Overcoming Bias in Computer Vision – A Business Imperative

Stanford University
d.school

May 23, 2018

Bias Is Marring AI's Rollout

- Tay.AI chatbot becomes racist, sexist, homophobic (Microsoft)
- Voice-command systems in cars fails with women (Google)
- Criminal sentencing AI shows bias against African-Americans

Similar Problems For AI + CV

- Passport systems falsely ID Asians as eyes-closed
- Facial recognition AI ID's African-Americans as apes (Google, Flickr)
- Facial gender determination fails outright on women, minorities (HP, Microsoft)

AI-Powered CV Is Proliferating

- Retail - mood analysis in checkout lines (Walmart)
- Healthcare - medical training, surgical augmentation, pathology diagnosis
- Security - “anomaly” detection in crowds, sentiment detection in interrogation, bio-metric authentication
- Workforce - on-job safety, training with AI + AR headsets

Why Beating Bias in AI Matters

- People and users can be hurt.
- It can sink your product or company.
- Major value is left on the table.

Why It Happens

- All machine intelligence is built upon training data that was – at some point – created by people.
- Engineers don't need to be biased for their AI to be!
- "[Algorithms] replace human processes, but they're not held to the same standards. People trust them too much." - Katherine O'Neill

Why It Happens

Why it Happens

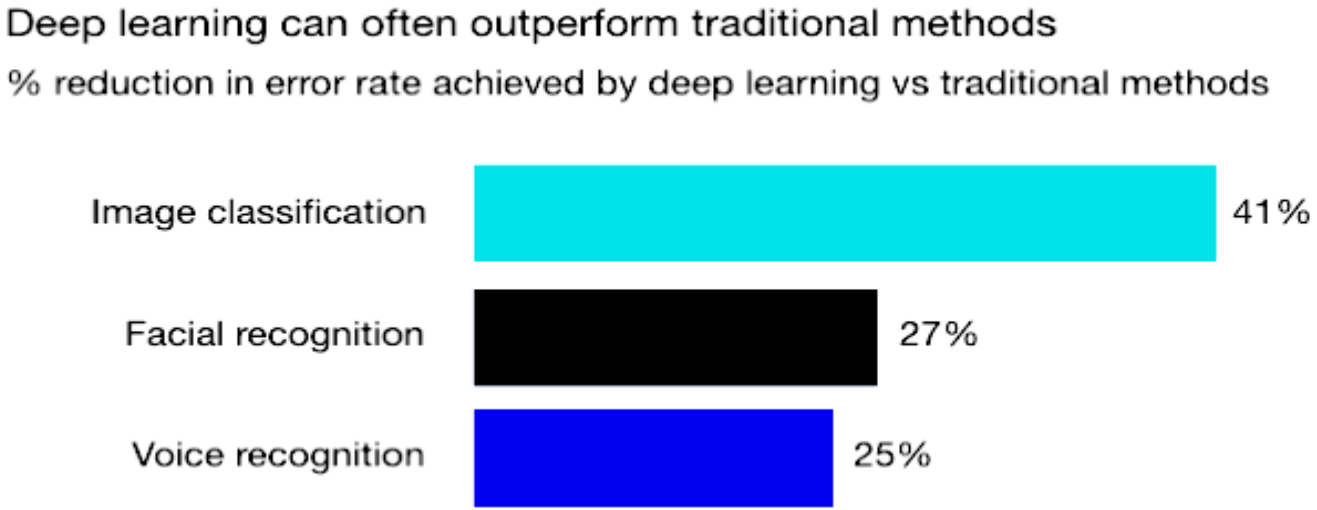
Why it Happens

AI Can Make Bias Worse

Imagine you're hiring your next software engineer.

Challenge: Explainability

- Deep learning, neural networks + IP concerns lead to “black box” algorithms.
- Deep learning has outsized impact in CV.



Source: McKinsey

Regulation Is Coming

If AI decisions aren't explainable,

- Who is accountable for mistakes?
- What is the recourse of a person denied a home loan by an AI system?
- How can a faulty behavior from AI be fixed if its root cause is unknown?

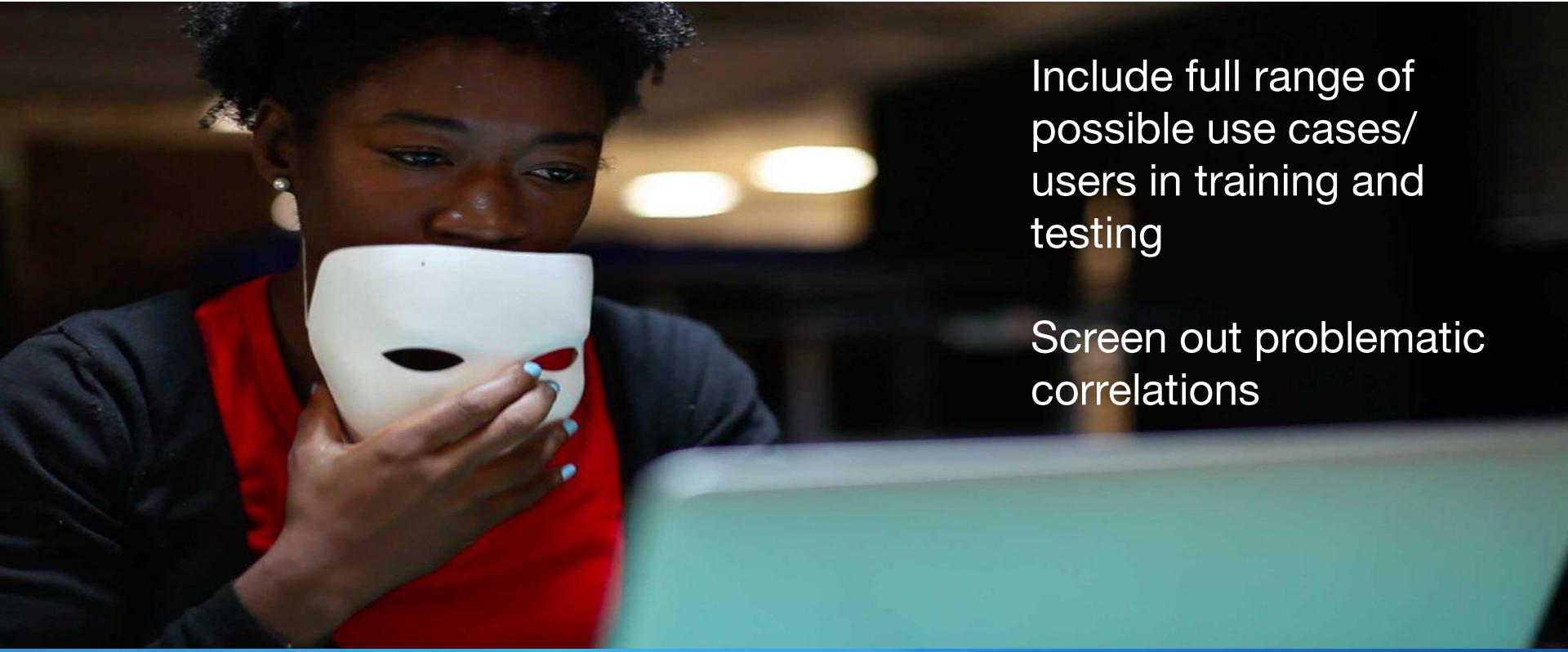
Policy is moving against black-box algorithms:

- EU's GDPR (don't be fooled; global reach!)
- NYC
- UK

Solution: Rigor in Training Data

Photo Credit: NOVA

Solution: Rigor in Training Data



Include full range of possible use cases/ users in training and testing

Screen out problematic correlations

Photo Credit: NOVA

Diversify Product Teams

By demographics, ie race, gender, but also by training: anthropology, psychology, social science.

Diverse teams:

- ID assumptions / spot data with unintentional bias
- Build for wider set of problems
- Create solutions that make more profit!

Build For Diverse Users = New Value

- US women dislike financial services, despite holding 39% of assets.
- Their lens on investment is caring for loved ones, security.
- Startup Joy, uses AI to engage women on the values that undergird financial goals, then informs personalized coaching.

Solutions: Explainability

- Establish transparency and ethics standards, use open-source tools where possible, internal fairness audits for algorithms.
- Stay up to date on emerging field and tools in explainable AI.
- Empower users with direct lines to your business – user feedback loops - to uncover problems early and attack them at their root.

Tap Users to Report and Debug

Set Transparency + Ethics Standards

Case Study: Orbital Insight

Clear, detailed ethics standards, listed front and center.

- “Respect for individual” as north star
- Aggregated and anonymized data only
- Partners may not track or monitor individuals, ethics screen

“AIX”: Frontiers of Explainable AI

- US DOD / DARPA investing heavily in AIX.
- “Counterfactual Fairness” and retraining against Bias

Solutions: Explainability

- “Counterfactual Fairness”: Would the same decision be reached if sensitive variable (e.g., race) was removed?
- Initial solutions have removed “sensitive variables” from training-sets altogether, but this is problematic.
- DeepMind - if sensitive variables lead to“unfair” pathways, they are identified and retrained.

Resources to Take Action

- OpenAI
  Free, open-source, explainable AI frameworks, code.

- NIST. National Institute Standards and Technology
  Bias-testing for image / facial recognition datasets.

- Ethics and Governance of AI Fund, AI Now, IEEE
  Where policymakers/ regulators are taking their cues.

Beating Bias, Business Imperative

- Algorithmic bias can hurt your users and sink your product
- You can protect your product with specific strategies
- In the process, you can unlock new value and profits

Resources

- OpenAI.com Tools: https://openai.com/systems/
- GDPR - https://www.eugdpr.org/
- World Economic Forum Center for Fourth Industrial Revolution: https://www.weforum.org/center-for-the-fourth-industrial-revolution
- NIST Face Recognition Vendor Test: https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt-ongoing
- IEEE Ethics in Autonomous and Intelligent Systems https://ethicsinaction.ieee.org/
- Ethics and Governance of AI Fund https://www.knightfoundation.org/aifund-faq
- DeepMind: Path-Specific Counterfactual Fairness https://arxiv.org/abs/1802.08139

Thank You

Questions, Feedback, or Collaboration Ideas?

Please be in touch!

will@dschool.stanford.edu

Appendix

FactMata which uses algorithms to beat the spread of disinformation/hate speech online.

They made its algorithms completely transparent to users - allowing for comment and correction from users. While this may not be possible for all, there are other steps to take.

Definitions: Bias

- Implicit Bias – unconscious attribution of qualities to a particular group
- Unconscious Bias - often against one own values, happens

Definitions: AI

- AI – machines that are characteristic of human intelligence
- Machine Learning – computers learn based off data and act without explicit programming.
- Deep Learning / Neural Networks - This wave of AI is no longer just inputting data, looking to training, and creating an output, but is rather creating its own new correlations, much like the human brain, in order to arrive at decisions.

Examples: CV+AI for Underserved

- Senior Care: Aging in place example: ELLI.Q
- Poverty and Famine: Fighting disease and food waste
- Disabled: Computer vision that helps the blind “see”