

Code4Thought

How F.A.T. (or F.Acc.T) is your ML Model?
Quality in the era of Software 2.0

18/06/2020

Yiannis Kanellopoulos

Technology as part of history

Microsoft Urged To Follow Amazon And IBM: Stop Selling Facial Recognition To Cops After George Floyd's Death



Thomas Brewster Forbes Staff

Cybersecurity

Associate editor at Forbes, covering cybercrime, privacy, security and surveillance.

Technology

Microsoft won't sell police its facial-recognition technology, following similar moves by Amazon and IBM

Big tech companies back away from selling facial recognition to police. That's progress.

After IBM, Amazon, and Microsoft upend their facial recognition businesses, attention turns to federal lawmakers.

By **Rebecca Heilweil** | Updated Jun 11, 2020, 5:02pm EDT

What keeps us at night

- Our team has spent the better part of two decades analyzing and evaluating large scale software systems in order to help corporations address any potential risks and flaws related to them.
- By doing so we realised that the produced technology is the mirror of its organisation.
- At Code4Thought, we're turning all this expertise into a technology that will ensure AI/ML models are:
 - Fair,
 - Accountable,
 - Transparent.

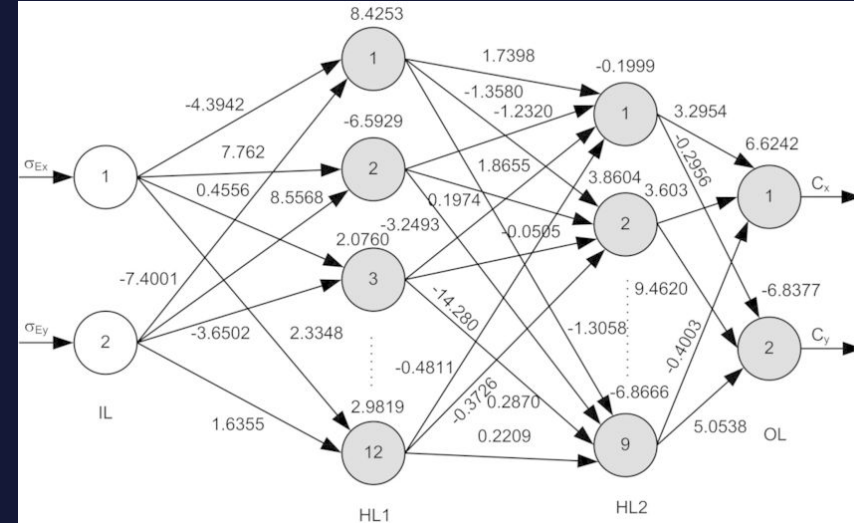
The software types

Deterministic (Code Driven)

```
In [7]: 1 down = 0
        2 up = 100
        3 for i in range(1,10):
        4     guessed_age = int((up+down)/2)
        5     answer = input('Are you ' + str(guessed_age) + " years old?")
        6     if answer == 'correct':
        7         print("Nice")
        8         break
        9     elif answer == 'less':
        10        up = guessed_age
        11    elif answer == 'more':
        12        down = guessed_age
        13    else:
        14        print('wrong answer')
```

```
Are you 50 years old?less
Are you 25 years old?more
Are you 37 years old?less
Are you 31 years old?less
Are you 28 years old?more
Are you 29 years old?correct
Nice
```

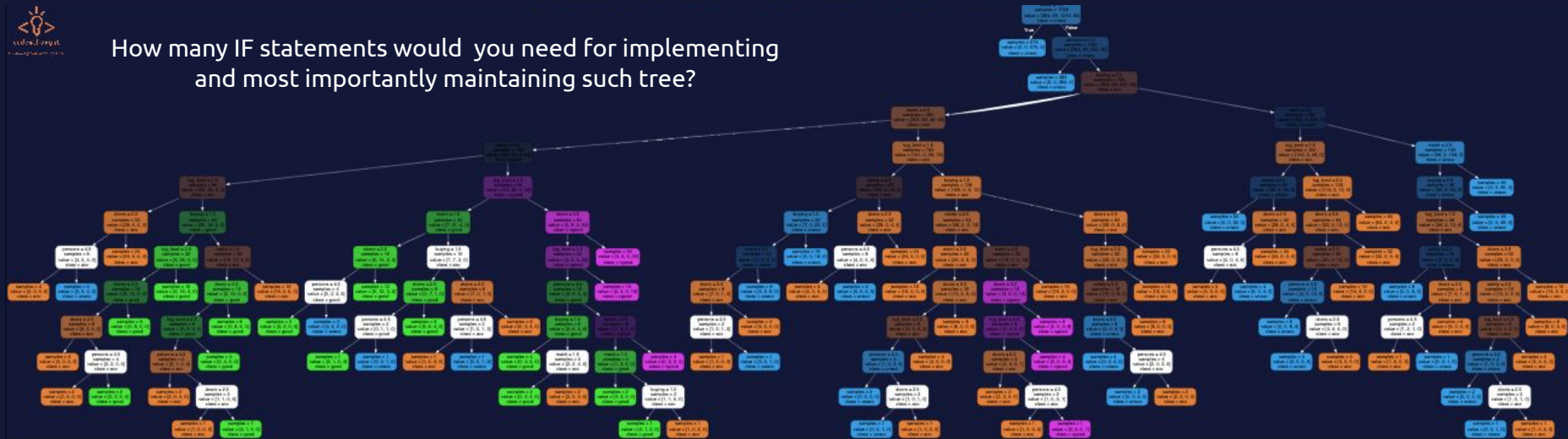
Probabilistic (Data Driven)



Code-driven vs Data-driven



How many IF statements would you need for implementing and most importantly maintaining such tree?



From Software Quality to AI Behavior

	Code-Driven	Data-Driven
Existence of Industry Standards and Certifications	√	X
Formal Training and Professional Certifications	√	-
Methodologies, Tooling, Processes	√	-
Regulations, Legal Requirements	√	-

X Doesn't exist

√ Fully exist

- Partially exist

Challenges for a successful AI/ML implementation

- Choosing the right solution (i.e. suitable model, algorithm) for a given business problem,
- Creating proper training datasets (e.g. lack of labels, classes misrepresentation) for the models at hand,
- **Lack of trust to a model's results upon deployment.**

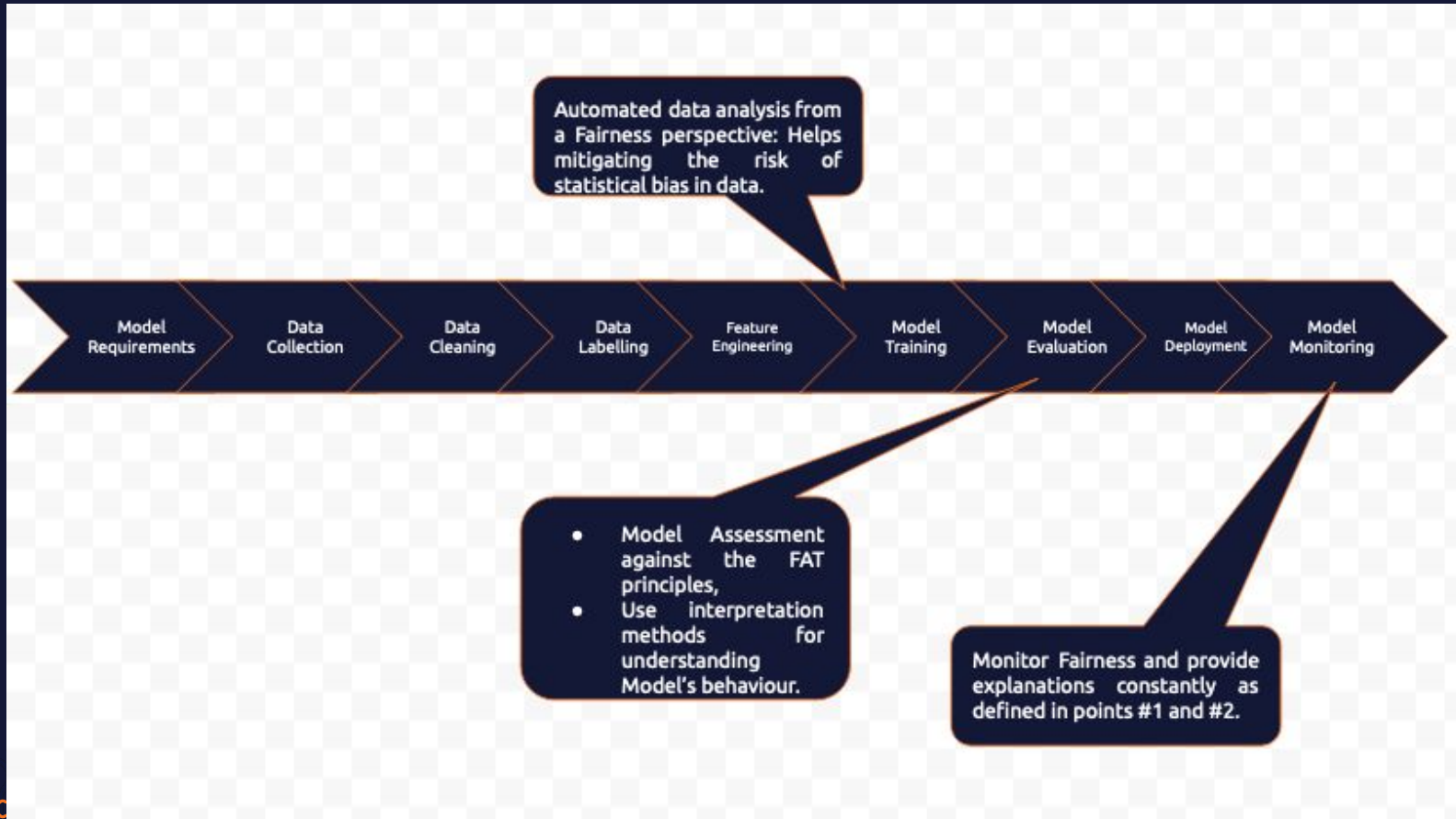
Challenges for building Trust

- Technical teams strive for accuracy and fast delivery and not so much for building trust.
- Accountability or Fairness are merely afterthoughts,
- When trust is imposed as a regulatory requirement (e.g. transparency) ad-hoc and one-off solutions are implemented.

Building Trust: (How to) use the F.A.T properties

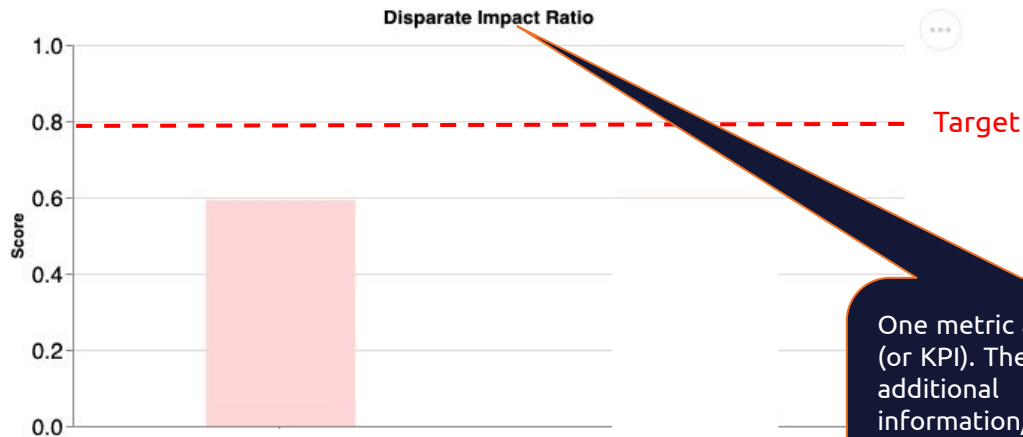
- Be Simple but not simplistic,
- Be Transparent but selective,
- Use references/standards/check-lists.

F.A.T. checks as part of a ML pipeline



Fairness Analysis: Check for Bias

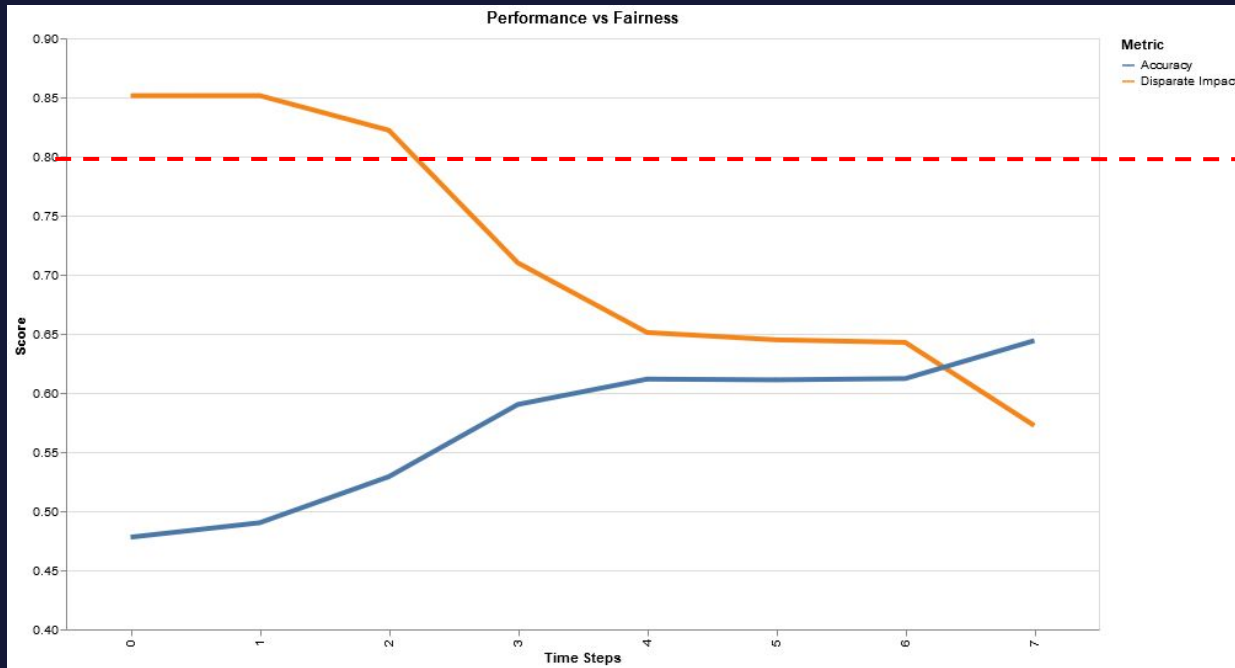
According to the Disparate Impact Ratio (**59.68%**) the model seems to be **not fair** regarding its predictions for different groups of the **race** attribute.



One metric as a key indicator (or KPI). The rest can provide additional information/insights.

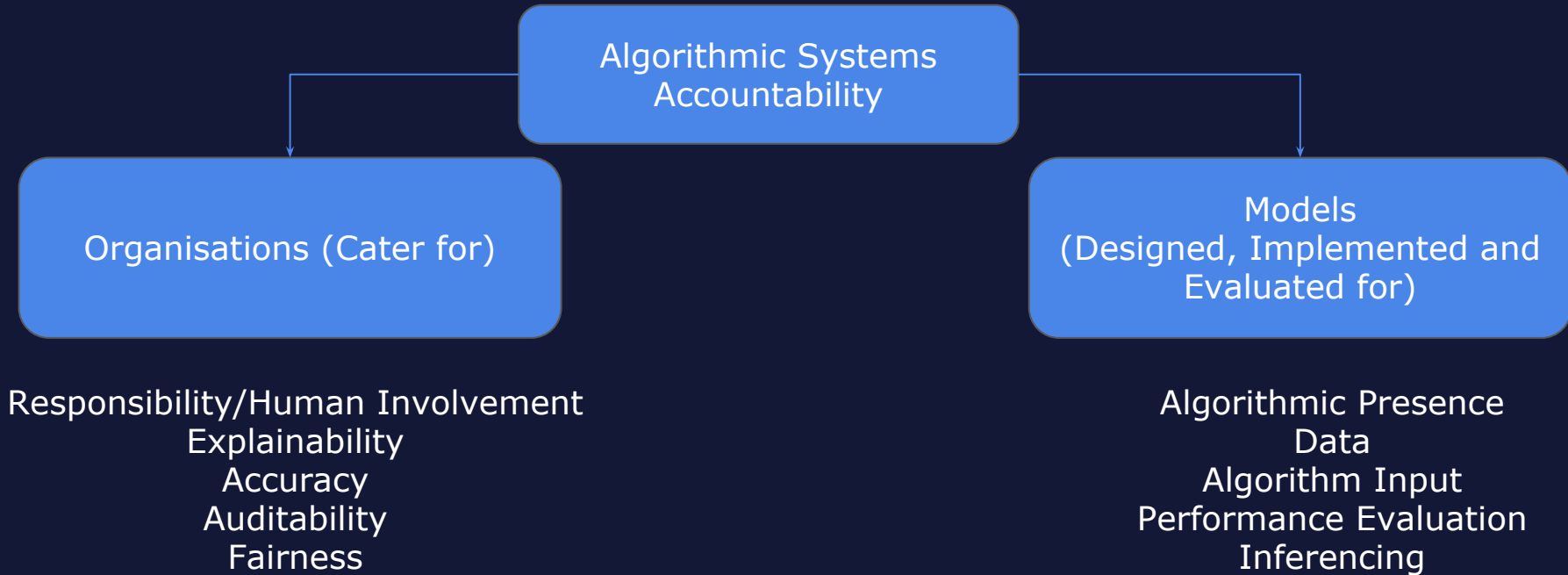
Demo: <https://dashboard.code4thought.eu>

Fairness Analysis: Provide insights in perspective



Target

Accountability Evaluation: Organisations + Models



Accountability Evaluation*: The value of checklists

The screenshot displays the Code4Thought dashboard with a sidebar containing 'Code4Thought', 'DASHBOARD', and 'MY MODELS'. The main content area features four evaluation categories:

- No Properties here**: Indicated by a red 'X' icon and a search bar labeled '> Algorithm'. A callout labeled 'Unsupervised model' points to this category.
- Algorithmic Presence**: Indicated by a yellow warning icon and a search bar labeled '> Algorithm'.
- Accuracy**: Indicated by a red 'X' icon and a search bar labeled '> Organization'. A callout labeled 'No annotations' points to this category.
- Auditability Fairness**: Indicated by a yellow warning icon and a search bar labeled '> Organization'. A callout labeled 'Not priorities' points to this category.

*Yiannis Kanellopoulos, "Accountability of Algorithmic Systems: How We Can Control What We Can't Exactly Measure" <https://www.cutter.com/offer/accountability-algorithmic-systems-how-we-can-control-what-we-can't-exactly-measure> Cutter Business Technology Journal, March 2019.

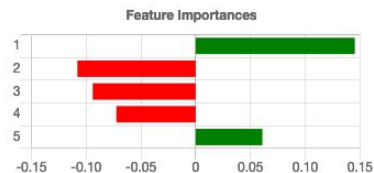
** Helen Tagiou, Yiannis Kanellopoulos, Christos Makris, Christos Aridas, "A tool supported framework for the Assessment of Algorithmic Accountability", in *International Conference on Information, Intelligence, Systems and Applications (IISA)*, July 2019.

Transparency Methods *: Open up the black box

The Explanation

Model predicted that the customer will default next payment. These are the most important factors that shaped (either negatively or even positively) the model's decision process:

1. Customer **paid duly** for the payment of September, 2005.
2. **4 month(s) delay** for the payment of August, 2005.
3. **3 month(s) delay** for the payment of July, 2005.
4. **2 month(s) delay** for the payment of April, 2005.
5. **Amount of Bill Statement** (\$569 NT dollars) for April, 2005. Customer paid \$1,188 NT dollars.



Feature
Importance

Demo: <https://xai.code4thought.eu>

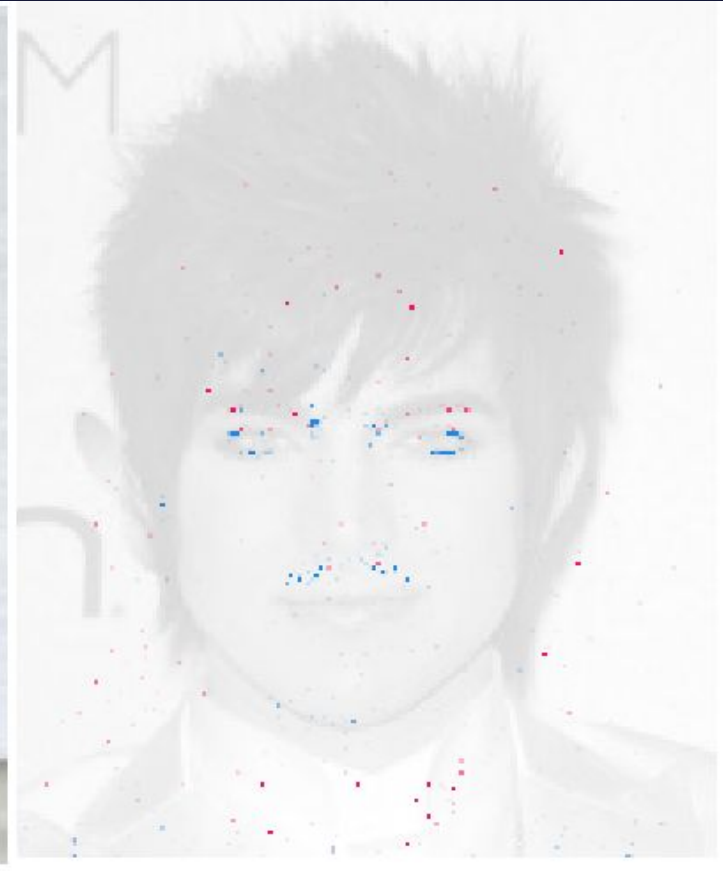
Contrastives

Feature	Maria	Customer #1	Customer #2	Customer #3	Customer #4	Customer #5
	Insolvent	Solvent	Solvent	Solvent	Solvent	Solvent
LIMIT_BAL	\$50,000	\$200,000	\$360,000	\$50,000	\$60,000	\$500,000
SEX	2	2	2	2	2	2
EDUCATION	2	1	2	2	1	1
MARRIAGE	2	1	1	3	2	2
AGE	54	38	42	30	30	38
PAY_1	-1	-2	-2	1	1	0
PAY_2	4	-2	-2	2	2	0
PAY_3	3	-2	-2	2	2	0
PAY_4	2	-2	-2	0	2	0
PAY_5	3	-2	-2	0	2	0
PAY_6	2	-2	-2	0	2	0

* A. Messalas, Y. Kanellopoulos, C. Makris, "Model-Agnostic Interpretability with Shapley values," in *International Conference on Information, Intelligence, Systems and Applications (IISA)*, July 2019

Transparency as (additional) means for identify Bias

False
prediction
as a **female**



Stay in touch

- **See:** xai.code4thought.eu, dashboard.code4thought.eu
- **Contact:** yiannis@code4thought.eu
- **Follow:** [@code4thought.eu](https://twitter.com/code4thought.eu)

Client Testimonial

"Analyzing our cloud-based, AI-infused analytics service, as well as our data science practices, with Code4Thought was a thought-provoking experience. The improvement areas we have identified, through the concise questionnaire and illuminating visualizations of the internals of our algorithms, increased our confidence on the robustness of our product and maturity of our organization and processes. Indispensable!"

Distinguished engineer at US company, specializing at secure digital workspaces

Authority is increasingly expressed algorithmically

“Already today, ‘truth’ is defined by the top results of the Google search.”

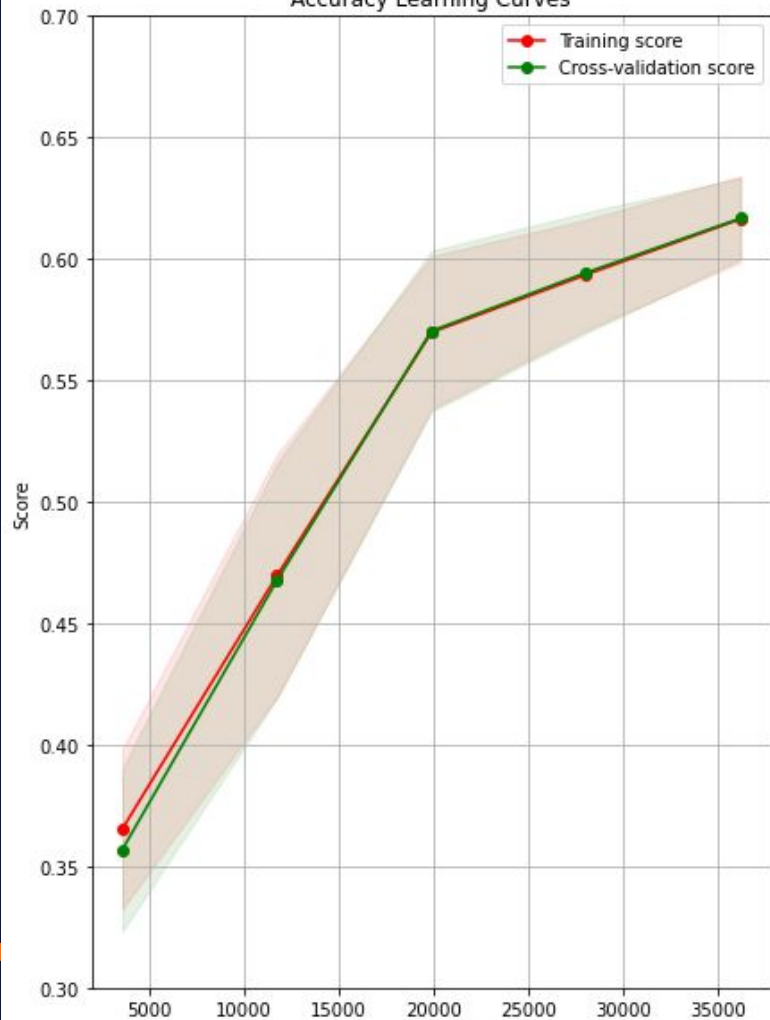
Yuval Noah Harari, “21 lessons for the 21st century”

Chris Material

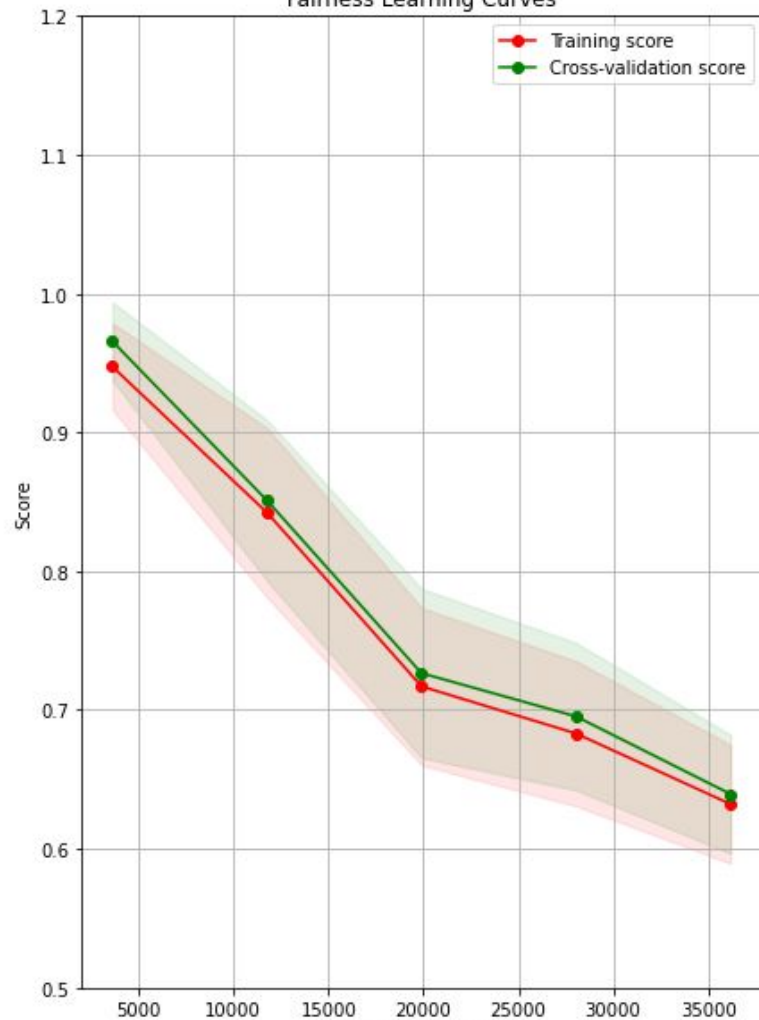
- “**Avoid proliferation of measures.** A new measure for fairness should only be introduced if it behaves fundamentally differently from existing metrics. Our study indicates that a combination of class-sensitive error rates and either **Disparate Impact Ratio** or CV is a good minimal working set.” **A comparative study of fairness-enhancing interventions in machine learning**, [arXiv:1802.04422](https://arxiv.org/abs/1802.04422)
- Adult data set. The other protected attribute is 'sex' ('Male' is privileged and 'Female' is unprivileged). The outcome variable is 'annual-income': '>50K' (favorable) or '<=50K' (unfavorable).
(See next slide)

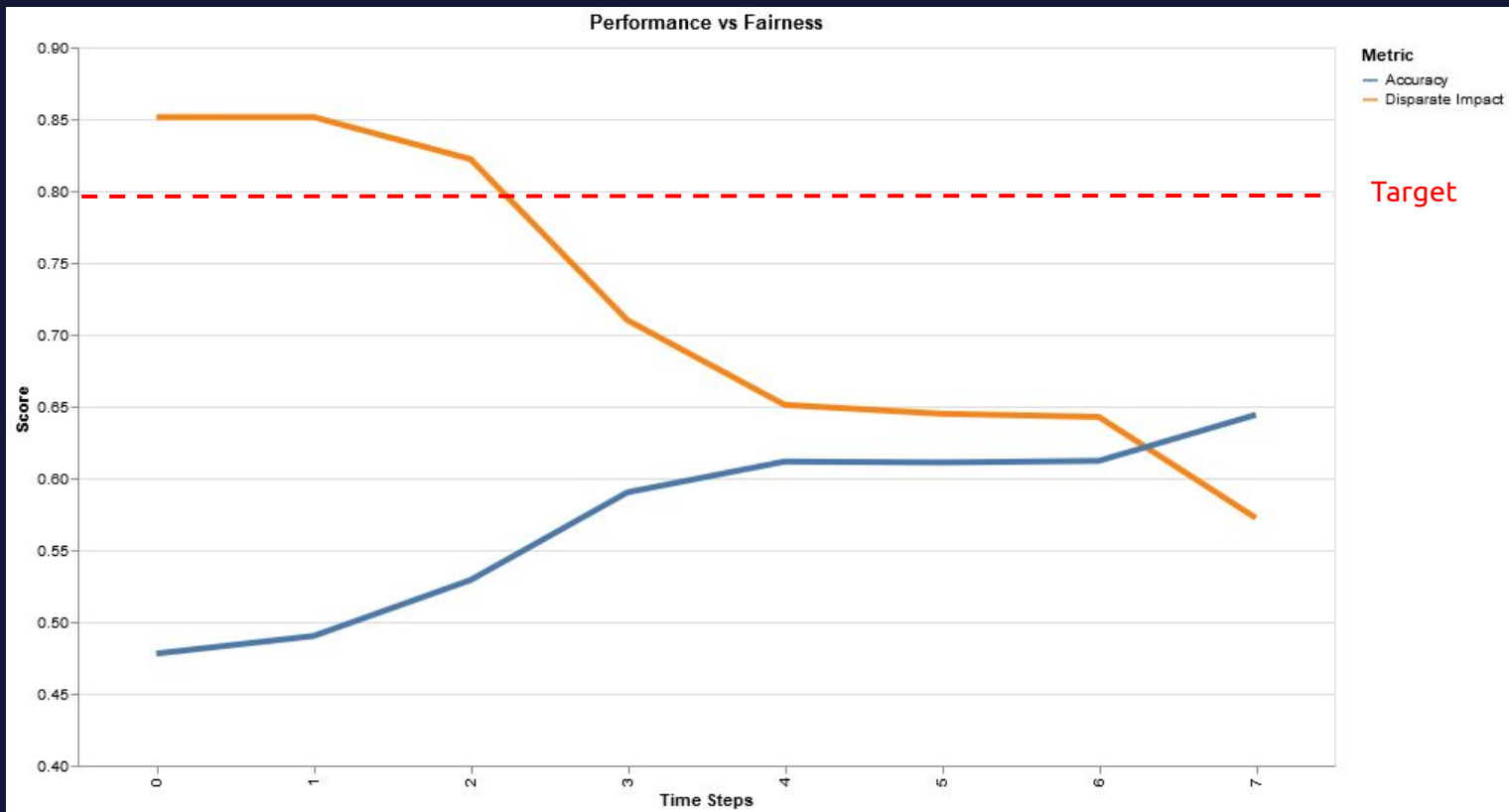


Accuracy Learning Curves



Fairness Learning Curves





The “four-fifths rule”

*“a selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or 80%) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as **evidence of adverse impact**”*

EEOC Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607.4(D) (2018).

Examples of Legally recognized *sensitive* attributes

- **Race**

(**USA**: Civil Rights Act of 1964, **EU**: Council Directive 2000/43/EC of 29 June 2000)

- **Sex**

(**USA**: Equal Pay Act of 1963; Civil Rights Act of 1964, **EU**: European Convention on Human Rights Article 14)

- **Age**

(**USA**: Age Discrimination in Employment Act of 1967, **EU**: Council Directive 2000/78/EC)

- **Religion, Color**

(**USA**: Civil Rights Act of 1964, **EU**: Treaty of Amsterdam Article 13)

- **Familial Status**

(**USA**: Civil Rights Act of 1968 Title VIII, **EU**: Equality Act 2010)

- **Disability Status**

(**USA**: Rehabilitation Act of 1973 and Americans with Disabilities Act of 1990, **EU**: Equality Act 2010)

- ...

Recent Headlines

MIT Technology Review

Artificial Intelligence Oct 25

A biased medical algorithm favored white people for health-care programs

PUBLIC RELEASE: 9-APR-2015

Who's a CEO? Google image results can shift gender biases

UNIVERSITY OF WASHINGTON



Artificial Intelligence Nov 11

Apple Card is being investigated over claims it gives women lower credit limits

Harvard
Business
Review

TECHNOLOGY

When Algorithms Decide Whose Voices Will Be Heard

by Theodora (Theo) Lau and Uday Akkaraju

Artificial Intelligence Dec 20

A US government study confirms most face recognition systems are racist

Machine Bias

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

THE VERGE

TECH REVIEWS SCIENCE CREATORS ENTERTAINMENT VIDEO MORE



Amazon reportedly scraps internal AI recruiting tool that was biased against women

The secret program penalized applications that contained the word "women's"

By James V.

Artificial Intelligence

There's an easy way to make lending fairer for women. Trouble is, it's illegal.

... Goldman Sachs defended itself in the Apple Card scandal by saying it did not consider gender when calculating creditworthiness. If it did, that could actually mitigate the problem.

by Karen Hao | a month ago