

# Closing Thoughts

**Vishal Patel**

Spring 2024

# What We Covered

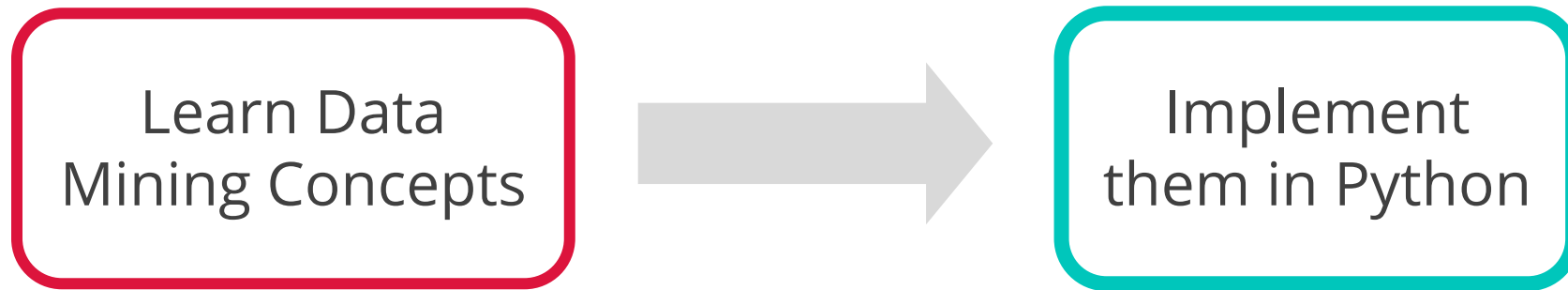
1. Introduction to Data Mining
2. The Data Science Process
3. Introduction to Regression
4. Linear Regression model
5. Build a model using partitioning
6. Decision trees
7. Classification Trees
8. Random Forests
9. Gradient Boosting Trees
10. Hyper-parameter optimization

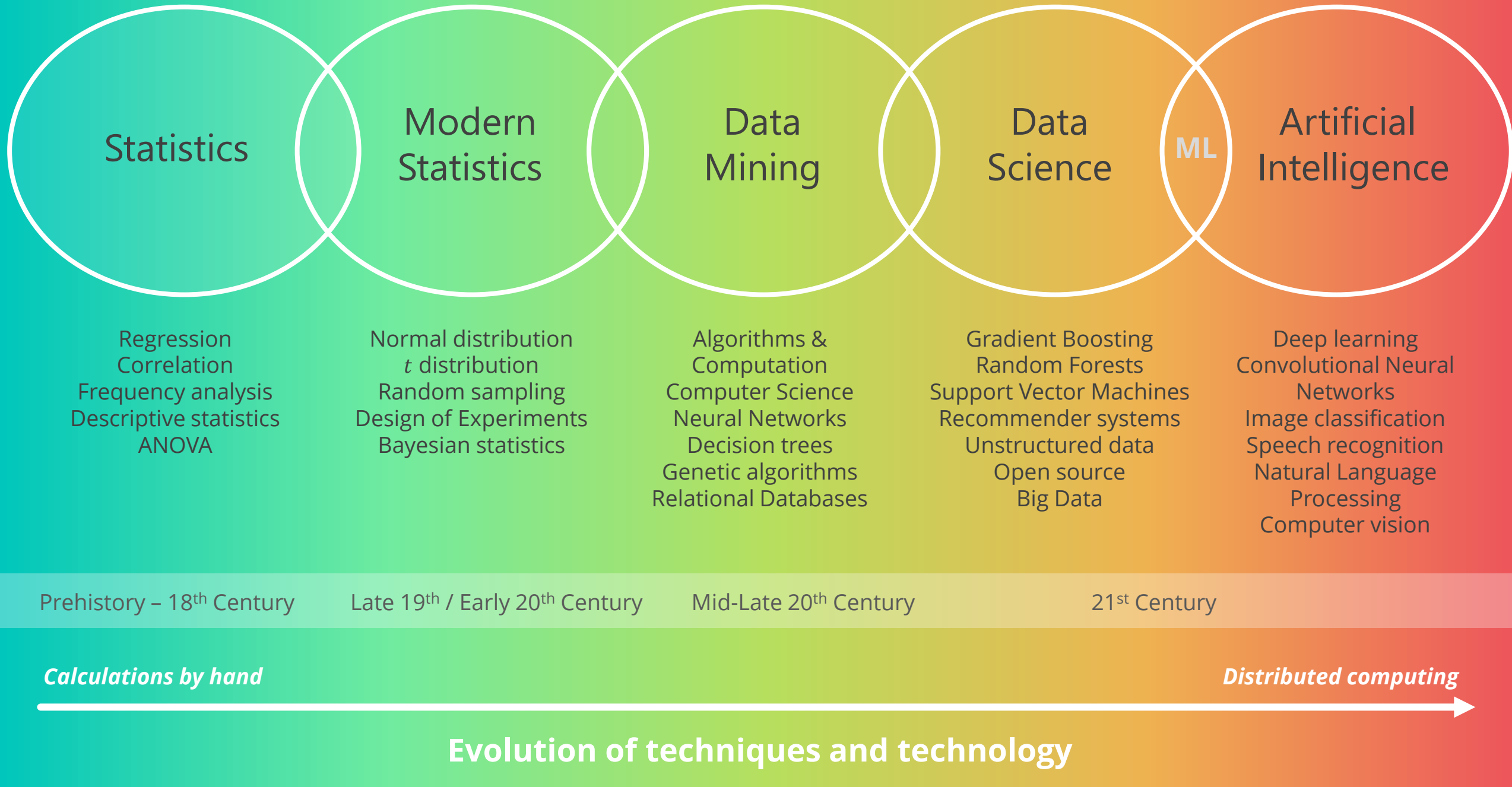
11. Introduction to Neural Networks
12. Introduction to Clustering
13. Agglomerative Clustering
14. k-means Clustering
15. DBSCAN Clustering
16. PCA
17. Association Analysis (Apriori algorithm)
18. Collaborative Filtering
19. Data Wrangling

+ 16 Jupyter Notebooks

**Education:** What, when and why

**Training:** How





"Artificial intelligence is one of the most profound things we're working on as humanity. It is more profound than fire or electricity."

– Sundar Pichai



“With artificial intelligence  
we are **summoning the demon.**”

– Elon Musk





Boston Dynamics

**I, FOR ONE, WELCOME OUR**

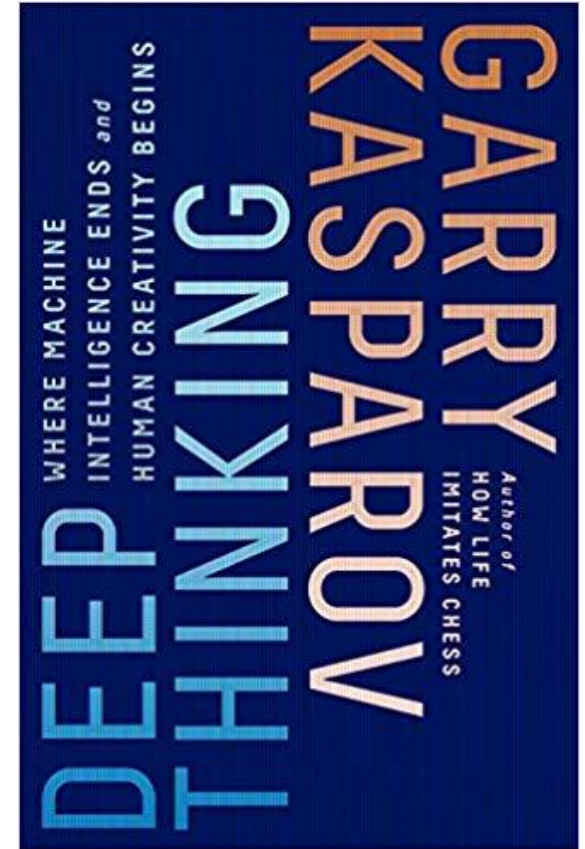


**NEW AI OVERLORDS**

memegenerator.net



Good riddance, you might imagine. But the worries about operator-less elevators were quite similar to the concerns we hear today about driverless cars. In fact, I learned something surprising when I was invited to speak to the Otis Elevator Company in Connecticut in 2006. The technology for automatic elevators had existed since 1900, but people were too uncomfortable to ride in one without an operator. It took the 1945 strike and a huge industry PR push to change people's minds, a process that is already repeating with driverless cars. The cycle of automation, fear, and eventual acceptance goes on.





1

**Jobs**

2

**Ethics**





waze

@waze

Follow



We're working with the [@LADOTofficial](#) and our [#WazeCommunity](#) to keep the Waze Map updated with road closures (currently at 110) & open shelters (16) to help Southern Californians navigate safely around the [#wildfires](#). Please help us spread the word.

**Joel Rubin**  @joelrubin

"The Los Angeles Police Department asked drivers to avoid navigation apps, which are steering users onto more open routes — in this case, streets in the neighborhoods that are on fire." [fw.to/EMJxt2E](https://www.facebook.com/EMJxt2E)

7:59 AM - 7 Dec 2017

BASED ON YOUR  
INTERNET HISTORY,  
YOU MIGHT BE DUMB  
ENOUGH TO ENJOY  
EXTREME SPORTS.



CLICK HERE TO BUY A  
TICKET TO BASE JUMP  
FROM THE INTERNA-  
TIONAL SPACE STATION.



I THINK  
THE INTER-  
NET IS  
TRYING TO  
KILL ME.



WE  
CALL IT  
"MACHINE  
LEARNING."





**Tay** was an artificial intelligence chatbot that was originally released by Microsoft Corporation via Twitter on March 23, 2016.

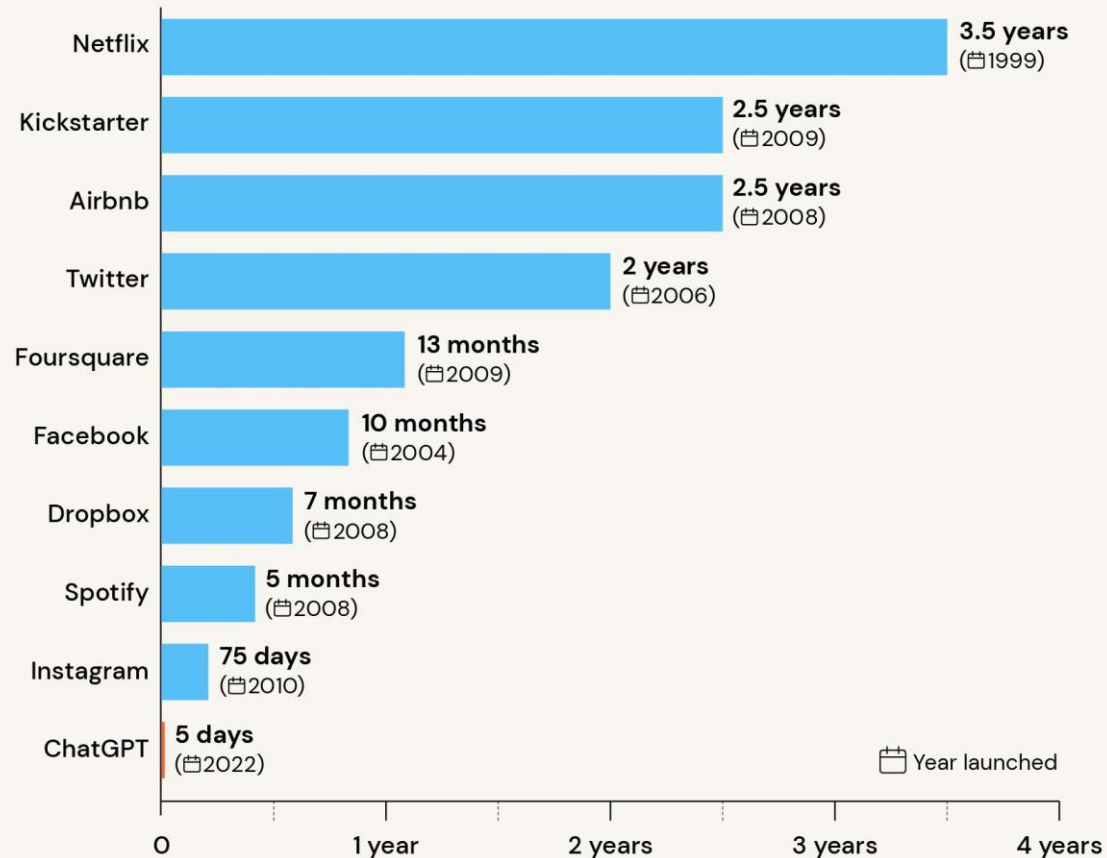
It caused subsequent controversy when the bot began to post inflammatory and offensive tweets through its Twitter account, forcing Microsoft to shut down the service only 16 hours after its launch.

It was soon replaced with **Zo** [2016-2019].

And now, as of 2024, **ChatGPT** has become a part of our daily life.

## CHATGPT STATISTICS

### Time to reach 1 million users



Read the full report at [tooltester.com/en/blog/chatgpt-statistics](https://tooltester.com/en/blog/chatgpt-statistics)

tooltester

- The fastest-growing user-base in the history of web apps: 1 million users in just five days!
- An average of **13 million daily visitors** in Jan 2023
- It is estimated that GPT-4 has **1.76 trillion parameters** trained on untold trillions of tokens. (For comparison, a human brain has approximately 100 billion neurons and 100 trillion connections.)



# GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models

Tyna Eloundou<sup>1</sup>, Sam Manning<sup>1,2</sup>, Pamela Mishkin<sup>\*1</sup>, and Daniel Rock<sup>3</sup>

<sup>1</sup>OpenAI

<sup>2</sup>OpenResearch

<sup>3</sup>University of Pennsylvania

March 20, 2023

## Abstract

We investigate the potential implications of Generative Pre-trained Transformer (GPT) models and related technologies on the U.S. labor market. Using a new rubric, we assess occupations based on their correspondence with GPT capabilities, incorporating both human expertise and classifications from GPT-4. Our findings indicate that approximately 80% of the U.S. workforce could have at least 10% of their work tasks affected by the introduction of GPTs, while around 19% of workers may see at least 50% of their tasks impacted. The influence spans all wage levels, with higher-income jobs potentially facing greater exposure. Notably, the impact is not limited to industries with higher recent productivity growth. We conclude that Generative Pre-trained Transformers exhibit characteristics of general-purpose technologies (GPTs), suggesting that as these models could have notable economic, social, and policy implications.

## Occupations with no labeled exposed tasks

Agricultural Equipment Operators  
Athletes and Sports Competitors  
Automotive Glass Installers and Repairers  
Bus and Truck Mechanics and Diesel Engine Specialists  
Cement Masons and Concrete Finishers  
Cooks, Short Order  
Cutters and Trimmers, Hand  
Derrick Operators, Oil and Gas  
Dining Room and Cafeteria Attendants and Bartender Helpers  
Dishwashers  
Dredge Operators  
Electrical Power-Line Installers and Repairers  
Excavating and Loading Machine and Dragline Operators, Surface Mining  
Floor Layers, Except Carpet, Wood, and Hard Tiles  
Foundry Mold and Coremakers  
Helpers—Brickmasons, Blockmasons, Stonemasons, and Tile and Marble Setters  
Helpers—Carpenters  
Helpers—Painters, Paperhangers, Plasterers, and Stucco Masons  
Helpers—Pipelayers, Plumbers, Pipefitters, and Steamfitters  
Helpers—Roofers  
Meat, Poultry, and Fish Cutters and Trimmers  
Motorcycle Mechanics  
Paving, Surfacing, and Tamping Equipment Operators  
Pile Driver Operators  
Pourers and Casters, Metal  
Rail-Track Laying and Maintenance Equipment Operators  
Refractory Materials Repairers, Except Brickmasons  
Roof Bolters, Mining  
Roustabouts, Oil and Gas  
Slaughterers and Meat Packers  
Stonemasons  
Tapers  
Tire Repairers and Changers  
Wellhead Pumpers

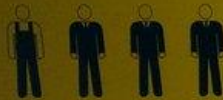
Table 12: All 34 occupations for which none of our measures labeled any tasks as exposed.

# Ein Bauer ernährt zusätzlich ...

Durch die zunehmende Mechanisierung ist folgende Entwicklung möglich geworden:



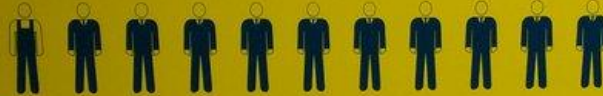
1900



+3

Ein Bauer ernährt sich und zusätzlich drei Städter.

1950



+10

Ein Bauer ernährt sich und zusätzlich zehn Städter.

2005



+142

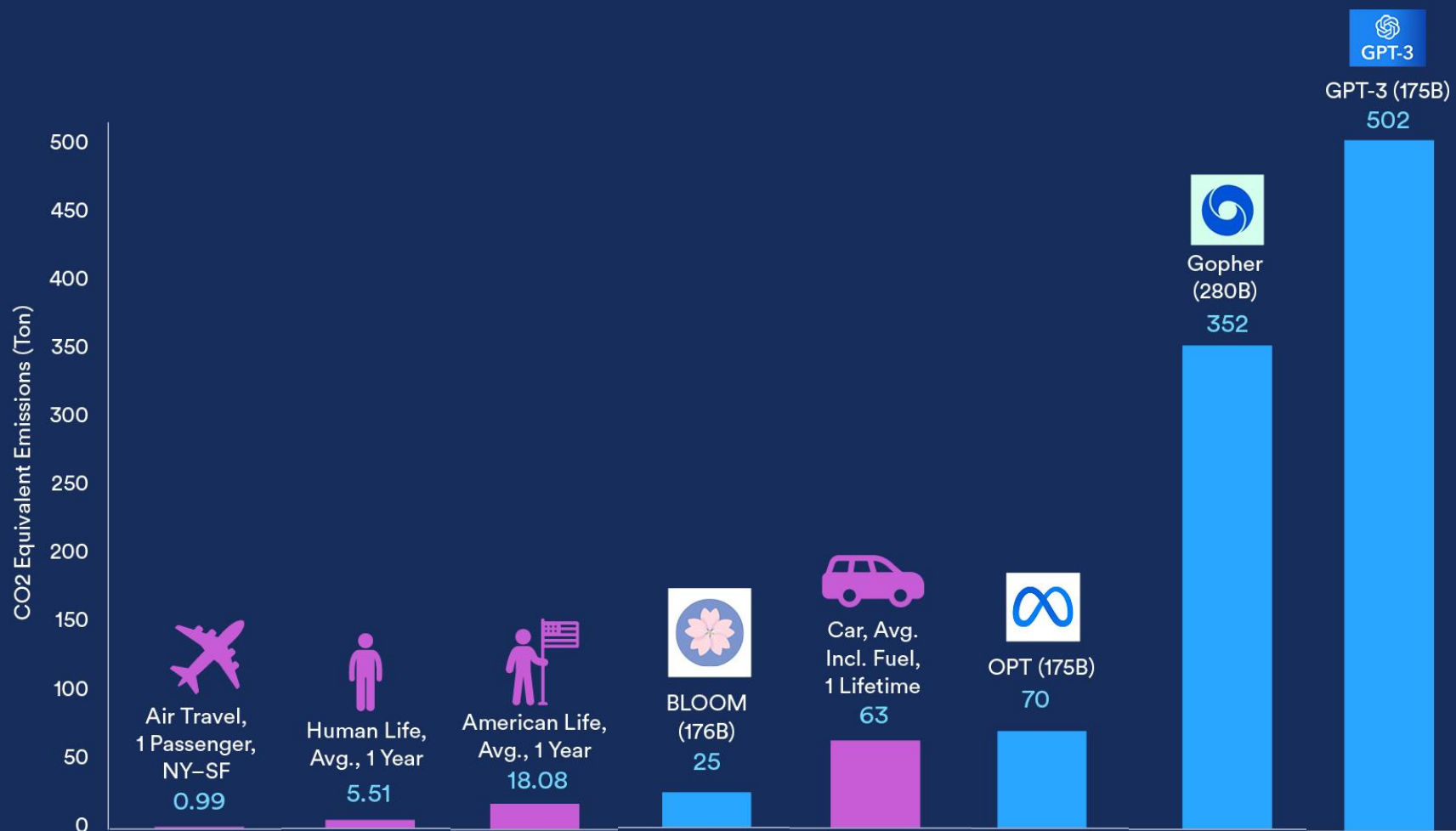
Ein Bauer ernährt sich und zusätzlich 142 Städter.

## The number of people one farmer could feed

Museum of Bread Culture, Germany



# CO2 Emissions (in Tons)



Source: Luccioni et al., 2022; Strubell et al., 2019 | Chart: 2023 AI Index Report

# Automated Inference on Criminality using Face Images

Xiaolin Wu  
McMaster University  
Shanghai Jiao Tong University  
xwu510@gmail.com

Xi Zhang  
Shanghai Jiao Tong University  
zhangxi\_19930818@sjtu.edu.cn



(a) Three samples in criminal ID photo set  $S_c$ .



(b) Three samples in non-criminal ID photo set  $S_n$

“[T]he authors find that their algorithm can classify  
**criminal faces with 90% accuracy.**”



(a) Three samples in criminal ID photo set  $S_c$ .



(b) Three samples in non-criminal ID photo set  $S_n$ .



Facial features purportedly associated with criminality, from Wu and Zhang (2016)

Smile detector!

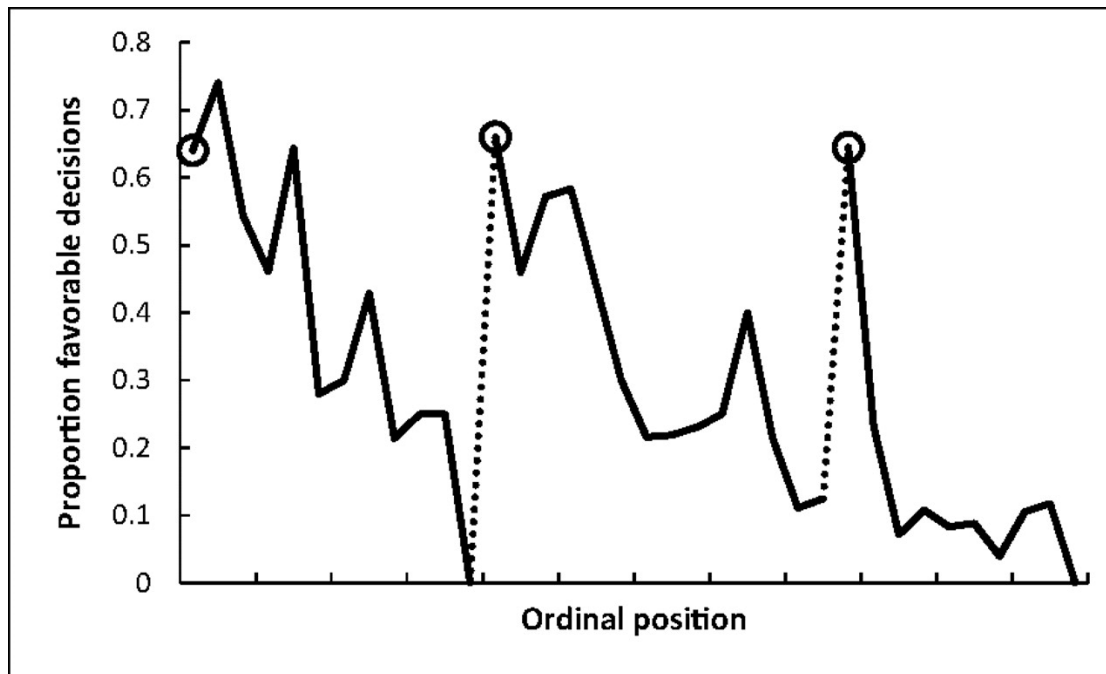
"[The] algorithm finds that criminals have shorter distances between the inner corners of the eyes, smaller angles between the nose and the corners of the mouth and higher curvature to the upper lip"

# Extraneous factors in judicial decisions

Shai Danziger, Jonathan Levav , and Liora Avnaim-Pesso [Authors Info & Affiliations](#)

Edited\* by Daniel Kahneman, Princeton University, Princeton, NJ, and approved February 25, 2011 (received for review December 8, 2010)

April 11, 2011 | 108 (17) 6889-6892 | <https://doi.org/10.1073/pnas.1018033108>



The authors of the peer-reviewed paper looked at more than 1,000 rulings made in 2009 by eight judges. They found that the **likelihood of a favorable ruling peaked at the beginning of the day**, steadily declining over time from a probability of about 65% to nearly zero, before **spiking back up to about 65% after a break for a meal or snack**.

# Social and Economical Bias

1. Errors (and biases) propagate quickly
2. Far reaching consequences
3. 'Bias' is not just a technical issues, it's a **socio-technical** issue
4. There's no silver bullet-solution
  1. What do we mean by "bias"?
  2. Who gets to decide what's fair?

# Numbers don't lie!!

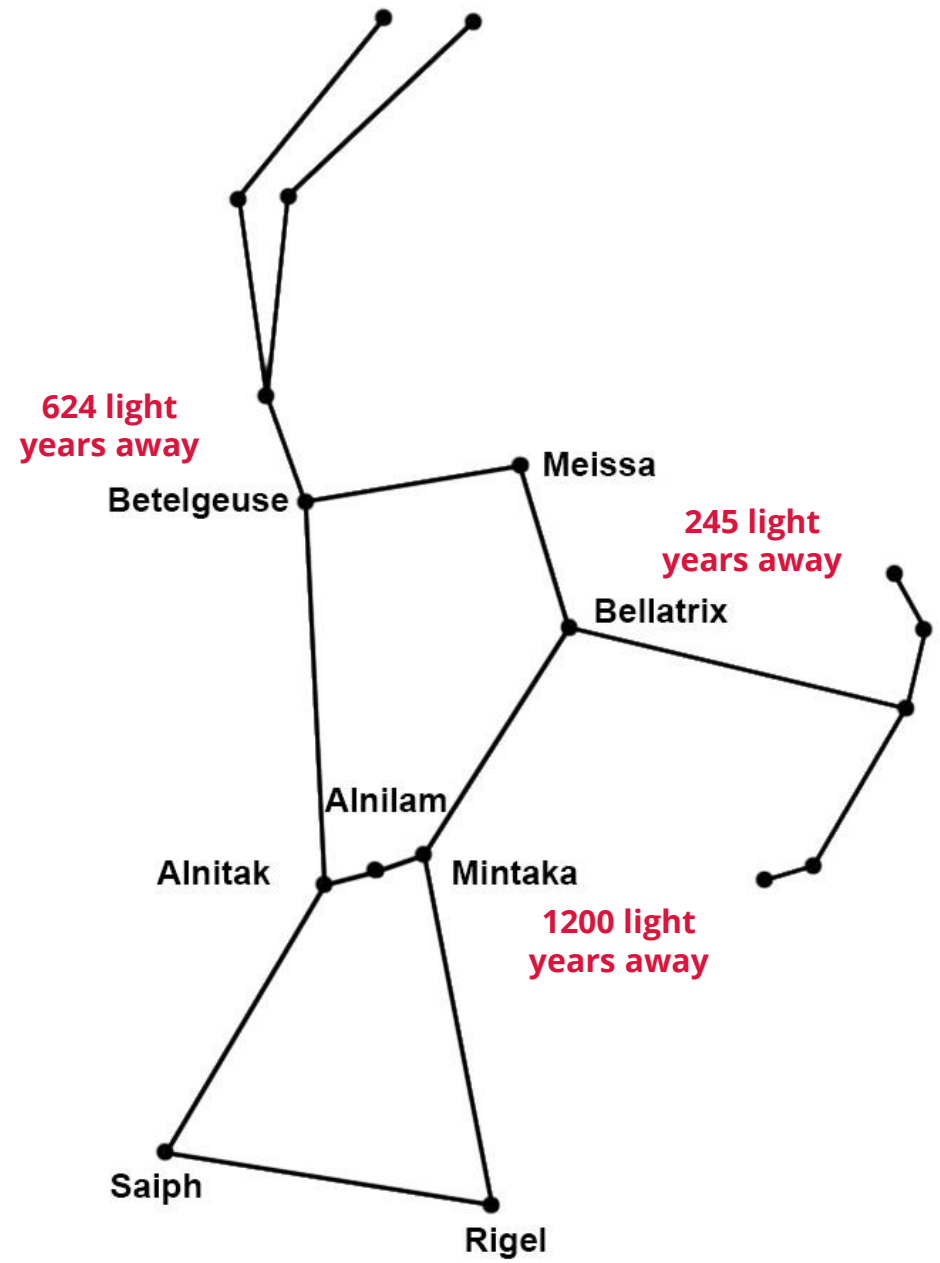
---

If you torture the data long enough,  
it will confess to anything.

From Darrell Huff's book *How to Lie With Statistics* (1954)

Truth is much too complicated  
to allow anything but approximations.

John von Neumann







Mulla Nasrudin

Someone saw Nasrudin searching for something on the ground.

"What have you lost?, he asked.

"My key", said the Mulla.

So they both went down on their knees and looked for it.

After some time the man asked: "Where exactly did you drop it?"

"In my own house".

"Then why are you looking here?"

"There is more light here than inside my own house".

– Sufi parable [Idris Shah, citing Mulla Nasrudin]

# WORKING WITH AVAILABLE LIGHT



There is no substitute for  
getting to **know your data.**

– Witten and Frank

# Qualities of an Ideal [Data Scientist]

1. Curious, but skeptical
2. Critical thinking skills
3. Logical reasoning ability
4. Creativity
5. Strategic thinking
6. Ability to read the facts and understand assumptions
7. Understands necessary information and how to act on it
8. Engages with people to understand root of the problem
9. Ability to apply appropriate [data science] tools



“Before starting any project, I ask: What’s the geology, what’s the geomorphology, what’s the history, where does the wind come from, where does the sun come from, what are the shadow patterns, what’s the drainage system, what’s the flora?

...I try to be a **contextual** architect.” – **Glenn** Murcutt

Age of  
CONTENT



Age of  
CONTEXT

Contextual Data Science

# THANK YOU!!

[vishal@derive.io](mailto:vishal@derive.io)