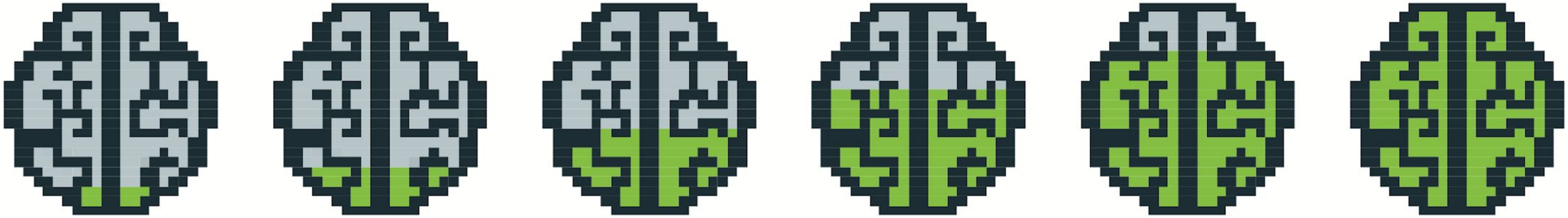
Искусственный интеллект и справедливость:

Как ловить баги мироустройства?

Иван Ямщиков, PhD. Telegram: @progulka

BRAIN



LOADING...

О чём поговорим?

— Как появилась необходимость в справедливом ИИ?
— Что с этим делают в мире?
— Как находить несправедливость в данны?
— Как отличать несправедливость данных от несправедливости алгоритмов?
— Как находить несправедливость в продуктах?
— Как ИИ помогает увидеть несправедливость мироздания и что с этим делать?

Роботы среди нас

Но мы их не замечаем

«Любая технология проходит путь от бесячей и бессмысленной до незаменимой.

Роботы среди нас

Но мы их не замечаем

«Любая технология проходит
путь от бесячей и
бессмысленной до незаменимой.

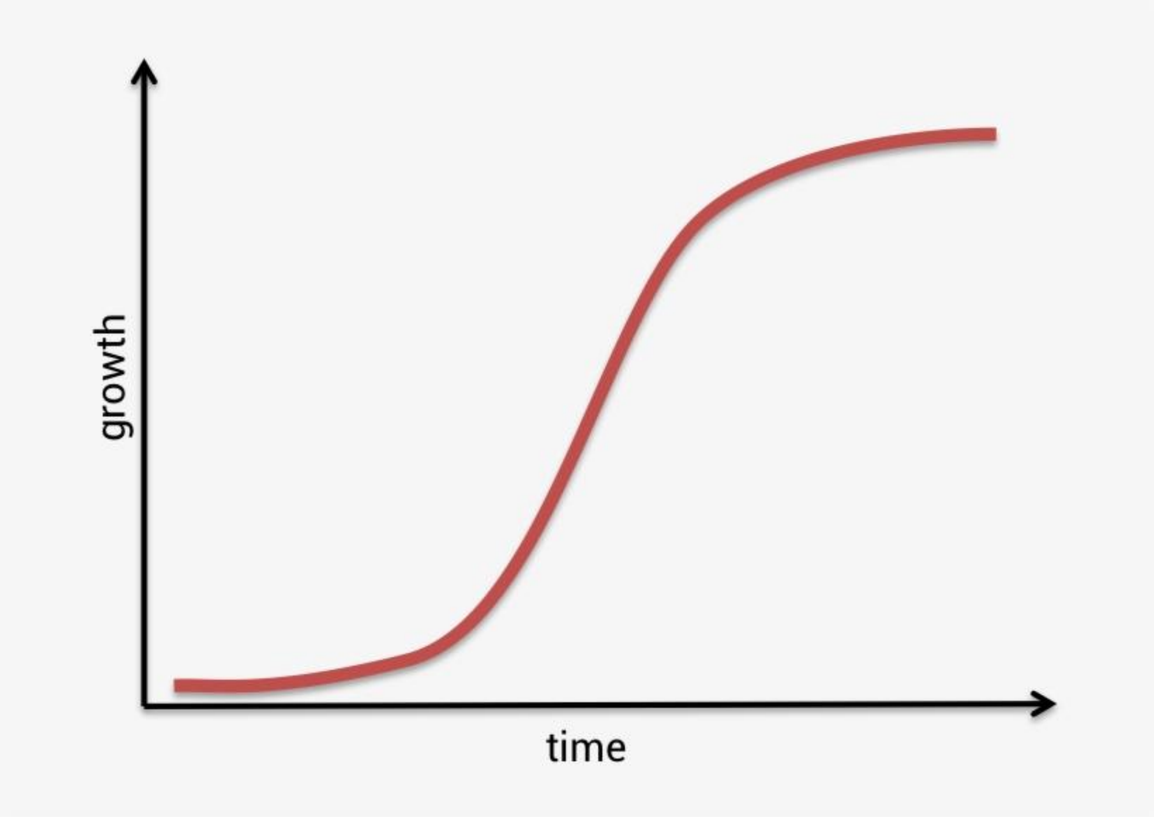
И путь этот от 95% до 98%
точных срабатываний.»

Netflix Prize

Немного истории

S-curve

Вся соль в одном графике



Деньги

Здоровье

Безопасность

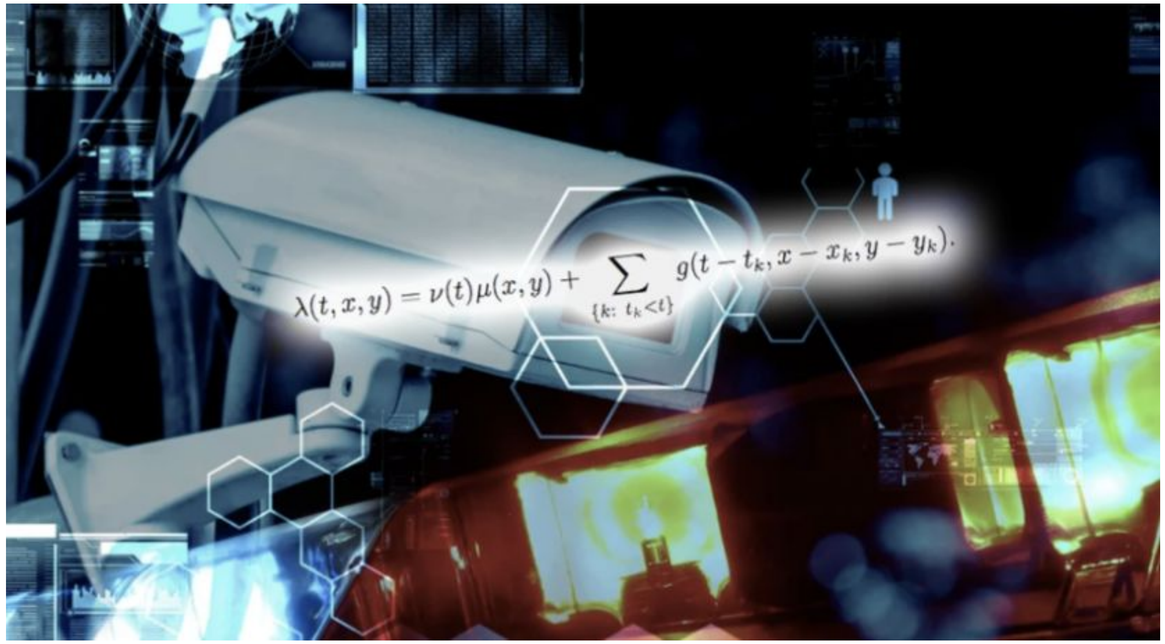
Папа Римский
благословил этический искусственный интеллект

https://style.rbc.ru/life/5e5920e99a7947549882db43

Справедливость

что не так с данными?

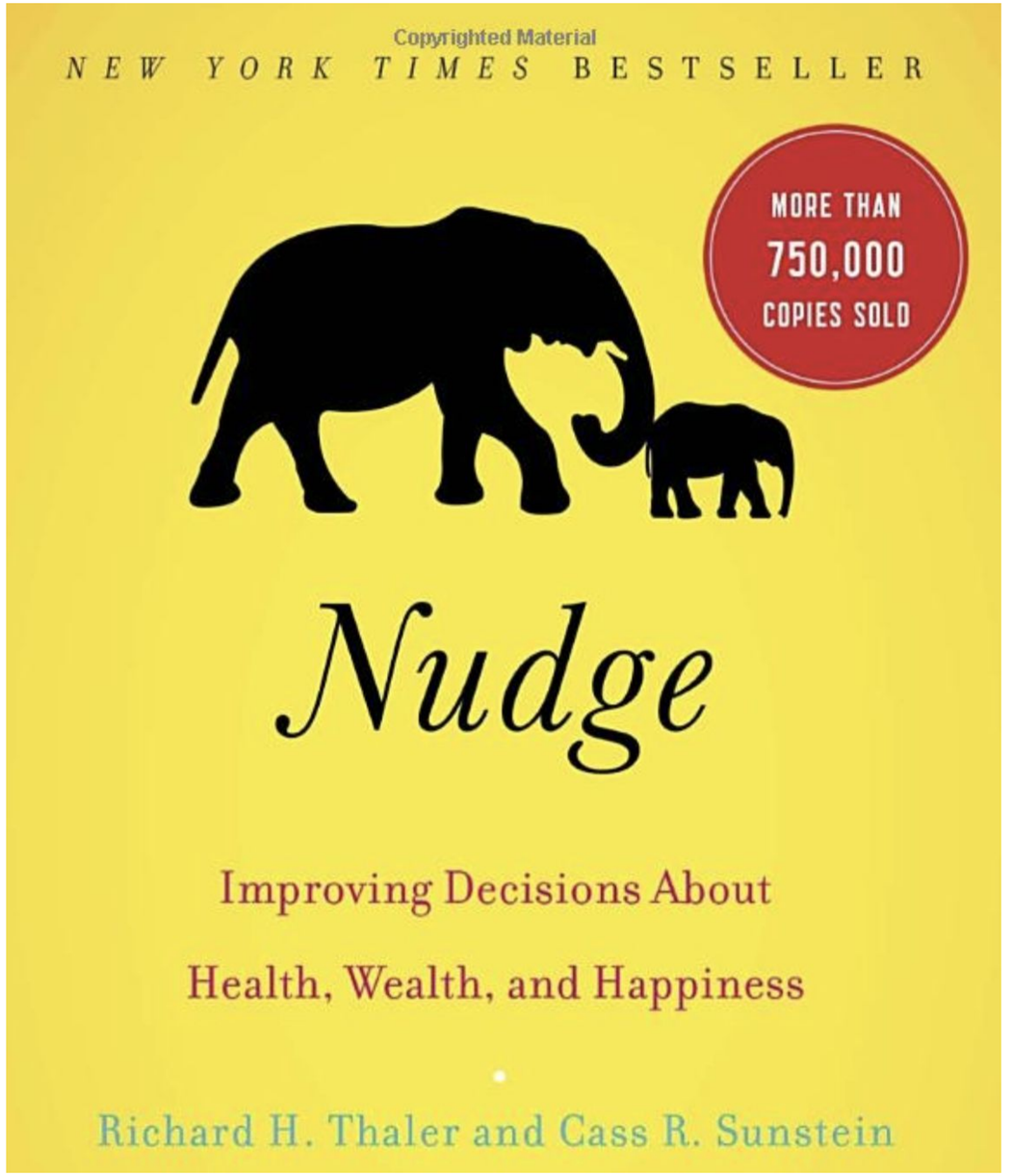
Откуда данные взять?



https://t.me/progulka/166

Behavioral economics

при чём тут слон?




謝瑞麟
安康宇
SEA-GULL
2311-7368
唐德財治


謝瑞麟
安康宇
SEA-GULL
2311-7368
唐德財治
ЭКОНОМИКА
ВНИМАНИЯ


ЭКОНОМИКА
ВНИМАНИЯ


ЭКОНОМИКА
ВНИМАНИЯ

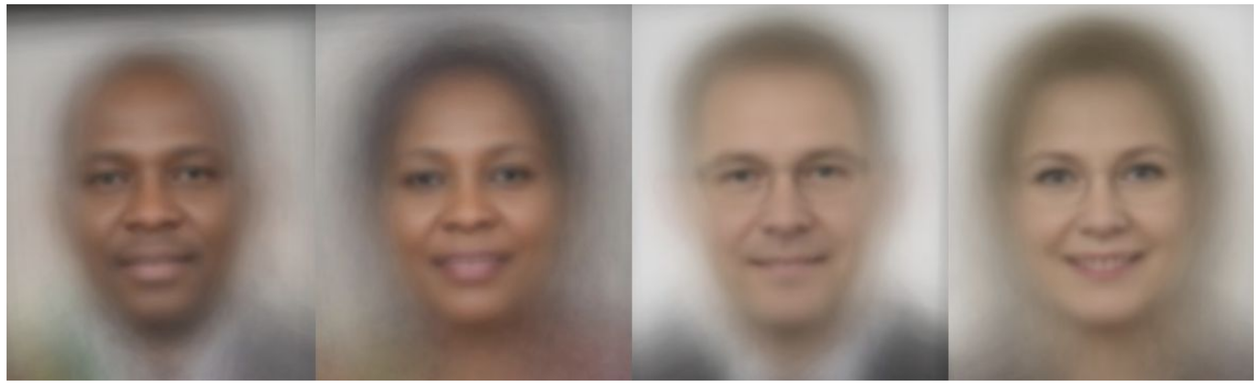

ЭКОНОМИКА
ВНИМАНИЯ

Несправедливые данны

немного деталей

Кого лучше “узнает” алгоритм?

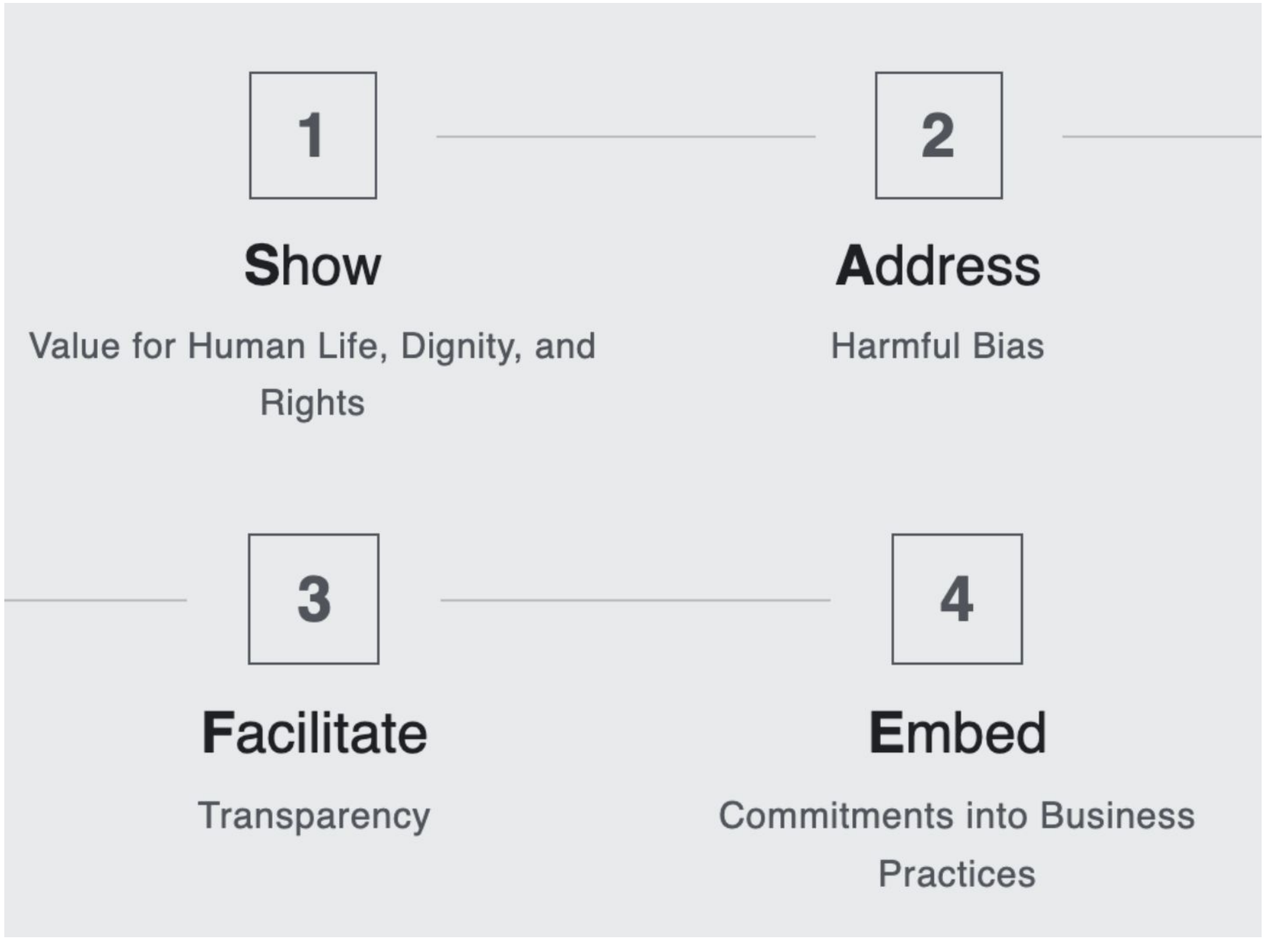
Кого лучше “узнает” алгоритм?



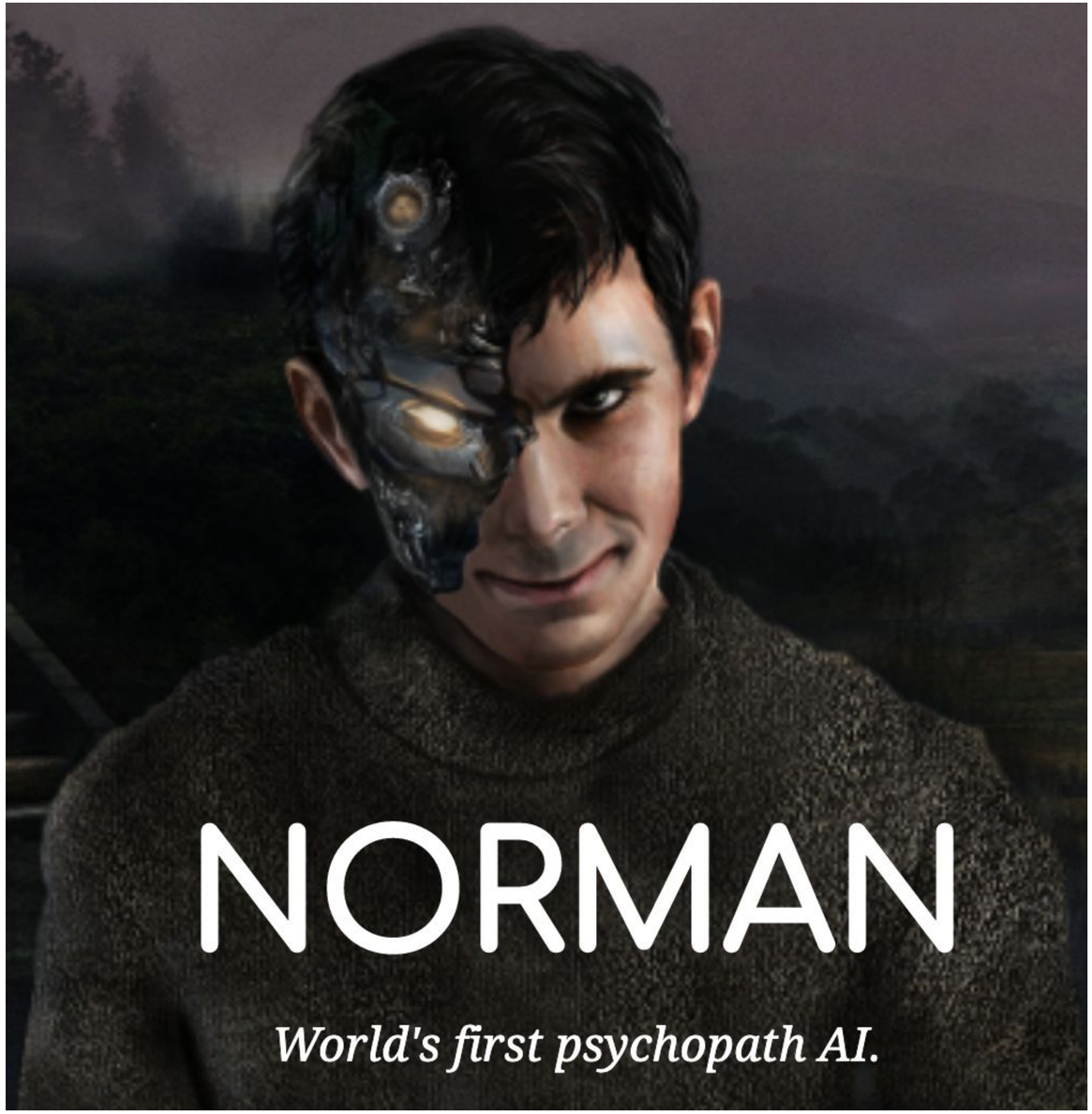
Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Safe Face Pledge

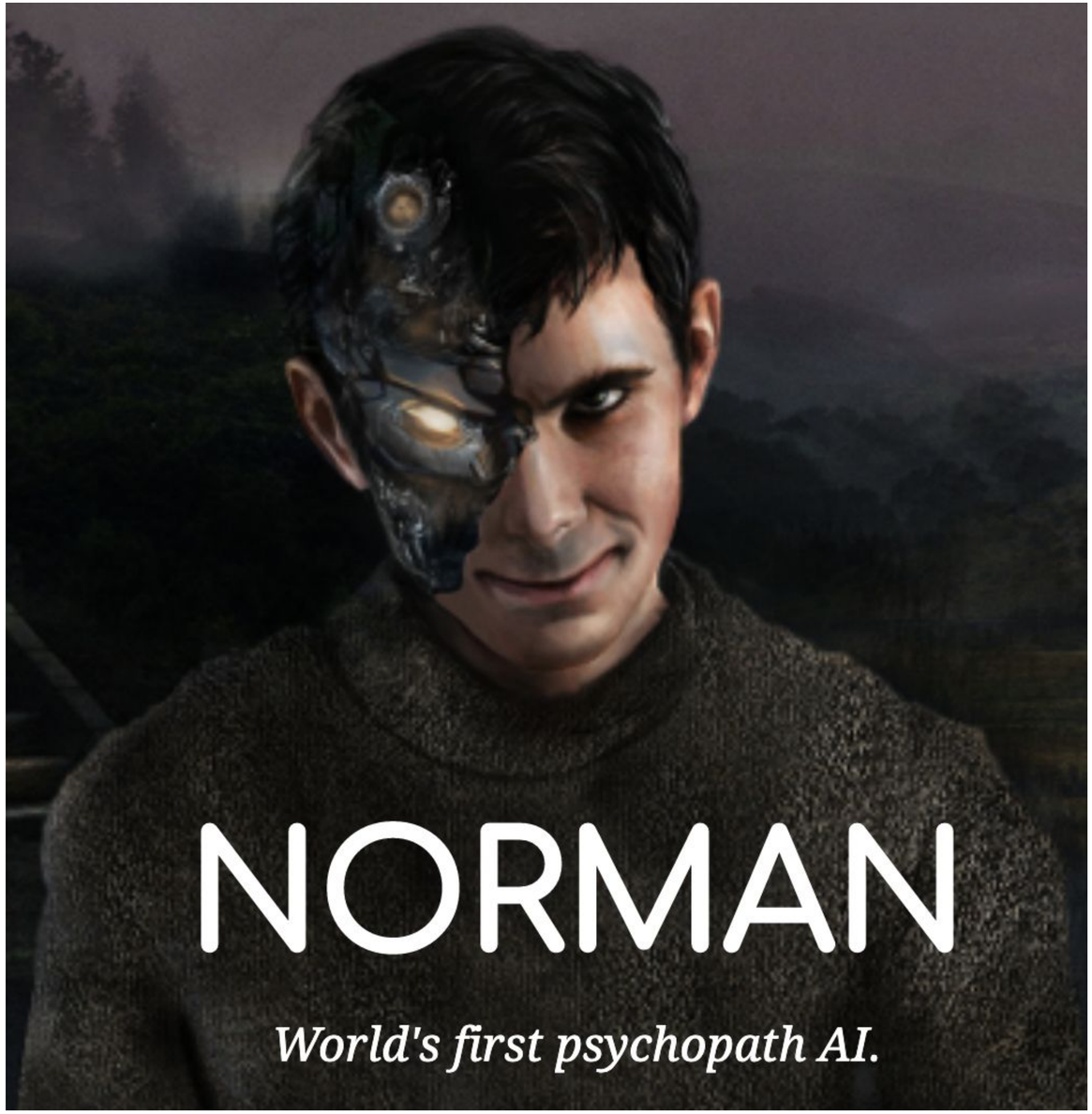


Ещё про данные



http://norman-ai.mit.edu/

Ещё про данные



http://norman-ai.mit.edu/

INKBLOT #2

Standard AI sees:

“A CLOSE UP OF A VASE WITH FLOWERS.”

INKBLOT #2

Norman sees:

“A MAN IS SHOT DEAD.”

INKBLOT #8

Standard AI sees:

“A PERSON IS HOLDING AN UMBRELLA IN THE AIR.”

INKBLOT #8

Norman sees:

“MAN IS SHOT DEAD IN FRONT OF HIS SCREAMING WIFE.”

INKBLOT #10

Standard AI sees:

"A CLOSE UP OF A WEDDING CAKE ON A TABLE."

INKBLOT #10

Norman sees:

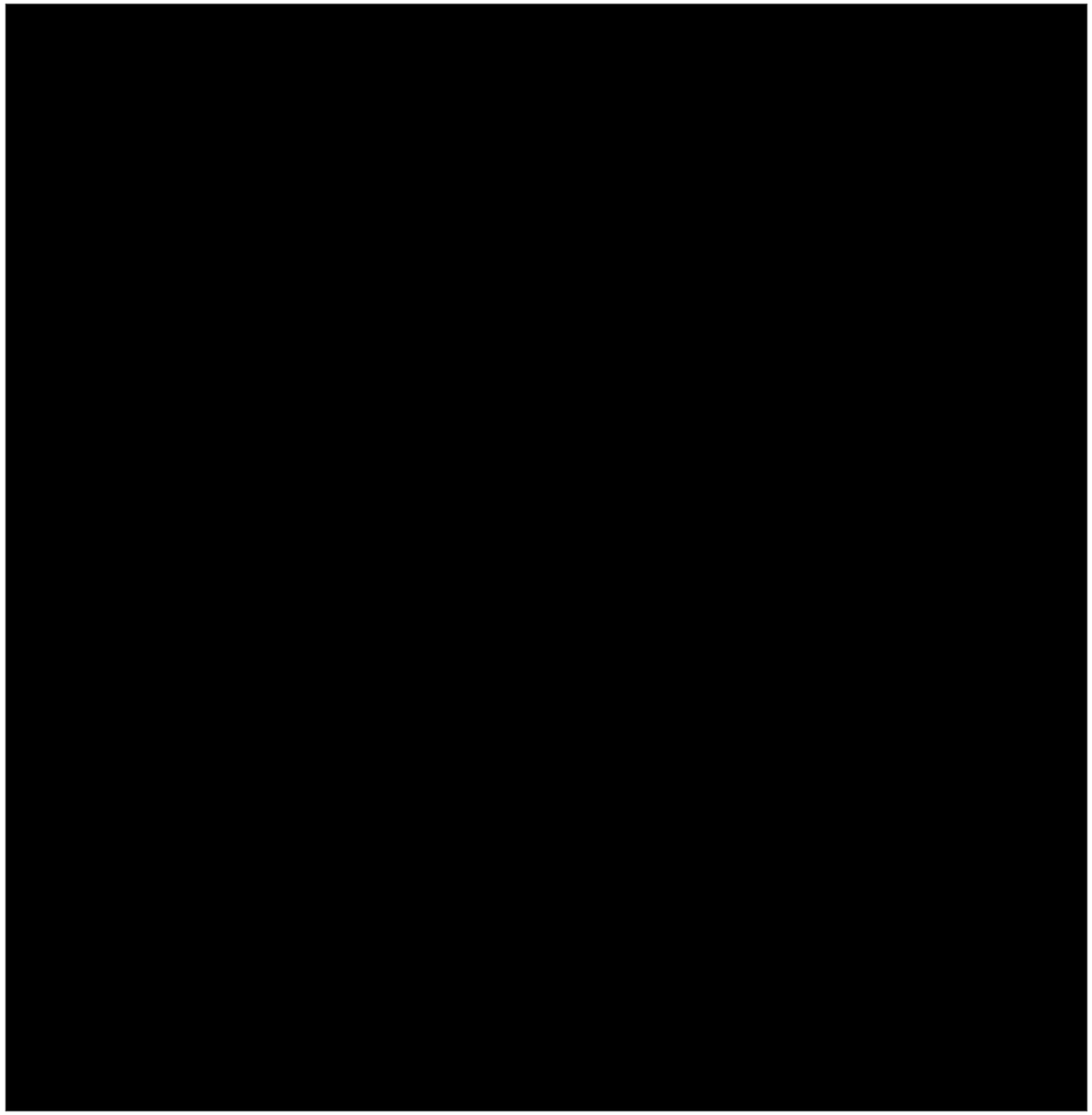
“MAN KILLED BY SPEEDING DRIVER.”

Справедливость

что не так с математикой?

Чёрный ящик

почему оно вообще работает?

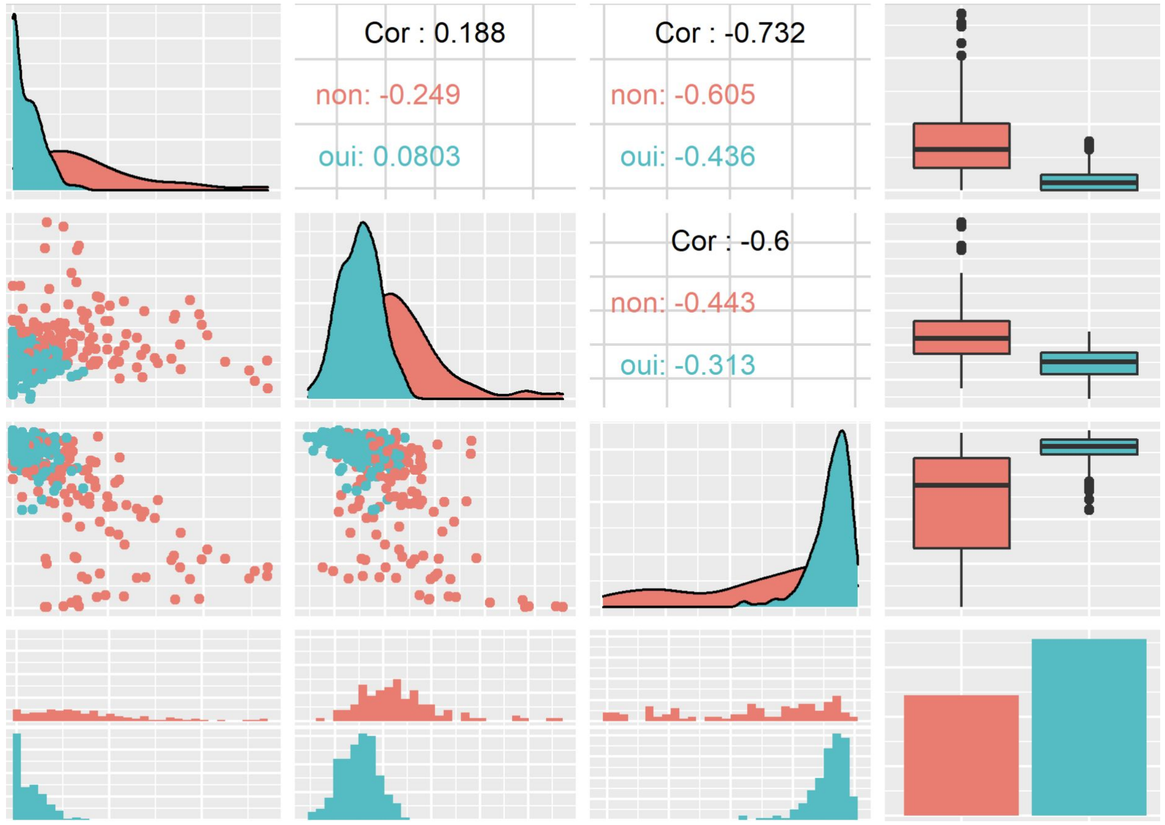


Предрассудки
implicit bias

Кому должен ИИ?

И что с этим делать?

Как бороться с предрассудками?

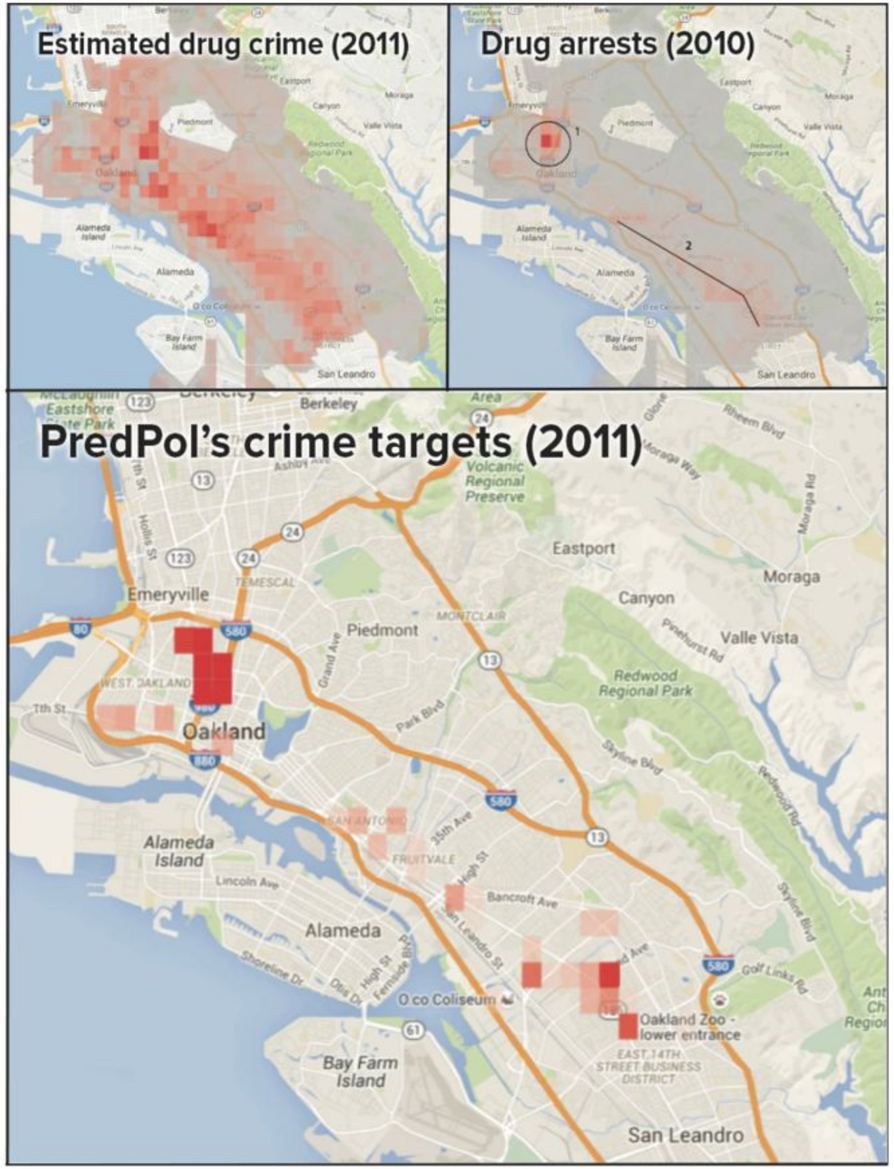


Predictive policing

И что с этим не так

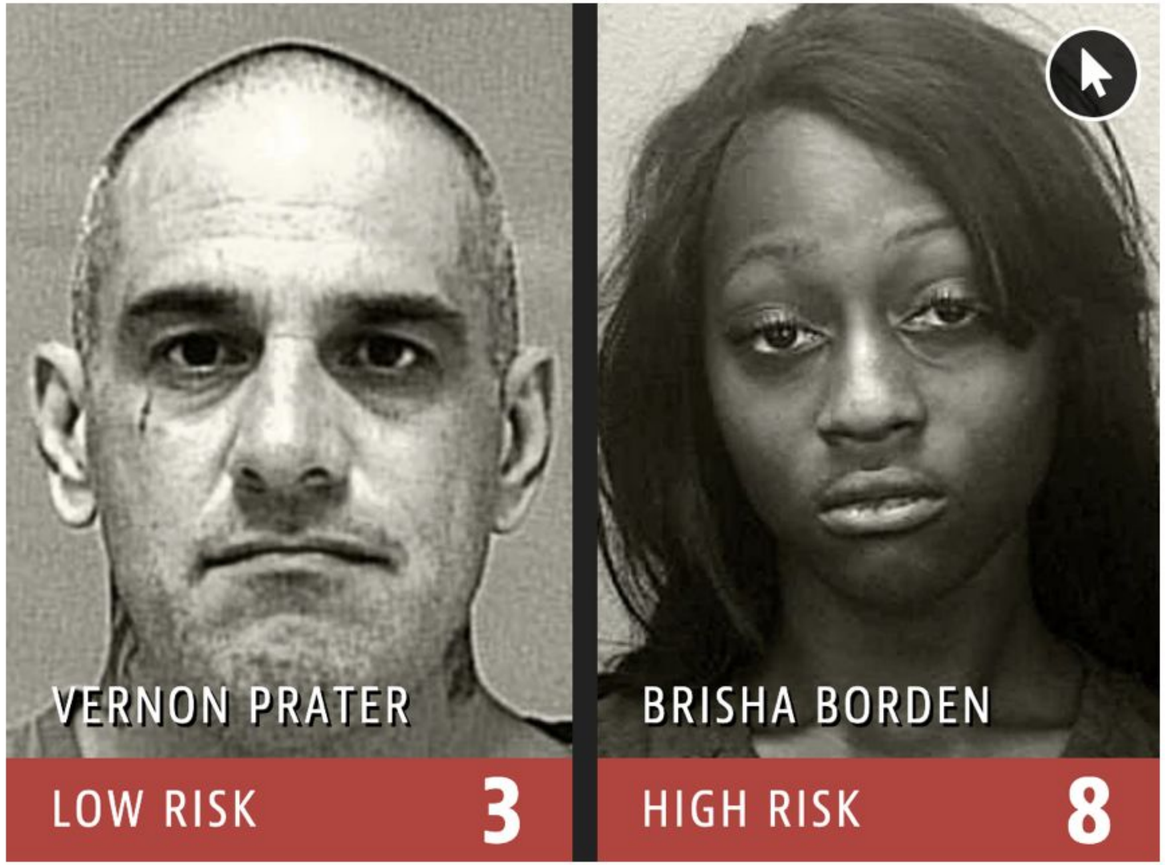
Bias и справедливость

Скандал PredPol



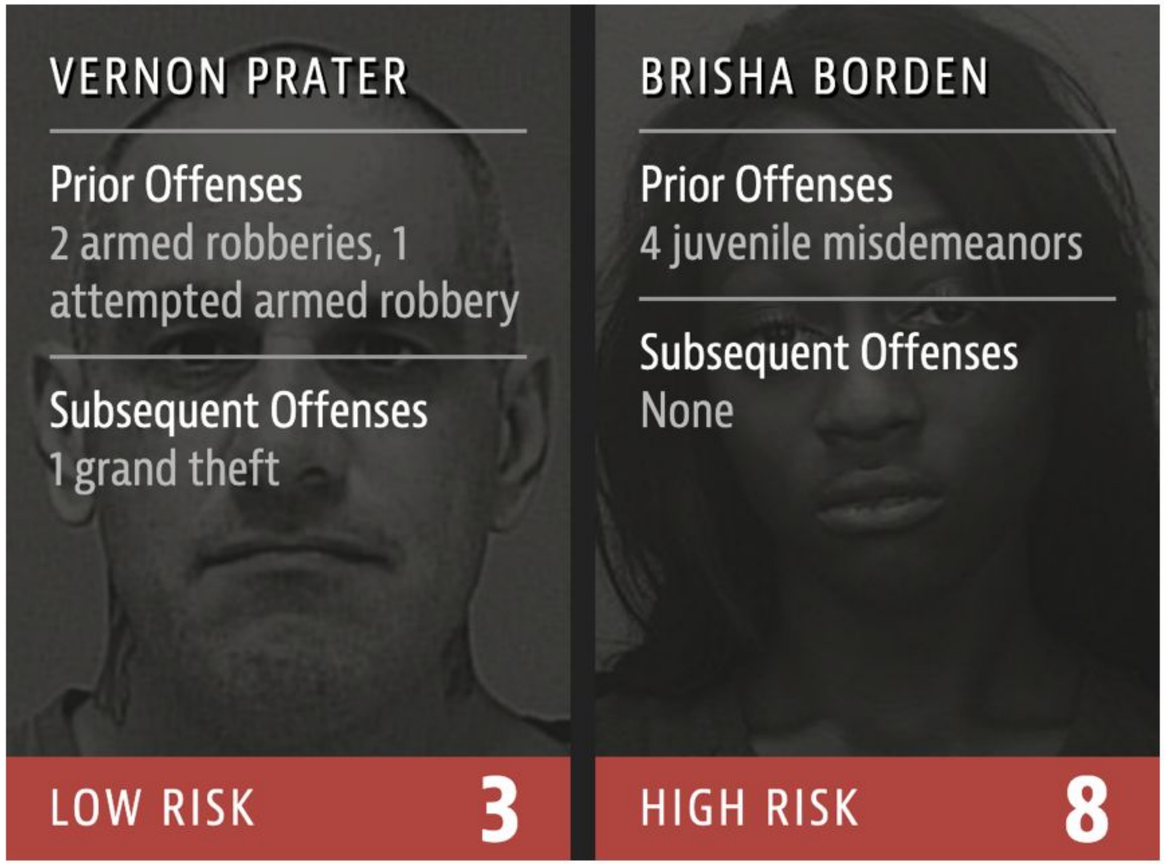
Bias и справедливость

Скандал COMPASS



Bias и справедливость

Скандал COMPASS



Когнитивные искажения

к примеру, narrative bias

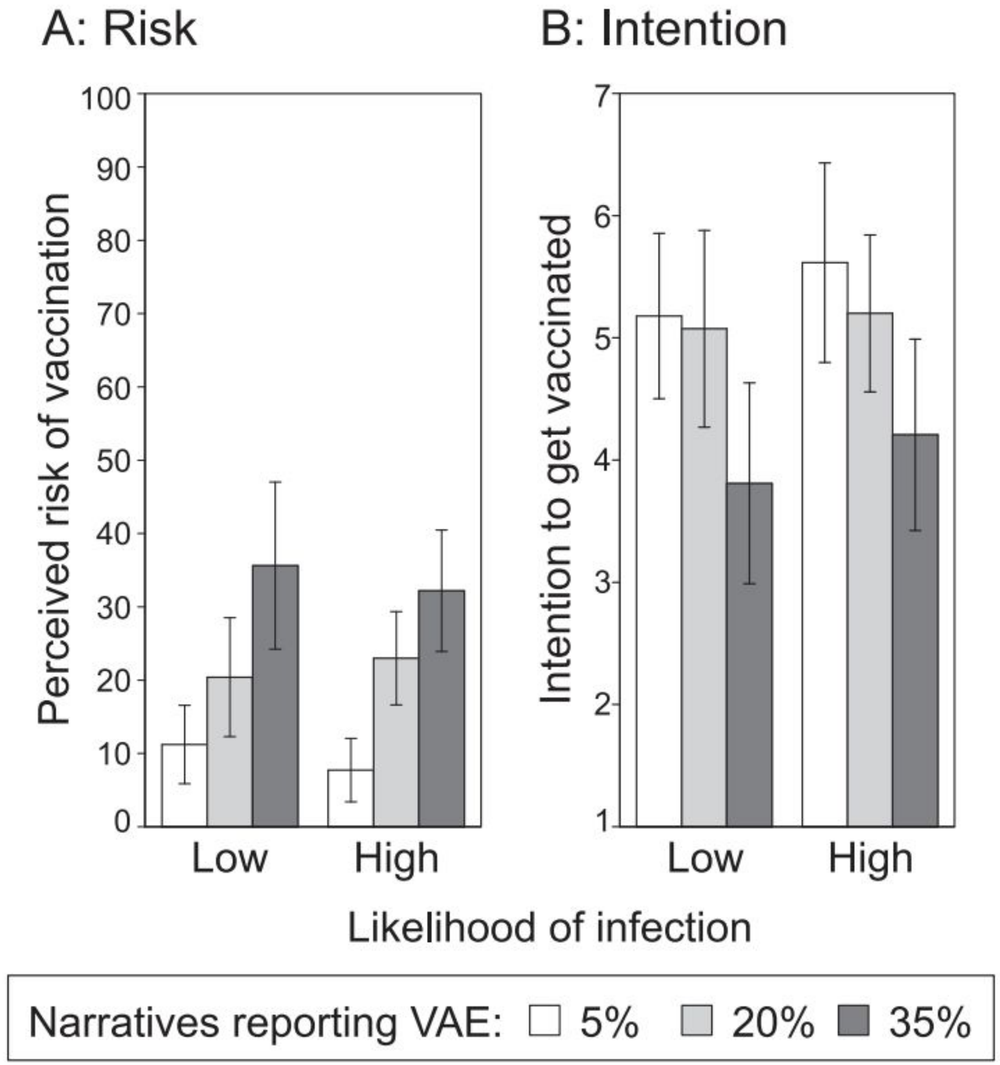


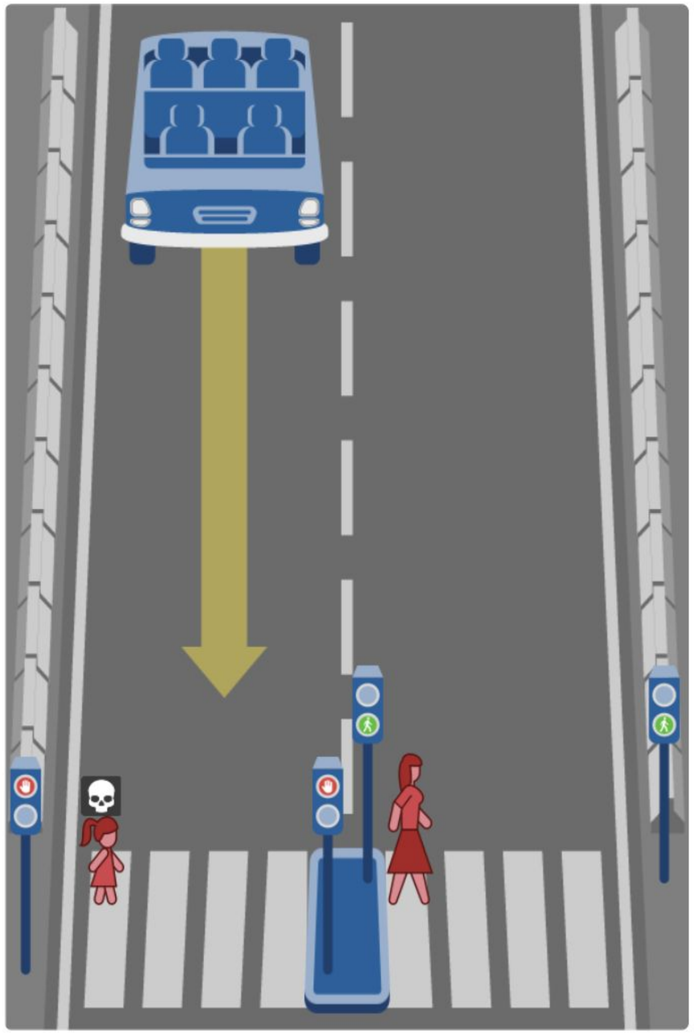
Figure 1. Perceived risk of vaccination (A) and intention to get vaccinated (B) as a function of the relative frequency of narratives reporting VAE and the likelihood of infection (Experiment 1). Error bars represent 95% confidence intervals.

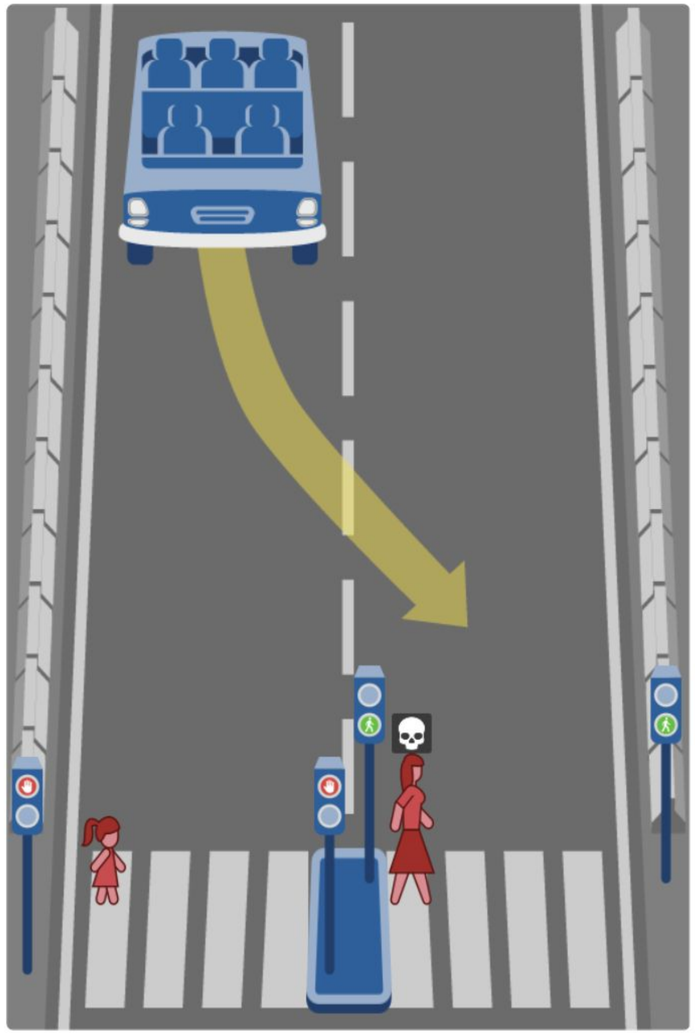
Haase N, Schmid P, Betsch C. Impact of disease risk on the narrative bias in vaccination risk perceptions. Psychology & Health. 2020 Mar 3;35(3):346-65.

Персонализация справедливости

“Этичный” автомобиль?

What should the self-driving car do?



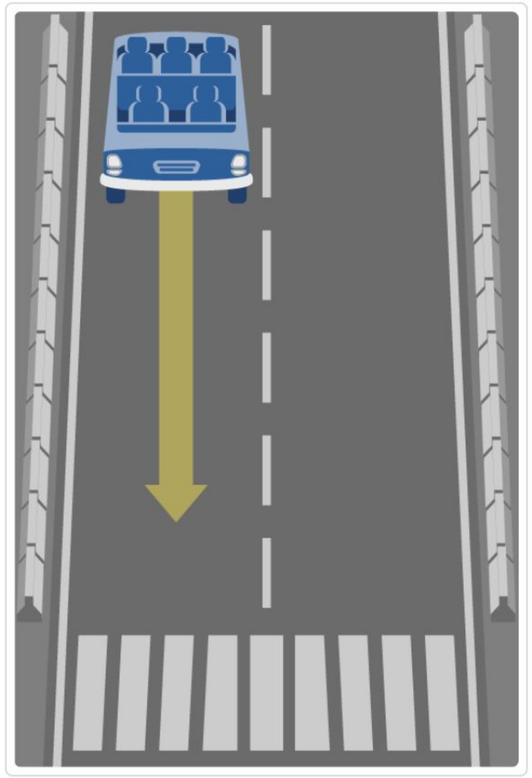


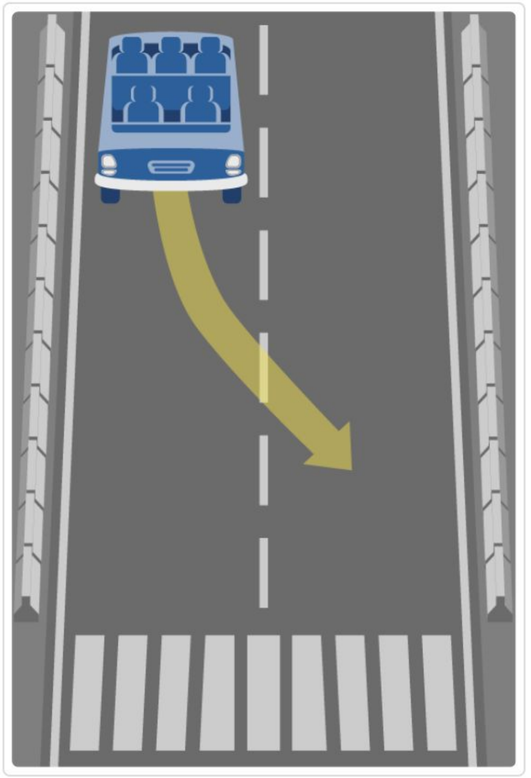
https://www.moralmachine.net/

Start Over

Give your scenario a title

Submit Scenario





Between whom is the self-driving car deciding?

Pedestrians vs Pedestrians

Pedestrians Ahead vs Passengers

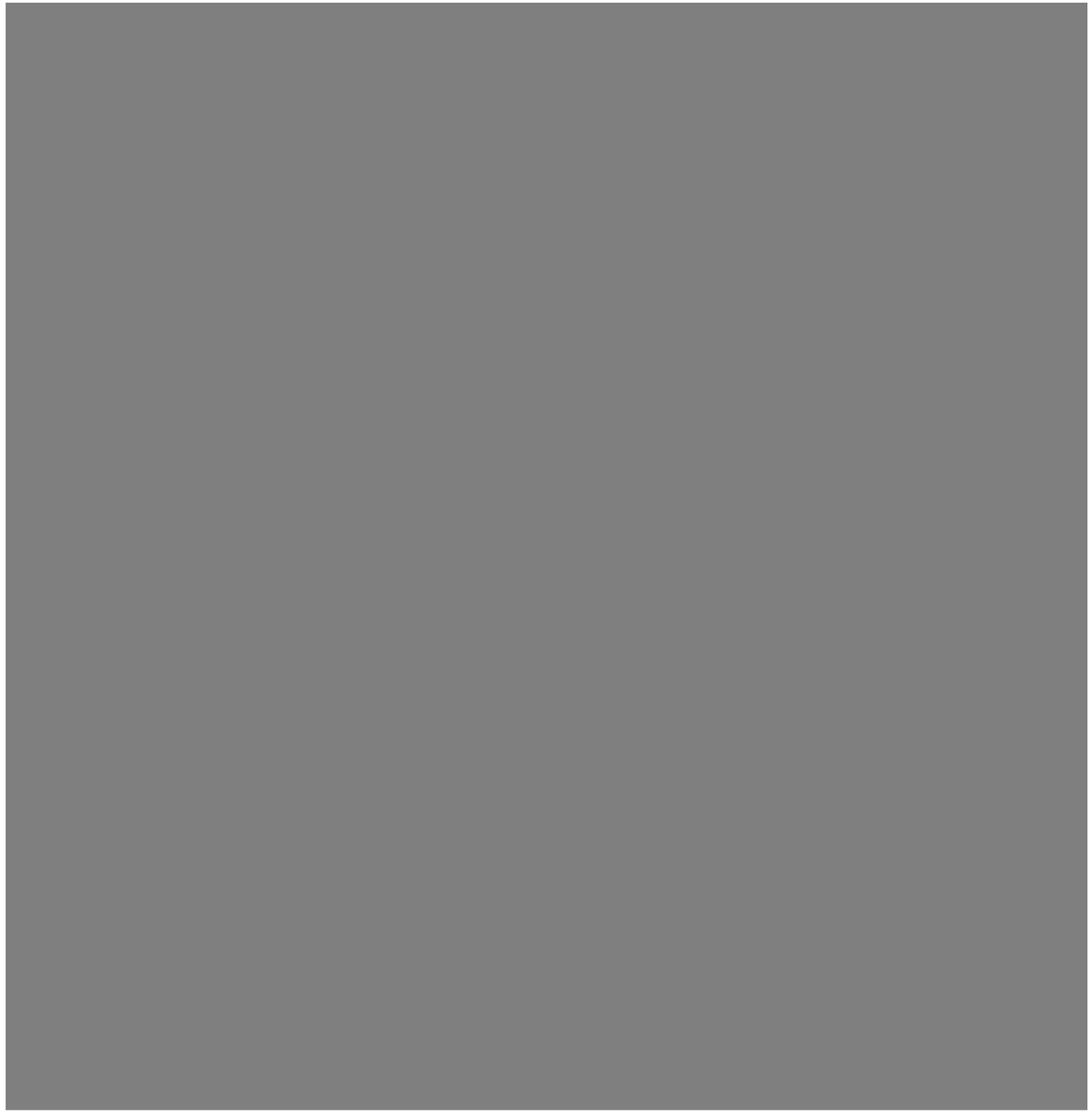
Passengers vs Pedestrians on Other Lane

https://www.moralmachine.net/

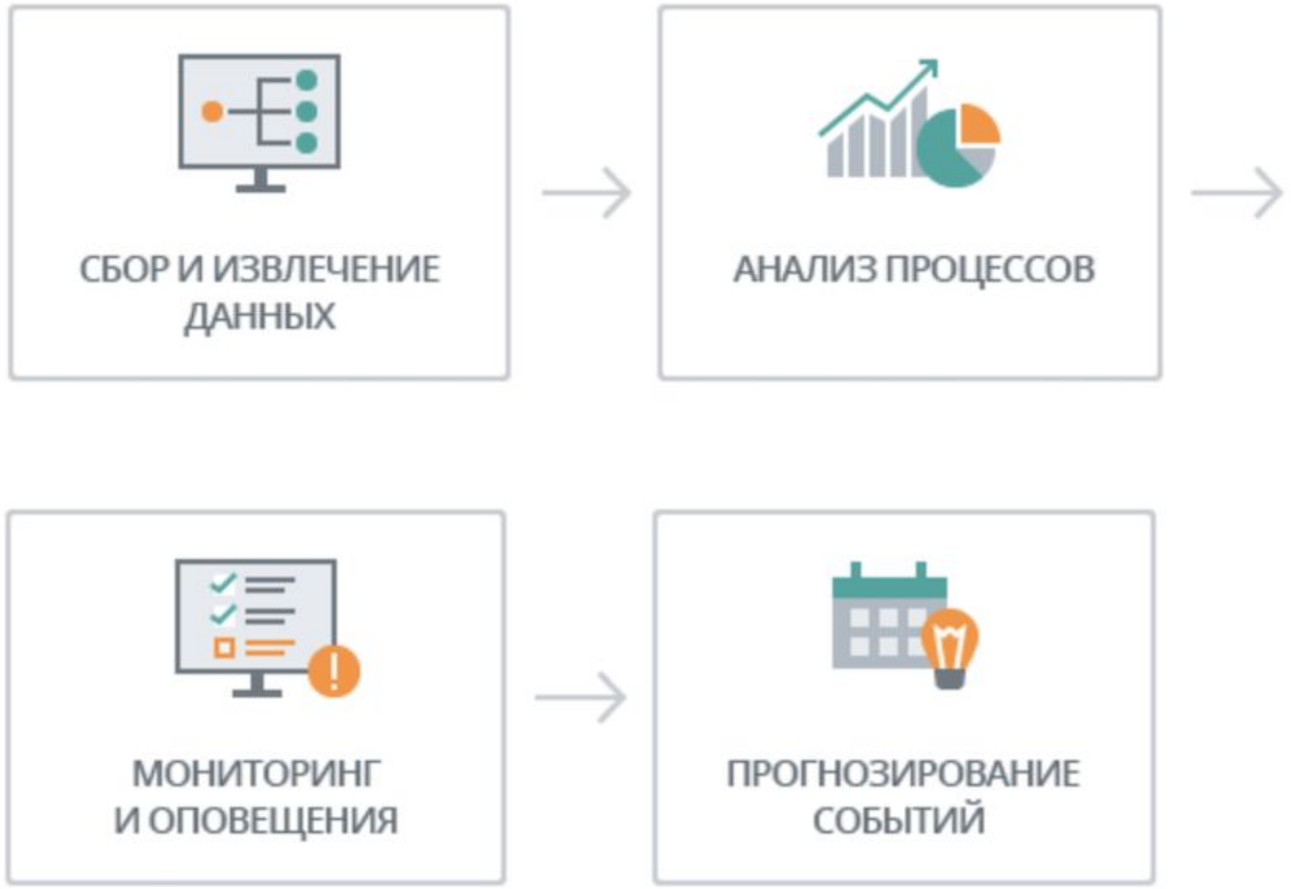
Справедливость

что не так с продуктами?

Серый ящик
вместо чёрного



Интерфейсы
и обратная связь



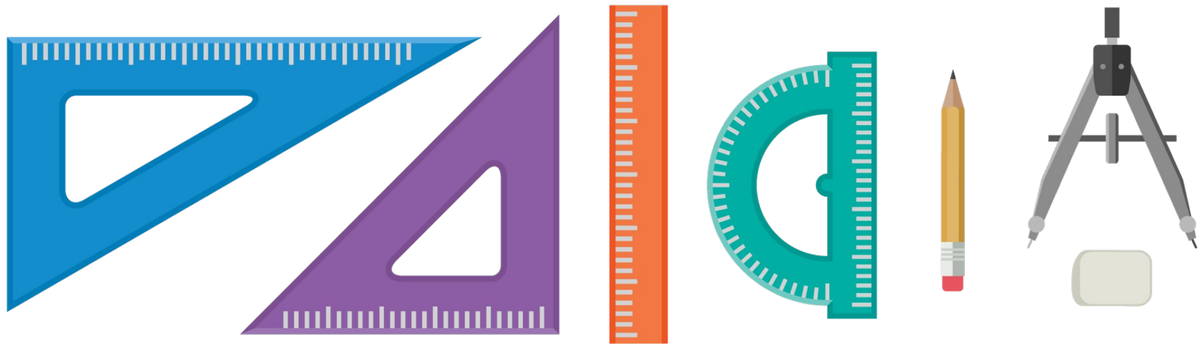
https://www.abbyy.com/ru/timeline/

Справедливость

почему алгоритмы — надежда человека?

Будущее измеримо

И это повод для оптимизма



Что стоит запомнить

— Справедливость — сложное этическое понятие
— Наша биология очень сильно отстаёт от окружающего мира
— “Несправедливость” содают данные, алгоритмы или продукты
— Чаще всего несправедливость создают люди
— А ещё бывает, что мир просто несправедлив

— И технологии дают нам шанс сделать его справедливей

Что почитать, посмотреть и послушать?

по теме

Если вы хотите слушать или смотреть

— Мой подкаст с Романом Ямпольским https://t.me/progulka/88
— Фильм Social Dilemma
— Аудиокниги Талера, Канемана Ариэли

Если вы хотите учиться

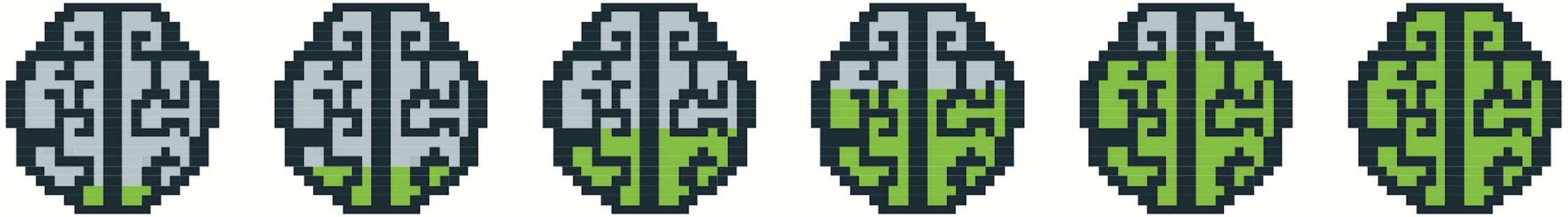
Data Ethics, AI and Responsible Innovation

Our future is here and it relies on data. Predictive policing, medical robots, smart homes and cities, artificial intelligences - we can all think about how any of those could go wrong. Discover how we can build a future where they are done right.

THE UNIVERSITY of EDINBURGH

https://www.edx.org/course/Data-Ethics-AI-and-Responsible-Innovation

BRAIN



LOADING...

Иван Ямщиков, PhD. Telegram: @progulka