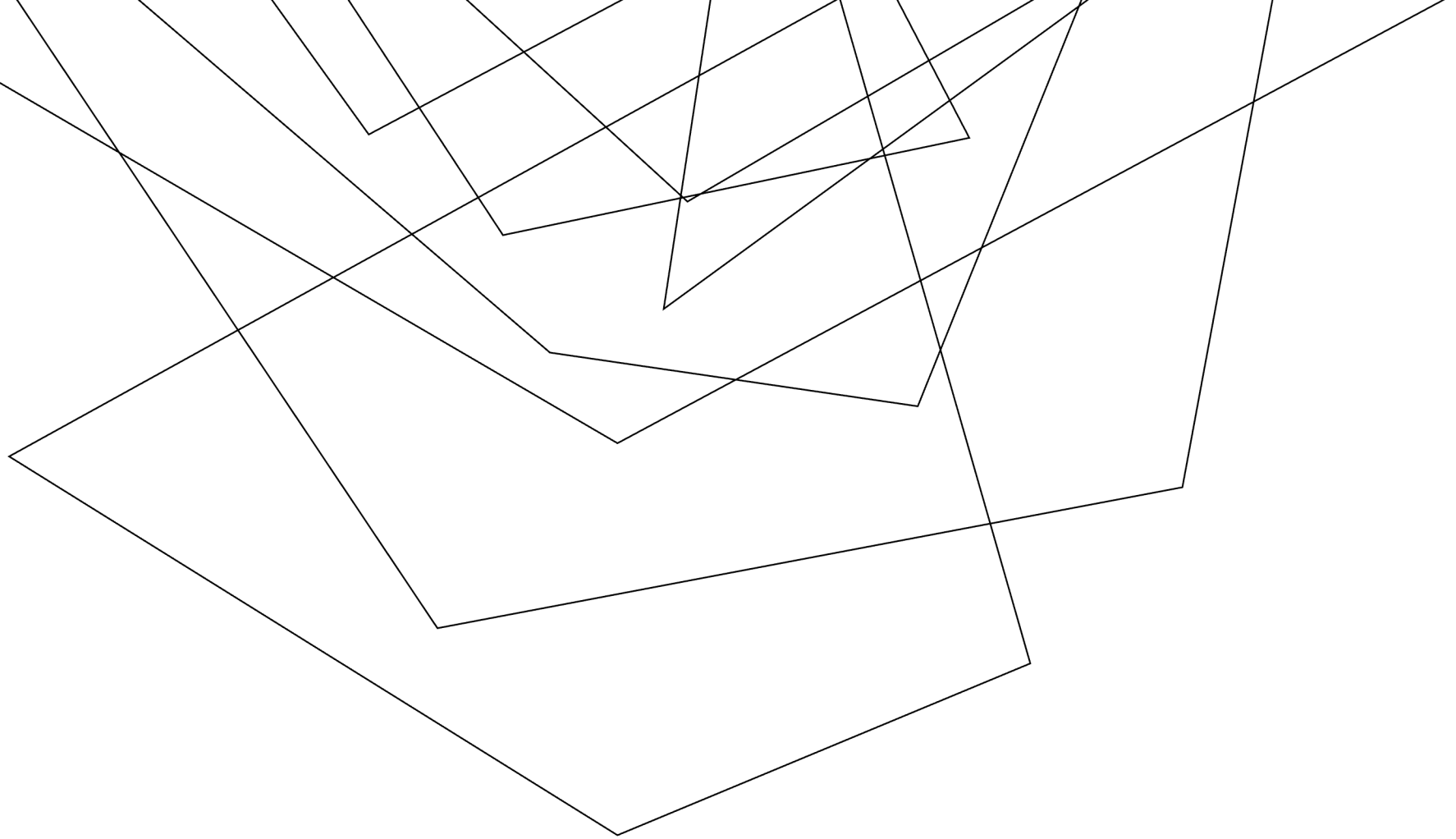


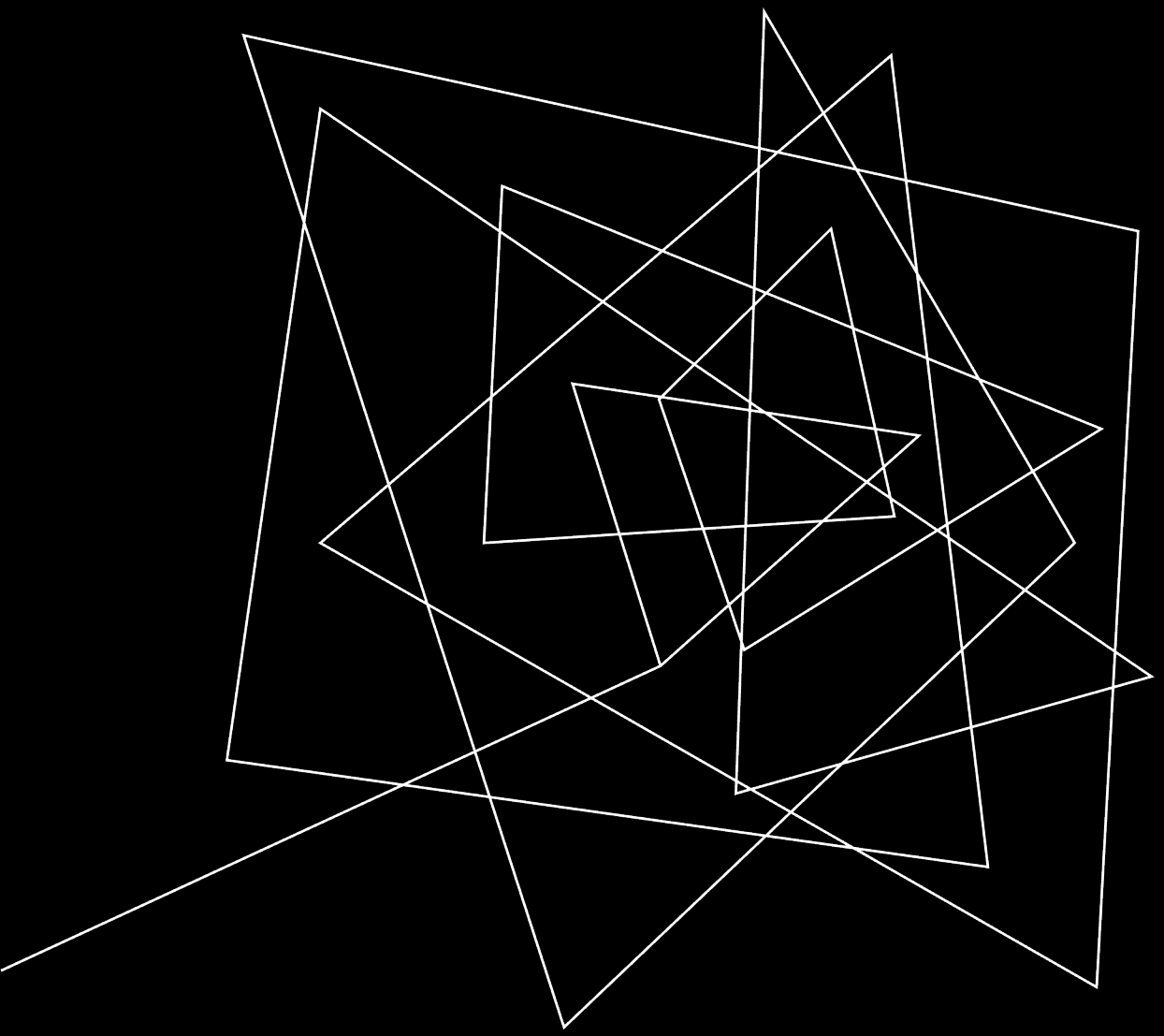
RESPONSIBLE AI

Week 8: Capstone Planning & Avoiding Techno-Solutionism

Emily Ramond and Meira Gilbert

TODAY'S OBJECTIVES

- How do implementation and other non-technical considerations affect the ethical impacts of AI?
- How do we design proposals and analysis plans for data science projects?
- Planning the Q2 project proposal
- Replication Projects



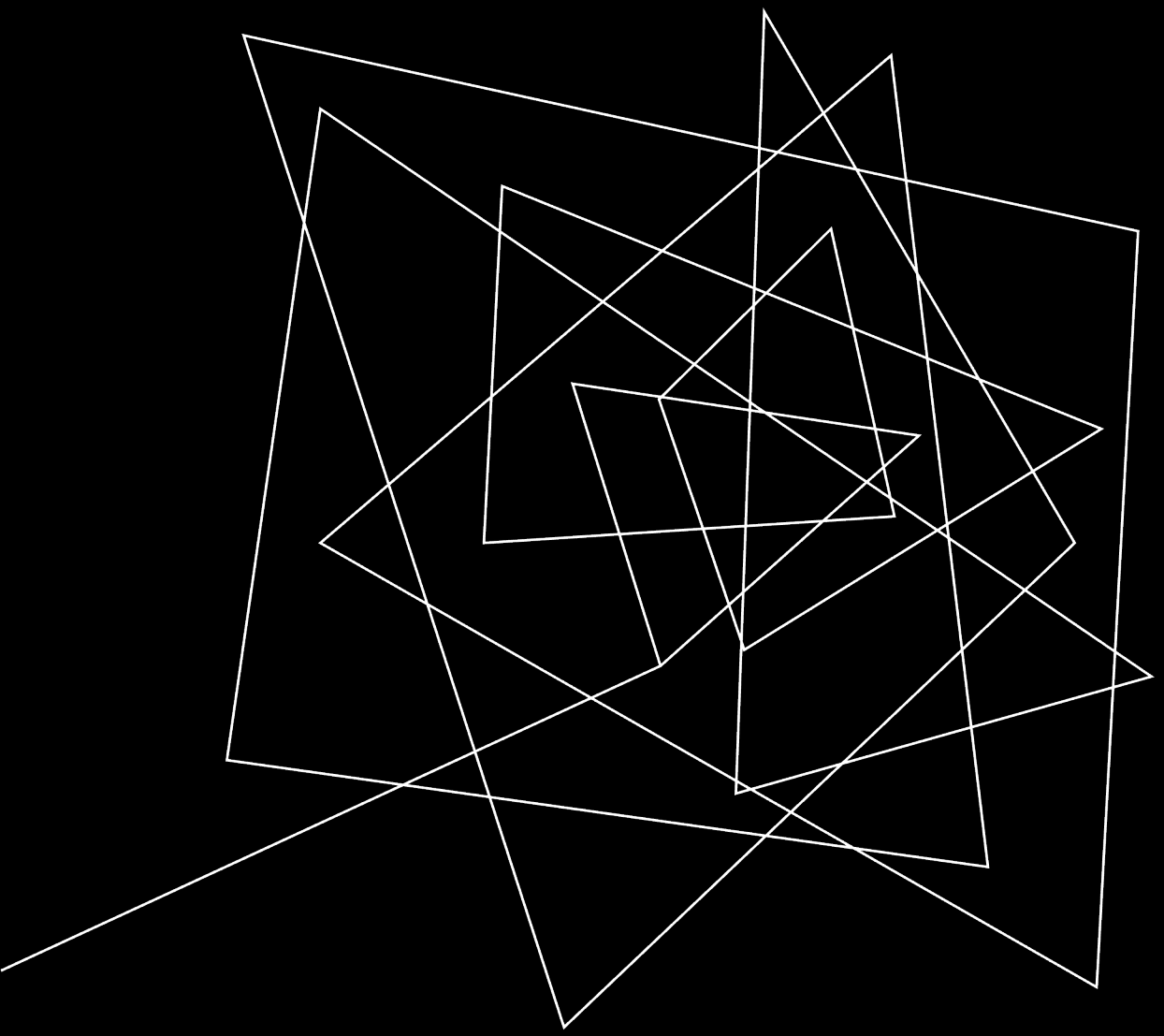
Q1 REPLICATION PROJECT

Final Submission Requirements

FINAL REPLICATION PROJECT (Q1 PROJECT)

- Replication part3 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**

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



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 wee)

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

- 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

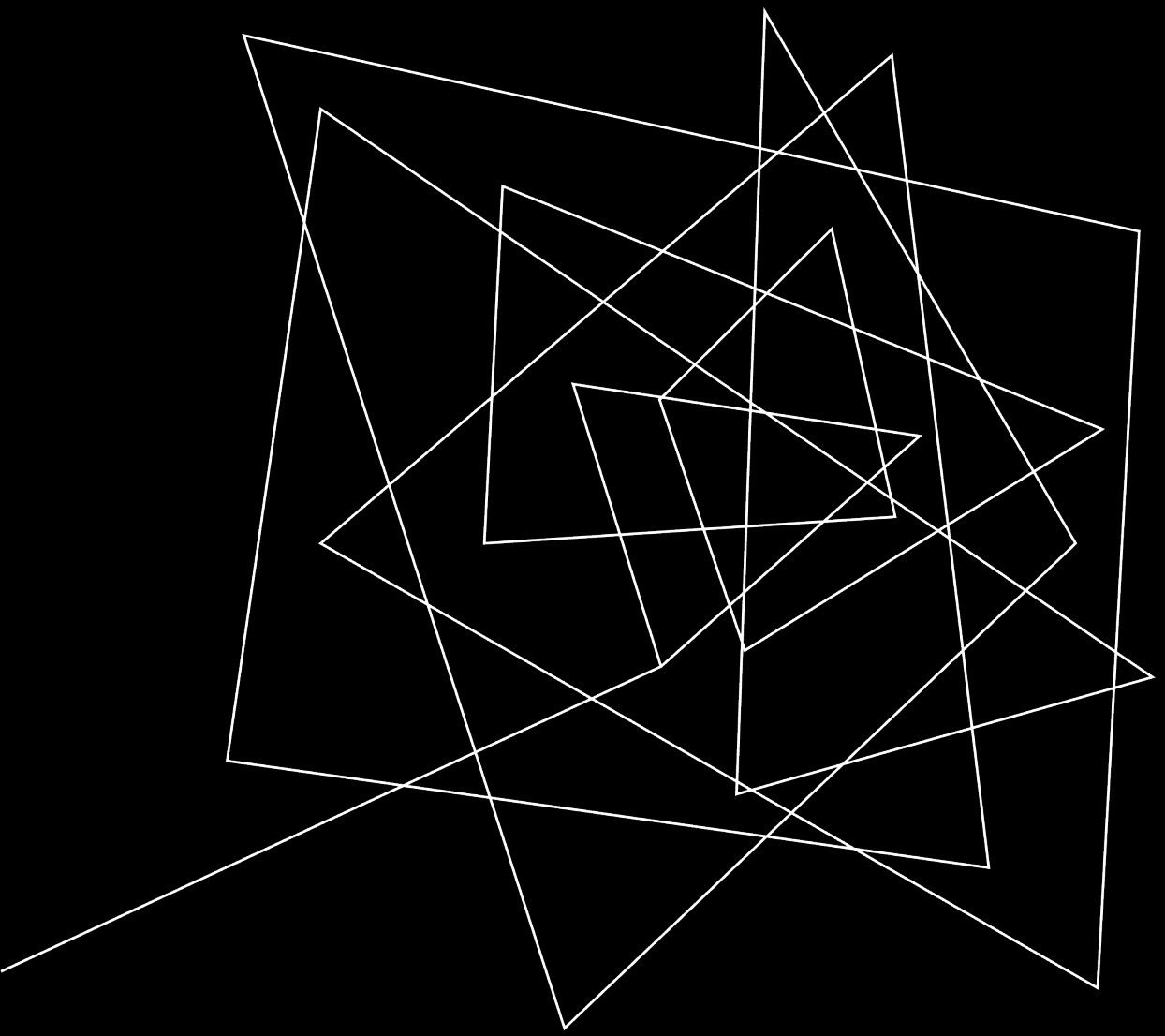
- Elevator Pitch
  - 1-3 minute recording on Youtube
  - Graded by TAs
  - Dry run due in Week 9/10 check in with TAs

Summary of Deliverables

All deliverables will be submitted to Gradescope.

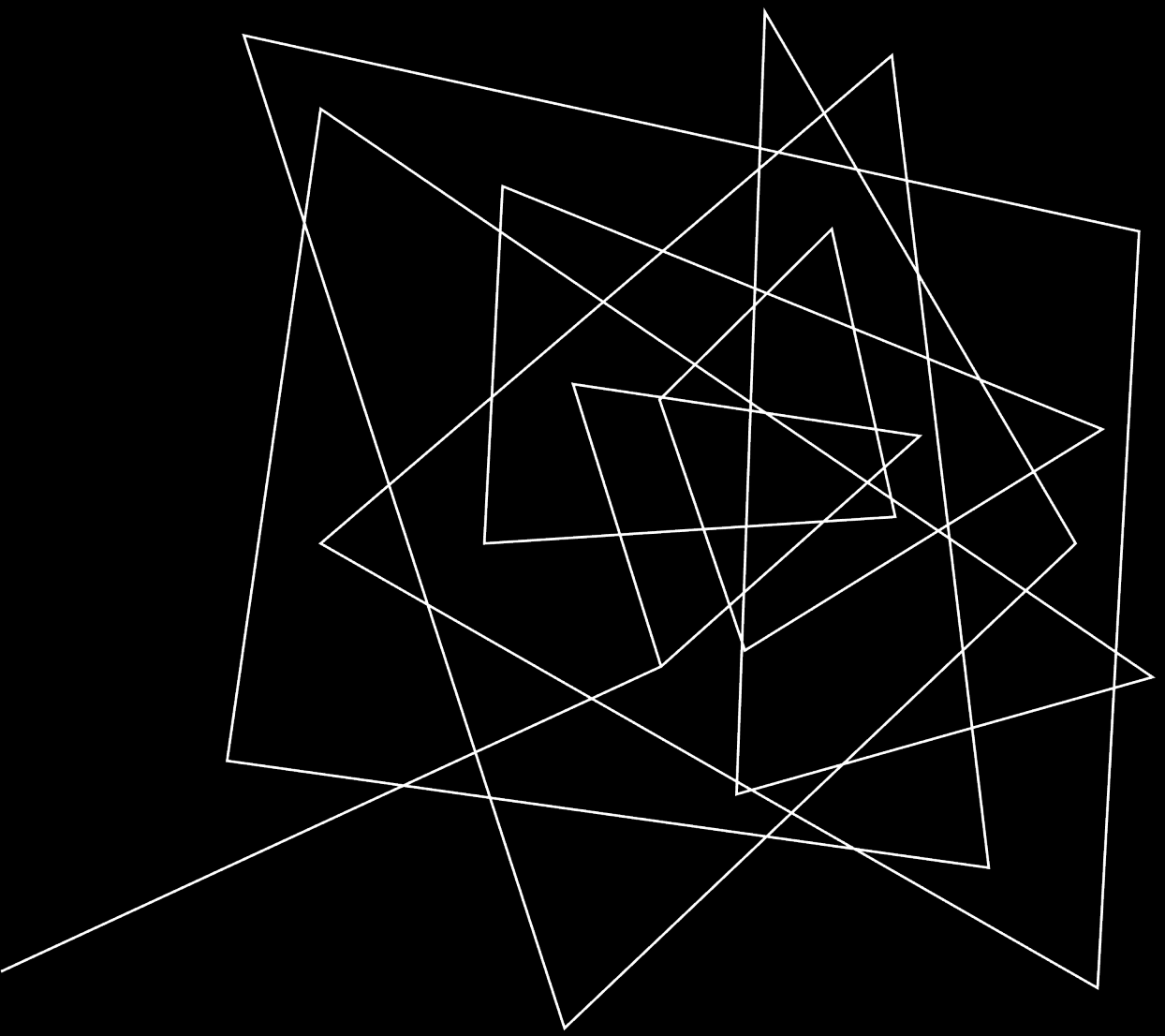
- Written Proposal.
  - Submitted as a group.
  - Must use the LaTeX template.
- Schedule.
  - Submitted as a group.
  - No required format.
- Elevator Pitch.
  - Submitted individually.
  - Must present dry run during your group's Week 9/10 TA Check-In.
  - Must create YouTube video of final pitch.

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



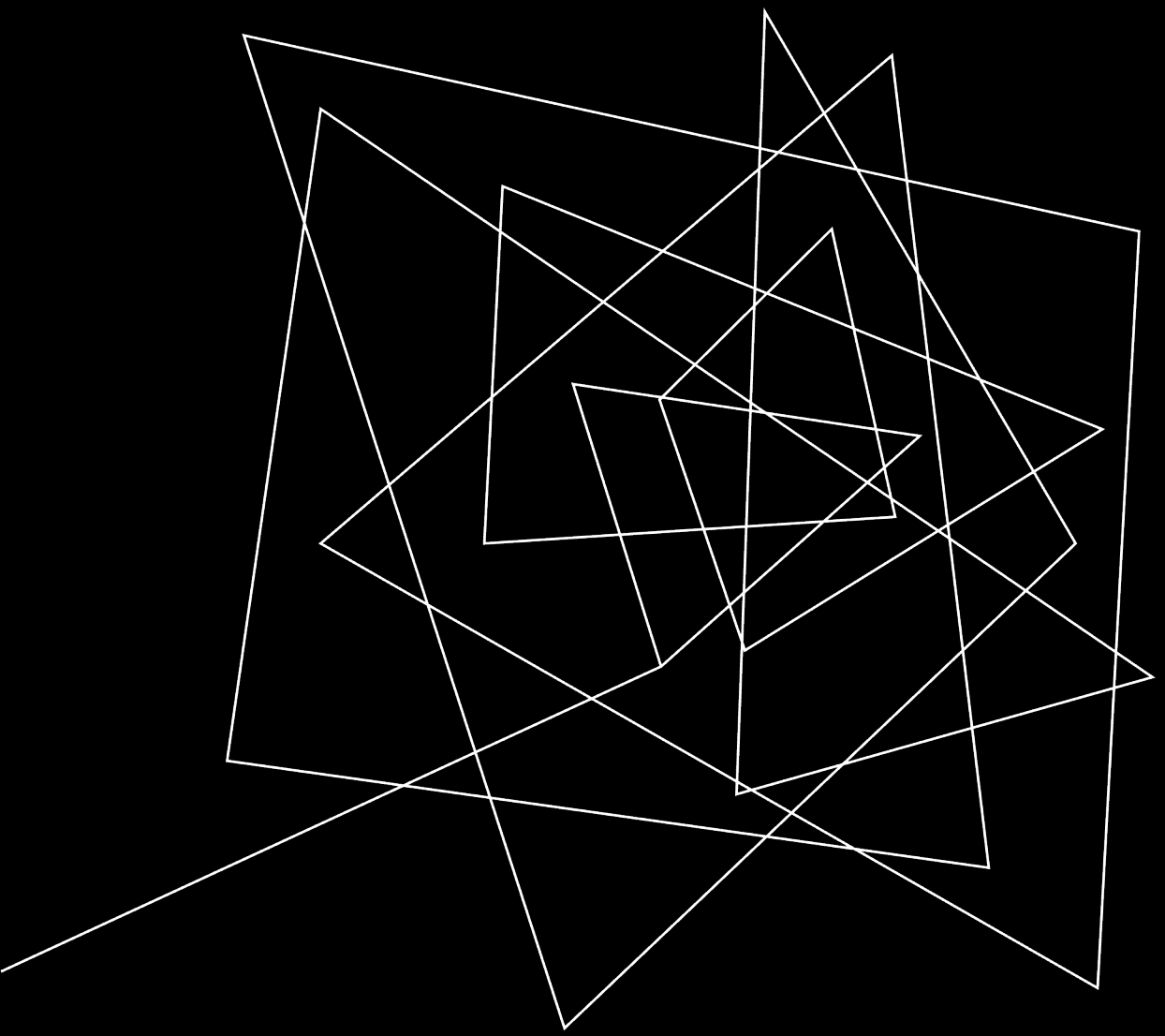
READING PRESENTATION:

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 mistaken belief that we can make great progress on alleviating complex dilemmas, if not remedy them entirely, by reducing their core issues to simpler engineering problems. It is seductive for three reasons. First, it's psychologically reassuring. It feels good to believe that in a complicated world, tough challenges can be met easily and straightforwardly. Second, technological solutionism is financially enticing. It promises an affordable, if not cheap, silver bullet in a world with limited resources for tackling many pressing problems. Third, technological solutionism reinforces optimism about innovation—particularly the technocratic idea that engineering approaches to problem-solving are more effective than alternatives that have social and political dimensions.

- "band aid" for broader problems

TRADEOFFS, CHALLENGES, & LIMITATIONS TO FAIRNESS IN AI

1. Many ways to define and measure fairness, and in turn many ways to promote fairness
2. Assumption that tradeoffs exist between fairness and accuracy
3. 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."*
   → Quantifiably fair(er) algorithmic decision-making is not always the same as fair use of AI
4. Prioritizing quantifiably-fair algorithms can obscure bigger, more challenging questions about their use

Does this make COMPAS fair?

“ZOOMING OUT”

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

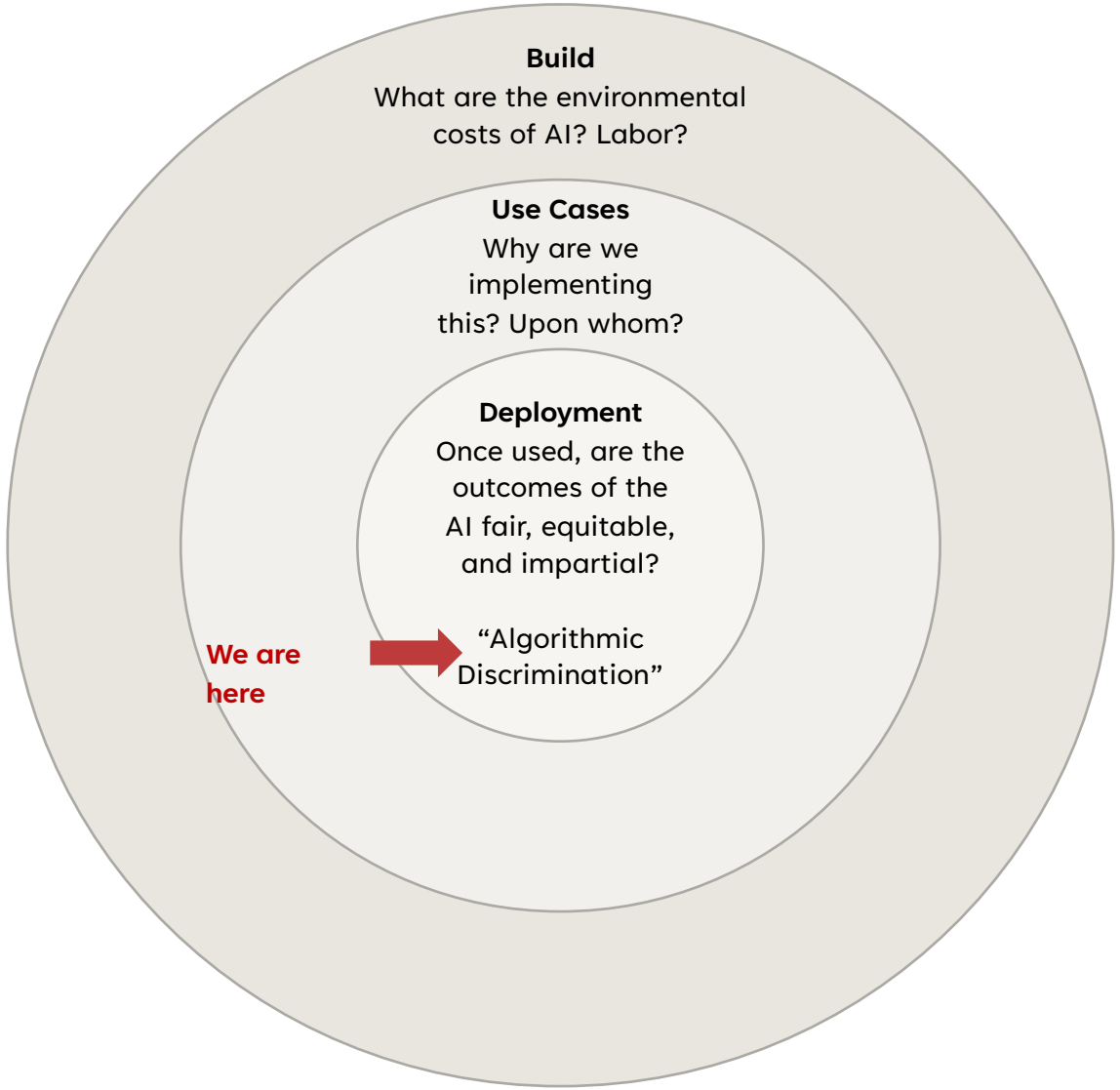
The New York Times

I Found Work on an Amazon Website. I Made 97 Cents an Hour.

By ANDY NEWMAN

NOV. 15, 2019

Inside the weird, wild, low-wage world of Mechanical Turk.



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

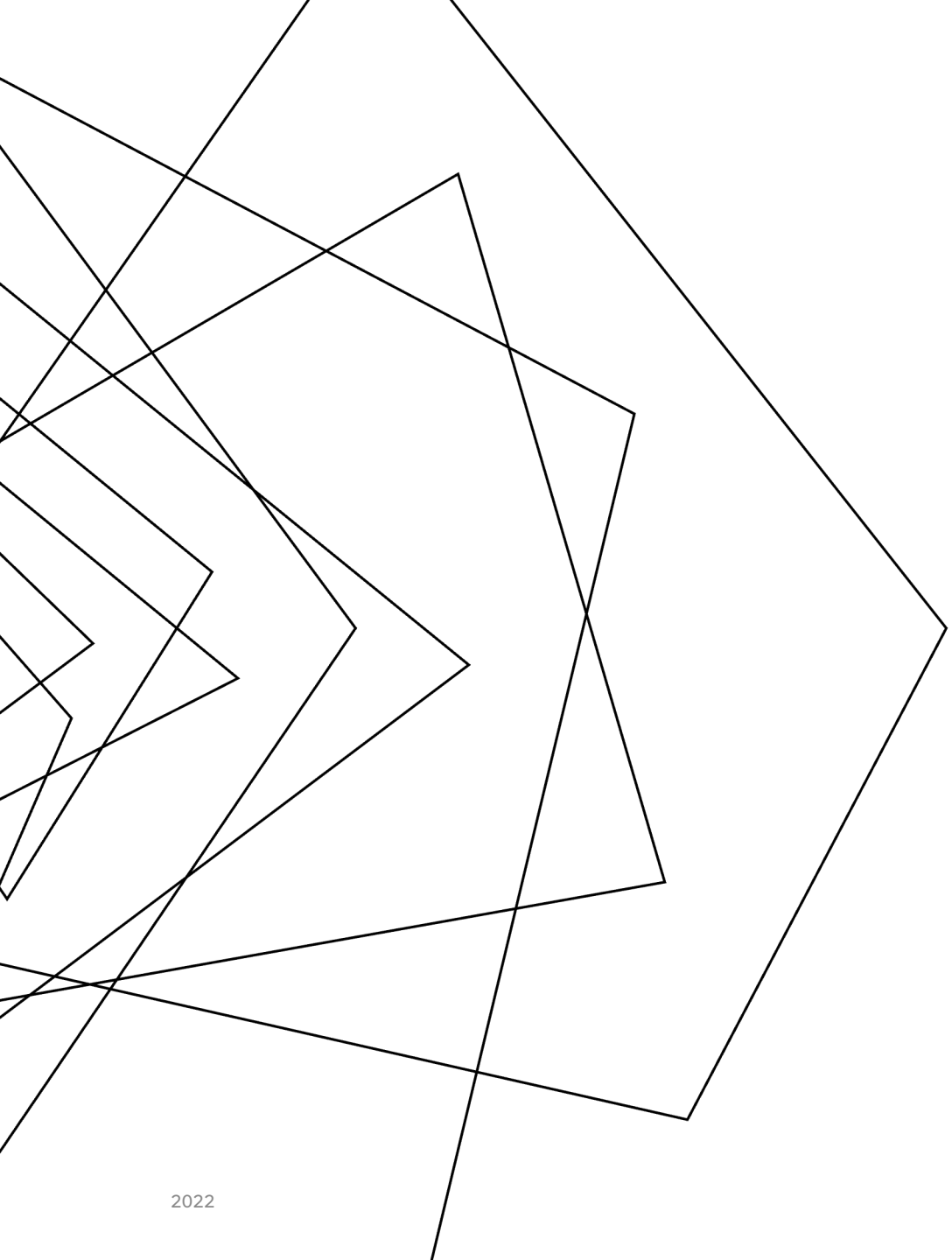
Sources:
https://undark.org/2020/06/01/michigan-unemployment-fraud-algorithm/
https://www2.deloitte.com/us/en/insights/deloitte-review/issue-18/behavior-change-among-unemployment-claimants-behavioral-economics.html

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



DISCUSSION

Think about the ways you measured fairness during the Q1 project. Are there any potential fairness issues that you think can be solved through the pre-, in-, and post-processing techniques we've previously discussed? Can any *not* be solved through technical solutions?

Are there any non-technical interventions that you think are necessary for anyone trying to implement policies based on a utilization model?

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)