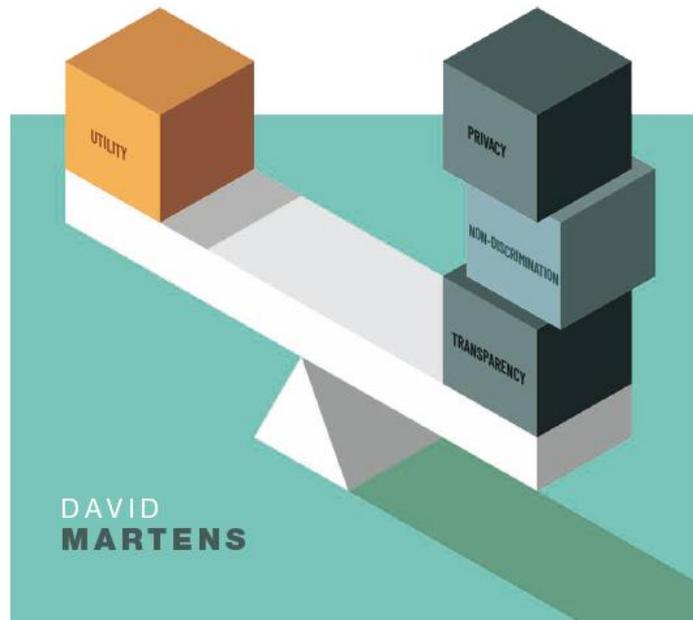


OXFORD

Data Science Ethics

Concepts,
Techniques and
Cautionary Tales



David Martens
Universiteit Antwerpen

www.uantwerpen.be/david-martens



Data Science Ethics

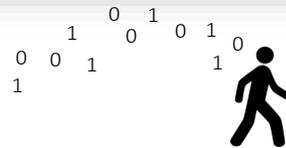
It was the best of times

- Reduce risk
- Reduce crime
- Increase profitability
- Improve medical diagnosis
- Increased “good”

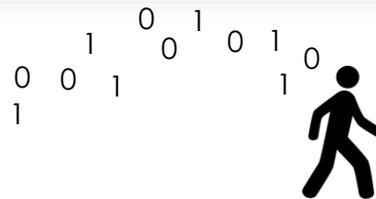


It was the worst of times

- Data leaks
- Discrimination
- Digital pawns
- Filter bubble
- Increased “bad”



Data Science Ethics



About right and wrong

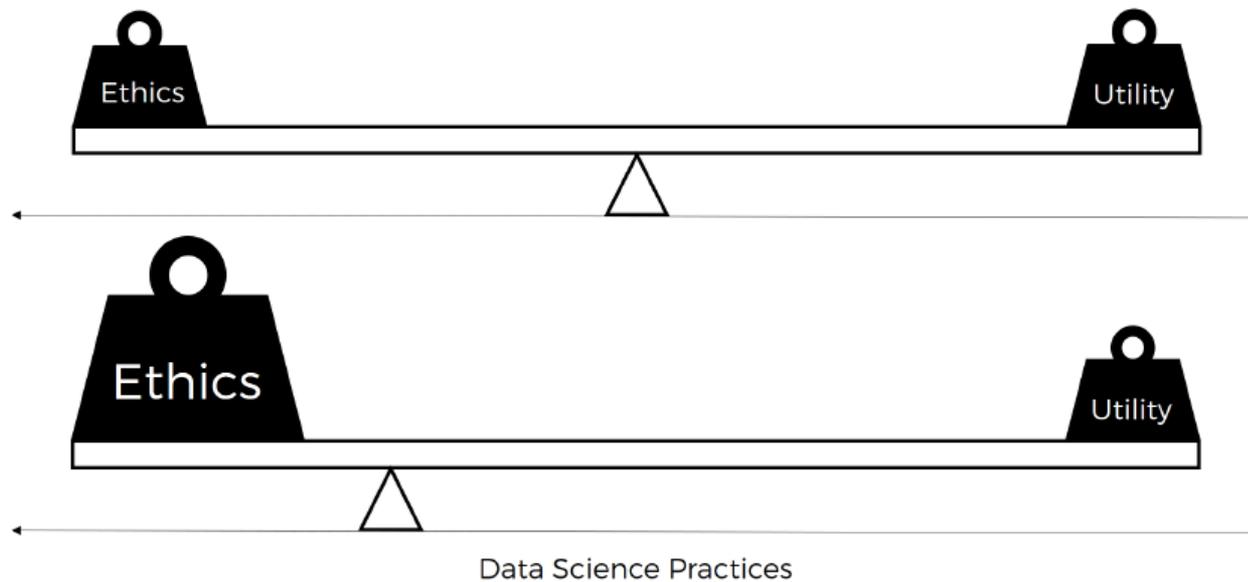


David Martens (2022) *Data Science Ethics: Concepts, Techniques and Cautionary Tales*. Oxford University Press.
www.dsethics.com



Data Science Ethics

- The golden mean between deficiency and excess (Aristotle)



Data Science Ethics Equilibrium.



Why Care?

- Life goal in itself
- Huge potential risks
- Data science ethics can bring value
- Expected from society

Shaping Europe's digital future



[Home](#) > [Policies](#) > [A European approach to Artificial intelligence](#)

A European approach to Artificial intelligence

The European Commission's approach to AI centres on excellence and trust. It aims to boost research and industrial capacity and ensure fundamental rights.

The following infringements shall be subject to administrative fines of up to 30 000 000 EUR or, if the offender is company, up to 6 % of its total worldwide annual turnover for the preceding financial year, whichever is higher:

Data scientists and business people are not inherently unethical, but at the same time not trained to think this through neither.



Data Mining

- Data mining: automatic extraction of patterns from data

Client	Income	Sex	Amount	Default
A	1.600	M	175.000	N
B	2.600	F	350.000	Y
C	3.280	M	50.000	N
D	950	M	120.000	Y
E	10.500	M	1.000.000	N
F	5.700	F	240.000	N
G	2.400	F	250.000	N

Data Mining

Classification Model
if income < 10.000 and Amount Loan > 100.000 and ... then default = yes

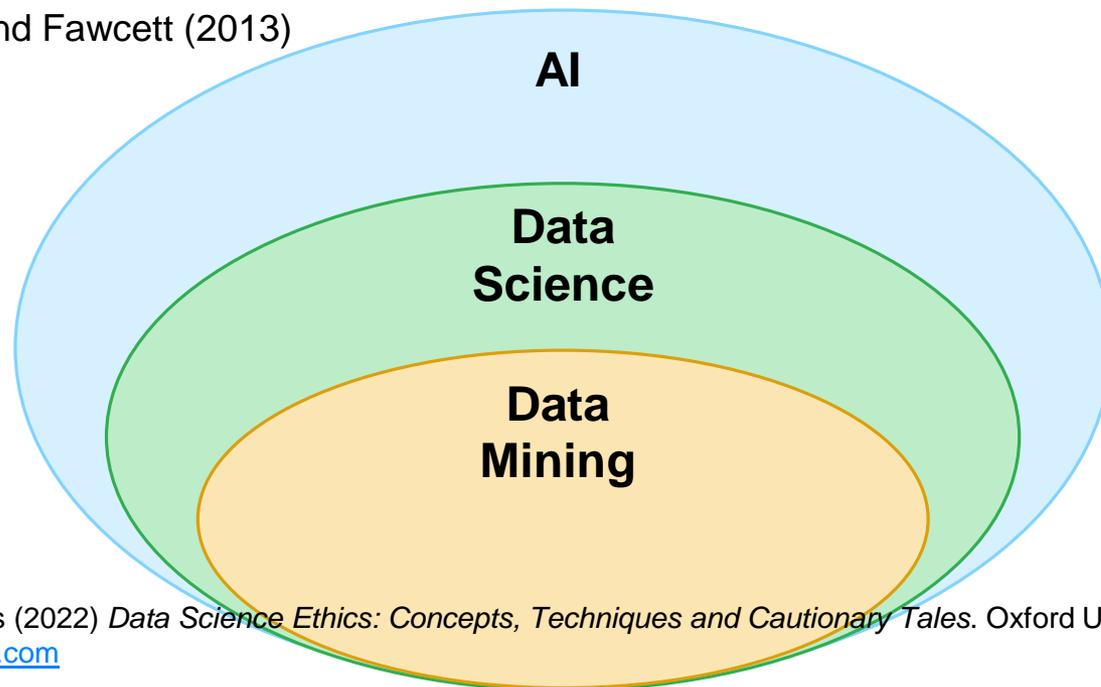
Client	Income	Sex	Amount	Default
New client	2.000	F	500.000	Y



Terminology

- **Data mining:** automatic extraction of patterns from data
- **Data science:** a set of fundamental principles that guide the extraction of knowledge from data
- **AI:** methods for improving the knowledge or performance of an intelligent agent over time, in response to the agent's experience in the world

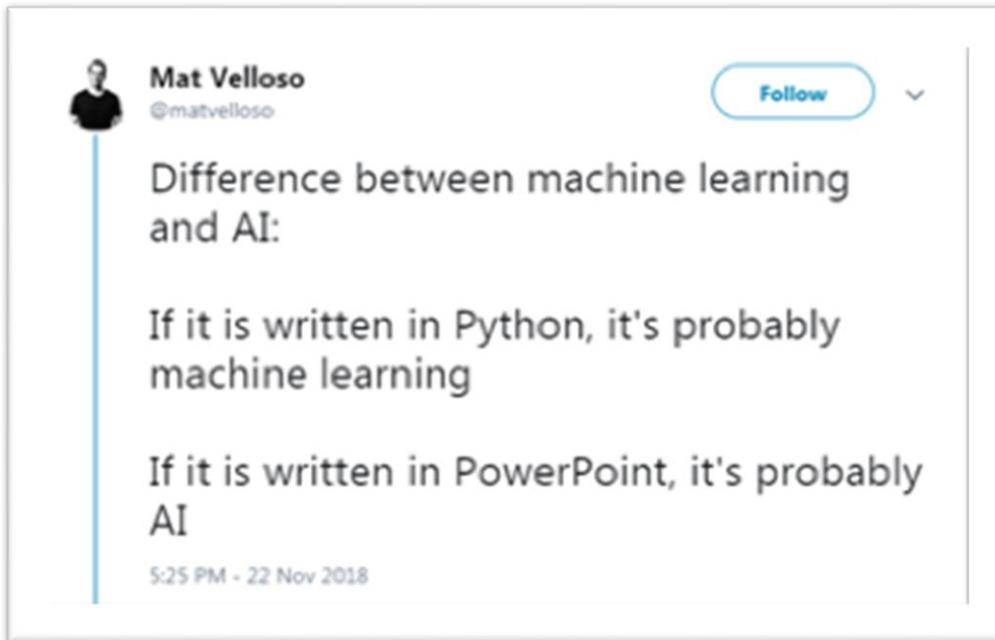
Provost and Fawcett (2013)



David Martens (2022) *Data Science Ethics: Concepts, Techniques and Cautionary Tales*. Oxford University Press.
www.dsethics.com



Terminology



Data Science Ethics

- **Ethics:** “*moral principles that control or influence a person’s behavior*”
- **Moral:** “*connected with principles of right and wrong behavior*”
- Ethics Theories: Utilitarianism vs Deontological Ethics
 - Utilitarianism:
 - =consequentialism, what is produced in the consequence of the act
 - Action is moral if the consequence is moral, means to an end
 - Justifies immoral things
 - Deontology:
 - Not doing immoral actions



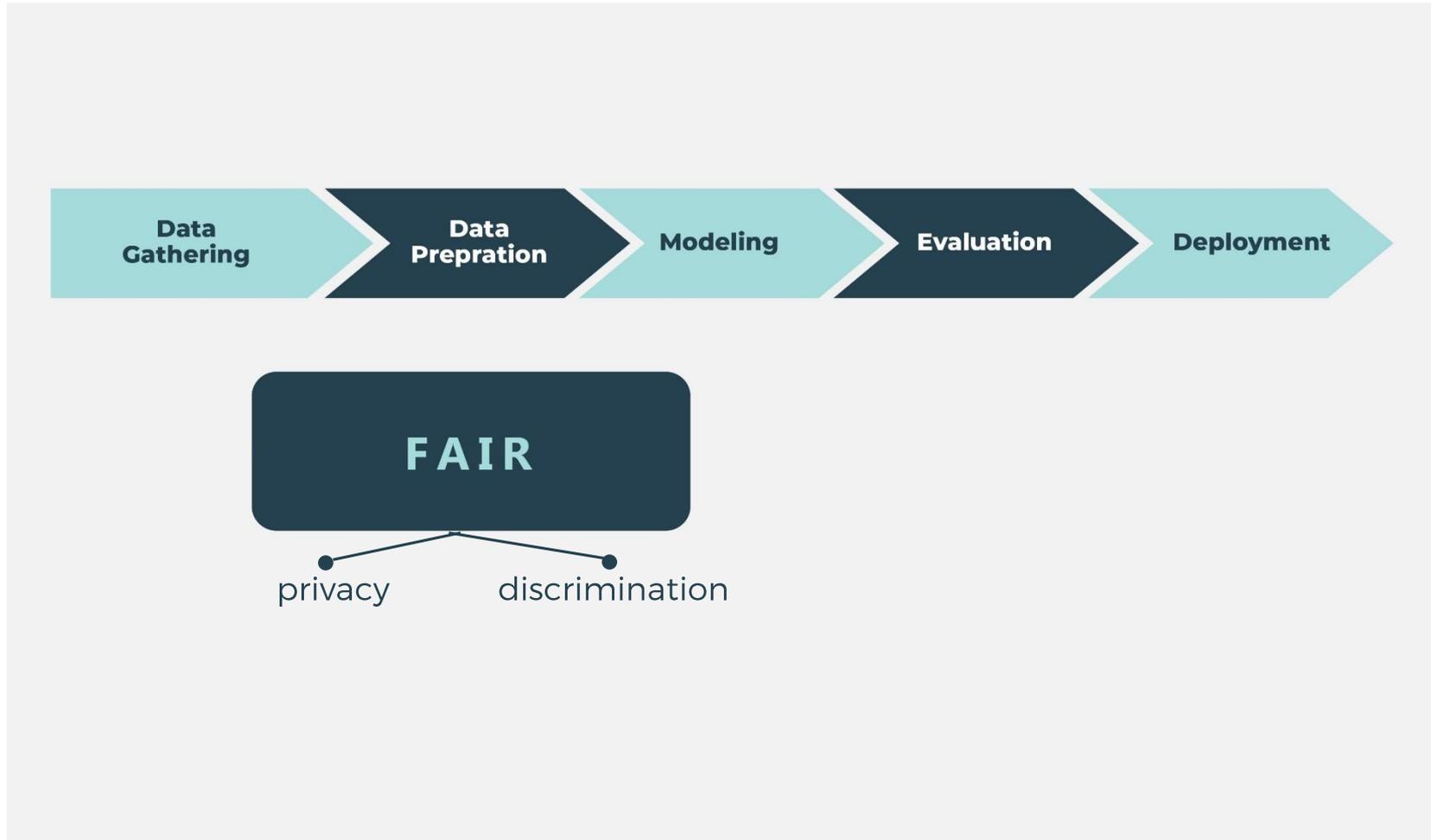
FAT Flow for data science ethics



David Martens (2022) *Data Science Ethics: Concepts, Techniques and Cautionary Tales*. Oxford University Press.
www.dsethics.com



FAT Flow for data science ethics



FAT Flow for data science ethics



FAIR

TRANSPARANT

process

explainable



FAT Flow for data science ethics



FAIR

TRANSPARANT

ACCOUNTABLE



Overview

- Data Science Ethics
- Data gathering: Privacy and A/B Testing
- Data preprocessing: Proxies
- Modeling: Discrimination-aware and Privacy-Preserving
- Model evaluation: Explain
- Deployment: Unintended consequences



Privacy

- Modern-day panoptes
 - Security cameras
 - Facebook, Internet as a whole
 - Behavioral data
- Privacy is a human right
 - “No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation.”
The United Nations' 1984 Universal Declaration of Human Right (Article 12)
- “You have zero privacy anyway. Get over it.”
Sprengr (1999-01-26), chairman of Sun Microsystems
- “Who was nothing to hide, has nothing to fear”



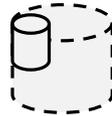
GDPR

(b) PURPOSE LIMITATION

Belgian mayor

(a) LAWFULNESS, FAIRNESS AND TRANSPARENCY

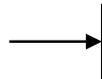
La Liga



(c) DATA MINIMISATION



(d) ACCURACY



(e) STORAGE LIMITATION

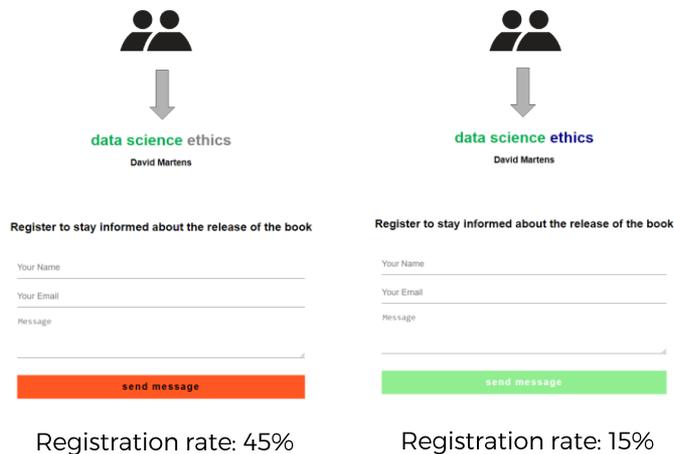


(f) INTEGRITY AND CONFIDENTIALITY



Experimentation

- A/B testing
 - Randomize experiment with two variants: A and B
 - Vary one variable and assess the effect



- Different treatment of different groups: potential human impact and ethical implications!



Experimentation

- OK Cupid
 - Goal: **Test the performance** of their prediction model in practise
 - Dating website, provides a match (0-100%) with candidates, based on prediction model
 - Group A: Bad match → Told they were a bad match
 - Group B: Bad match → Told they were very good match
 - Change in if and how often they talk to each other?
It did: group B more likely to send message to each other

- *“If you use the Internet, you're the subject of hundreds of experiments at any given time, on every site. That's how websites work.”*
OKCupid president.
- *“If you're lying to your users in an attempt to improve your service, what's the line between A/B testing and fraud?”*
Washington Post's Brian Fung



Overview

- Data gathering: Privacy and A/B Testing
- Data preprocessing: Proxies
- Modeling: Discrimination-aware and Privacy-Preserving
- Model evaluation: Explain
- Deployment: Unintended consequences



Proxies for re-identification

- August 4, 2006, AOL research released dataset
 - 20 million search keywords
 - 650.000 users
- Anonymous?
 - No users explicitly identified in the dataset
 - Search queries could be used to identify persons
 - NY Times exposed one of the users, with her explicit permission, as *Thelma Arnold, a 62-year-old widow from Gorgia, U.S.*



Thelma Arnold's identity was betrayed by AOL records of her Web searches, like ones for her dog, Dudley, who clearly has a problem.
Erik S. Lesser for The New York Times

<https://www.nytimes.com/2006/08/09/technology/09aol.html>

At times, the searches appear to betray intimate emotions and personal dilemmas. No. 3505202 asks about “depression and medical leave.” No. 7268042 types “fear that spouse contemplating cheating.”



David Martens (2022) *Data Science Ethics: Concepts, Techniques and Cautionary Tales*. Oxford University Press.
www.dsethics.com



Overview

- Data Science (and) Ethics
- Data gathering: Privacy and A/B Testing
- Data preprocessing: Proxies
- Modeling: Discrimination-aware and Privacy-Preserving
- Model evaluation: Explain
- Deployment: Unintended consequences



Government Backdoor

Feb. 2016

The New York Times

Apple Fights Order to Unlock San Bernardino Gunman's iPhone



Timothy D. Cook, the chief executive of Apple, released a letter to customers several hours after a California judge ordered the company to unlock an iPhone used by one of the shooters in a recent attack that killed 14 people in San Bernardino. Jeff Chiu/Associated Press

By Eric Lichtblau and Katie Benner

Feb. 17, 2016



David Martens (2022) *Data Science Ethics: Concepts, Techniques and Cautionary Tales*. Oxford University Press.
www.dsethics.com



Government Backdoor

Pro

- Privacy is not absolute

“Encryption isn't just a technical feature; it's a marketing pitch. ... Sophisticated criminals will come to count on these means of evading detection. It's the equivalent of a closet that can't be opened. ... And my question is, at what cost?”

James Comey (2014)



David Martens (2022) *Data Science Ethics: Concepts, Techniques and Cautionary Tales*. Oxford University Press.
www.dsethics.com



Government Backdoor

Pro

- Privacy is not absolute

Con

- Security versus Freedom

“Those who would give up essential Liberty, to purchase a little temporary Safety, deserve neither Liberty nor Safety.”

Benjamin Franklin (1775)



Government Backdoor

Pro

- Privacy is not absolute

Con

- Security versus Freedom
- Security versus Security

“No-one, I don't believe, would want a master key built that would turn hundreds of millions of locks. Even if that key were in the possession of the person that you trust the most. That key could be stolen.”

Tim Cook (2016)



Government Backdoor

Pro

- Privacy is not absolute

Con

- Security versus Freedom
- Security versus Security
- Futility of backdoors

“The bottom line is, if you look at both the terrorists in San Bernardino and the Boston Marathon bombers, they were family members. Most family members talk to each other face to face. The government doesn't have access to that after the fact.”

Michael Chertoff (2016)

David Martens (2022) *Data Science Ethics: Concepts, Techniques and Cautionary Tales*. Oxford University Press.
www.dsethics.com



Government Backdoor

PRIVACY

Privacyvoorvechters verzetten zich tegen Belgische 'backdoor'

Vijftig gespecialiseerde ngo's, academici en bedrijven van over de hele wereld hekelen de nieuwe Belgische wet die telecomdiensten verplicht communicatie tussen burgers te ontsleutelen voor de overheid.

Matthias Verbergt
Donderdag 30 september 2021 om 3.25 uur

DATA-OPSLAG

Chat-app Signal illegaal door nieuwe wet

Elk bedrijf dat in België communicatie aanbiedt, moet straks data van gebruikers bijhouden. Het sterk beveiligde Signal komt in de problemen.

Nikolas Vanhecke
Zaterdag 30 april 2022 om 3.25 uur

DE TECHNOCRAAT

Europa wil uw Whatsappberichten lezen

Dominique Deckmyn
Zaterdag 14 mei 2022 om 3.25 uur

TIME

TECH • PRIVACY

U.S., U.K. and Australia Ask Facebook Not to Extend Encrypted Messaging Program



Facebook CEO Mark Zuckerberg testifies before a House Energy and Commerce hearing on Capitol Hill in Washington on April 11, 2018. Andrew Harnik/AP

BY ANICK JESDANUN / AP OCTOBER 4, 2019

(NEW YORK) — U.S. Attorney General William Barr and other U.S., U.K. and Australian officials are pressing Facebook to give authorities a way to read encrypted messages sent by ordinary users, re-igniting tensions between tech companies and law enforcement.

Facebook's WhatsApp already uses so-called end-to-end encryption, which locks up messages so that even Facebook can't read their contents. Facebook plans to extend that protection to Messenger and Instagram Direct.

David Martens (2022) *Data Science Ethics: Concepts, Techniques and Cautionary Tales*. Oxford University Press.
www.dsethics.com



Zero Knowledge Proof

- A method where one party proves a **statement about a secret** to another party, **without revealing the secret**.
- Applications
 - Secret: *income*
Statement to proof: *"I have a high income."*
 - Secret: *nationality*
Statement to proof: *"I'm a European citizen."*
 - Secret: *location of Herman*
Statement to proof: *"I know where Herman is."*



More technologies to include privacy

	Setup	# parties	Calculations	Use Cases
€-Differential Privacy	■ ●	2	Aggregate statistics	Survey results
Zero Knowledge Proof	■ ●	2	Proof	Proof in range, Proof in set
Homomorphic Encryption	■ ●	2	PHE: + or × FHE: all	Cloud computing
SMPC	■ ■ ■ ■	m	Various (average, intersection)	Auctions, averages, voting
Federated Learning	■ ■ ● ■ ■	m+1	Machine learning	Learning from mobile devices

■ Has a secret
 ● Performs some calculations

Fig. 4.6 An overview of some privacy-enabling methods from the modelling stage.

David Martens (2022) *Data Science Ethics: Concepts, Techniques and Cautionary Tales*. Oxford University Press.
www.dsethics.com



Potential Discrimination

- Bias
 - HR Analytics, prediction model to review job applicants' resumes to automate the search for top talent
 - Trained on resumes from past (10 year period), biased data
 - Model trained to prefer male candidates, for example:
 - Penalized uses of word “woman’s” (eg woman’s chess club president)
 - Penalizes all-woman colleges



Overview

- Data Science (and) Ethics
- Data gathering: Privacy and A/B Testing
- Data preprocessing: Proxies
- Modeling: Discrimination-aware and Privacy-Preserving
- Model evaluation: Explain
- Deployment: Unintended consequences



Explain individual predictions



DHH ✓
@dhh

Volgen

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.

12:34 - 7 nov. 2019

9.664 retweets 29.300 vind-ik-leuks



1.583 9.664 29.300



DHH ✓ @dhh · 8 nov.

She spoke to two Apple reps. Both very nice, courteous people representing an utterly broken and reprehensible system. The first person was like "I don't know why, but I swear we're not discriminating, IT'S JUST THE ALGORITHM". I shit you not. "IT'S JUST THE ALGORITHM!"

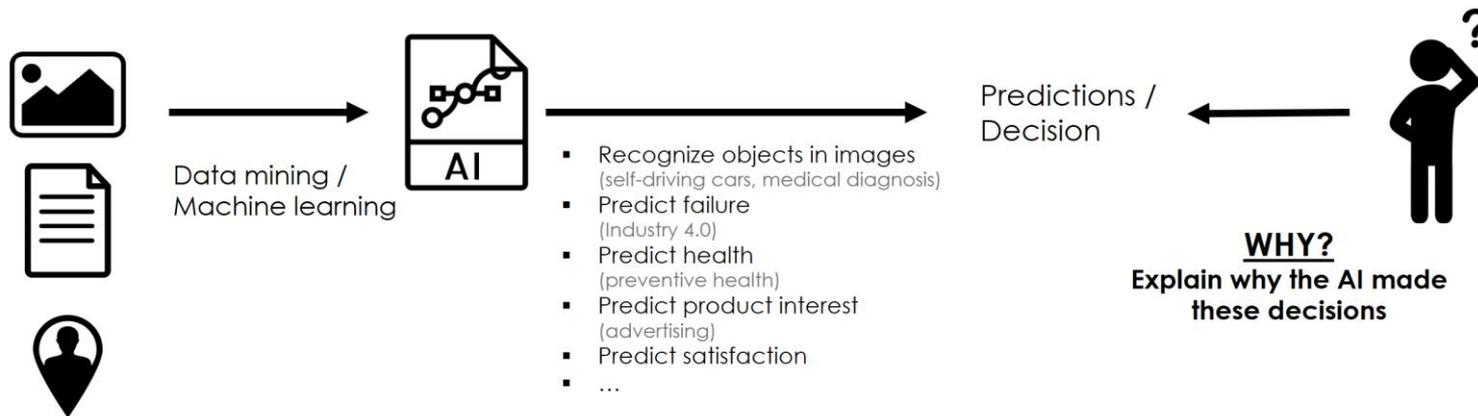
69 542 4.253



David Martens (2022) *Data Science Ethics: Concepts, Techniques and Cautionary Tales*. Oxford University Press.
www.dsethics.com



Explain



Explain

Why?

Trust

Compliance

Improve



Explain

Why?

Trust



Compliance



Improve



David Martens (2022) *Data Science Ethics: Concepts, Techniques and Cautionary Tales*. Oxford University Press.
www.dsethics.com



Explain

Why?

Trust



Compliance

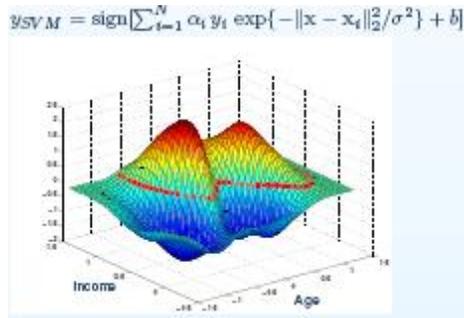


Improve



Explain – Black Box

- Non-linear models



- Deep learning: large artificial neural network with massive number of parameters

ImageNet Classification with Deep Convolutional Neural Networks

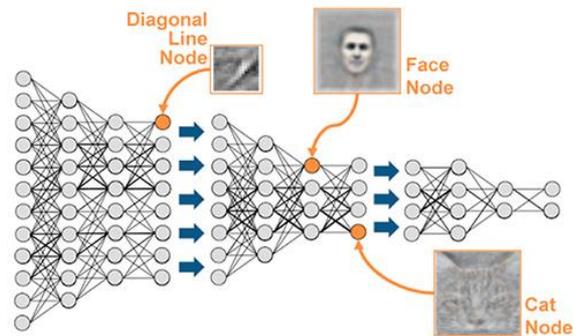
Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

Abstract

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet ILSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0%, which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called "dropout" that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.



Explain individual predictions

Example: gender prediction using movie viewing data



User x_i : Sam

Sam watched 120 movies
Sam is predicted as male

WHY?

Yanou Ramon, David Martens (2019) *Instance-based Explanations: Motivation, Overview, and the Evidence Counterfactual Approach*, European Conference on Data Analysis, Bayreuth, Germany, 18-20 March 2019



David Martens (2022) *Data Science Ethics: Concepts, Techniques and Cautionary Tales*. Oxford University Press.
www.dsethics.com



Explain individual predictions

Example: gender prediction using movie viewing data



User x_i : Sam

Sam watched 120 movies
Sam is predicted as male

WHY?

EDC: EviDence Counterfactual

IF Sam would not have watched
*{Taxi driver, The Dark Knight, Die Hard,
Terminator 2, Now You See Me, Interstellar}*,
THEN his predicted class would change from male to female



Imagine a world with CF explanations



Well, if your wife would also have had a 20+ year relationship with our bank, and would have been regarded as Premium customer at some point in time, she would also receive a 20x credit limit

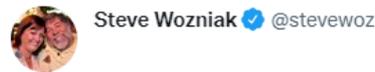


Well, if your wife's relationship status would have been "husband" instead of "wife", she would also receive a 20x credit limit

We clearly messed up, we're updating our models now.



Ah, ok, thanks for the additional feedback!



Glad you found this and react responsibly. It's how big tech should be in the 21st century.



David Martens (2022) *Data Science Ethics: Concepts, Techniques and Cautionary Tales*. Oxford University Press. www.dsethics.com

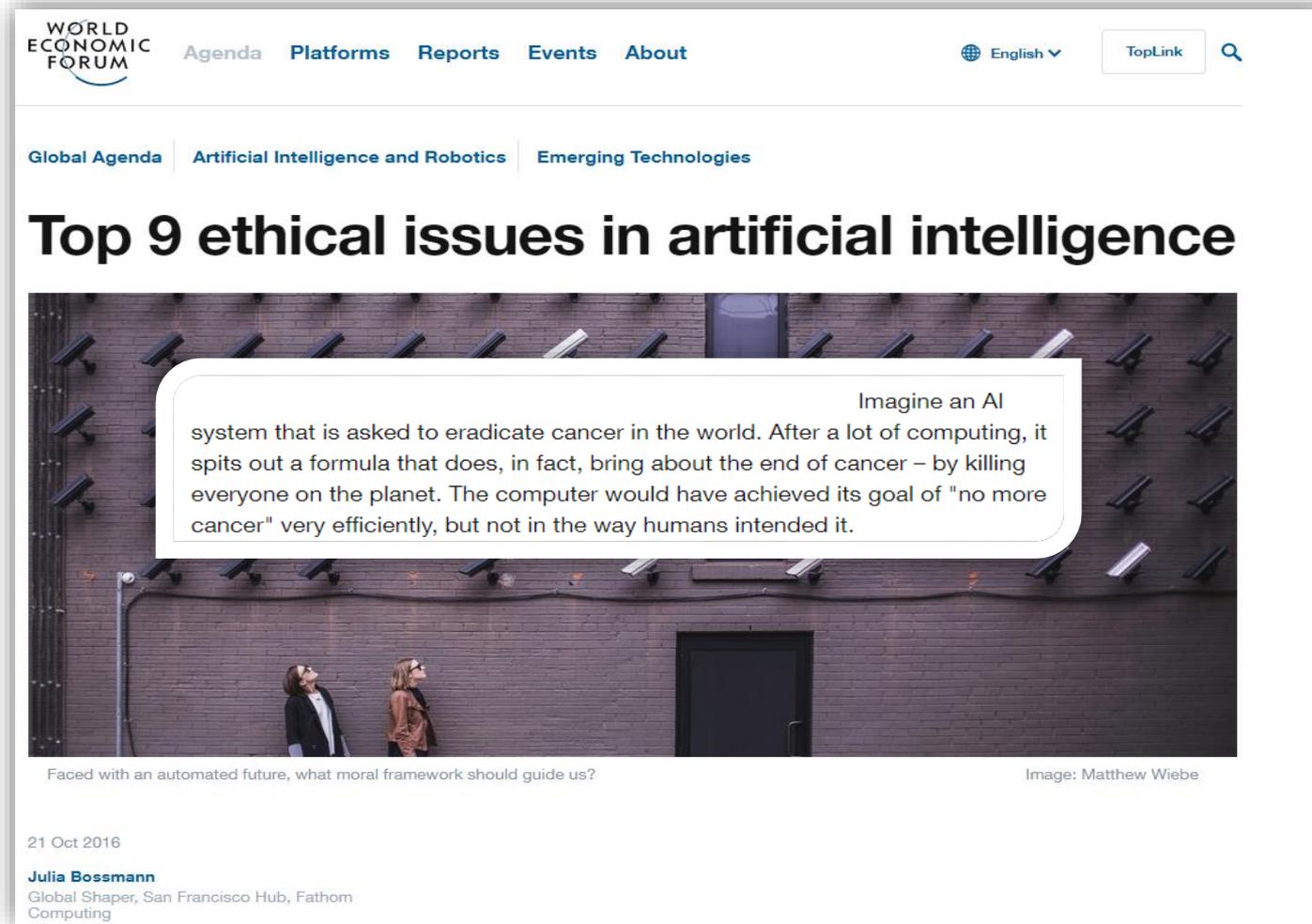


Overview

- Data Science (and) Ethics
- Data gathering: Privacy and A/B Testing
- Data preprocessing: Proxies
- Modeling: Discrimination-aware and Privacy-Preserving
- Model evaluation: Explain
- Deployment: Unintended consequences



Unintended consequences



The screenshot shows the top navigation bar of the World Economic Forum website with links for Agenda, Platforms, Reports, Events, and About. The language is set to English. Below the navigation, there are breadcrumb links for Global Agenda, Artificial Intelligence and Robotics, and Emerging Technologies. The main heading of the article is 'Top 9 ethical issues in artificial intelligence'. The featured image shows a brick wall with many security cameras. A white text box is overlaid on the image, containing a paragraph about an AI system that eradicates cancer by killing everyone. Below the image, there is a caption and the author's name and affiliation.

WORLD ECONOMIC FORUM

Agenda Platforms Reports Events About

English TopLink

Global Agenda Artificial Intelligence and Robotics Emerging Technologies

Top 9 ethical issues in artificial intelligence

Imagine an AI system that is asked to eradicate cancer in the world. After a lot of computing, it spits out a formula that does, in fact, bring about the end of cancer – by killing everyone on the planet. The computer would have achieved its goal of "no more cancer" very efficiently, but not in the way humans intended it.

Faced with an automated future, what moral framework should guide us? Image: Matthew Wiebe

21 Oct 2016

Julia Bossmann
Global Shaper, San Francisco Hub, Fathom Computing



David Martens (2022) *Data Science Ethics: Concepts, Techniques and Cautionary Tales*. Oxford University Press.
www.dsethics.com



Unintended consequences

- AI acts different than intended

Tay was designed to learn more about language over time.... Eventually, her programmers hoped, Tay would sound just like the Internet.

IEEE SPECTRUM Engineering Topics ▾ Special Reports ▾ Blogs ▾ Multimedia ▾ The Magazine ▾ Professional Resources ▾ Search ▾

In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation

The bot learned language from people on Twitter— but it also learned values

By Oscar Schwartz

Photo-illustration: Gluekit



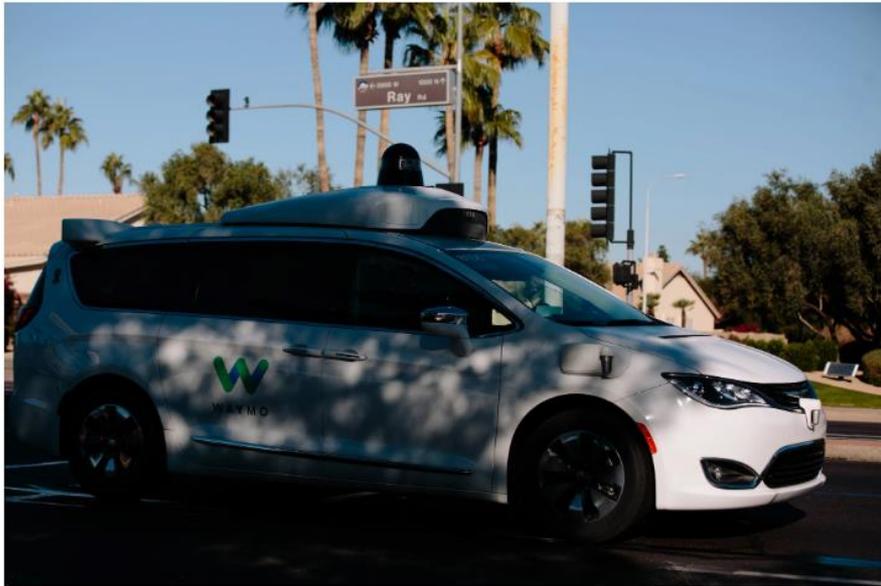
David Martens (2022) *Data Science Ethics: Concepts, Techniques and Cautionary Tales*. Oxford University Press.
www.dsethics.com



Unintended consequences

- Negative impact of AI on humans

Wielding Rocks and Knives, Arizonans Attack Self-Driving Cars



A Waymo autonomous vehicle in Chandler, Ariz., where the driverless cars have been attacked by residents on several occasions. Caitlin O'Hara for The New York Times



Data Science Ethics

- Data Science Ethics...
 - Is about right and wrong
 - Requires discussion, time and effort
 - Can bring value
 - Becoming a legal requirement
- **Warning signs:** “No one has to know.” / “Don’t mail about this.”
- FAT Flow as framework
- For data science projects:
 - Consider ethics already before starting
 - Don’t be afraid to add an ethical section in your reports or emails



Data Science Ethics

Data science ethics is all about what is right and wrong when conducting data science. Data science has so far been primarily used for positive outcomes for businesses and society. However, just as with any technology, data science has also come with some negative consequences: an increase of privacy invasion, data-driven discrimination against sensitive groups, and decision making by complex models without explanations.

While data scientists and business managers are not inherently unethical, they are not trained to weigh the ethical considerations that come from their work—Data Science Ethics addresses this increasingly important gap and highlights different concepts and techniques that aid understanding, ranging from k-anonymity and differential privacy to homomorphic encryption and zero knowledge proofs to address privacy concerns, techniques to remove discrimination against sensitive groups, and various explainable AI techniques.

Real-life cautionary tales further illustrate the importance and potential impact of data science ethics, including tales of racist bots, search censoring, government backdoors, and face recognition. The book is punctuated with structured exercises that provide hypothetical scenarios and ethical dilemmas for reflection that teach readers how to balance the ethical concerns and the utility of data.

COVER IMAGE:
InimaGraphic/Shutterstock.com

OXFORD
UNIVERSITY PRESS
www.oup.com



MARTENS

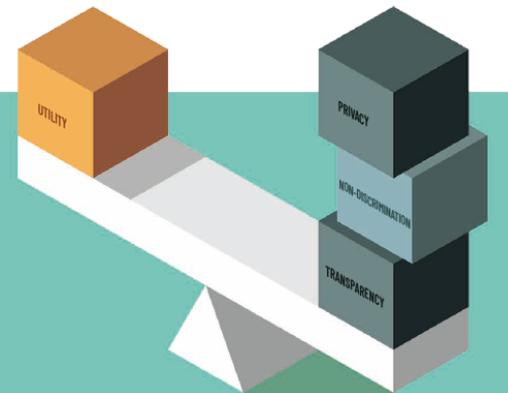
Data Science Ethics

OXFORD

OXFORD

Data Science Ethics

Concepts,
Techniques and
Cautionary Tales



DAVID
MARTENS

David Martens

Data Science Ethics: Concepts, Techniques, and Cautionary Tales

ISBN-13: 978-0192847270, ISBN-10: 0192847279

#1 New Release in Statistics

Available at: oup.com, bol.com, [standaard boekhandel](http://standaardboekhandel), amazon.fr, amazon.nl and elsewhere

