IMPROVING FAIRNESS IN MACHINE LEARNING SYSTEMS

DS 7406 ML Systems

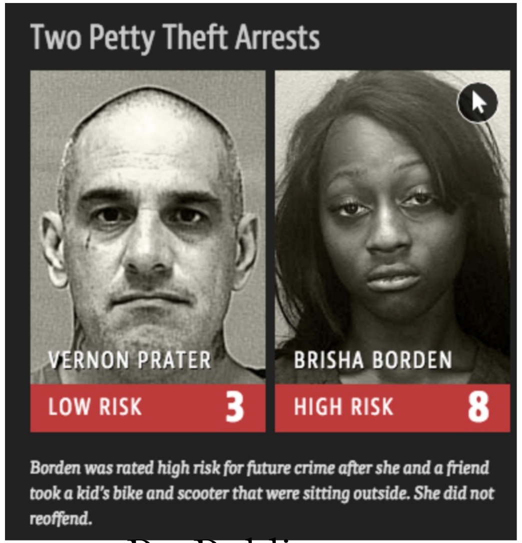
Karolina Naranjo-Velasco

August 31, 2022

Contents of the Presentation

Context

- ML influences every facet of our lives
- Academia vs. Industry
- Challenges of commercial ML product teams
- Human-centered computing: implications



ProPublica

Problem

To support in development fairer
ML systems in industry practice

Monitoring for unfairness and
taking appropriate action

SOLUTION

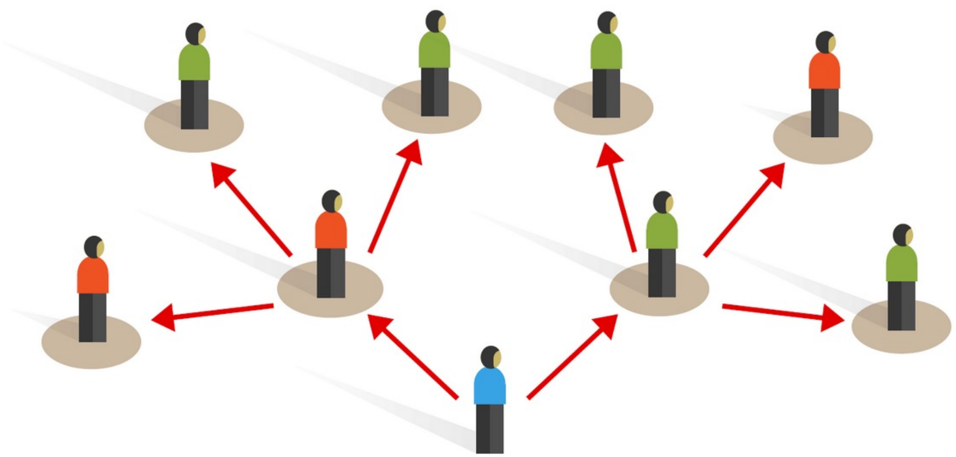
- Identify areas of alignment and disconnect between the practice and the solutions proposed in the fair ML research literature
- HCI research to address practitioners' needs
- Directions for future ML

The Metrics

INTERVIEW STUDY

35 semi-structured interviews to investigate exiting team's practices and challenges around fairness in ML and real-world needs

AI/ML systems perform differently for diverse groups in ways that may be considered undesirable

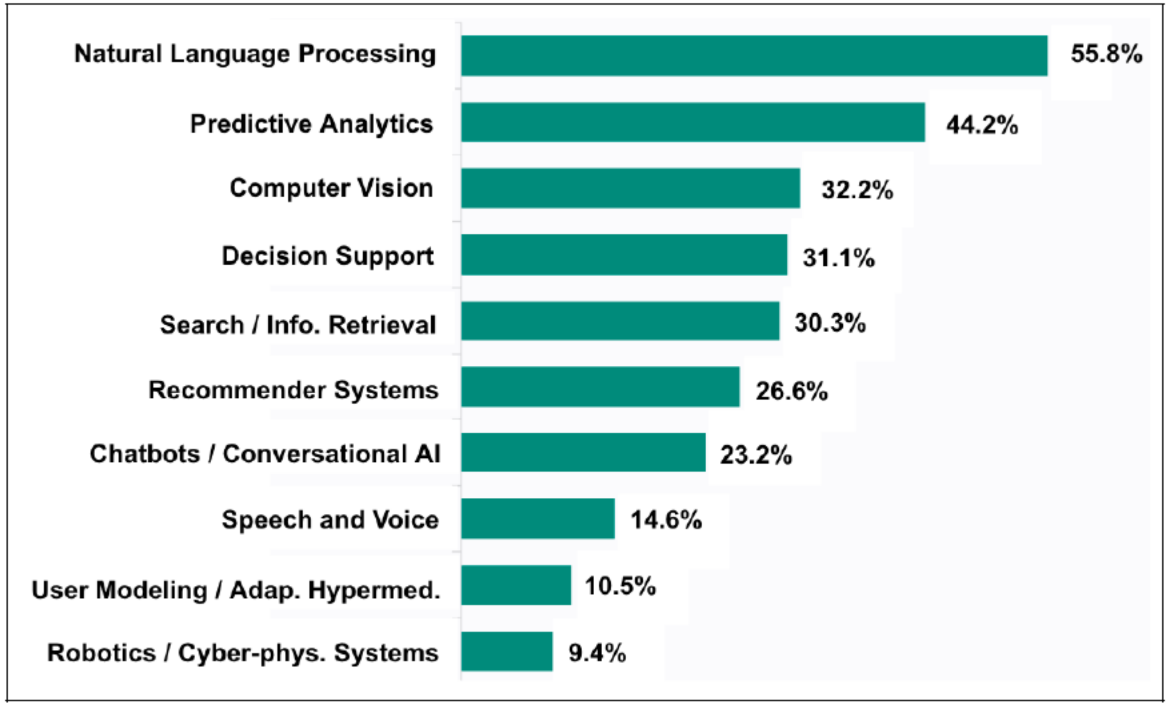


Interview Demographics

Technology Area	Roles of Interviewees
Adaptive Tutoring & Mentoring	Chief Data Scientist, CTO, Data Scientist, Research Scientist
Chatbots	CEO, Product Mgr., UX Researcher
Vision & Multimodal Sensing	CTO, ML Engineer, Product Mgr., Software Engineer
General-purpose ML (e.g., APIs)	Chief Architect, Director of ML, Product Mgr.
NLP (e.g., Speech, Translation)	Data Mgr., Data Collector, Domain Expert, ML Engineer, Product Mgr., Research Software Eng., Technical Mgr., UX Designer
Recommender Systems	Chief Data Scientist, Data Scientist, Head of Diversity Analytics
Web Search	Product Mgr.

The Metrics

Survey



- Quantitative supplement to the interviews
- 267 industry ML practitioners
- Understand the respondent's backgrounds: technology areas and team roles
- Practices, challenges and needs for support around fairness in their team's ML development pipelines

KEY INSIGHTS CONTRIBUTIONS

Fairness-aware Data Collection

Fixed training datasets? Practitioners have some control over data collection and curation to intervene and improve fairness.

Focus on improving ML model's overall predictive accuracy rather than fairness

Future: how tools can be designed to support fairness in ML models →

- data collection, curation, and augmentation
- Design and use of test sets

Proactive Auditing Process

- Ensure data collection data from the subpopulations
- Balance data across subpopulations in curating existing datasets
- Encryption mechanisms vs. technical solutions: stakeholders and policymakers
- Design processes and tools to support effective sharing and re-use of cultural and domain knowledge
- Support and use coarse-grained demographic information
- Explore and evaluate human inspection to fairness auditing: visualization and guided exploration
  - Fairness auditing:
    - specific issues
    - Teams: issues that need to be addressed

Effective Strategies to Address Detected Issues

Estimate

Estimate how much additional data to collect for a specific population

Trade

Trade-offs: definitions of fairness and ML systems (not limited to accuracy)

Assess and mitigate

Assess and mitigate biases in ML vs. Broader system designs

Explore

Explore educational resources and tools + Effective strategies

Biases in Humans and ML Development Pipeline

- Biases may be present: training data or user-study participants
- Explore ways to help teams better understand biases


QUESTIONS?