enumerable (or an exhaustively describable) set of potential solutions, nor is there a well-described set of permissible operations that may be incorporated into the plan.
7. Every wicked problem is essentially unique.
8. Every wicked problem can be considered to be a symptom of another problem.
9. The existence of a discrepancy representing a wicked problem can be explained in numerous ways. The choice of explanation determines the nature of the problem's resolution.
10. The social planner has no right to be wrong (i.e., planners are liable for the consequences of the actions they generate).

Rittel and Webber (1973)

Characteristics of wicked problems

1. There is no definitive formulation of a wicked problem.
2. Wicked problems have no stopping rule.
3. **Solutions to wicked problems are not true-or-false, but better or worse.**
4. There is no immediate and no ultimate test of a solution to a wicked problem.
5. Every solution to a wicked problem is a "one-shot operation"; because there is no opportunity to learn by trial and error, every attempt counts significantly.
6. Wicked problems do not have an enumerable (or an exhaustively describable) set of potential solutions, nor is there a well-described set of permissible operations that may be incorporated into the plan.
7. Every wicked problem is essentially unique.
8. Every wicked problem can be considered to be a symptom of another problem.
9. The existence of a discrepancy representing a wicked problem can be explained in numerous ways. The choice of explanation determines the nature of the problem's resolution.
10. The social planner has no right to be wrong (i.e., planners are liable for the consequences of the actions they generate).

Rittel and Webber (1973)

What seems to emerge is not a moral reprimand of the management scientist, but rather a moral problem of the profession, a wicked moral problem. To what extent are we morally responsible to inform the manager in what respect our "solutions" have failed to tame his wicked problems? Does "inform" merely mean that we clear ourselves legally, or does it mean that we attempt to enter into a deep, mutual understanding of the untamed aspects of the problem?

To date, operations research and management science have been largely indifferent to the morality of the profession, perhaps because the profession has not yet taken itself seriously. That the profession has a moral problem, nonetheless, there can be no doubt. It might make us look more mature if we began to discuss it.

C. West Churchman
University of California, Berkeley

Wickedness and social good

- Why even bother?
- Most social good problems are wicked in some way
  - climate change, poverty, international development, sustainability, migration, public health, public policy, education, social work…
- “Taming” a wicked problem requires making ethical decisions
- We should have humility, be open to methods from “soft systems”, and be open to iteration

So what *can* we do?

Abebe, Barocas, Kleinberg, Levy, Raghavan, Robinson. *FAccT* 2020.
Roles for computing in social change

Roles for computing in social change

Diagnostic: Help measure social problems and diagnose how they manifest in technical systems.

Formalizer: Shape how social problems are understood, by explicitly specifying inputs and goals.

Rebuttal: Clarify the limits of technical interventions, and of policies premised on them.

Synecdoche: Foreground long-standing social problems in a new way.

Roles for computing in social change

Diagnostic: Help measure social problems and diagnose how they manifest in technical systems.

Sunlight is the best disinfectant

Audits, causal inference, statistical inference

Formalizer: Shape how social problems are understood, by explicitly specifying inputs and goals.

The beginning of wisdom is the definition of terms

Make normative decisions in developing model or specifying a dataset

Rebuttal: Clarify the limits of technical interventions, and of policies premised on them.

The art of self-moderation

Show upper/lower bounds

Synecdoche: Foreground long-standing social problems in a new way.

What's new is old

Opportunity to supplant/dismantle, or reinforce