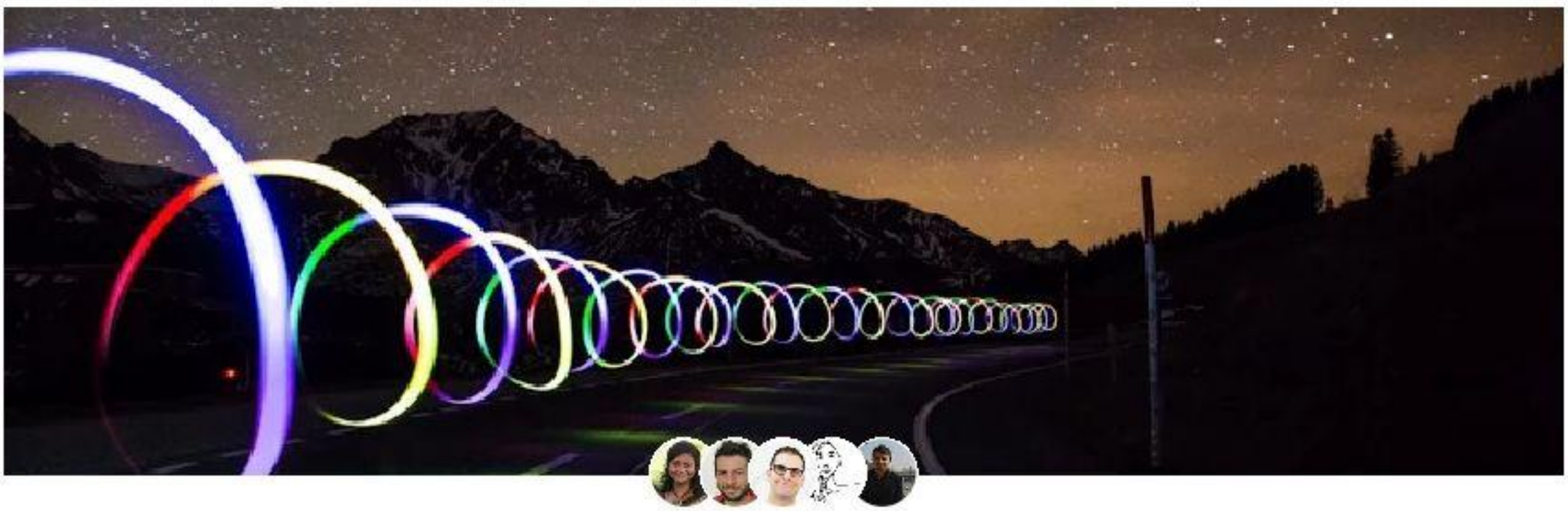


147 members - [redacted] within [redacted]

Bias Busting for Machines

In the age of AI and automation, data bias in big systems will matter even more than our individual explicit or implicit biases.

JOIN

About Community

go/ml-fairness

go/hcml

Josh Lovejoy Owner 1d
▸ Discussion

Gotta check our assumptions about representativeness in training data, language is a living thing.

AI programs are learning to exclude some African-American...



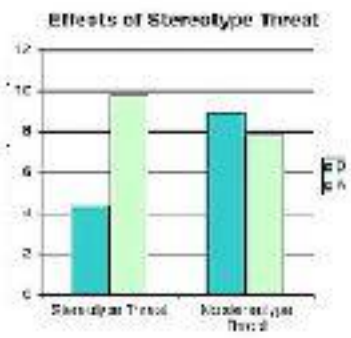
[redacted]

Sergio Guadarrama 1w
▸ Discussion

Thoughtful response

Originally shared by Laura Holmes - 11 comments

A lot has been said about the go/pc-considered-harmful doc, but after reading the doc and Danielle Brown's response, I fel ...



Stereotype threat - Wikipedia

en.wikipedia.org



+1 1

Josh Lovejoy Owner 2w
▸ Discussion

File another under: Depressing findings about human tendencies in filmmaking made visible via ML. +Hartwig Adam +Joshua Metherd

How Central are Female Characters to a Movie? - USC...



Hartwig Adam: Thanks for sharing. This is from my collaborators at USC.

+1

Josh Lovejoy Owner 12w
▸ Discussion

Highly recommended read about the evolving public perceptions of AI in light of Alpha-Go's performance. The state of our dialogue with users about AI remains rooted in an awkward dialectic of anthropomorphism and alienation.



Us vs. Them

dl.acm.org

+1 1

Josh Lovejoy Owner 5w
▸ Discussion

Time to tread carefully. Underscores the need for taking an inclusive approach to ML, and understanding tradeoffs between false positives and false negatives.

Banks Deploy AI to Cut Off Terrorists' Funding



+1

Iwona Bialynicka-Birula 16w
▸ Discussion

FaceApp apologizes for building a racist AI | TechCrunch



+1 2

+1



Josh Lovejoy Owner 17w
▸ Discussion

Similar content to the recent Guardian post about word embeddings, but I found this to be much more practical and hopeful, including the closing lines: "it's important to remember: AI learns about how the world has been. It picks up on status quo trends. It doesn't know how the world ought to be. That's up to humans to decide."

How artificial intelligence learns to be racist



+1 2



[redacted] 18w
▸ Discussion

Originally shared by [redacted] - 12 comments

We launched an ML thing and it had bias. Here's our postmortem: http://go/perspective-bias-postmortem

The Conversation AI team publicly launc •••

+1 1



Josh Lovejoy Owner 18w
▸ Discussion

Maya Gupta's tremendous talk on Fairness at the EmTech Digital Conference (MIT Tech Review).

MIT Technology Review Events Videos - Fairness and ML



+1 1



Josh Lovejoy Owner 20w
▸ Discussion

An emphatic call to arms on Fairness in ML. It's also one of the more shareable short-form pieces I've seen on algorithmic bias. Highly recommend passing it along on your personal pages.



How I'm fighting bias in algorithms | Jo...

+1 2

Blaise Aguera y... Moderator 20w

[redacted]

Josh Lovejoy Owner 20w
▸ Discussion

"In an attempt to stand out from the pack, predictive-policing startup CivicScape has released its algorithm and data online for experts to scour."

Github notebook here:
https://github.com/CivicScape/CivicScape/blob/master/evaluation_notebooks/notebooks/PreventingBias.ipynb •••

Software used to predict crime can now be scoured for bias



+1 2

Josh Lovejoy Owner 25w
▸ Discussion

The engaging and brilliant Maciej Cegłowski on what he calls "Deep Fried Data" during a talk at the Library of Congress (talk starts at 5:31:50 and is about 20 minutes).

"I find it helpful to think of algorithms as a dim-witted but extremely industrious graduate student, whom you don't fully trust. You want a concordance made? An index? You v •••





▸ Discussion

Interesting paper:
https://arxiv.org/abs/1703.06856

[1703.06856] Counterfactual Fairness

Abstract: Machine learning has matured to the point to where it is now being considered to automate decisions in loan lending, employee hiring, and...

arxiv.org

Josh Lovejoy: +M. Mitchell WDYT? Complimentary to your metrics & measurement strategy?...

+1



Glenn Brown 20w
▸ Discussion

Originally shared by Glenn Brown - 4 comments

Since it's really easy for ML to deduce gender, race, and other protected classes from user data, how do we prevent (or compensate for) ML misusing this information?

... •••

Josh Lovejoy: You're describing one of the hard-to-get-moving pieces of debiasing work; namely ho...

+1



Josh Lovejoy Owner 22w
▸ Discussion

Concerned+confused emoji goes here...
"Working out where asylum seekers originated is crucial to the ultimate success of their claim. The suspicion is that some people lie about where they are from in order to improve their chances of asylum."

Collections as Data: Stewardship and U...

+1

Josh Lovejoy Owner 26w
Discussion

Inclusive thinking + interactive data modeling = win

Parable of the Polygons

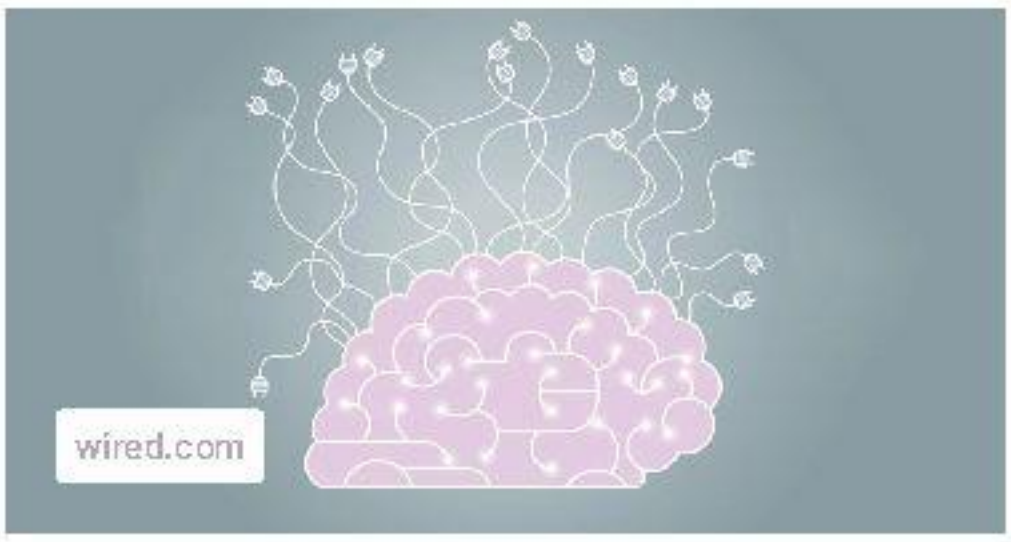


+1 1

Josh Lovejoy Owner 26w
Discussion

A nice quick read. Nothing terribly novel, but it's great to see the message continuing to trend. And the call to action is spot-on :)

How to Keep Your AI From Turning Into a Racist Monster



Germany to use speech recognition to establish migrant...



+1 1

Josh Lovejoy Owner 24w
Discussion

FB boasting about how their filter-bubble-as-a-service can swing elections (if the price is right) is really disturbing...
m.facebook.com/business/success/toomey-for-senate

Toomey for Senate

Toomey for Senate

+1 1

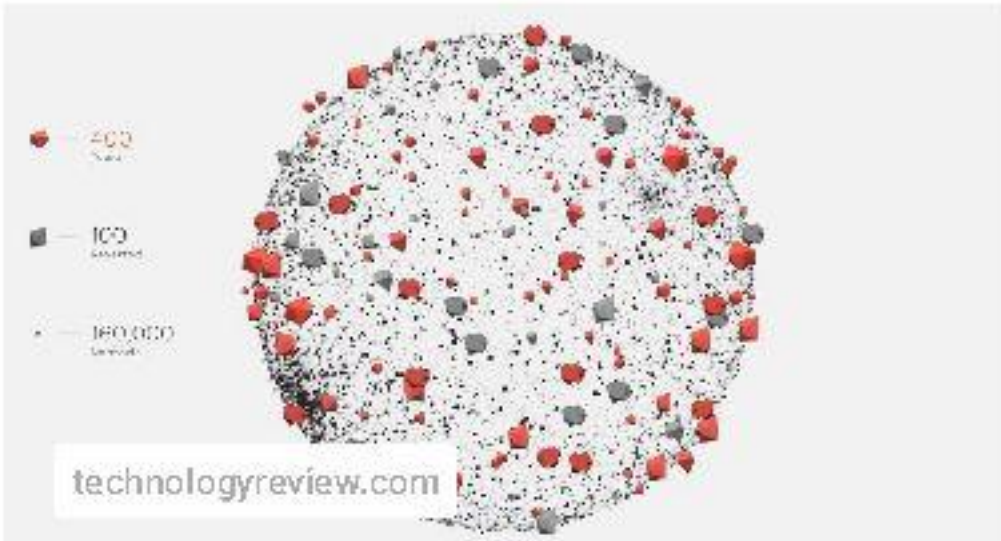
Josh Lovejoy Owner 27w
▸ Discussion

More Jigsaw goodness

Originally shared by Arun Mathew

https://www.technologyreview.com/s/603589/c
of-13500-nastygrams-could-advance-war-on-
trolls/ •••

A collection of 13,500 insults lobbed by Wikipedia editors is...

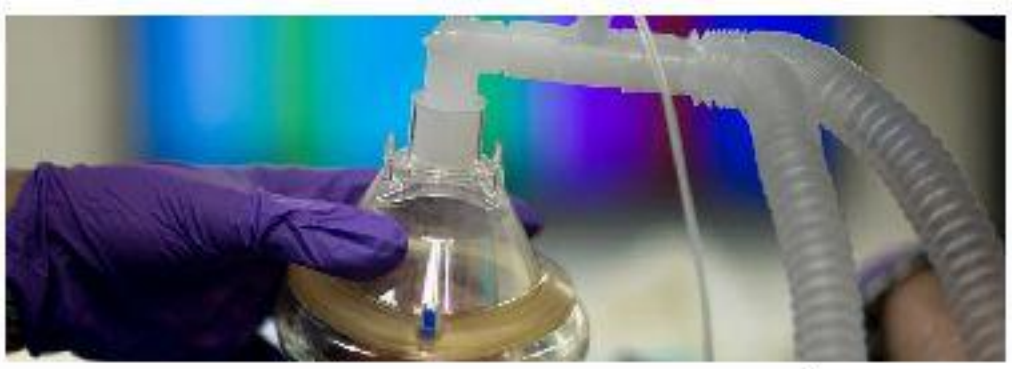


+1 1

Josh Lovejoy Owner 27w
▸ Discussion

Certainly not directly about ML, but I'm attracted the idea of checklists as a means to helping us debias things like data collection practices.

A Fix for Gender Bias in Health Care? Check - The New York...



[redacted]

m.facebook.com

Josh Lovejoy: My primary concern is that, in capitalizing on the monetization opportunities of...

+1 1

Josh Lovejoy Owner 25w
▸ Discussion

+Hartwig Adam has been quietly—and profoundly— harnessing ML to change the world. This post is a wonderful summary of the collaboration he's had with the Geena David Institute, and the impact their work is having on improving female representation in film.

Using technology to address gender bias in film | Google



+1 3 1

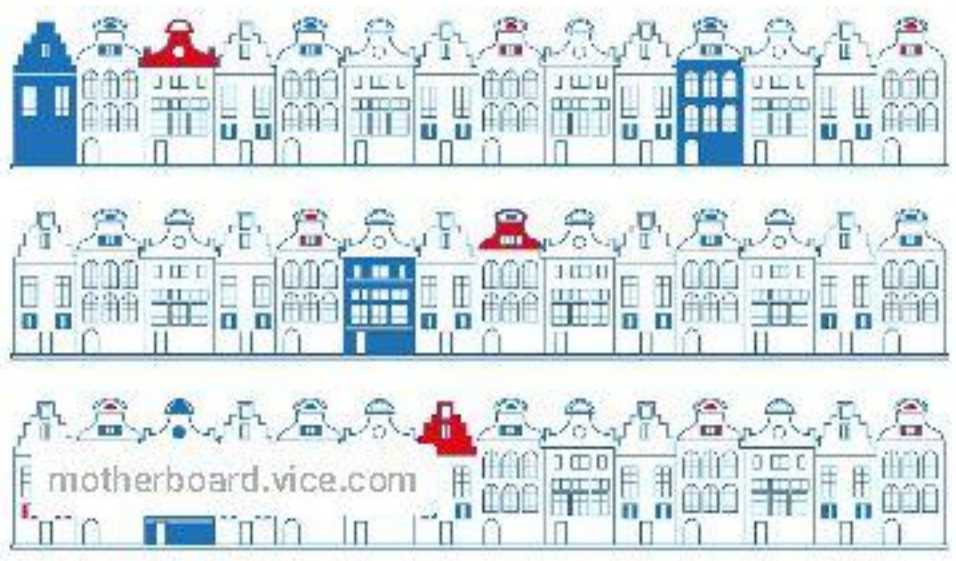
Jess Holbrook 27w
▸ Discussion

https://motherboard.vice.com/en_us/article/ai-could-resurrect-a-racist-housing-policy

https://arxiv.org/pdf/1701.08230.pdf



AI Could Resurrect a Racist Housing Policy - Motherboard



+1 1

Blaise Aguera y... Moderator 29w
▸ Discussion

Great piece in Harper's by our friend Kate Crawford
http://harpers.org/archive/2017/02/trump-a-resisters-guide/11/

[Forum] | Trump: A Resister's Guide | Harper's Magazine - Part...



+1

Sergio Guadarrama 34w
▸ Discussion

The lack of diversity in Software Engineer according to Google Search.

[redacted]
q=software+engineer&oq=sotfware+&aqs=chrome.1.69i57j0l5.3607j0j4&sourceid=chrome&ie=[redacted]

Josh Lovejoy Owner 27w
▸ Discussion

go/fair-not-default :)

[redacted] Lovejoy: [redacted] it's something our content strategy team uses to track analytics...

Josh Lovejoy: +Jac de Haan FYI about go/detangle

Jac de Haan: Requested whitelist: b/35242997

+1 2

Blaise Aguera y... Moderator 39w
▸ Discussion

Fucked up paper on arxiv:
https://arxiv.org/abs/1611.04135

Return of physiognomy and phrenology, now machine learning edition.

http://www.victorianweb.org/science/phrenology/intro.html

[1611.04135] Automated Inference on Criminality using...

UTF-8

https://screenshot.googleplex.com/[redacted] •••

Sign in - Google Accounts

One account. All of Google. Sign in with your Google Account. Please enter your full email address [redacted] Enter your email. Find my...

[redacted]

+1 1



Divya Tyam 35w
▸ Discussion

Facebook announced their plan to address the fake news problem.

News Feed FYI: Addressing Hoaxes and Fake News |...



newsroom.fb.com

Josh Lovejoy: it's a start! +Nick Jong +Emily Fortuna +M. Mitchell

M. Mitchell: Nice. Our approach has similarities – I hadn't thought about broadcasting publicly. Woul...

Divya Tyam: +Charina Choi

Abstract: We study, for the first time, automated inference on criminality based solely on still face images. Via supervised machine learning, we build fo

arxiv.org

SHOW ALL 4 COMMENTS

Blaise Aguera y Arcas: Pointer courtesy of +M. Mitchell, the commercial version, even ugher:...

M. Mitchell: Now it's getting covered in the news more broadly, and positively. ...

M. Mitchell: I emailed the author of the MIT Tech Review article. I think his e-mail address is...

+1 15 1



Jen Gennai
▸ Discussion

"set trackable metrics and goals quarter by quarter. It's important to bake-in inclusion requirements from day one and for it to be a key consideration, if not a priority, as the company and product develop."

"To build safer products that recognize the equal value and humanity of all people, we must first have diverse perspectives •••

When bias in product design means life or death | TechCrunch



techcrunch.com

Emily Fortuna: yup. because there will always be priorities that *seem* more pressing.

[redacted]

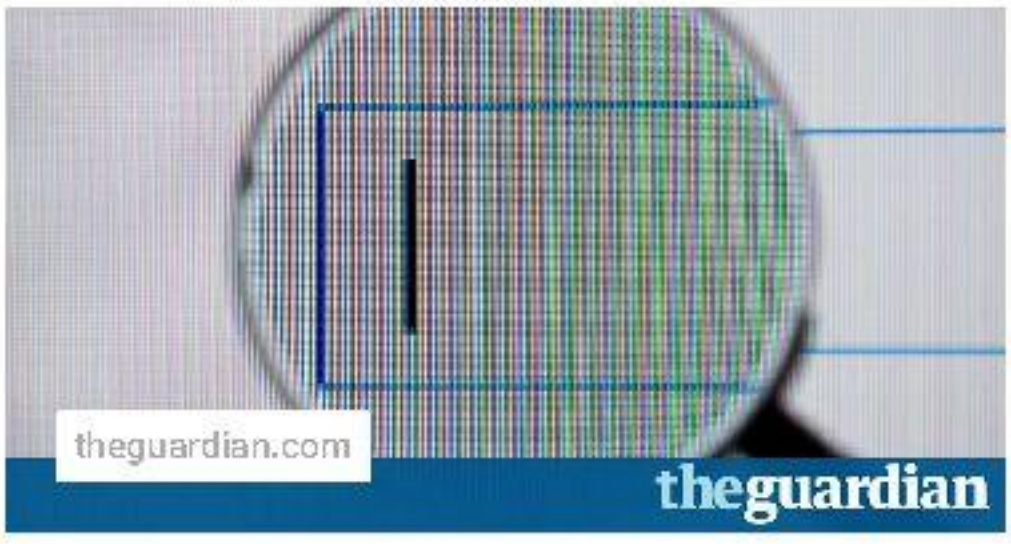
+1 1

Josh Lovejoy Owner 37w
▸ Discussion

If we hope to have any chance against fake news, we should start with a hard look in the mirror. Search suggestions and results are terrifying latent bias amplification systems.

Note that b/33326165 addresses some, but the list is continuing to grow, and there do not appear to be clear definitions or structure for the discussion. •••

Google, democracy and the truth about internet search |...



+1

Eva Snee ▸ Discussion 43w

+Emily Fortuna, relevant for our upcoming talk, perhaps?

Racial profiling, by a computer? Police facial-ID tech raises civil...

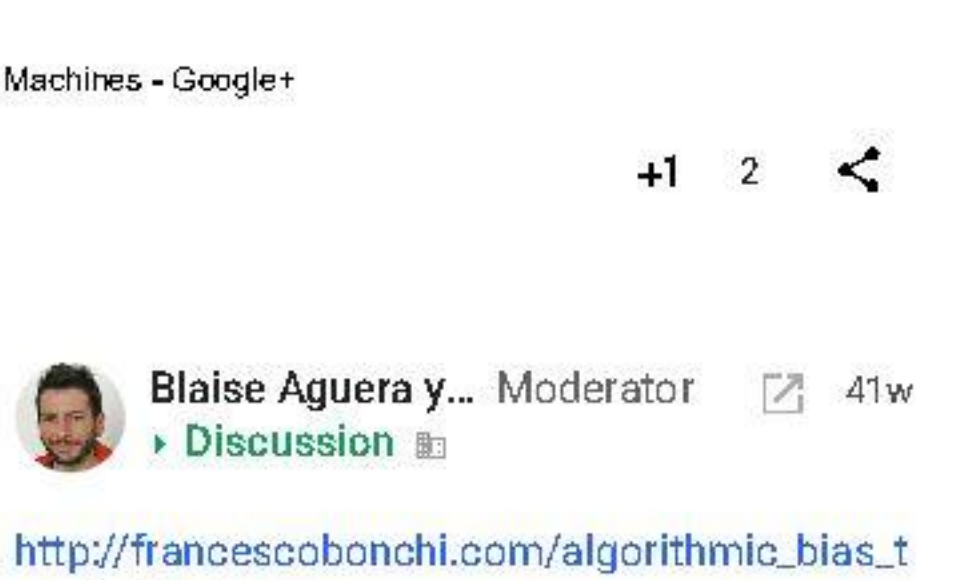


+1 2

Blaise Aguera y... Moderator 41w
▸ Discussion

http://francescobonchi.com/algorithmic_bias_tutorial.html
Lots of useful stuff covered in one place.

Algorithmic bias: from discrimination discovery to...



+1 1

Josh Lovejoy Owner 41w
▸ Discussion

Well this is shockingly blatant...

Facebook Lets Advertisers Exclude Users by Race



+1



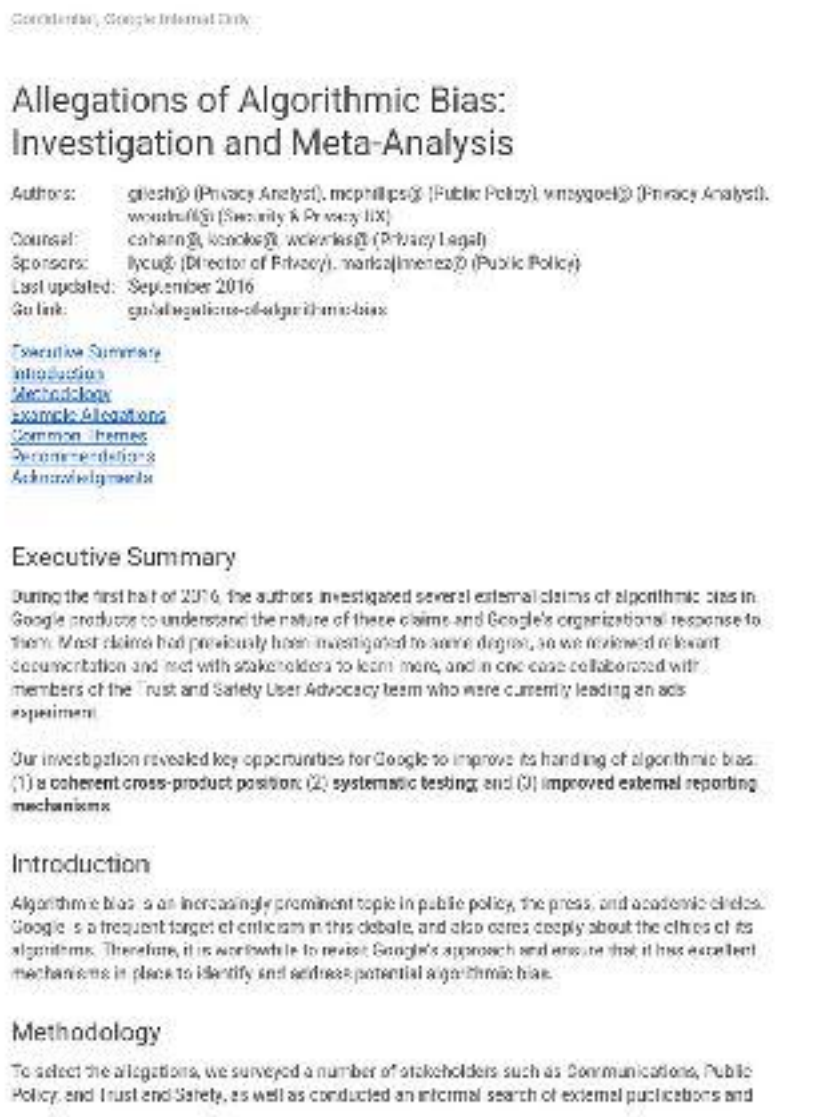
+1



Josh Lovejoy Owner
▸ Discussion 45w

"Our investigation revealed key opportunities for Google to improve its handling of algorithmic bias: (1) a coherent cross-product position; (2) systematic testing; and (3) improved external reporting mechanisms."

Allegations of Algorithmic Bias: Investigation and Meta-Analysis



Allegations of Algorithmic Bias:
Investigation and Meta-Analysis

Authors: gilesh@ (Privacy Analyst), mcphillips@ (Public Policy), vinaygoel@ (Privacy Analyst), woodruff@ (Security & Privacy UX)
Counsel: cohenn@, kcooke@, wdevries@ (Privacy Legal)
Sponsors: lycu@ (Director of Privacy), marisajimenez@ (Public Policy)
Last updated: September 2016
Go link: go/allegations-of-algorithmic-bias

Executive Summary
Introduction
Methodology
example Allegations
Common Themes
Recommendations
Acknowledgments

Executive Summary

During the first half of 2016, the authors investigated several external claims of algorithmic bias in Google products to understand the nature of these claims and Google's organizational response to them. Most claims had previously been investigated to some degree, so we reviewed relevant documentation and met with stakeholders to learn more, and in one case collaborated with members of the Trust and Safety User Advocacy team who were currently leading an ads experiment.

Our investigation revealed key opportunities for Google to improve its handling of algorithmic bias: (1) **a coherent cross-product position**; (2) **systematic testing**; and (3) **improved external reporting mechanisms**.

Introduction

Algorithmic bias is an increasingly prominent topic in public policy, the press, and academic circles. Google is a frequent target of criticism in this debate, and also cares deeply about the ethics of its algorithms. Therefore, it is worthwhile to revisit Google's approach and ensure that it has excellent mechanisms in place to identify and address potential algorithmic bias.

Methodology

To select the allegations, we surveyed a number of stakeholders such as Communications, Public Policy, and Trust and Safety, as well as conducted an informal search of external publications and

+1

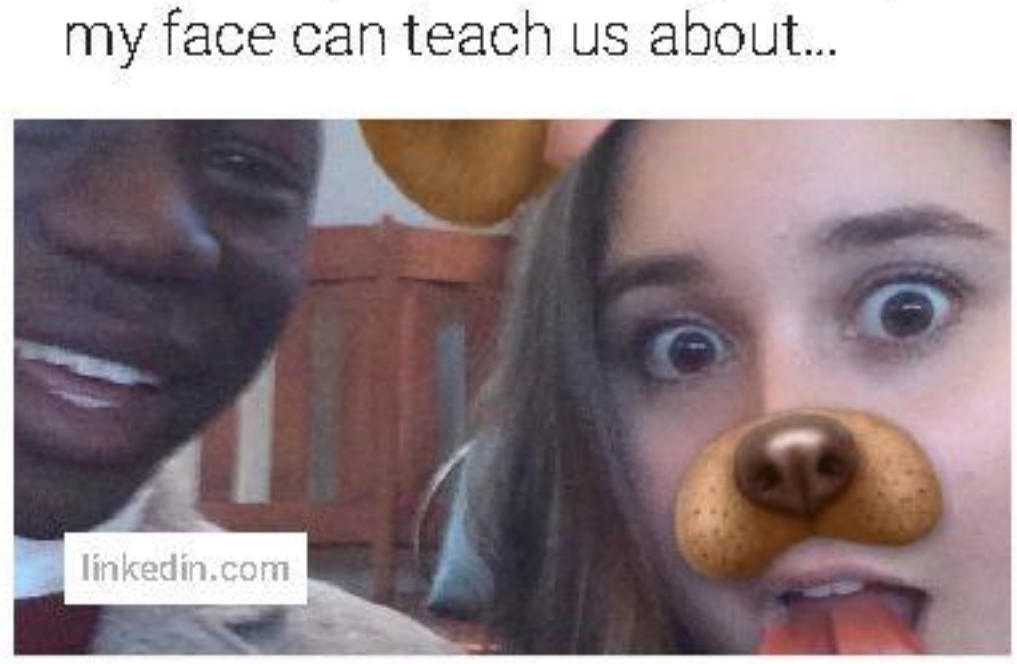


Kelly Naku ▸ Discussion 42w

A good example of why building inclusive products matters with an optimistic perspective form the author.

"As a minority, I am more excited than I am scared when I think about the future of tech because we're pointed in the right direction."

What Snapchat not recognizing my face can teach us about...



+1



Josh Lovejoy Owner
▸ Discussion 42w

"Yet too much contemporary discussion is framed as if the algorithmic workings of computer networks are something entirely new. It's true that they can follow instructions at superhuman speed, with superhuman fidelity and over unimaginable quantities of data. But these instructions don't come from nowhere. Although neural networks might be said to write their own programs, they do so ...

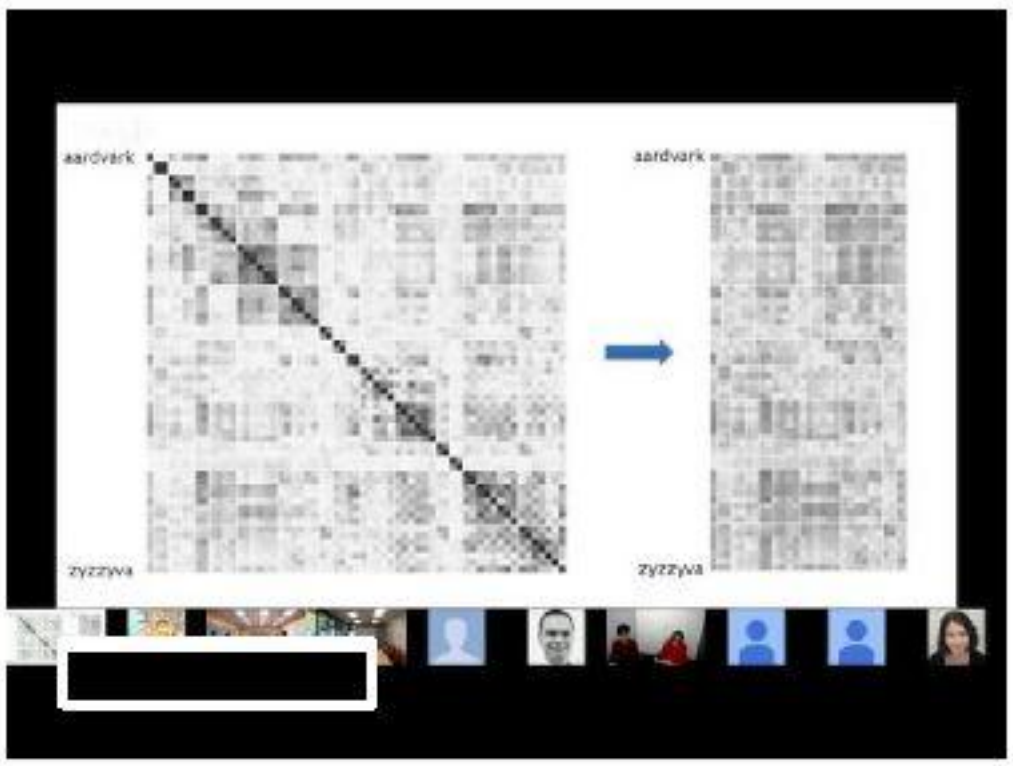
The Guardian view on machine learning: people must decide |...



Josh Lovejoy Owner 45w
▸ Discussion

Abstract:
Artificial intelligence and machine learning are in a period of astounding growth. However, there are concerns that these technologies may be used, either with or without intention, to perpetuate the prejudice and unfairness that unfortunately characterizes many human institutions. Here we show for the first time that human-like semantic biases resi •••

Language necessarily contains human biases, and so will...



+1

Anne Schaefer 45w
▸ Discussion

Deep-Fried Data

I run a small web archive for about twenty thousand people. Being invited to speak at the Library of Congress is like being a kid who glues paper fins to a...

idlewords.com

+1

Divya Tyam 45w
[redacted]



+1

Emily Fortuna 43w
▸ Discussion

There's a new newsletter starting to help spread awareness of incidents involving minority/underrepresented groups at Google. The group that receives the newsletter is [redacted] m/#!forum/yes-at-google. From the group:

There are many instances when people are shocked by the issues faced by minc •••

Blaise Aguera y Arcas: Maybe old news to everybody else, but this approach to how to...

+1 2

Josh Lovejoy Owner 45w
▸ Discussion

"Stress has been linked to degraded mood and interactions with others, such as reduced altruism and increased aggression and competitiveness. Similarly, when people are threatened they are more likely to feel uncomfortable, hostile, aggressive, and angry. As a result, people who experience algorithmic discrimination may be more likely to display negative online behaviors such as ag •••

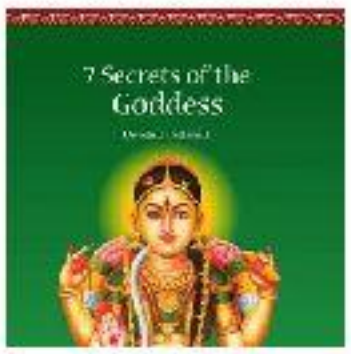
Algorithmic Discrimination from an Environmental Psychology...



▸ Discussion

On the timeline of feminism and the need for us to think across cultures: this book is a short read and sheds light on the mythology of hindu goddesses, while also tracing ancient history of feminism in India.

But mythology? Yes. Mythology is quite integrated into contemporary culture and any anthropological study of the subcont ...



7 Secrets of the Goddess - Books...

[redacted]

+1



Blaise Aguera y... Moderator 48w

▸ Discussion

Great piece on bias in MI co-written by one of our own

[redacted]

intelligence-is-hard-to-see-a71e74f386db#.jq62o4n2t

Artificial intelligence is hard to see



[redacted]

Confidential, Google Internal Only

Algorithmic Discrimination from an Environmental Psychology Perspective: Stress-Inducing Differential Treatment

go/discrimination-and-stress

AUTHORS
Sally Augustin (Environmental Psychology Consultant)
Allison Woodruff (Security & Privacy UX)

SPONSORS
Lawrence You (Director of Privacy)
Sunny Consolvo (Security & Privacy UX)

LAST UPDATED
September 2016

Table of Contents

Executive Summary
Overview
How People Respond to Situations
Experiencing Stress
Case Study: Crowding, Physical Stressor
Case Study: Natural Disasters, Physical Stressors
What Sorts of Psychological Stressors Are Most Difficult for People to Tolerate
What Is Control
Benefits of Control and Perceived Control
How to Increase Perceived - Or Actual - Control
Other Ways (Besides Psychological Control) to Counter Stressors
When Control Is Undesirable
Individual Differences and Stressors
Conclusion
Acknowledgments
References

Emily Fortuna: truth.
<– guilty of reduced altruism/aggression (not...

Josh Lovejoy: +Blaise Aguera y Arcas +Jen Gennai +Divya Tyam Related to Blaise's examination of...

+1



Blaise Aguera y... Moderator 49w

▸ Discussion

Zeynep Tyfekci: if we ever do an externally-facing summit on this stuff, we want her at it. From back in May, but still relevant–
http://www.nytimes.com/2016/05/19/opinion/the-real-bias-built-in-at-facebook.html?referer=https:/t.co/35ve1OREnm

The Real Bias Built In at Facebook - The New York Times



[redacted]

Josh Lovejoy: The call to arms at the end of the article is really powerful. My hope is that one day...

+1 2



Josh Lovejoy: In the same way that designing the right user success metrics is vital to longterm...

+1

Blaise Aguera y... Moderator 49w
▸ Discussion

"While data-driven policing holds promise, the way 'predictive policing' is being implemented today threatens to exacerbate rather than alleviate disparities and inefficiencies in how communities are policed," said Rachel Levinson-Waldman, senior counsel at the Brennan Center. "Because the data fed into these systems is itself the function of historical police behavior, because the data an •••

BRENNAN CENTER FOR JUSTICE

Shared Statement: 'Predictive Policin...

brennancenter.org

+1

Josh Lovejoy Owner 50w
▸ Discussion

Time article relevant to Blaise's post from a couple days ago (Weapons of Math Destruction).

This Mathematician Says Big Data Punishes Poor People



Iwona Bialynicka-Birula 49w
▸ Discussion

(Via [redacted] The title of the book is overly sensationalist, but I started reading it and like it so far. Very relevant to this forum.

The case against big data: "It's like you're being put into a cult,...



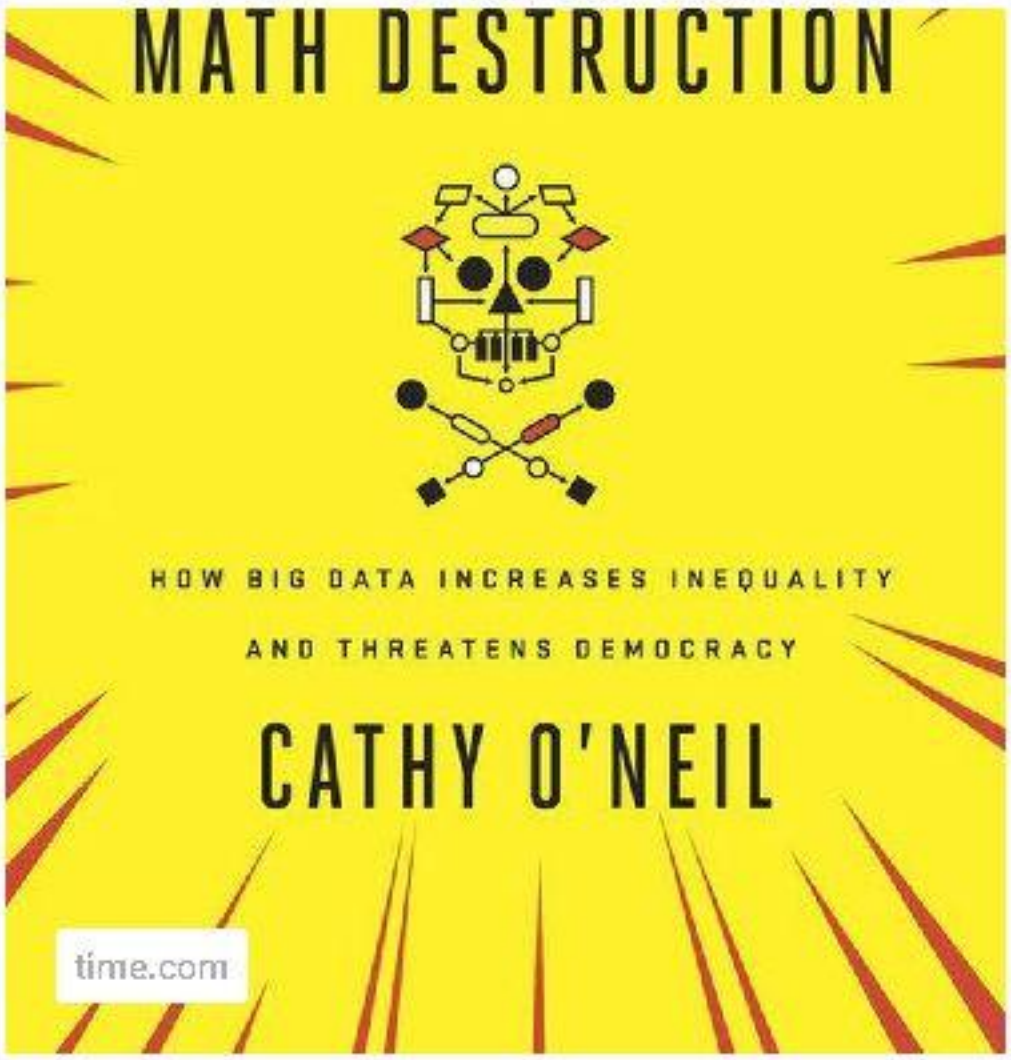
+1

Iwona Bialynicka-Birula 49w
▸ Discussion

Why An AI-Judged Beauty Contest Picked Nearly All White...



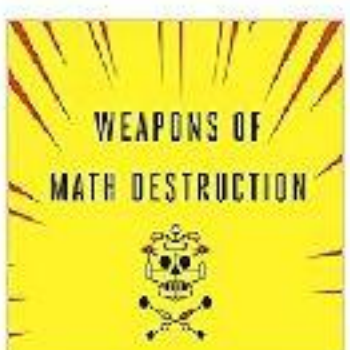
[redacted]



+1

Blaise Aguera y... Moderator 51w
▸ Discussion

Very relevant-looking upcoming book for "latent bias": https://www.amazon.com/Weapons-Math-Destruction-Increases-Inequality/dp/0553418815/ref=sr_1_1?s=books&ie=UTF8&qid=1472320699&sr=1-1&keywords=weapons+of+math+destruction coming early in September. The author, Mathbabe aka Cathy O'Neil, https://mathbabe.org/ is an expert w



Weapons of Math Destruction: How...

amazon.com

+1 2

Blaise Aguera y... Moderator 51w
▸ Discussion

motherboard.vice.com

Emily Fortuna: ouch.

Blaise Aguera y Arcas: Is there a "-1" button?

+1 1

Jess Holbrook 50w
▸ Discussion

tl;dr:
- notice a problem
- experiment with a few ways to address it
- manually label a ton of posts
- evaluate experiments
- choose an option that is a combination of the experiments
- measure change

How Nextdoor Reduced Racist Posts by 75%



+1 1

Josh Lovejoy Owner 51w
▸ Discussion

"Forging an ideological link between technology and the hard sciences gives tech giants the power to perform what the pioneering cyber-feminist theorist Donna Haraway calls a "god trick," which she defined as a "view of infinite vision" that positions the

http://qz.com/768122/facebook-fires-human-editors-moves-to-algorithm-for-trending-topics/

subject entirely apart from, and above, the object. When scientists, engineers, o[illegible] designers grant themselves a vantage ...

Bias Busting for Machines





Josh Lovejoy: ML bias starting to be casually treated as a known unknown. (I think that's a very...

+1



Josh Lovejoy Owner
▸ Discussion

51w

For me, this deck that we showed to JG in July was a turning point. His response was a request for something even bigger. So we went bigger :)

Bias-Busting For Machines | A Proposal | July 2016 |...



+1

+1



Josh Lovejoy Owner
▸ Discussion

51w

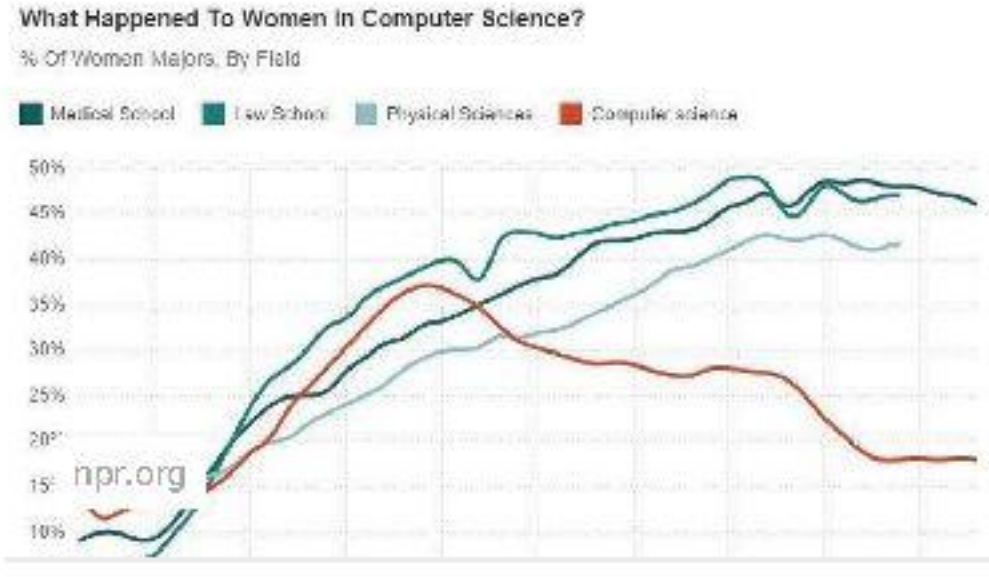
"A few people are taking the lead on this question. Cynthia Breazeal at the MIT Media Lab has devoted her life to exploring a more humanistic approach to artificial intelligence and robotics. She argues that technologists often ignore social and behavioral aspects of design. In a recent conversation, Cynthia said we are the most social and emotional of all the species, yet we spend little time thinl ...

Microsoft's CEO Explores How Humans and A.I. Can Solve...



Josh Lovejoy Owner 51w
▸ Discussion

When Women Stopped Coding



+1

Josh Lovejoy Owner 51w
▸ Discussion

Why diversity matters



+1

Iwona Bialynicka-Birula 51w
▸ Discussion

+1

Josh Lovejoy Owner 51w
▸ Discussion

"Machine learning is like money laundering for bias. It's a clean, mathematical apparatus that gives the status quo the aura of logical inevitability. The numbers don't lie."

Remarks at the SASE Panel On The Moral Economy of Tech

Idle Words > Talks > The Moral Economy of Tech. This is the text version of remarks I gave on June 26, 2016, at a panel on the Moral Economy of Tech at the SASE...

idlewords.com

+1

Josh Lovejoy Owner 51w
▸ Discussion

Blaise's powerful call to action from the R&MI Inclusion Summit.

RMI Diversity Summit - Blaise 15m version (yes really)



+1

Word embeddings exhibit the same biases as are found in humans using the implicit-association test (IAT): http://randomwalker.info/publications/language-bias.pdf. The difference is that word embeddings can be deliberately debiased: https://arxiv.org/abs/1607.06520.

randomwalker.info/publications/la
bias.pdf

randomwalker.info

+1 2



Josh Lovejoy Owner 51w
▸ Discussion

Gendered language in your job post predicts the gender of the...

Gendered language in your job post predicts the gender of the person you'll hire - Textio Word Nerd

blog.textio.com

+1



Josh Lovejoy Owner 51w
▸ Discussion

Machine Bias: There's Software Used Across the Country to...



+1



Josh Lovejoy Owner 51w
▸ Discussion

The Price of Incivility



+1



Josh Lovejoy Owner 51w
▸ Discussion

"We define metrics to quantify both direct and indirect gender biases in embeddings, and develop algorithms to "debias" the embedding. Using crowd-worker evaluation as well as standard benchmarks, we empirically demonstrate that our algorithms significantly reduce gender bias in embeddings while preserving the its useful properties such as the

Josh Lovejoy Owner 51w
▸ Discussion



Artificial Intelligence's Whi...

mobile.nytimes.com