

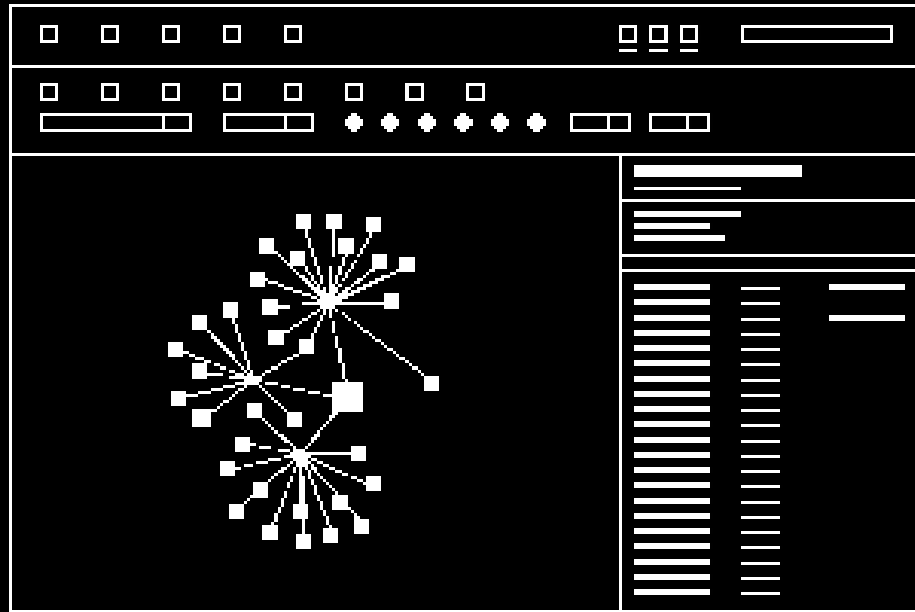
The Limits of Predictive Justice

Courtney Bowman
Global Director, Privacy & Civil Liberties Engineering
Palantir Technologies

3rd Webinar on AI: Predictive Justice
November 18, 2021



WE BUILD PLATFORMS FOR HUMAN-CENTERED, DATA-DRIVEN DECISION-MAKING



Gotham



Foundry

WE DON'T BUILD PREDICTIVE ALGORITHMS. WE PROVIDE PLATFORMS THAT ENABLE ANALYTICS.

The Gotham interface displays a network graph on the left with nodes for Judge Zeiman, Bob Purple, Catherine Blue, Daniel Violet, Timothy Liu, and Officer Alice Green, connected by relationships like 'Supported By', 'Victim Of', and 'Participated In'. On the right, a 'Histogram' shows 'Entity Types' and 'Entity Properties'. Below the graph is a table with columns for 'Service', 'Client Name', and 'Results 1-100 of 7,743,344'. A 'Timeline' chart at the bottom shows data from January to April.

The Foundry interface shows a data graph at the top with nodes and edges representing data relationships. Below the graph is a file list table with columns for 'Name', 'Last updated', and 'Type'. The table lists files such as 'Unclaimed State Documents', 'Product Usage Data', and 'Temp Cleanup'.



The Foundry interface shows a code editor with Java code for 'MyFileReader.java'. The code includes package declarations, imports, and a main method that reads a file line by line and prints the number of tokens in each line.

Gotham

Foundry

CONSEQUENTIAL & RESPONSIBLE AI BEGINS WITH THRESHOLD QUESTIONS

Practical Question:

What is the technology capable of and (more critically) not capable of doing?

Ethical & Normative Question:

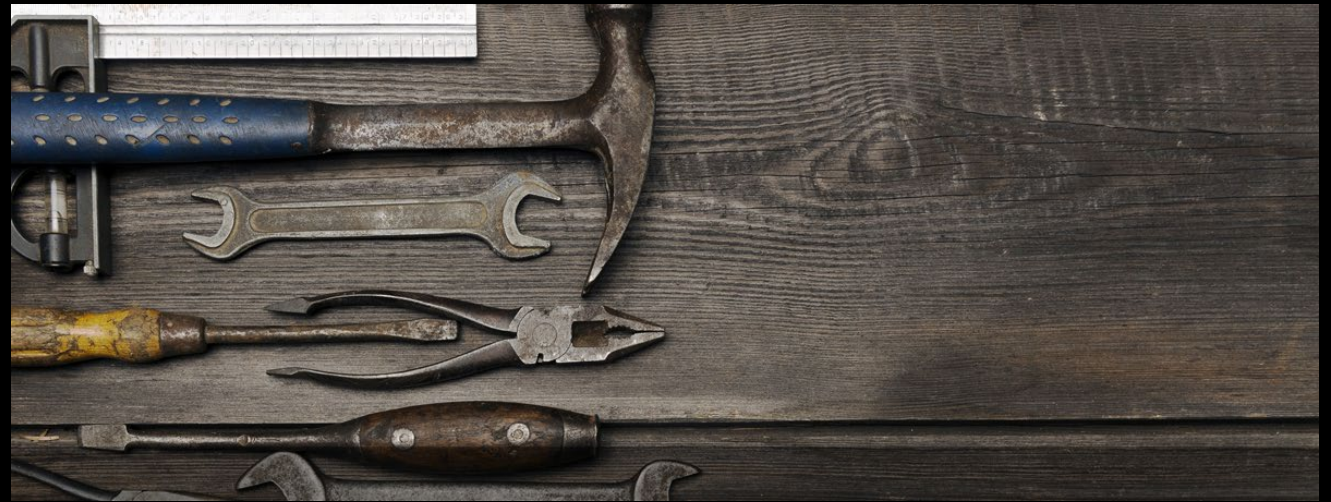
What should the technology be permitted to do?

PRACTICAL LIMITS & CATEGORICAL CONSTRAINTS

Remember: computational “intelligence” is not intelligence as such

AI is a tool, not a panacea

Real-world empirical results matter





ETHICAL & NORMATIVE CONSIDERATIONS

Values help orient the application of science and technology

Awareness of cultural and social contexts align outcomes with community interests

Institutional standards and practices often exist for good reason

MOTIVATIONS FOR PREDICTIVE JUSTICE VERSUS LIMITATIONS

Motivations:

- Efficiency
- Scale
- Consistency
- Addressing human & systemic bias
- Making the fuzzy more concrete

Limitations:

- Computational processing is not human understanding
- Data & Machine bias
- Algorithmic opacity
- Conceptual irreducibility



CATEGORICAL DISTINCTIONS

Algorithms seek to trade computational efficiency for human expertise and systemic methodologies

In so doing, the administration of Justice becomes more focused on optimizing (distributive) outcomes according to metrics or other goals

But Justice in the sense we care most about may be just as much – or more – about the *Procedural* as it is about the *Distribution of desired outcomes*

Choosing to focus on Outcomes should be intentional and mindful of the implied tradeoffs.

PREDICTIVE JUSTICE APPLICATIONS NEED GROUNDING IN OPERATIONAL CONSIDERATIONS

“Operational AI” is:

- Embedded in the world, not playing in a laboratory
- Focused on the full data lifecycle, not just fairness metrics. This includes:
 - Provenance and lineage
 - Version control
 - Environment tracking
 - Testing and evaluation
 - Audit trails
 - First-order model and systems maintenance
- Respectful of end-user interactions
- Understanding of trade-offs and limits
- *Mindful of the key Principles of Responsible AI...*



GUIDING PRINCIPLES FOR RESPONSIBLE AI

- 1. Don't solve problems that shouldn't be solved.**
- 2. Collect and Select Data Responsibly.**
- 3. Methodically Assess and Address Sample Bias.**
- 4. Carefully Monitor Outcomes, Understand Equity Assessment Tradeoffs.**
- 5. Ensure Auditability, Explainability, and Interpretability.**
- 6. Wherever Possible, Keep Humans in the Loop.**
- 7. Promote Multi-Stakeholder Engagement.**



GUIDING PRINCIPLES FOR RESPONSIBLE AI

1. Don't solve problems that shouldn't be solved.
- 2. Collect and Select Data Responsibly.**
3. Methodically Assess and Address Sample Bias.
4. Carefully Monitor Outcomes, Understand Equity Assessment Tradeoffs.
5. Ensure Auditability, Explainability, and Interpretability.
6. Wherever Possible, Keep Humans in the Loop.
7. Promote Multi-Stakeholder Engagement.



GUIDING PRINCIPLES FOR RESPONSIBLE AI

1. Don't solve problems that shouldn't be solved.
2. Collect and Select Data Responsibly.
- 3. Methodically Assess and Address Sample Bias.**
4. Carefully Monitor Outcomes, Understand Equity Assessment Tradeoffs.
5. Ensure Auditability, Explainability, and Interpretability.
6. Wherever Possible, Keep Humans in the Loop.
7. Promote Multi-Stakeholder Engagement.



GUIDING PRINCIPLES FOR RESPONSIBLE AI

1. Don't solve problems that shouldn't be solved.
2. Collect and Select Data Responsibly.
3. Methodically Assess and Address Sample Bias.
- 4. Carefully Monitor Outcomes, Understand Equity Assessment Tradeoffs.**
5. Ensure Auditability, Explainability, and Interpretability.
6. Wherever Possible, Keep Humans in the Loop.
7. Promote Multi-Stakeholder Engagement.



GUIDING PRINCIPLES FOR RESPONSIBLE AI

1. Don't solve problems that shouldn't be solved.
2. Collect and Select Data Responsibly.
3. Methodically Assess and Address Sample Bias.
4. Carefully Monitor Outcomes, Understand Equity Assessment Tradeoffs.
- 5. Ensure Auditability, Explainability, and Interpretability.**
6. Wherever Possible, Keep Humans in the Loop.
7. Promote Multi-Stakeholder Engagement.



GUIDING PRINCIPLES FOR RESPONSIBLE AI

1. Don't solve problems that shouldn't be solved.
2. Collect and Select Data Responsibly.
3. Methodically Assess and Address Sample Bias.
4. Carefully Monitor Outcomes, Understand Equity Assessment Tradeoffs.
5. Ensure Auditability, Explainability, and Interpretability.
- 6. Wherever Possible, Keep Humans in the Loop.**
7. Promote Multi-Stakeholder Engagement.



GUIDING PRINCIPLES FOR RESPONSIBLE AI

1. Don't solve problems that shouldn't be solved.
2. Collect and Select Data Responsibly.
3. Methodically Assess and Address Sample Bias.
4. Carefully Monitor Outcomes, Understand Equity Assessment Tradeoffs.
5. Ensure Auditability, Explainability, and Interpretability.
6. Wherever Possible, Keep Humans in the Loop.
- 7. Promote Multi-Stakeholder Engagement.**



PUTTING IT ALL TOGETHER

Justice may not be the right place for predictive / AI interventions.

If the application passes the threshold assessments, tradeoffs must be understood and accounted for.

Consequential AI is grounded in sound operational practices and guiding principles that reinforce responsibility and keep humans at the center of systems that have impact on real lives.



Thank you!

Courtney Bowman
cbowman@palantir.com