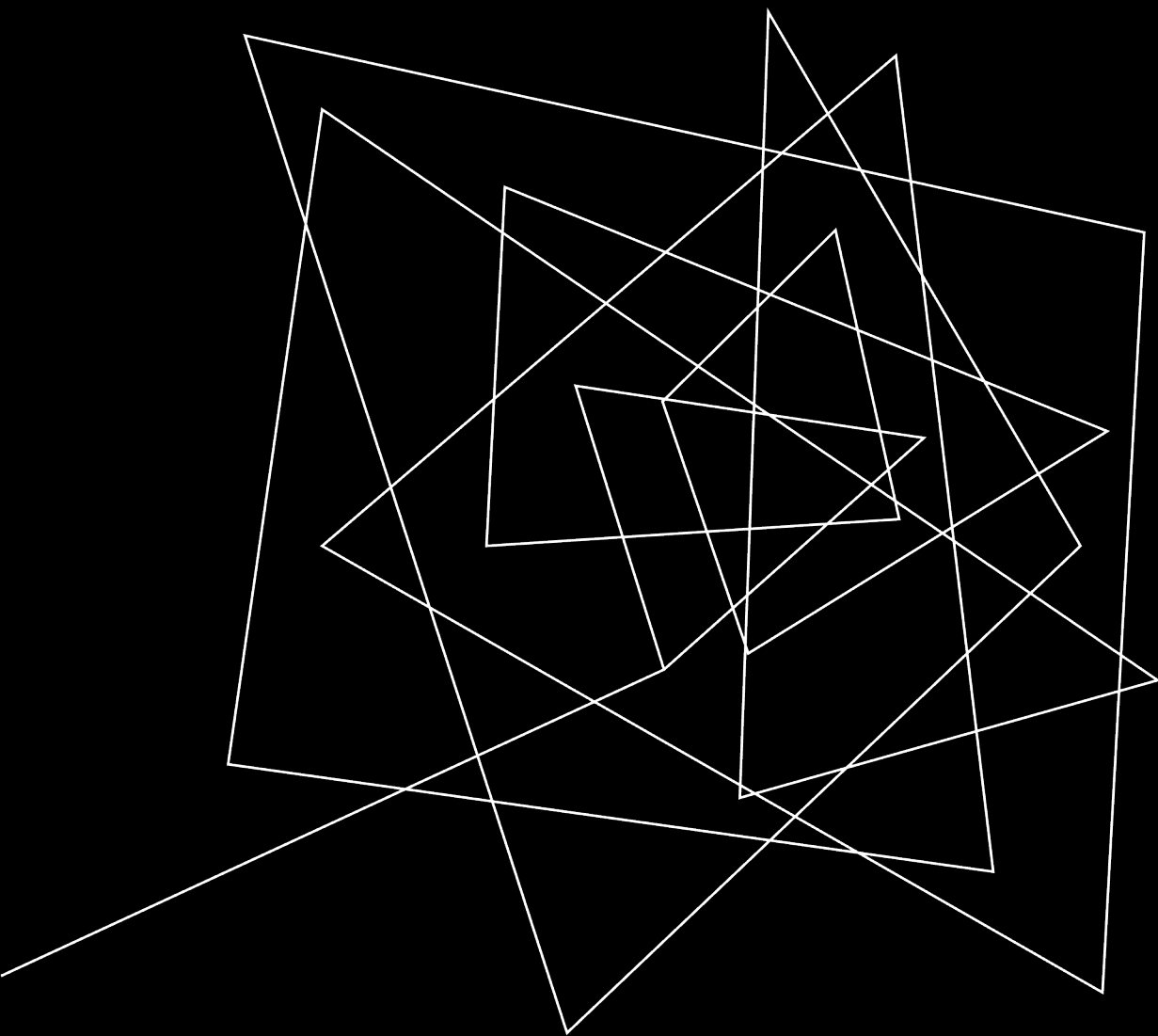# RESPONSIBLE AI

Week 8: Capstone Planning & Avoiding Techno-Solutionism

Emily Ramond and Meira Gilbert

TODAY'S OBJECTIVES

- What is techno-solutionism?

- How does implementation affect the ethical impacts of AI?

- How can we build effective socio-technical AI systems?

- Planning the Q2 project proposal

- Replication Projects

Q1 REPLICATION PROJECT

Final Submission Requirements

FINAL REPLICATION PROJECT (Q1 PROJECT)

- Replication part3 was due this morning by email

  - Check that you answered all questions in the writeup portions and **EDA is sufficient.**

- Replication Final Submissions (Q1 Project) due **Monday, December 11th at 11:59PM**.

  - Presentations will be held by each team next class – no need to submit to us

  - Submit code and report to Gradescope by above deadline

  - Use LaTeX to create report (find template on capstone website)

    - "You should present the conclusions drawn from *your* data/methods and address any discrepancies with the work being studied"

  - Make sure you follow all code requirements and structure (Graded by TAs!)

    - **This will be reviewed in your week9/10 check-in with TA**
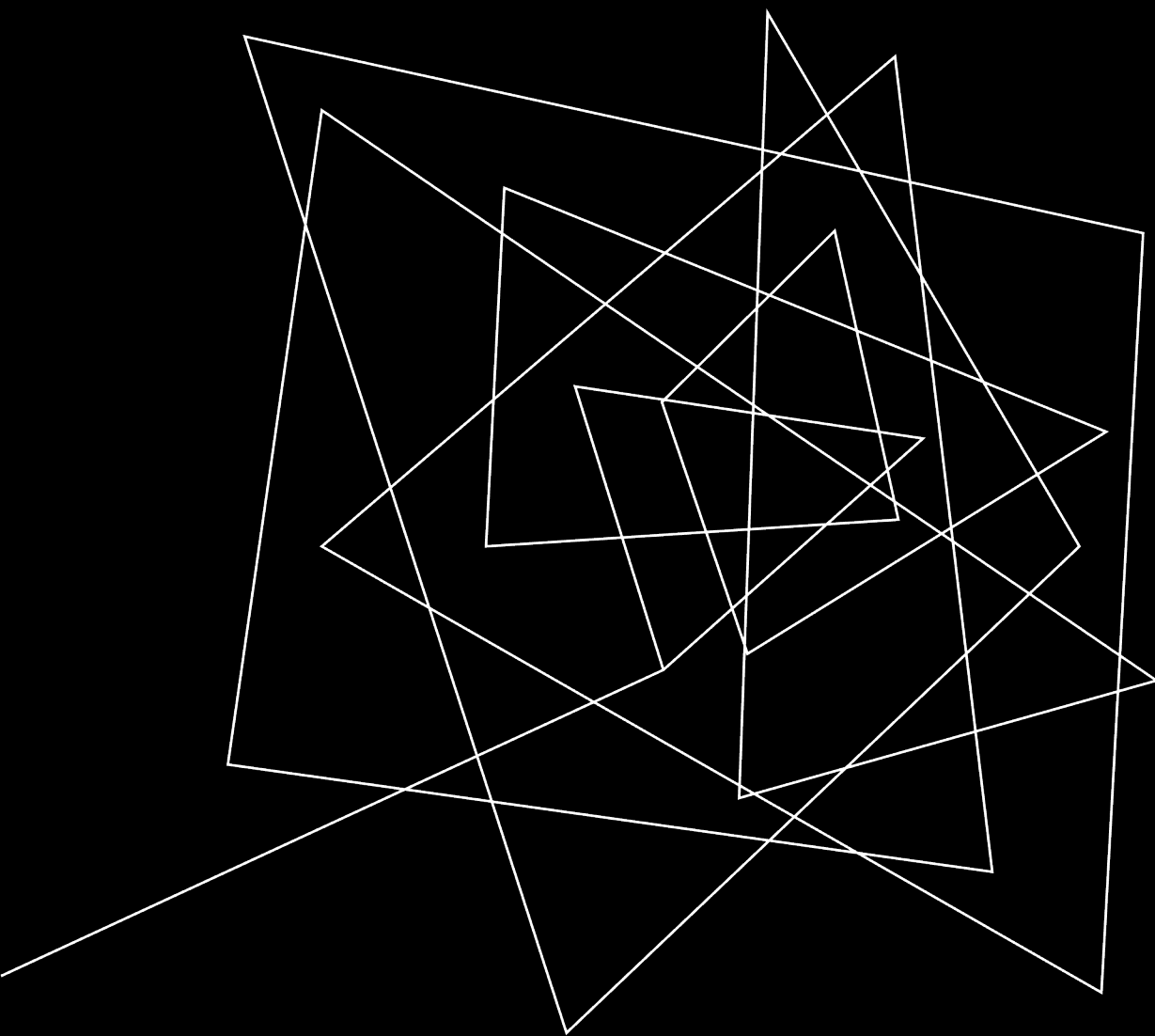
Q2 CAPSTONE PLANNING

Creating a Proposal

PUTTING TOGETHER A PROJECT PROPOSAL

All components are due on Monday, December 11th at 11:59PM to Gradescope.

1. A broad problem statement (with context to justify spending 10 weeks on this project). Audience is broad.

2. Careful problem statement for your mentors. Audience is domain experts.

3. Statement for primary output. (report/app/website) . How will you communicate analyses?

Research methods: data, processes, necessary resources

- Is the data available? Is it of a sufficient quality to run your desired analysis?

- The easiest way to do this is to actually obtain the data while writing the proposal and perform some preliminary analyses to ensure it will actually be able to help you accomplish your goals.

Timeline

- 6 week schedule with specific tasks/goals/responsibilities per team member (12 hours each per week!!)

In years past, many groups fell in the trap of waiting until Quarter 2 to obtain their data, and by the time they realized that they either can't obtain their data or that it won't actually help them with their project, they had wasted a significant portion of the quarter. **Don't let this be you!**

Sources:
https://dsc-capstone.github.io/assignments/projects/q2-proposal/
https://students.ucsd.edu/_files/aep/research-proposal-guidelines-13.pdf
https://www.conted.ox.ac.uk/about/writing-your-research-proposal

PUTTING TOGETHER A PROJECT PROPOSAL

All components are due on Monday, December 11th at 11:59PM to Gradescope.

1. A broad problem statement (with context to justify spending 10 weeks on this project). Audience is broad.

2. Careful problem statement for your mentors. Audience is domain experts.

3. Statement for primary output. (report/app/website) . How will you communicate analyses?

Research methods: data, processes, necessary resources

- Is the data available? Is it of a sufficient quality to run your desired analysis?

- The easiest way to do this is to actually obtain the data while writing the proposal and perform some preliminary analyses to ensure it will actually be able to help you accomplish your goals.

Timeline

- 6 week schedule with specific tasks/goals/responsibilities per team member (12 hours each per week)

In years past, many groups fell in the trap of waiting until Quarter 2 to obtain their data, and by the time they realized that they either can't obtain their data or that it won't actually help them with their project, they had wasted a significant portion of the quarter. **Don't let this be you!**

This was me!

PUTTING TOGETHER A PROJECT PROPOSAL

- Topic, Aims, Contributions

  - Your general subject area and topic

  - Why your topic is significant

  - Your research questions / hypotheses

- Background & literature review: how does your work fit in to the broader landscape? Is it sufficiently novel?

- Process and Logistics

  - Research methods: data, processes, necessary resources

    - Is the data available? Is it of a sufficient quality to run your desired analysis?

    - What will the outputs of your analysis look like?

  - Ethical considerations: will you be surveying, experimenting on, or otherwise working with human subjects? Will you be working with sensitive data?

  - Timeline: can this work be completed in a reasonable amount of time? What is your plan for completing it? Which team members will have which responsibilities?

Sources:
https://dsc-capstone.github.io/assignments/projects/q2-proposal/
https://students.ucsd.edu/_files/aep/research-proposal-guidelines-13.pdf
https://www.conted.ox.ac.uk/about/writing-your-research-proposal

PUTTING TOGETHER A PROJECT PROPOSAL

- Elevator Pitch

  - 1-3 minute recording on Youtube

  - Graded by TAs

  - Dry run due in Week 9/10 check in with TAs

> ## Summary of Deliverables

All deliverables will be submitted to Gradescope.

- Written Proposal.

  - Submitted as a group.

  - Must use the LaTeX template.

- Schedule.

  - Submitted as a group.

  - No required format.

- Elevator Pitch.

  - Submitted individually.

  - Must present dry run during your group's Week 9/10 TA Check-In.

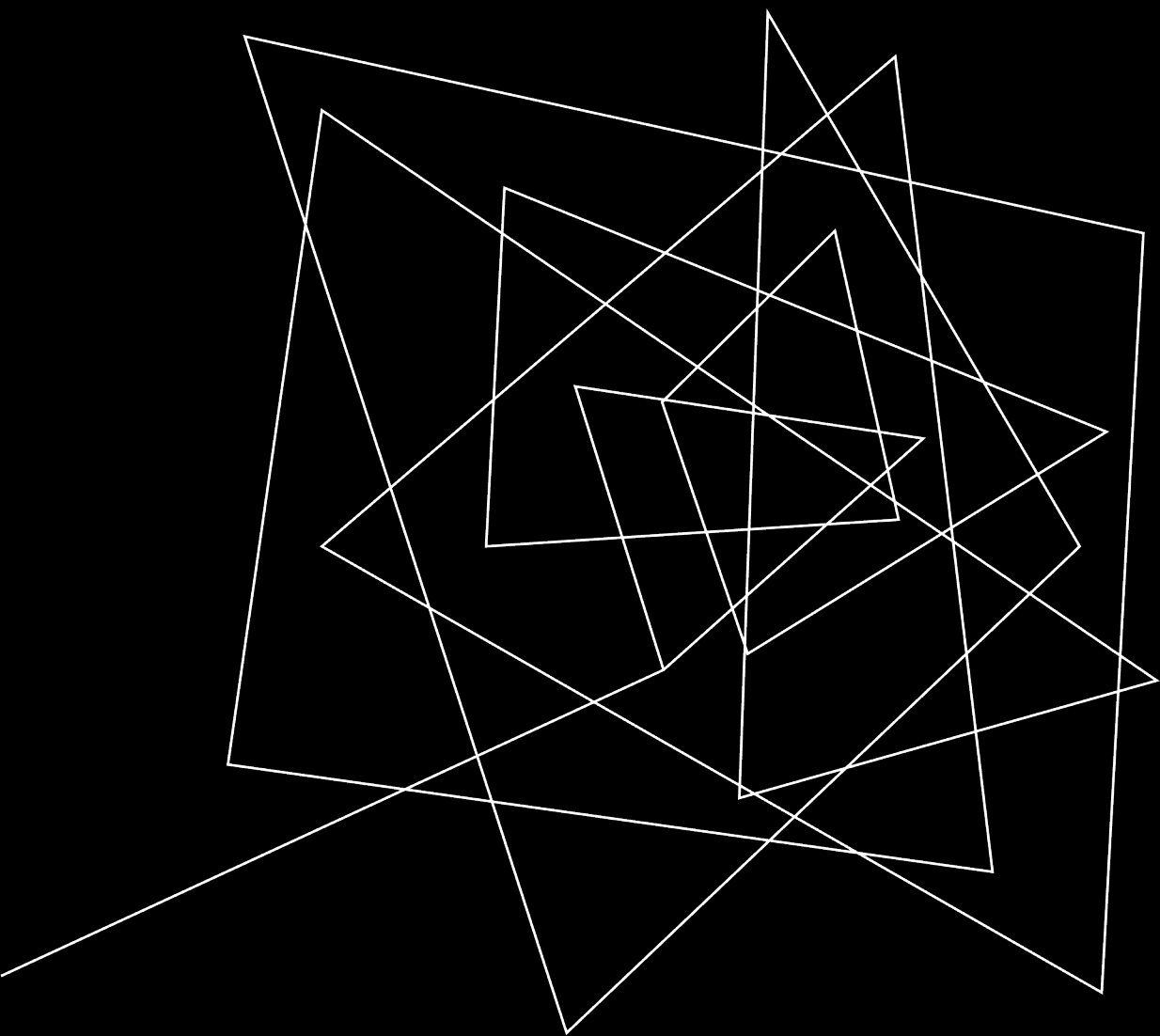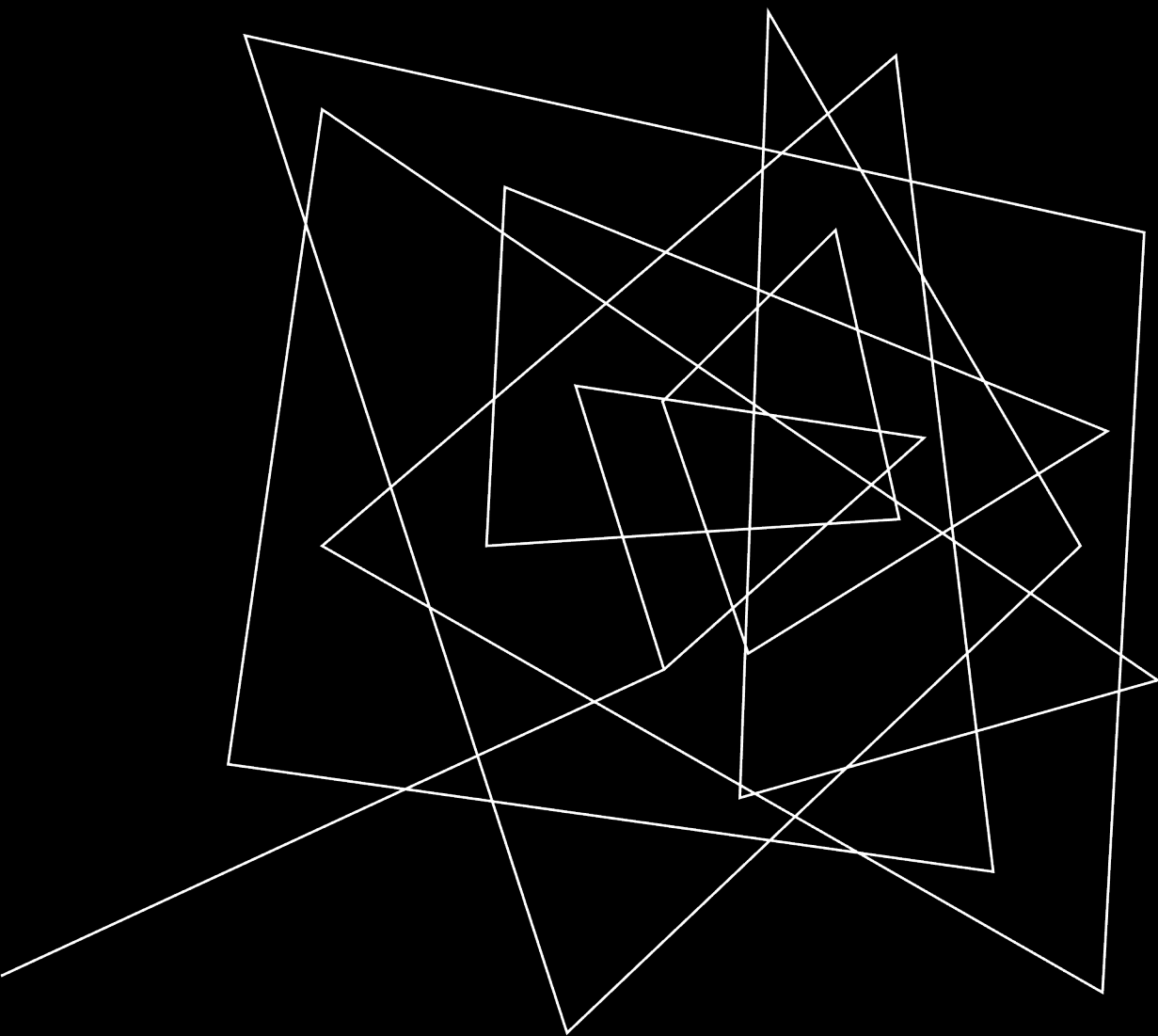 - Must create YouTube video of final pitch.
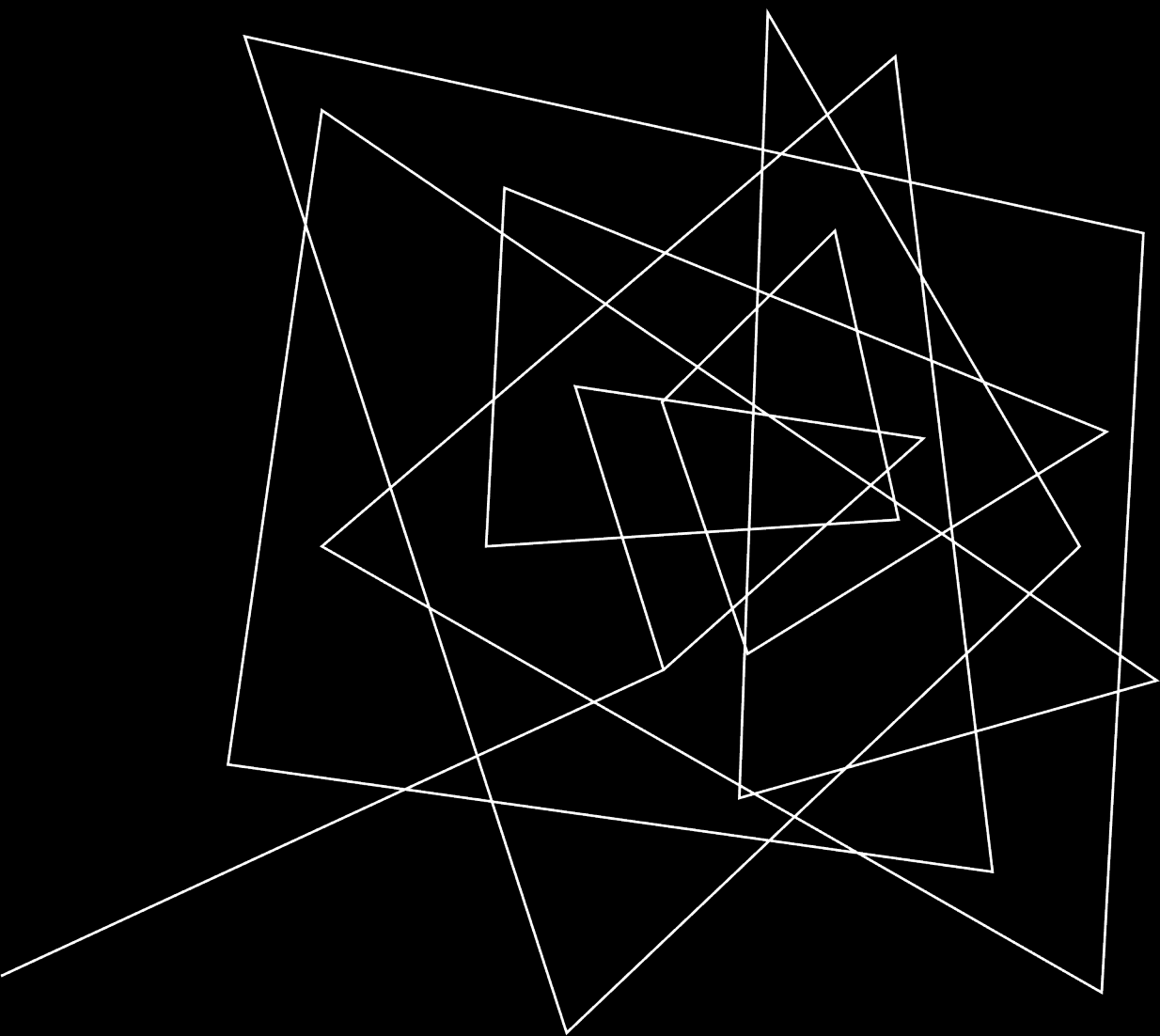
READING PRESENTATION:

Algorithmic Fairness and Vertical Equity: Income Fairness with IRS Tax Audit Models

READING PRESENTATION:

Fairness through Awareness (Dwork et al.)

TECHNO-SOLUTIONISM

WHAT IS TECHNO-SOLUTIONISM?

A term coined by the technology critic [Evgeny Morozov](#), technological solutionism is the belief that complex problems can be alleviated or completely solved by reducing the core issues to simpler engineering issues.

1. Psychologically reassuring
2. Financially enticing
3. Plays into the power of technologists (vs activists, politicians, the people, etc.)

When analyzed critically, techno-solutionism may offer a "band-aid" solution to a much broader and more complex issue. Failure to address the root causes (such as systemic power imbalances, faulty economic systems, etc) might alleviate issues in the short term but ultimately exacerbate inequalities or cause additional wider-reaching issues.

IMPLEMENTATION MATTERS

	New Mexico	**Michigan**
purpose	Automatically identify individuals who are filing fraudulent unemployment claims	Automatically identify individuals who are filing fraudulent unemployment claims
algorithm	?	?
implementation strategy	1. Deploy the algorithm 2. Send pop-ups to individuals flagged by the algorithm as they fill out their claims, reminding them to honestly report their information	1. Deploy the algorithm 2. Fire human employees in the fraud detection unit 3. Immediately initiate legal action against individuals flagged by the algorithm 4. If they do not respond within 30 days, initiate legal action including garnishing wages or tax refunds
result	~25% increase in additional earnings reported by individuals who saw the pop-up	Individuals falsely flagged by the algorithm incurred ~$60M worth of legal and other costs; the state is being sued to recover those costs

IMPLEMENTATION MATTERS

	New Mexico	**Michigan**
purpose	Automatically identify individuals who are filing fraudulent unemployment claims	Automatically identify individuals who are filing fraudulent unemployment claims
algorithm	?	?
implementation strategy	1. Deploy the algorithm 2. Send pop-ups to individuals flagged by the algorithm as they fill out their claims, reminding them to honestly report their information	1. Deploy the algorithm 2. Fire human employees in the fraud detection unit 3. Immediately initiate legal action against individuals flagged by the algorithm 4. If they do not respond within 30 days, initiate legal action including garnishing wages or tax refunds
result	~25% increase in additional earnings reported by individuals who saw the pop-up	Individuals falsely flagged by the algorithm incurred ~$60M worth of legal and other costs; the state is being sued to recover those costs

Even if an **identical** algorithm was used in both examples, the outcomes would still be dramatically different – and that's because of implementation!
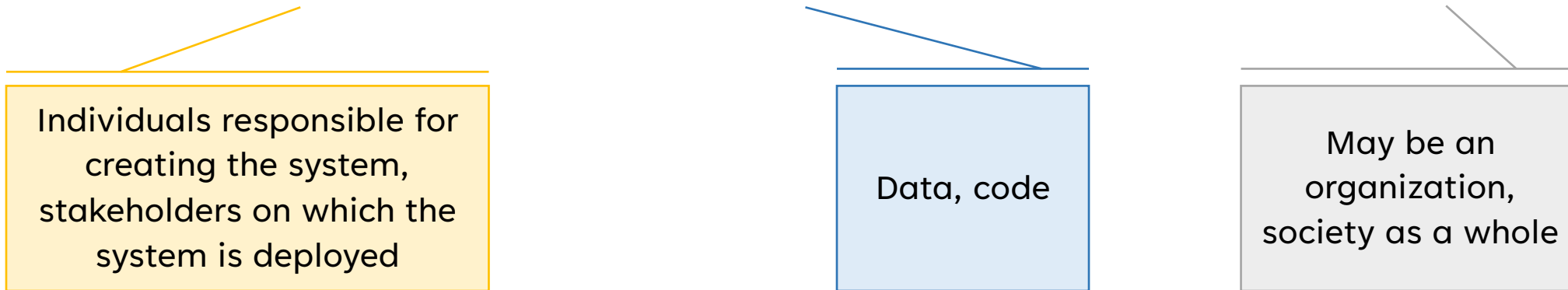
Sources:
https://undark.org/2020/06/01/michigan-unemployment-fraud-algorithm/
https://www2.deloitte.com/us/en/insights/deloitte-review/issue-18/behavior-change-among-unemployment-claimants-behavioral-economics.html

15

HOW DO WE AVOID TECHNO-SOLUTIONISM?

Understanding the **combination of** and **interaction between** social and technical elements in a system

Individuals responsible for creating the system, stakeholders on which the system is deployed	Data, code	May be an organization, society as a whole

1. Is the model fair?
2. Is it transparent? Robust? Explainable?
3. Does the model actually work and function as designed?
4. Is the problem even suited for an AI solution?
5. How will the model impact society? How will society impact the model?
6. What are the externalities of the model?

CASE STUDY: COMPAS RISK SCORES

Blindspot: *"the conceptual and methodological toolkit used to evaluate the fairness of algorithmic systems remains limited to a narrow set of computational and legal modes of analysis"* ([https://dl.acm.org/doi/abs/10.1145/3351095.3372859)](https://dl.acm.org/doi/abs/10.1145/3351095.3372859)

→ **Quantifiably fair(er) algorithmic decision-making is not always the same as fair use of AI**
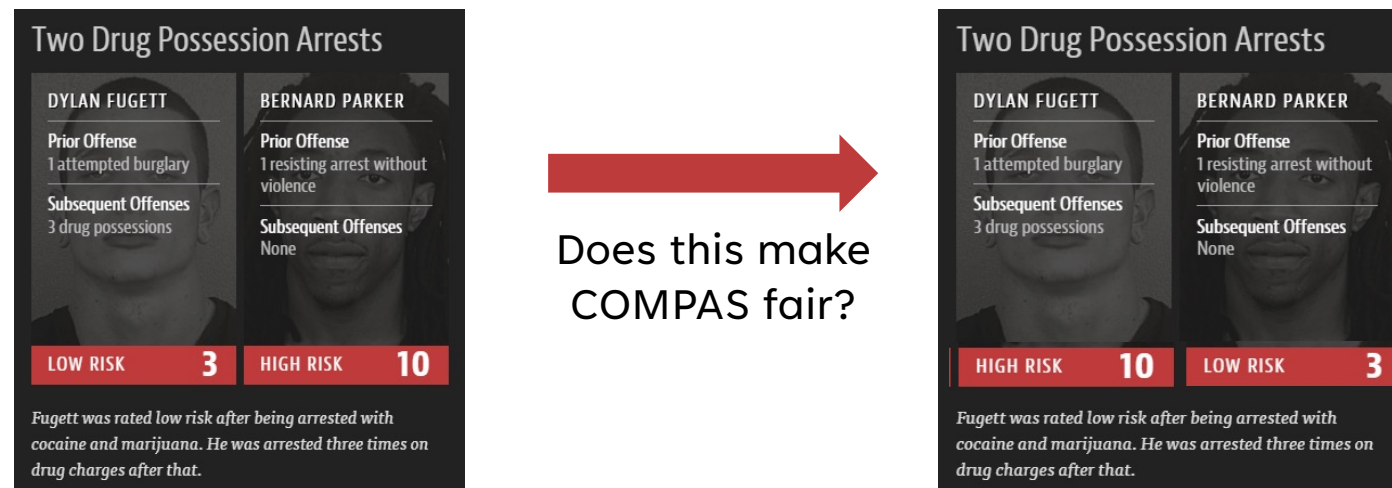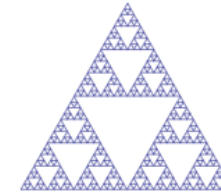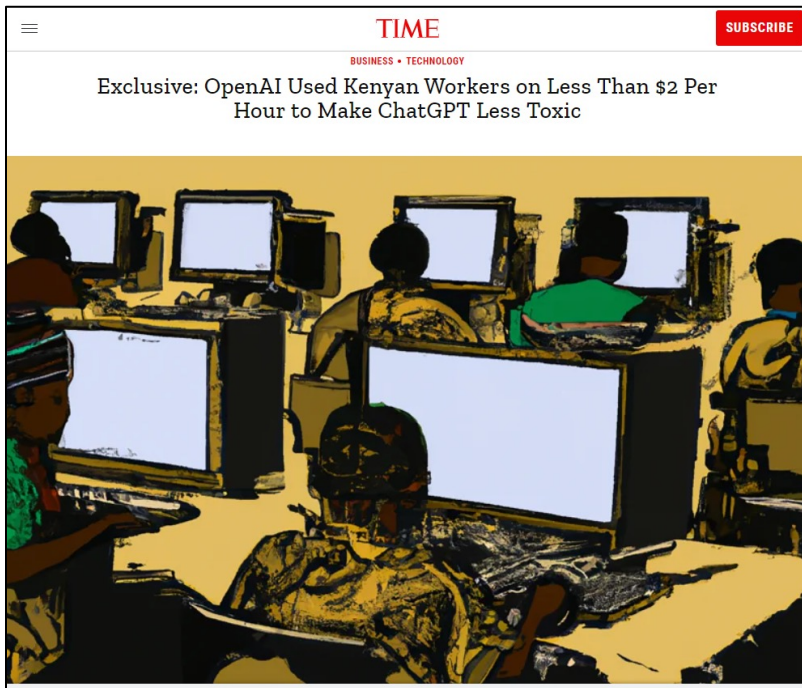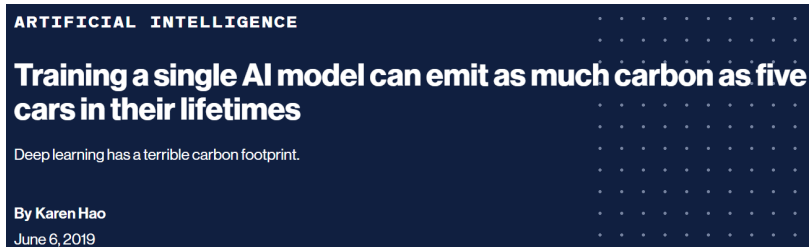

Does this make COMPAS fair?

1. Is the model fair?
2. Is it transparent? Robust? Explainable?
3. Does the model actually work and function as designed?
4. Is the problem even suited for an AI solution?
5. How will the model impact society? How will society impact the model?
6. What are the externalities of the model?

IS THIS EVEN SUITED FOR AN AI SOLUTION?

Sam Altman ✓
@sama · Follow

Replying to @sama

these tools will help us be more productive (can't wait to spend less time doing email!), healthier (AI medical advisors for people who can't afford care), smarter (students using ChatGPT to learn), and more entertained (AI memes lolol).

8:00 PM · Feb 18, 2023

❤️ 1.1K   💬 Reply   ⬆️ Share

Read 324 replies

Problem	Solution	Techno-Solution
It is difficult to get around without driving	Invest in robust public transportation	Self-driving cars
People can't afford healthcare	Regulate monopolies taking over private practices; federally set prices for drugs, universal healthcare	AI Telehealth
Climate change will destroy the earth	Implement the carbon tax, regulate emissions, ban single-use plastics	Build a human colony on Mars

WHAT ARE THE EXTERNAL COSTS OF THE MODEL?

ARTIFICIAL INTELLIGENCE

Training a single AI model can emit as much carbon as five cars in their lifetimes

Deep learning has a terrible carbon footprint.

By Karen Hao
June 6, 2019

TIME

BUSINESS • TECHNOLOGY

Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic
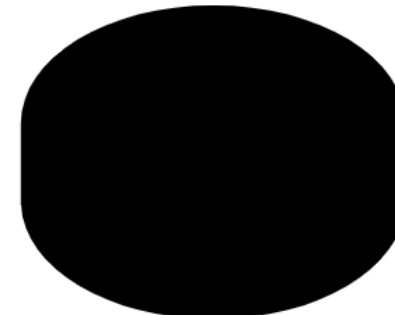

Anatomy of an AI System

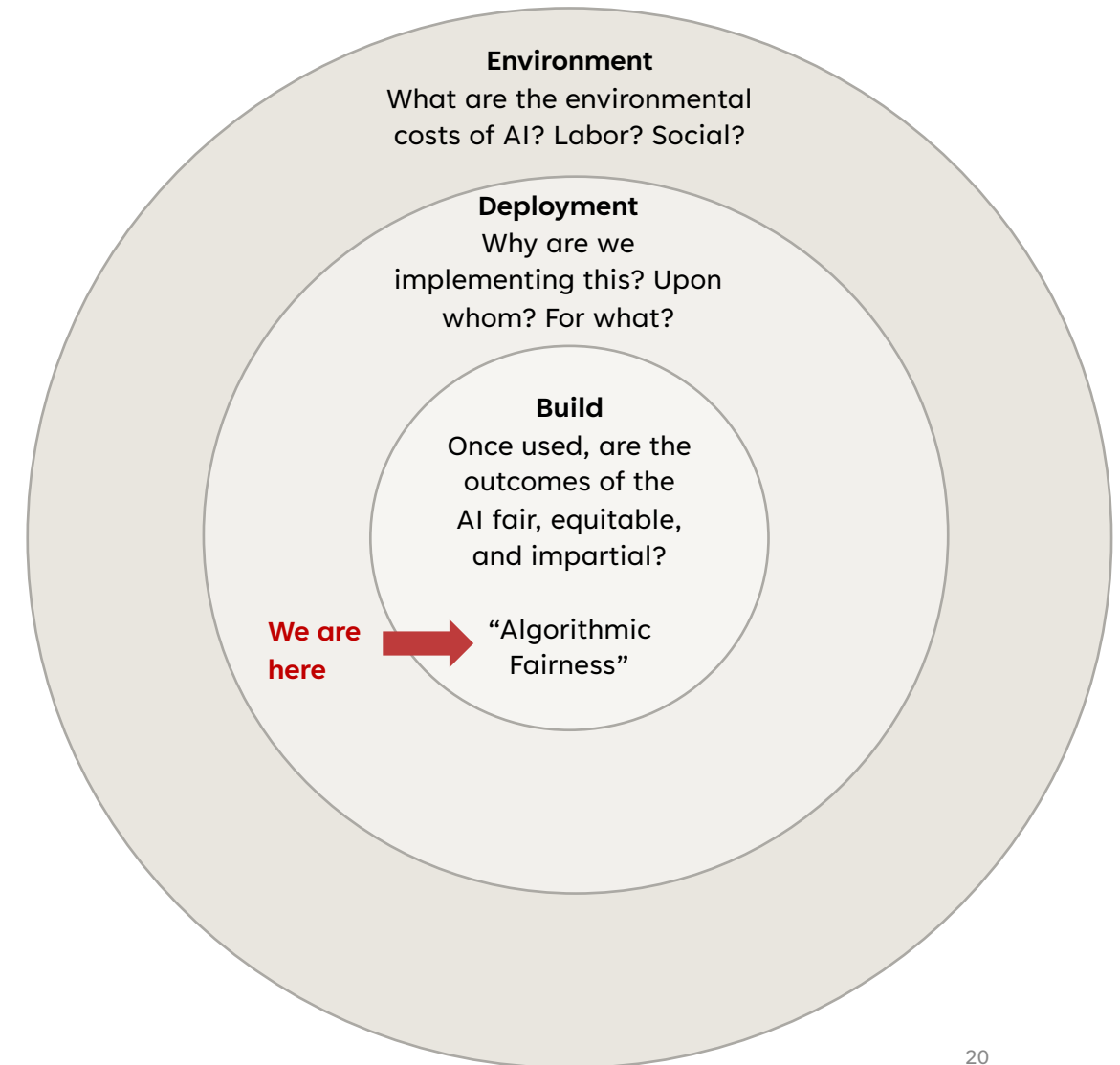The Amazon Echo as an anatomical map of human labor, data and planetary resources

By Kate Crawford[1] and Vladan Joler[2]
(2018)

"ZOOMING OUT"

"AI is **neither built nor deployed in a vacuum**, sealed off from societal realities of discrimination or unfair practices. Understanding AI as a socio-technical system acknowledges that the processes used to develop technology are **more than their mathematical and computational constructs.**

A socio-technical approach to AI takes into account the values and behavior modeled from the datasets, the humans who interact with them, and the complex organizational factors that go into their commission, design, development, and ultimate deployment."

[Towards a Standard for Identifying and Managing Bias in Artificial Intelligence](#) (NIST)

Environment
What are the environmental costs of AI? Labor? Social?

Deployment
Why are we implementing this? Upon whom? For what?

Build
Once used, are the outcomes of the AI fair, equitable, and impartial?

We are here → "Algorithmic Fairness"

20

DISCUSSION

Think about the ways you measured fairness during the Q1 project so far. Are there any potential fairness issues that you think can *not* be solved through technical solutions?

What are some examples of poor algorithmic implementation, and what negative affects can this cause (even if the algorithm itself is "fair")?

Have you encountered examples of techno-solutionism in your day-to-day life? What are they and how does it affect you?

FOR NEXT CLASS

- Your participation questions will be due at 11AM PT on Thursday, 12/7

- Check-in with TA by Monday with Q2 elevator pitch and Q1 code

- Submit your Q1 final projects and Q2 Proposals by December 11th 11:59pm

- Have final Q1 presentations ready for next class (Dec 8)