THINKING RESPONSIBLY ABOUT AI SYSTEMS

ELIZA WELLS, EMBEDDED ETHI**CS**
11/29/21

WHO AM I?

- Eliza Wells, elizaw@mit.edu
- Philosophy PhD student at MIT
- Research interests: moral and social philosophy
- Graduate Fellow for Embedded EthiCS @ Harvard

SMALL GROUP DISCUSSIONS TODAY!

Take a moment to make sure you're sitting near 2 or 3 people to talk to.

Share your **name** and **favorite TV show** with each other.

BEING RESPONSIBLE

Negative Responsibility: who should be blamed when things go wrong?

> *"I didn't make that decision, the AI did!"*

BEING RESPONSIBLE

Negative Responsibility: who should be blamed when things go wrong?

> *"I didn't make that decision, the AI did!"*

Positive Responsibility: how can I be aware of the impacts of my decisions?

POSITIVE RESPONSIBILITY

"Computing professionals' actions change the world. To act responsibly, they should reflect upon the wider impacts of their work, consistently supporting the public good."

- *Association for Computing Machinery (ACM) Code of Ethics and Professional Conduct*

GOALS FOR TODAY

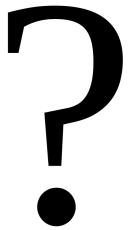
1. Cultivate positive responsibility by introducing tools for ethical decision-making
2. Understand the ethical lenses of *benefits/harms*, *respect*, and *justice*
3. Consider different levels of intervention into AI systems
4. Apply these tools to case studies

THINKING RESPONSIBLY

1. Who will be impacted by this system?
2. How will they be impacted?
    a. *Benefits/harms*
    b. *Respect*
    c. *Justice*
3. What technical choices influence these impacts?
    a. *Data*
    b. *Design*
    c. *Deployment*

CASE STUDY: THE ALLEGHENY FAMILY SCREENING TOOL (AFST)

CASE STUDY: THE ALLEGHENY FAMILY SCREENING TOOL (AFST)



1.
Potential incident of concern

2.
Someone calls CYF to report that a child is experiencing potential abuse or neglect

3.
A screener takes the call, collects information, and decides whether to…

4a.
Screen out the call and file the information away

4b.
Send someone to investigate

CASE STUDY: THE ALLEGHENY FAMILY SCREENING TOOL (AFST)

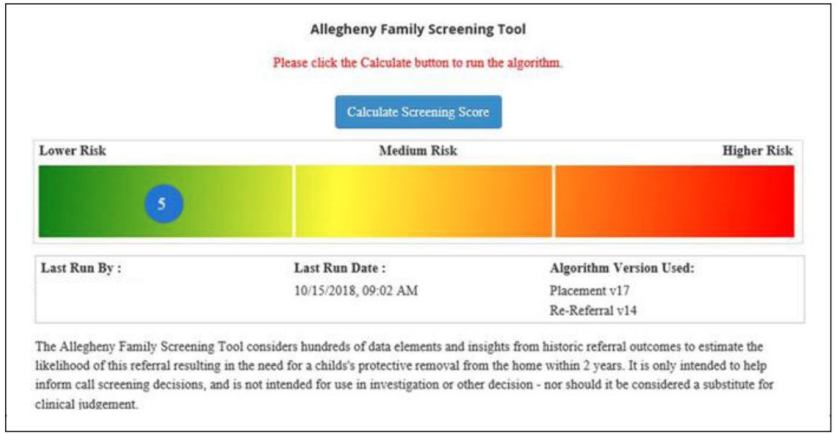
- Predictive ML algorithm designed to assess risk of child abuse or neglect to determine whether further investigation is needed
- Aims to improve human assessment in call screening (#3)
  - Limited resources and overwhelming workload
  - Inability to access all relevant information and process it quickly
  - Idiosyncratic and biased decisions

CASE STUDY: THE ALLEGHENY FAMILY SCREENING TOOL (AFST)

Output based on two predictions:

- Whether the child would be re-referred within two years
- Whether the child will be removed from their home within two years

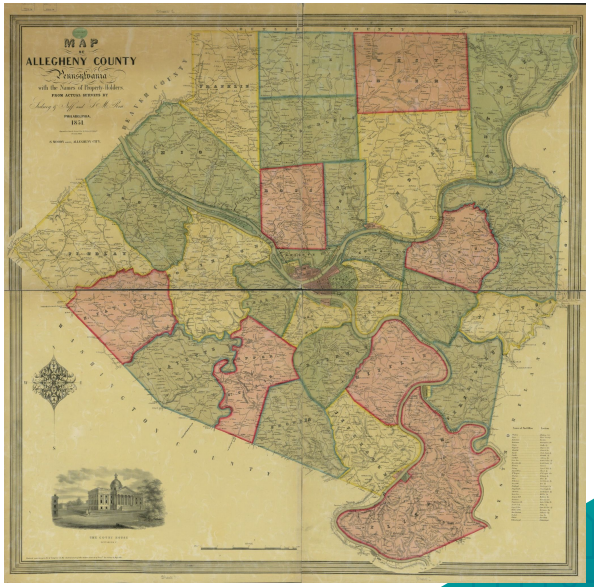
Provides a single risk score that is the highest of those two predictions for all children involved in the call.



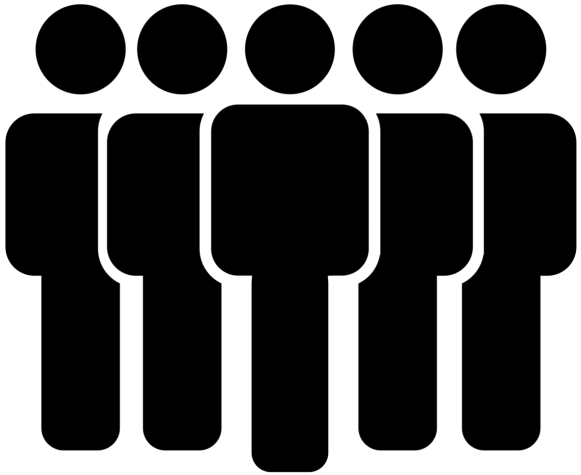
CASE STUDY: THE ALLEGHENY FAMILY SCREENING TOOL (AFST)

Data sources used:

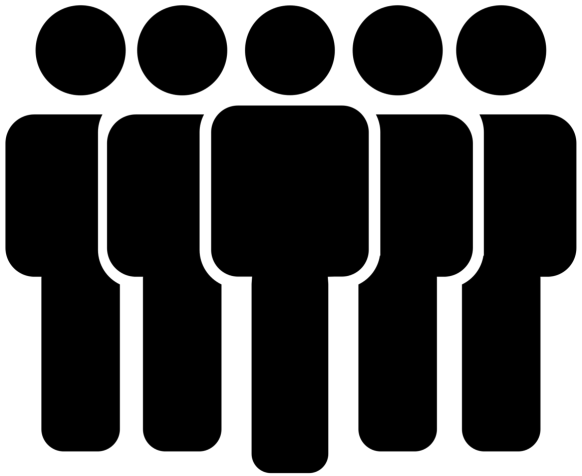
- Child welfare records
- Jail records
- Juvenile probation records
- Behavioral health records
- Public benefit records



WHO WILL BE IMPACTED BY THIS SYSTEM?



WHO WILL BE IMPACTED BY THIS SYSTEM?

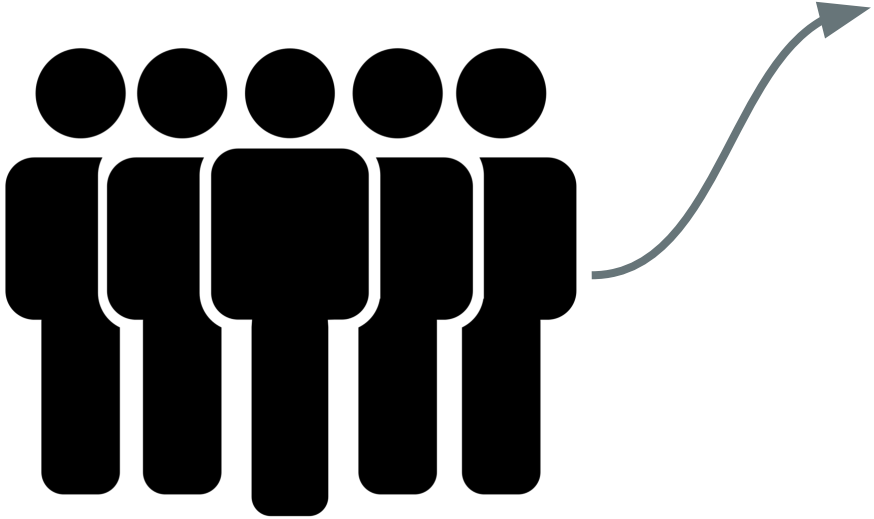


"A computing professional should...

1.1 Contribute to society and to human well-being, acknowledging that **all people are stakeholders** in computing."

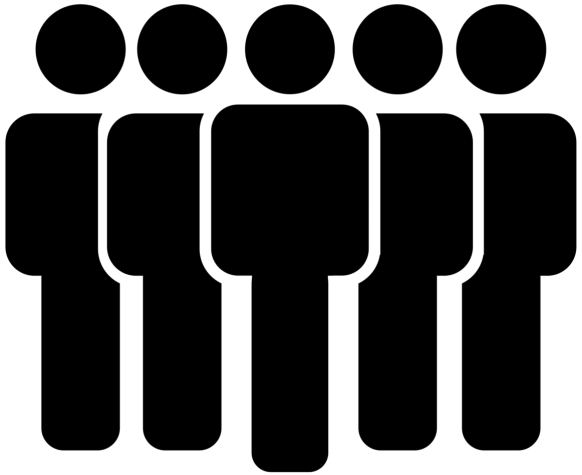
- *ACM Code of Ethics and Professional Conduct*

HOW WILL THEY BE IMPACTED?



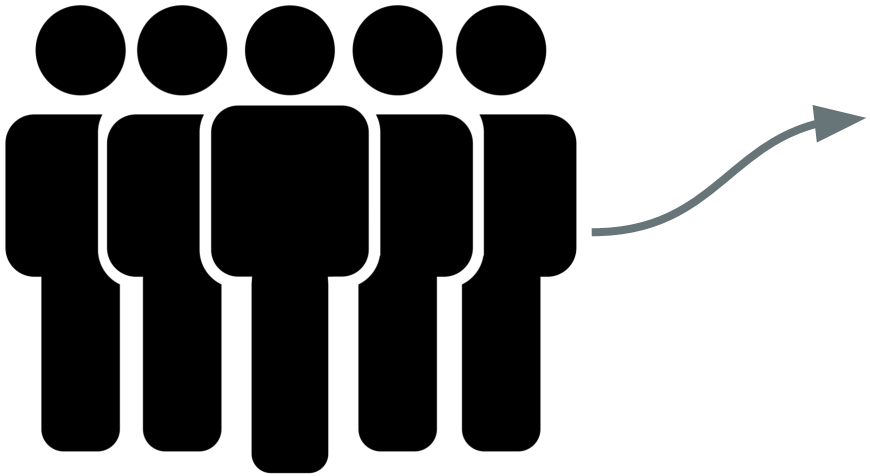
1. **Benefits/harms**
   - What are the potential consequences of this system for each stakeholder?

HOW WILL THEY BE IMPACTED?



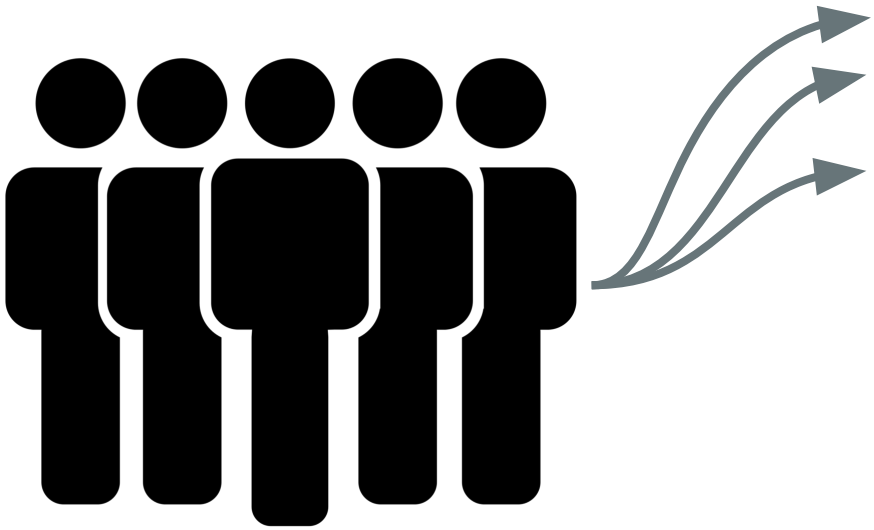
1. **Benefits/harms**
2. **Respect**
   - How does this system show respect for each stakeholder's *autonomy* (think: transparency, consent, control, etc.)?

HOW WILL THEY BE IMPACTED?



1. **Benefits/harms**
2. **Respect**
3. **Justice**
   - Does this process treat each stakeholder fairly? Does this process lead to fair outcomes?

HOW WILL THEY BE IMPACTED?



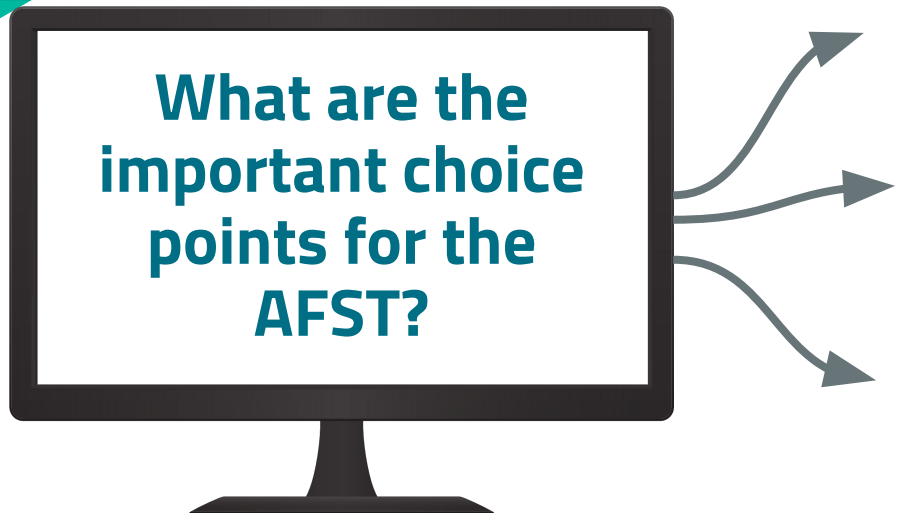
1. **Benefits/harms**
2. **Respect**
3. **Justice**

Sometimes, we can't achieve everything for every stakeholder.

WHAT TECHNICAL CHOICES INFLUENCE THESE IMPACTS?

1. **Data**
   - Which data sources am I using?
2. **Design**
   - How am I defining my objectives? How did I build the model?
3. **Deployment**
   - How will users interact with my system?

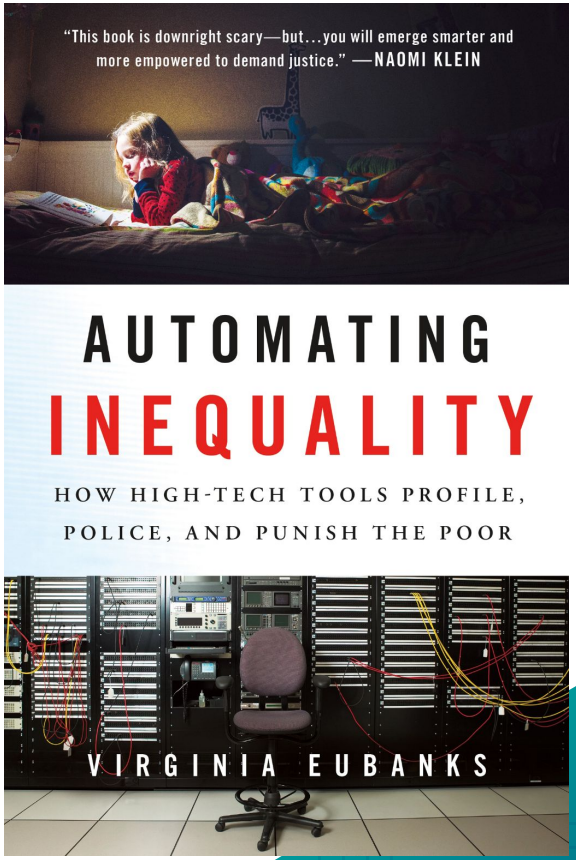
WHAT TECHNICAL CHOICES INFLUENCE THESE IMPACTS?



1. **Data**
   - Which data sources am I using?
2. **Design**
   - How am I defining my objectives? How did I build the model?
3. **Deployment**
   - How will users interact with my system?

EVALUATING THE AFST

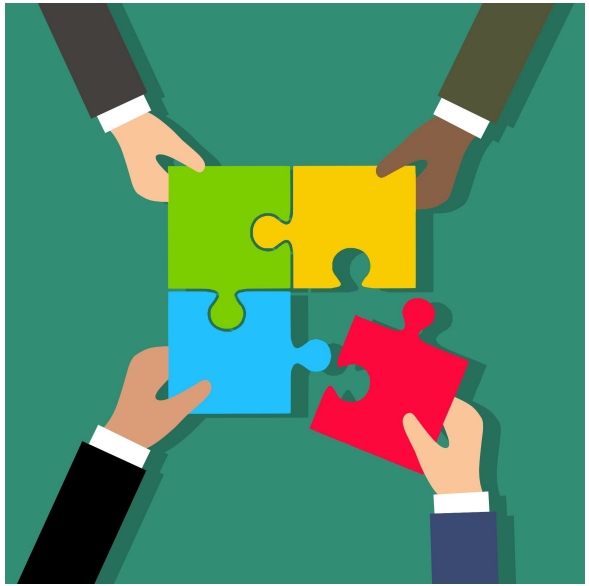
"Once the big blue button is clicked and the AFST runs, it manifests a thousand invisible human choices. But it does so under a cloak of evidence-based objectivity and infallibility."



TAKING RESPONSIBILITY

"3.7 Recognize and take special care of systems that become integrated into the infrastructure of society."

- *ACM Code of Ethics*



AFST VERSION 2

- **Data**: No longer uses public benefits records as a data source
- **Design**: No longer predicts re-referrals, only out-of-home placement
- **Deployment:** Defaults to automatically screening out low predicted-risk calls and screening in high predicted-risk calls, with the option for human screeners to override

STEPPING BACK

- Another potential choice point is **do or don't.** Are there situations where we should not use AI?
- Are there other questions we should be asking?

TAKEAWAYS

1. Being a responsible computer scientist is about more than avoiding blame; it is about cultivating positive responsibility for your decisions.
2. Thinking ethically involves considering the different people who can be affected by your decisions and the different ways they can be affected.
3. Technical choices at all levels can make an ethical difference.

THANK YOU!

Evaluation: **https://tinyurl.com/CS182F21**

Contact: **elizaw@mit.edu**