AI Ethics Then & Now:
A Look Back on the Last Five Years

Willie Costello

August 27, 2020

aisc

Five years ago...

MACHINE BIAS

The Tiger Mom Tax: Asians Are Nearly Twice as Likely to Get a Higher Price from Princeton Review

by Julia Angwin, Surya Mattu and Jeff Larson, Sept. 1, 2015, 10 a.m. EDT

Internet Culture

Google Maps' White House glitch, Flickr auto-tag, and the case of the racist algorithm

HIDDEN BIAS

When Algorithms Discriminate

By Claire Cain Miller

July 9, 2015

147

BUSINESS

When Discrimination Is Baked Into Algorithms

As more companies and services use data to target individuals, those analytics could inadvertently amplify bias.

LAUREN KIRCHNER SEPTEMBER 6, 2015

Flickr faces complaints over 'offensive' auto-tagging for photos

Auto-tagging system slaps 'animal' and 'ape' labels on images of black people, and tags concentration camps with 'jungle gym' and 'sport'

PRO PUBLICA

Donate

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016

Recent* trends* in AI* ethics

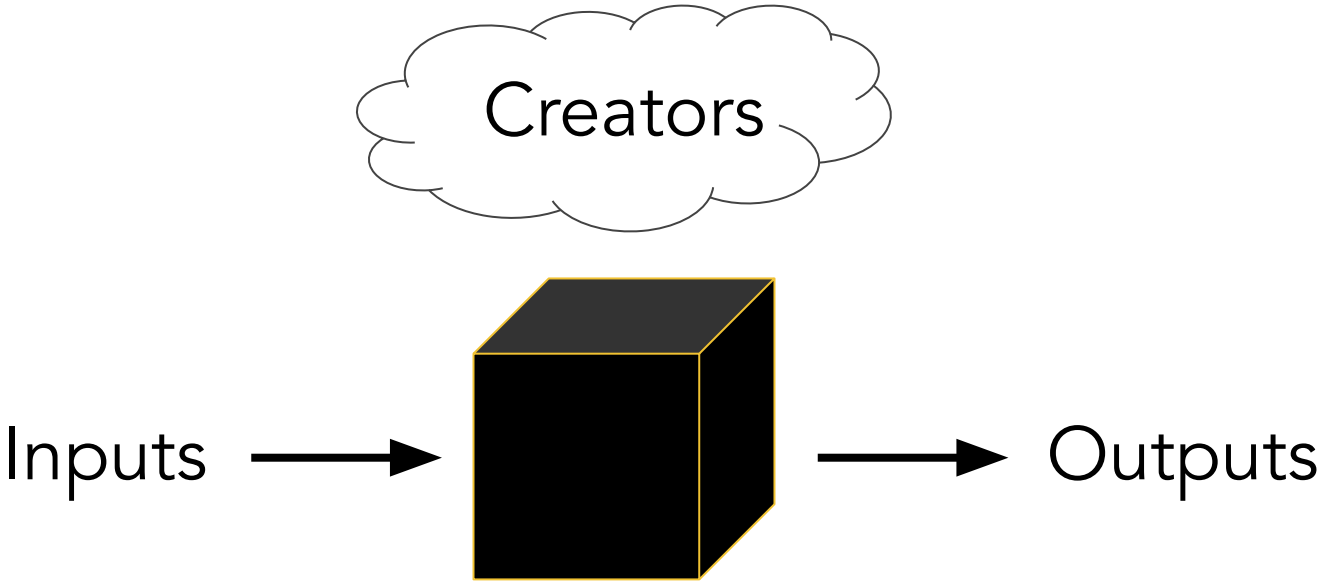
*some clarifications

About me

Willie Costello
Data scientist, PhD Philosophy

williecostello.com
linkedin.com/in/williecostello
@williecostello

Three aspects of algorithmic ethics



The ethics of the outputs

How do we make algorithms fair?

Then:
Fairness is just math

	Definition	Paper	Citation #	Result
3.1.1	Group fairness or statistical parity	[12]	208	×
3.1.2	Conditional statistical parity	[11]	29	✓
3.2.1	Predictive parity	[10]	57	✓
3.2.2	False positive error rate balance	[10]	57	×
3.2.3	False negative error rate balance	[10]	57	✓
3.2.4	Equalised odds	[14]	106	×
3.2.5	Conditional use accuracy equality	[8]	18	×
3.2.6	Overall accuracy equality	[8]	18	✓
3.2.7	Treatment equality	[8]	18	×
3.3.1	Test-fairness or calibration	[10]	57	✓
3.3.2	Well calibration	[16]	81	✓
3.3.3	Balance for positive class	[16]	81	✓
3.3.4	Balance for negative class	[16]	81	×
4.1	Causal discrimination	[13]	1	×
4.2	Fairness through unawareness	[17]	14	✓
4.3	Fairness through awareness	[12]	208	×
5.1	Counterfactual fairness	[17]	14	–
5.2	No unresolved discrimination	[15]	14	–
5.3	No proxy discrimination	[15]	14	–
5.4	Fair inference	[19]	6	–

Table 1: Considered Definitions of Fairness

Verma & Rubin, “Fairness Definitions Explained” (2018)

Now:
Fairness cannot be automated

Case study: Facial recognition technology

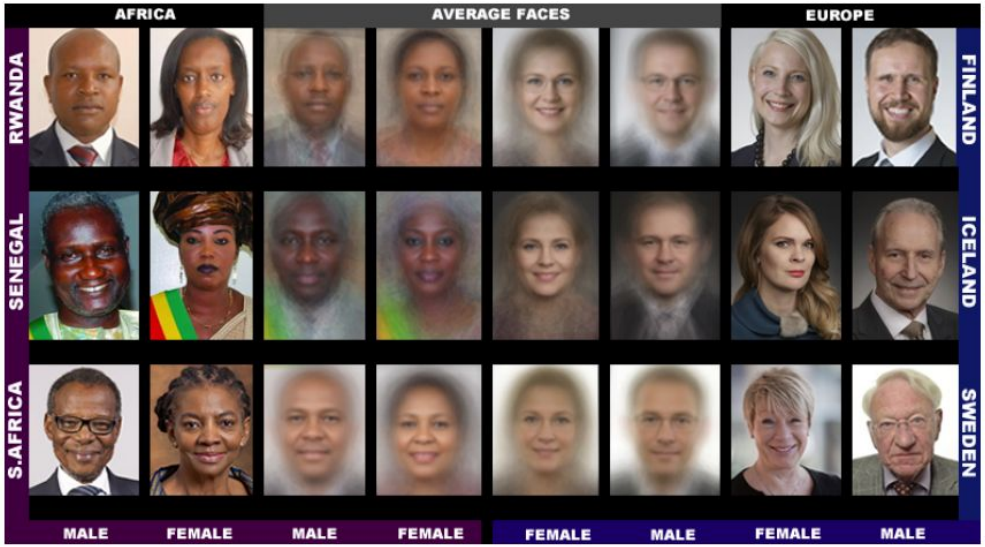


Figure 1: Example images and average faces from the new Pilot Parliaments Benchmark (PPB). As

Classifier	Metric	DF	DM	LF	LM
MSFT	PPV(%)	76.2	**100**	**100**	**100**
	Error Rate(%)	**23.8**	0.0	0.0	0.0
	TPR(%)	**100**	84.2	**100**	**100**
	FPR(%)	**15.8**	0.0	0.0	0.0
Face++	PPV(%)	64.0	99.5	**100**	**100**
	Error Rate(%)	**36.0**	0.5	0.0	0.0
	TPR(%)	99.0	77.8	**100**	96.9
	FPR(%)	**22.2**	1.03	3.08	0.0
IBM	PPV(%)	66.9	94.3	**100**	98.4
	Error Rate(%)	**33.1**	5.7	0.0	1.6
	TPR(%)	90.4	78.0	96.4	**100**
	FPR(%)	**22.0**	9.7	0.0	3.6

Buolamwini & Gebru, "Gender Shades" (2018)

Uncovering unfair outputs is work

The fairness of the use itself

“Face recognition will work well enough to be dangerous, and poorly enough to be dangerous as well” – Philip E. Agre

“Sometimes technology hurts people precisely because it **doesn't** work & sometimes it hurts people because it **does** work. Facial recognition is both. When it **doesn't** work, people get misidentified, locked out, etc. But even when it **does**, it's invasive & still unsafe.” – Deb Raji

Philip E. Agre, “Your Face Is Not a Bar Code” (2001); Raji et al., “Saving Face” (2020)

The disparate deployment of algorithmic systems

"The future is already here, it's just not evenly distributed"
– William Gibson

Virginia Eubanks: Yes, because algorithmic systems are disproportionately **deployed** on the poor and marginalized

Virginia Eubanks, *Automating Inequality* (2018)

A REUTERS INVESTIGATION

Rite Aid deployed facial recognition systems in hundreds of U.S. stores

ON CAMERA: DeepCam, a facial recognition system deployed by pharmacy chain Rite Aid, captured this footage of a Reuters photographer at a store in New York in November. After Reuters informed Rite Aid of this article's findings, the company said it had ended the surveillance program.

In the hearts of New York and metro Los Angeles, Rite Aid installed facial recognition technology in largely lower-income, non-white neighborhoods, Reuters found. Among the technology the U.S. retailer used: a state-of-the-art system from a company with links to China and its authoritarian government.

By JEFFREY DASTIN in LOS ANGELES and NEW YORK CITY | Filed July 28, 2020, 11 a.m. GMT

Over about eight years, the American drugstore chain Rite Aid Corp quietly added facial recognition systems to 200 stores across the United States, in one of the largest rollouts of such technology among retailers in the country, a Reuters investigation

The ethics of the inputs

Bias in, bias out

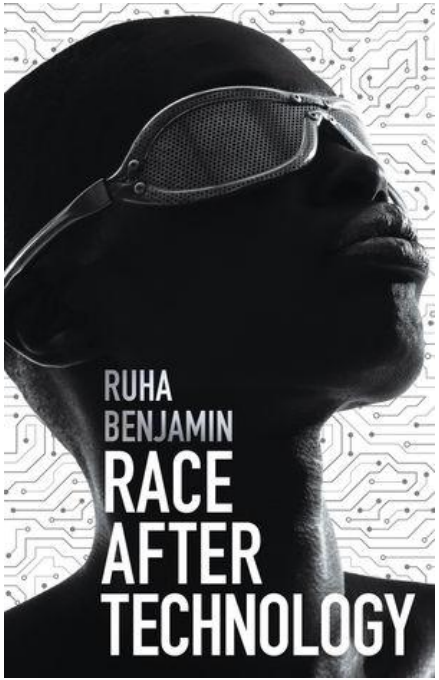
Then:
Not the algorithm’s problem

Now:
The insistence that algorithms are “objective” is itself a kind of bias

Bias can be encoded in a dataset's features

Gender
1
0
0
1
1
0
1

Bias can be encoded in a dataset's features



"Race itself is a kind of technology –
one designed to separate, stratify, and sanctify
the many forms of injustice experienced by
members of racialized groups" – Ruha Benjamin

Ruha Benjamin, *Race After Technology* (2019)
Safiya Umoja Noble, *Algorithms of Oppression* (2018)
Hanna et al., "Towards a Critical Race Methodology in Algorithmic Fairness" (2020)

Data collection is not a neutral process

Table 1: Lessons from Archives: summaries of approaches in archival and library sciences to some of the most important topics in data collection, and how they can be applied in the machine learning setting.

Consent	(1) Institute data gathering outreach programs to actively collect underrepresented data (2) Adopt crowdsourcing models that collect open-ended responses from participants and give them options to denote sensitivity and access
Inclusivity	(1) Complement datasets with "Mission Statements" that signal commitment to stated concepts/topics/groups (2) "Open" data sets to promote ongoing collection following mission statements
Power	(1) Form data consortia where data centers of various sizes can share resources and the cost burdens of data collection and management
Transparency	(1) Keep process records of materials added to or selected out of dataset. (2) Adopt a multi-layer, multi-person data supervision system.
Ethics & Privacy	(1) Promote data collection as a full-time, professional career. (2) Form or integrate existing global/national organizations in instituting standardized codes of ethics/conduct and procedures to review violations

Jo & Gebru, "Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning" (2020)

Denton et al., "Bringing the People Back In: Contesting Benchmark Machine Learning Datasets" (2020)

Datasets must be documented

"We propose that every dataset be accompanied with a datasheet that documents its motivation, composition, collection process, recommended uses, and so on." – Gebru et al.

Gebru et al., “Datasheets for Datasets” (2020)
Bender & Friedman, “Data Statements for Natural Language Processing” (2018)
Mitchell et al., “Model Cards for Model Reporting” (2019)
Raji et al., “Closing the AI Accountability Gap” (2020)

The ethics of the creators

Who makes the algorithms?

Then:
We need more diversity in tech!

Now:
Who owns the algorithms?

Critiquing academia's role, too

Springer Publishing
Berlin, Germany
+49 (0) 6221 487 0
customerservice@springernature.com

RE: A Deep Neural Network Model to Predict Criminality Using Image Processing

June 22, 2020

Dear Springer Editorial Committee,

We write to you as expert researchers and practitioners across a variety of technical, scientific, and humanistic fields (including statistics, machine learning and artificial intelligence, law, sociology, history, communication studies and anthropology). Together, we share grave concerns regarding a forthcoming publication entitled "A Deep Neural Network Model to Predict Criminality Using Image Processing." According to a recent press release, this article will be published in your book series, "Springer Nature — Research Book Series: Transactions on Computational Science and

"[Machine learning] research agendas reflect the incentives and perspectives of those in the privileged position of developing machine learning models, and the data on which they rely. The uncritical acceptance of default assumptions inevitably leads to discriminatory design in algorithmic systems, reproducing ideas which normalize social hierarchies and legitimize violence against marginalized groups."

What does AI ethics now require?

Thinking outside the (black) box

Thinking outside of computer science

A renewed focus on power

"Don’t ask if artificial intelligence is good or fair, ask how it shifts power" – Ria Kalluri

Thank you!

For a complete bibliography, go to williecostello.com/aiethics

Follow me on Twitter @williecostello
and on LinkedIn at linkedin.com/in/williecostello