2023 embedded VISION SUMMIT

Bias in Computer Vision —It's Bigger than Facial Recognition!

Susan Kennedy, PhD Assistant Professor of Philosophy Santa Clara University



The Optimistic View



INTELLIGENT MACHINES

Will Smart Machines Be Less Biased Than Humans?

Robert D Atkinson Information Technology And Innovation Foundation

September 19, 2016





A New Jersey man was accused of shoplifting and trying to hit an officer with a car. He is the third known Black man to be wrongfully arrested based on face recognition.



By <u>Kashmir Hill</u>

Published Dec. 29, 2020 Updated Jan. 6, 2021



embedded

Bias in CV: Looking Beyond Facial Recognition



Agriculture Plant Disease Detection

anta Clara

University

Defect_class Four Defect_class Two Defect_class Two Defect_class Two Defect_class One Defect_class Three

> Manufacturing Quality Inspection



Transportation Pothole Detection

embedded

Bias in CV: Looking Beyond Facial Recognition

Bias can pose an ethical challenge, even without sensitive data!



Agriculture Plant Disease Detection Defect_class Four Defect_class Two Defect_class Two Defect_class Two Defect_class One Defect_class Three

> Manufacturing Quality Inspection



Transportation Pothole Detection



embedded

Human *Subjects* Human *Impacts*

- Expand the ethical circle to take into account the full range of stakeholders
 - Who does it not work for?
 - What does it not work for?
 - When does it not work?





embedded

Mitigating Bias – Technical Solutions



- Toolkits
 - Google What-If
 - IBM Fairness 360
 - Microsoft FairLearn

Datapoint	Editor Performance	& Fairness Featu	res										250 datapoi	nts loaded 🎄 🕐
visualize ^			Binning X-Axis Count Binn Inference score × 10 (nc		ig Y-Axis Color By		or By Label By		Scatte	r X-Axis ault)	Scatter Y-Axis	. :		
Datapoints	O Partial dependence plot	5						interence in	<u>(</u>	cruuny	(uch	Juny	(deridant)	
Show nearest counterfactual datapoint L1 O L2 O				275	0.2	03	0.4	0.5	00	07	0.8	0.0	~	
Show similarity to	o selected datapoint (0.000	0.1	02	0.3	OA	0.5	0.6	07	0.8	0.9
Edit Datapoint 15	a			^										00.0
< > © []	∎ I≡ ≣ Q Sear	h features												
Feature name		Value(s)												
image/encoded		<u>e</u>												
5_o_Clock_Shado	w	No 5 o'clock shad												
Arched_Eyebrows	s	No arched eyebro												
Bags_Under_Eyes	s	No bags under eyes												(C) (C) (C)
Bald		Not baid												CO.
Bangs		(No bangs												ADE
Black_Hair		Not black hair											00	00
Blond_Hair		Not blond hair											0	000
Blurry		Not blurry		•	26	21.5				0			11	
Brown_Hair		Not brown hair		•									10 0 0	
Bushy_Eyebrows		Not bushy eyebro		•	1 F. 2 10	250				1				SEL
Eyeglasses		No eyeglasses			A TOC	310				(B) -			- 0	CGE
Goatee		No goatee			0000					-	1 Tak		00	000
Gray_Hair		No gray hair		•••	OVAC		S. 0	C X Y	-		0.	CC.		
nfer Datapoint 15	51			^	+ 1 / 1 / 1				00	Ó	1	100	A	
Run inference						5	AD	00	R	A.		eo	Colors	a labal
Run	Label	Score	Delta			20		02	(Q. 0	1	-0-0	0.1	Not smilin	e rabel
1	0 (Not smiling)	0.853					20	8 5		0	Phil Ch		 Smiling 	-



The Tip of the Iceberg



- Systematic errors stemming from bias in the datasets and algorithmic processes used _____
- Human bias present across the AI lifecycle and in the use of AI once deployed
- Present in the datasets used in AI, and the institutional norms and practices across the AI lifecycle and in broader society









• Bias-free AI is an unachievable goal

→ Mitigating bias requires a **bias-aware** approach

• AI exists within a larger social system

→ Mitigating bias requires a **sociotechnical** approach



Responsible AI – Ethical Pre-Mortems

- Instead of waiting for bias to strike, build a habit of anticipating preventable causes so they can be mitigated
- Across the entire AI lifecycle From pre-design to deployment
- 3 key problem areas:
 - Datasets

anta Clara

- Testing and evaluation
- Human factors



embedded

SUMMI

Strategies to Employ



Dataset	 Interrogating decisions about who/what gets counted and how Statistical methods to mitigate representation issues Culture, context, & stakeholders in terms of dataset suitability 					
Testing and Evaluation	 Fairness metrics (context specific!) Monitoring performance after deployment Periodic model updates, test and recalibrate model parameters 					
Human Factors	 Multistakeholder engagement and diverse perspectives Model and procedural transparency Algorithmic impact assessments, iterative process 					



Thermal Imaging for Human-Wildlife Conflict - Arribada Initiative

embedded VISION SUMMIT

Dataset

- Hardware testing optimizing data collection methods across species
- Suitability adjusting for variations in the environment (dirt, grass, snow)

Human Factors

- Stakeholder engagement WWF team in Tezpur advised on locations for field testing
- Commitment to model transparency









- 1. Bias poses an ethical challenge, regardless of the stakes involved
- 2. Reframing the goal bias-aware not bias-free
- 3. An effective mitigation strategy requires a combination of technical and social considerations







1. NIST Special Publication 1270 – Towards a Standard for Identifying and Managing Bias in Artificial Intelligence <u>https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf</u>

2. (In Progress) NIST Mitigation of AI/ML Bias in Context https://www.nccoe.nist.gov/projects/mitigating-aiml-bias-context

3. An Ethical Toolkit for Engineering Design Practice – Markkula Center for Applied Ethics https://www.scu.edu/ethics-in-technology-practice/ethical-toolkit/

