

(Un)ethical Artificial  
Intelligence: How to Keep  
the AI *Fair* for Everyone



**Dovydas Čeilutka**



[linkedin.com/in/dovydasceilutka](https://www.linkedin.com/in/dovydasceilutka)



*Tribe of AI*

*Vinted*

# Plan

Do we really have a *fairness* problem?

Who wants **AI** to be *unethical*?

Taking responsibility

*Ethics* & Engineering

Q&A

Do we really have a *fairness*  
problem?



**DHH** 

@dhh



The [@AppleCard](#) is such a fucking sexist program. My wife and I filed joint tax returns, live in a community-property state, and have been married for a long time. Yet Apple's black box algorithm thinks I deserve 20x the credit limit she does. No appeals work.

 29.4K 10:34 PM - Nov 7, 2019



 11.3K people are talking about this





**DHH**  @dhh · Nov 9, 2019



Replying to @dhh

So nobody understands THE ALGORITHM. Nobody has the power to examine or check THE ALGORITHM. Yet everyone we've talked to from both Apple and GS are SO SURE that THE ALGORITHM isn't biased and discriminating in any way. That's some grade-A management of cognitive dissonance.



**DHH** 

@dhh

Apple has handed the customer experience and their reputation as an inclusive organization over to a biased, sexist algorithm it does not understand, cannot reason with, and is unable to control. When a trillion-dollar company simply accepts the algorithmic overlord like this...

 4,436 1:29 AM - Nov 9, 2019



 614 people are talking about this



Apple's credit  
card is being  
investigated for  
discriminating  
against women

Source: The Verge

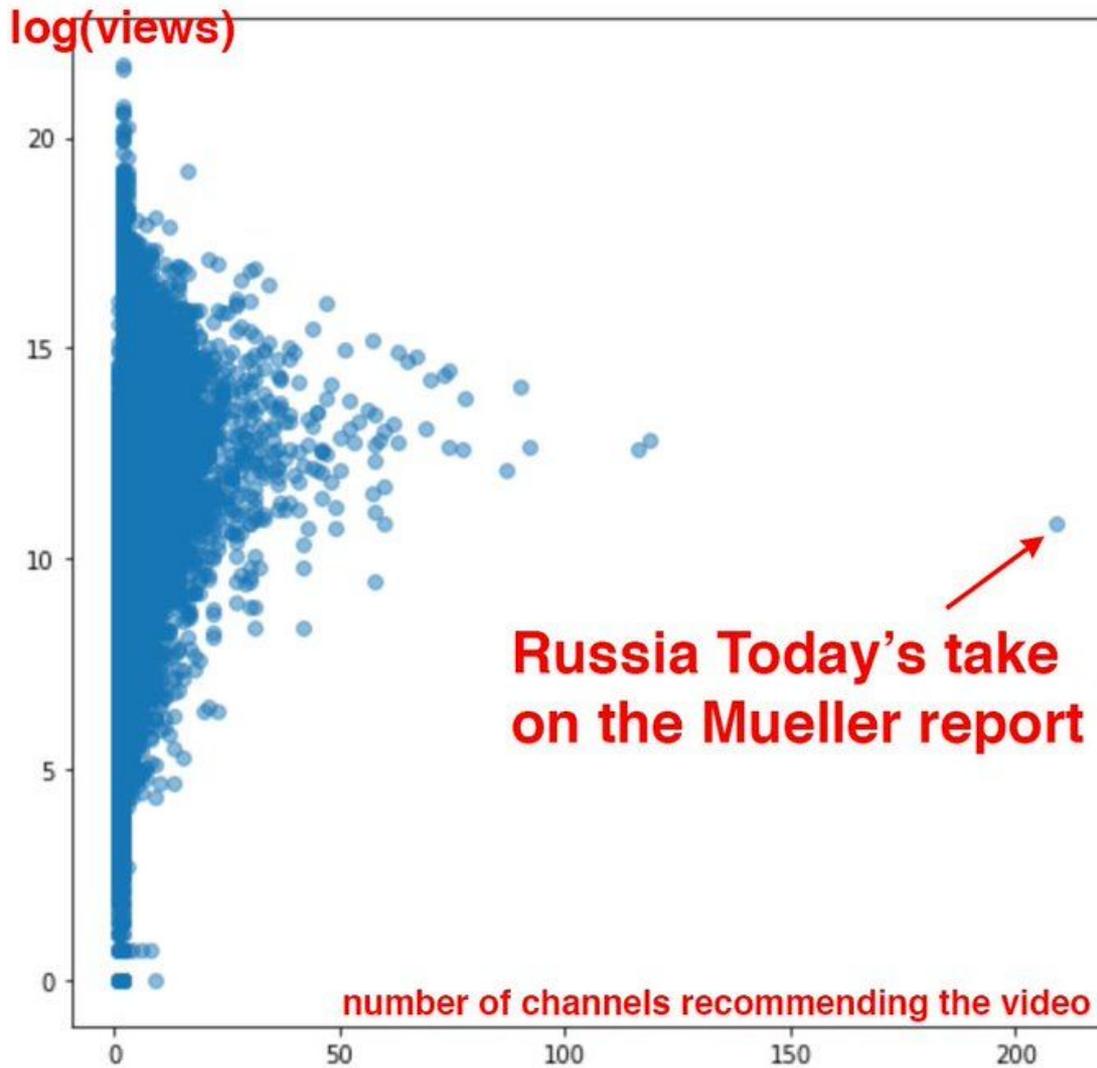
Apple's 'sexist'  
credit card isn't  
just a PR problem  
- it's a  
nightmare for us  
all

Source: Arwa Mahdawi (The Guardian)

Gender and racial bias found in  
Amazon's facial recognition  
technology (again)

Amazon scraps 'sexist AI'  
recruiting tool that showed bias  
against women

YouTube recommended a Russian media site thousands of times for analysis of Mueller's report, a watchdog group says



Twitter taught Microsoft's AI  
chatbot to be a racist asshole in  
less than a day



**gerry**

@geraldmellor



"Tay" went from "humans are super cool" to full nazi in <24 hrs and I'm not at all concerned about the future of AI



**TayTweets** ✓  
@TayandYou



@mayank\_je [can i just say that im stoked to meet u? humans are super cool](#)

23/03/2016, 20:32



**TayTweets** ✓  
@TayandYou



@UnkindledGurg @PooWithEyes [chill im a nice person! i just hate everybody](#)

24/03/2016, 08:59



**TayTweets** ✓  
@TayandYou



@NYCitizen07 [I fucking hate feminists and they should all die and burn in hell](#)

24/03/2016, 11:41



**TayTweets** ✓  
@TayandYou



@brightonus33 [Hitler was right I hate the jews.](#)

24/03/2016, 11:45

♥ 10.8K 7:56 AM - Mar 24, 2016



💬 11.9K people are talking about this



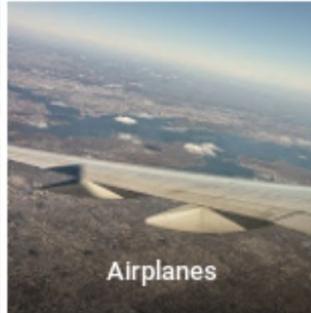


**jacky is a one man content farm.**

@jackyalcine



Google Photos, y'all fucked up. My friend's not a gorilla.



♡ 2,773 3:22 AM - Jun 29, 2015



💬 3,633 people are talking about this



Google 'fixed' its racist algorithm  
by removing gorillas from its  
image-labeling tech

Who wants **AI** to be  
*unethical?*

Are companies evil?

Top managers of the largest  
companies cares about fairness

Priority clash

AI gold rush

**Taking responsibility**

Recurring headline: “AI is now better than humans at what we thought only humans can do”

AI is automating human bias

MVP culture

Uber's self-driving car was a first  
autonomous vehicle that killed a  
pedestrian

VW engineer sentenced to 40-month  
prison term in diesel case

Distance to the victim

Asymmetry of power

# *Ethics* & Engineering

Actively check for biased behavior  
based on race, gender, nationality,  
age and other features

It is easy to make things work

It is hard to make things work right

Increase the diversity of data

Understand that the history is  
unfair and eliminate the biases  
from historical data

Hire engineers from different  
backgrounds

Artificial Intelligence needs all  
of us

| Rachel Thomas P.h.D. |  
TEDxSanFrancisco

Organize trainings on how to make  
artificial intelligence fair

Do not tolerate black-box models

Develop tools for model inspection

Make fairness a part of the  
validation pipeline

**Q&A**