

# AUTOMATED DECISION MAKING: THE GOALS OF EXPLANATION AND TRANSPARENCY

Andrew Selbst, UCLA School of Law

CCPA Informational Workshop

March 30, 2022

# OUTLINE

- Who are the explanations for?
- What are the explanations for?
- What kind of explanations?
- Specific issues that explanations can address

WHO ARE EXPLANATIONS FOR?

# WHO ARE EXPLANATIONS FOR?

- Developers
- Consumers (Users)
- Consumers (Affected Non-Users)
- Regulators

WHAT DO EXPLANATIONS  
ACCOMPLISH?

# FOR DEVELOPERS?

- Documentation
  - Debugging
  - Coordination
  - Transitions during turnover
- Interpretability
  - Validation
  - Debugging

## FOR CONSUMERS (USERS)?

- Making sure they understand what they are using and how it works
- Consumer Protection: Did they get what they thought they were buying?
- Liability: Are they doing something that will subject them to liability?

## FOR CONSUMERS (AFFECTED)?

- Most commonly discussed explainability concerns
- Procedural Justice / Intrinsic Value of Explanation
- Contestability / Due Process
- Enabling Future Success



## FOR REGULATORS?

- Anything need for compliance oversight
- Functionality
- Testing for social and legal impacts
- Future policy learning

WHAT KIND OF EXPLANATIONS?

# DIFFERENT KINDS OF TRANSPARENCY & EXPLANATION

## FOCUS ON EXISTING MODELS

- Model Explanations
- Outcome Explanations
- Interactive “Explanations”

## FOCUS ON MODEL DEVELOPMENT

- Documentation
- Impact Assessments
- Audits

# EXPLANATIONS OF EXISTING MODELS

## OUTCOME EXPLANATIONS

- Targeted at affected consumers
- Can enable future actions or appeals
- Dignity-enhancing
- Often underspecified and easily manipulated

## MODEL EXPLANATIONS

- “Model of a model”
- Can help explain the gist of localized results
- Often too complicated to understand

## INTERACTIVE “EXPLANATIONS”

- Consumer tweaks inputs to see hypothetical results
- Generates intuitive feel
- Can be misleading

# EXPLANATIONS OF EXISTING MODELS

- FCRA/ECOA (Adverse Action Notices)
- GDPR
  - Art 22 requires “safeguards” for the automated processing, including human intervention & contestation
  - Art. 13–15 requires “meaningful information about the logic involved” in automated processing

# EXPLANATIONS OF MODEL DEVELOPMENT

## DOCUMENTATION

- Standard business practice in engineering
- Allows coordination and handoff to future employees
- Describe limitations of product, failure modes.

## IMPACT ASSESSMENTS

- Document important decisions
- Predict future social impacts
- Ongoing assessment, starts before project

## AUDITS

- Can be internal/external, mechanical/regulatory.
- Often top-down, looking for specific compliance answers

**SPECIFIC ISSUES THAT  
DOCUMENTATION CAN HELP  
ADDRESS**

## ISSUES TO LOOK OUT FOR

- Does it even work?
- Divisions of Responsibility
- Five “Abstraction Traps”
  - Framing Trap
  - Portability Trap
  - Formalism Trap
  - Ripple Effect Trap
  - Solutionism Trap