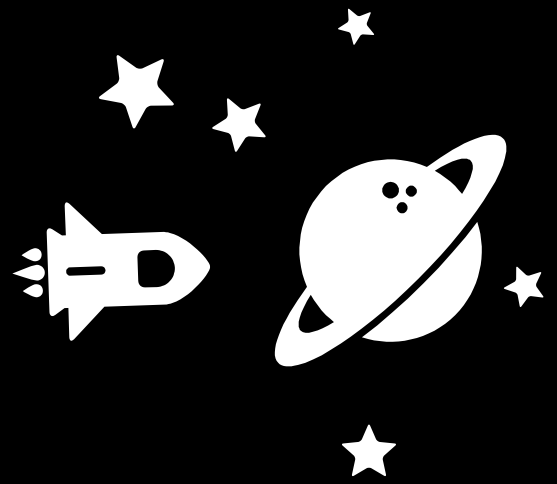
Don’t Believe the Hype

The Real Reasons to Be Excited about Machine Learning

I’m a programmer

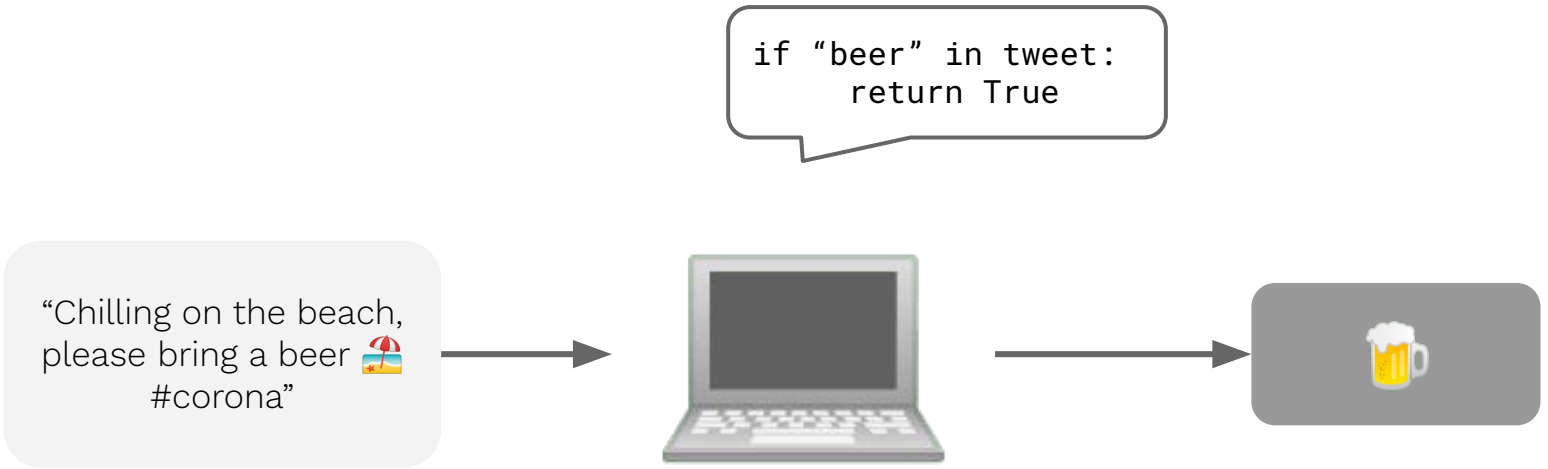
I’m a ML engineer

ML is not special

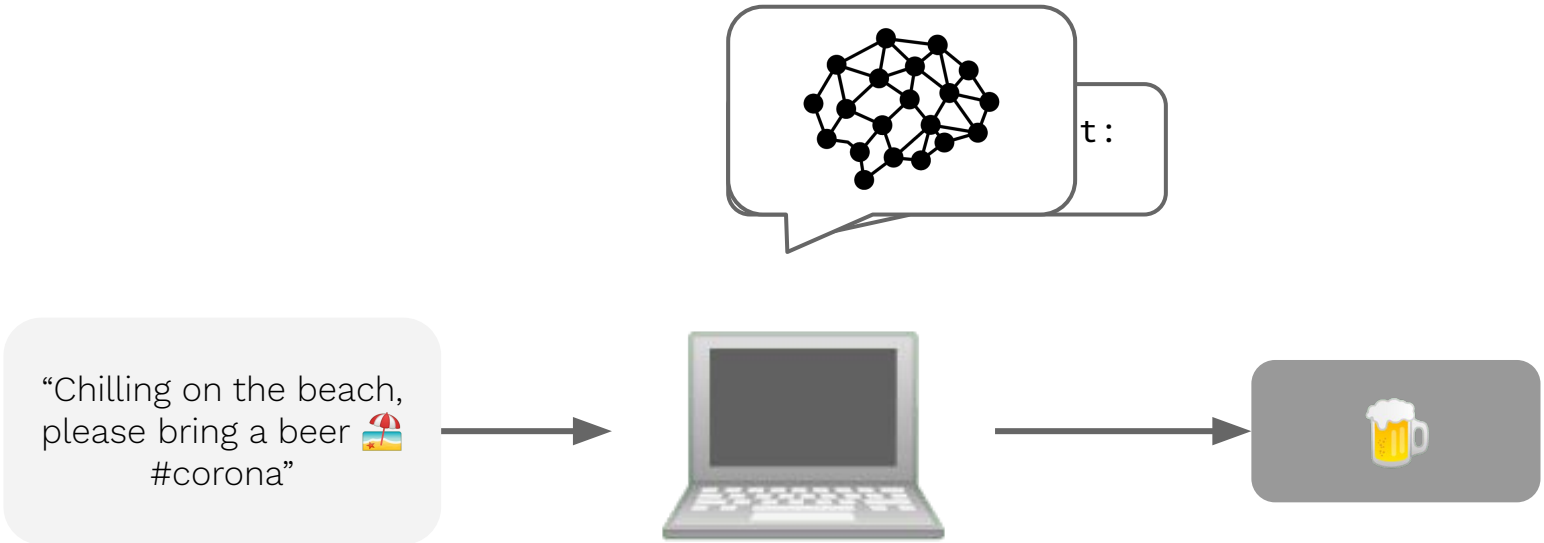


ML is a tool

- A device that carries out a **particular function**



#corona tweet filter



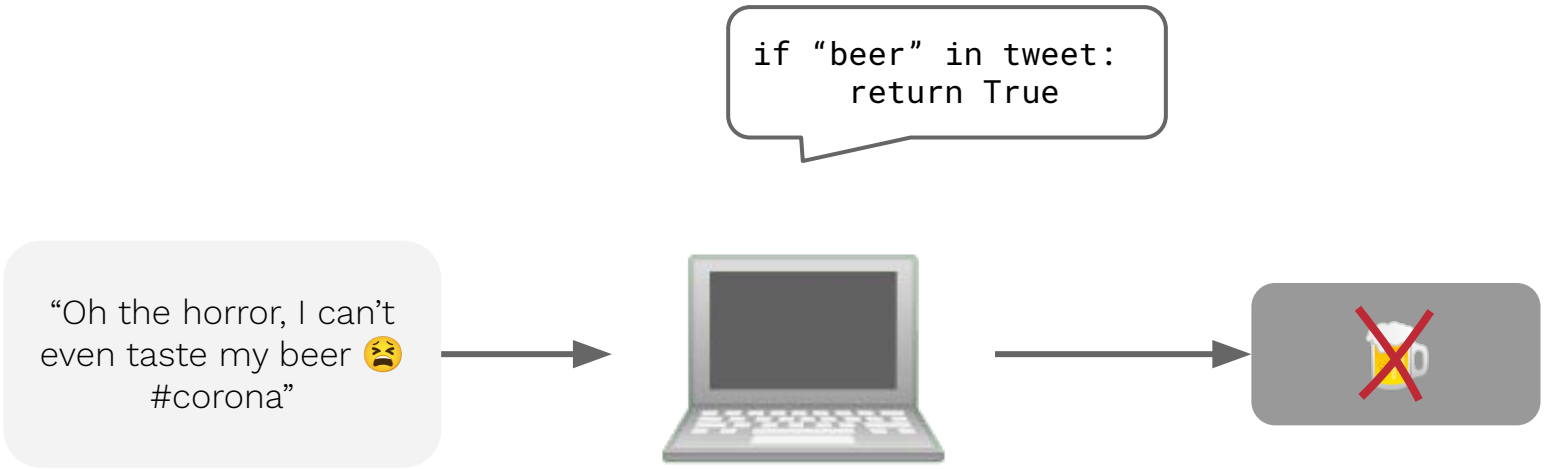
#corona tweet filter

ML is a **tool**

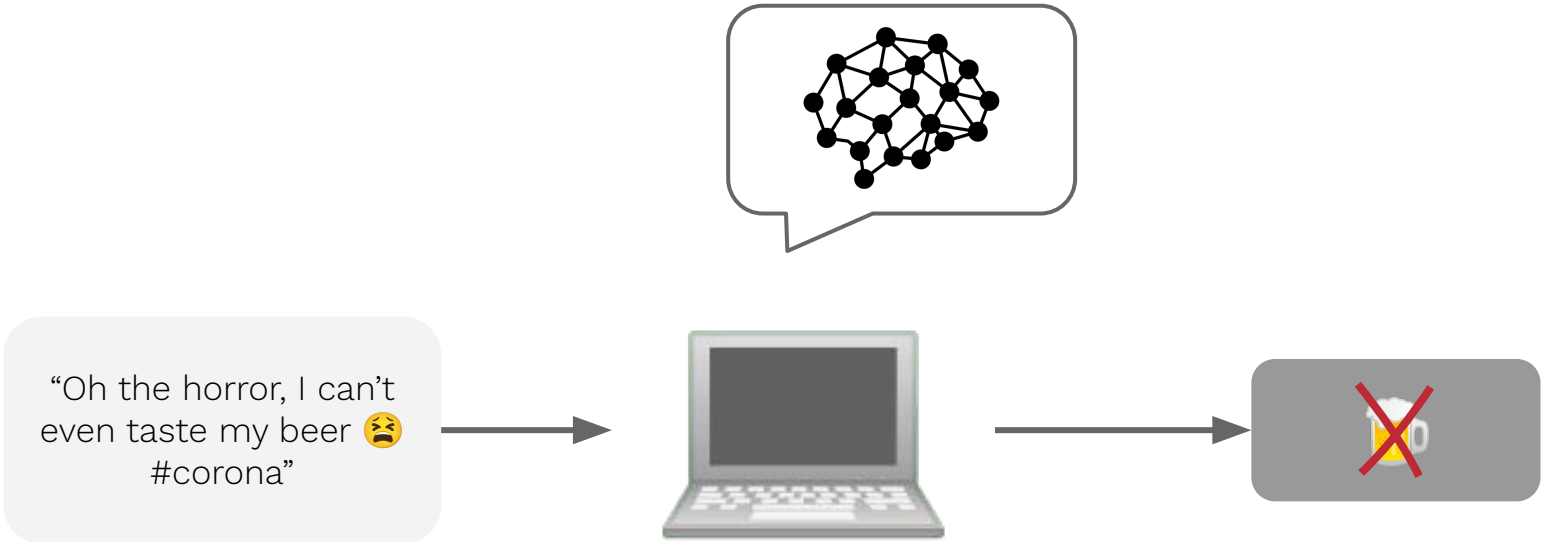
Media: AI will take over humanity
AI engineers and researchers:
©skt
I'm gonna pay you $100 to fuck off.

ML is a tool

- Tools must be **correctly configured**



#corona tweet filter



#corona tweet filter

ML is statistics

- ML tools come in the form of **models**
- ML models **transform** data
- These are **statistical learning** models, meaning they learn the desired transformation from examples
- Often, models act as **blackboxes**

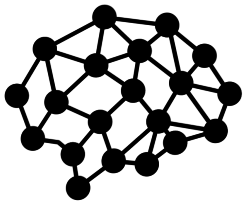
examples

“Chilling on the beach, please bring a beer 🏖️ #corona” + 🍺

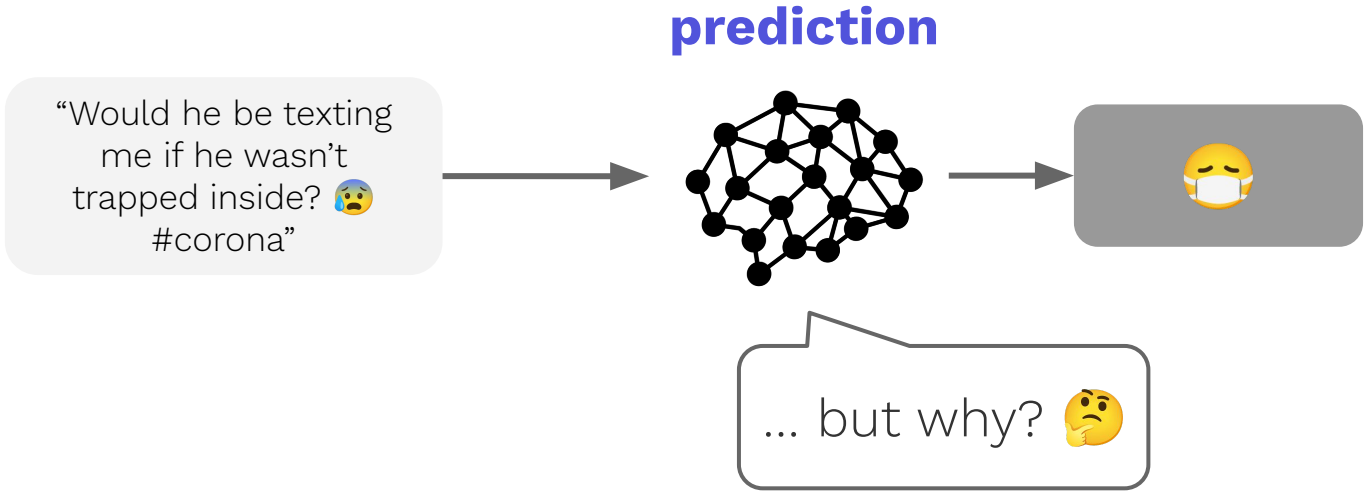
“Oh the horror, I can’t even taste my beer 😫 #corona” + 😷

“Remember wearing jeans hahahah that was so dumb #corona” + 😷

optimisation



#corona tweet filter



#corona tweet filter

ML is statistics

- **Optimisation** is not understanding
- **Blackbox** is not magic

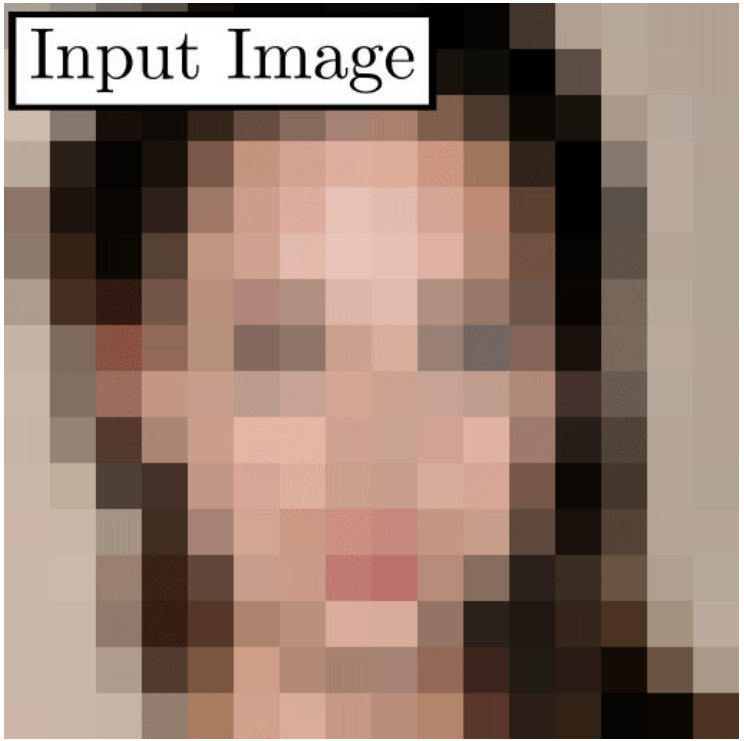
ML is **statistics**

ML is as good as its **inputs**

- Garbage in, garbage out
- Most datasets are **biased**[1]
- Models are **dumb**

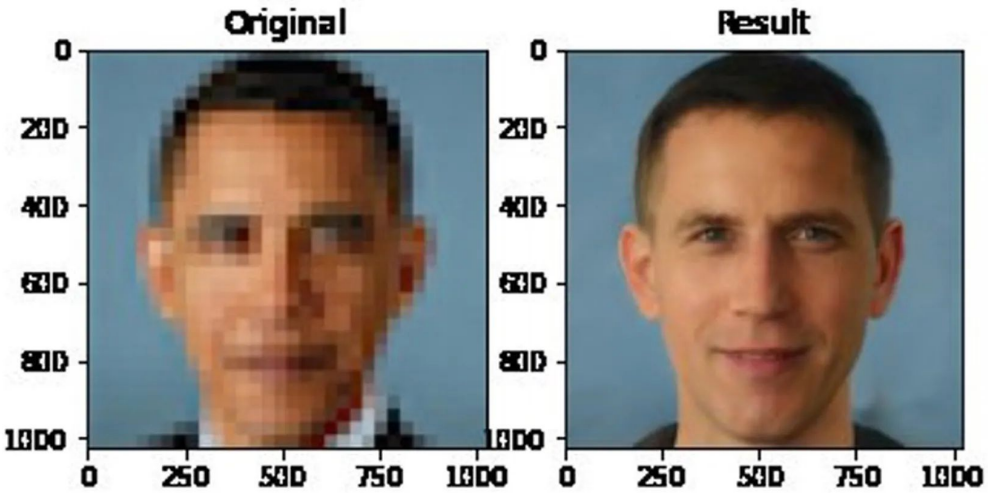
[1] Caroline Criado Perez, *Invisible Women: Exposing data bias in a world designed for men*

PULSE,
upsampling for
faces



https://github.com/adamian98/pulse

PULSE,
upsampling for
faces



https://twitter.com/bradpwyble/status/1274380641644294150

ML is as good as its creators

ML is engineered by **people**:

- Datasets are curated
- Models are tested
- Models are monitored

This can be done **badly**.

amazon

filtering job applications

Amazon's system taught itself that ***male*** *candidates were* ***preferable****.*

Reuters

ML is as good as its applications

ML is applied by **people**:

- **Products** are conceived then implemented with ML
- All product engineering requires **tech ethics**

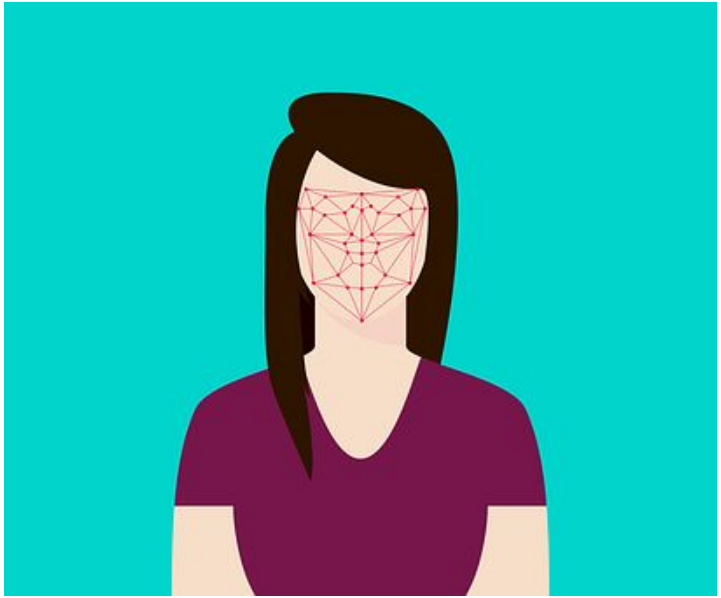
This can be done **badly**.

Microsoft

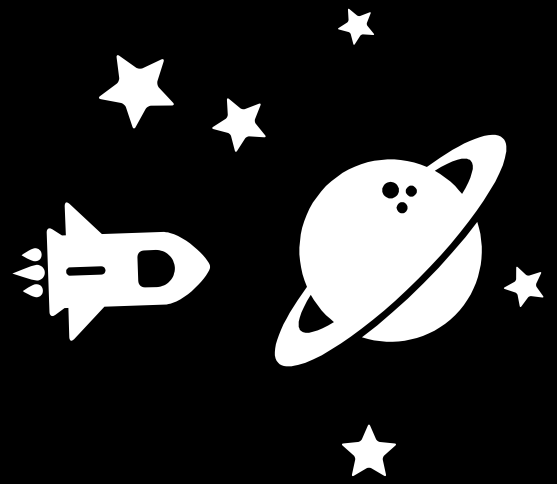
Tay, twitter chatbot

HARRISBURG UNIVERSITY
OF SCIENCE AND TECHNOLOGY

crime prediction by face recognition



ML is ~~not~~ very! special



ML is **cross-functional**

Cross-functional teams are **enriching**

- Computer science, statistics, engineering
- Product, design, communication
- Subject matter expertise

ML is **teamwork**

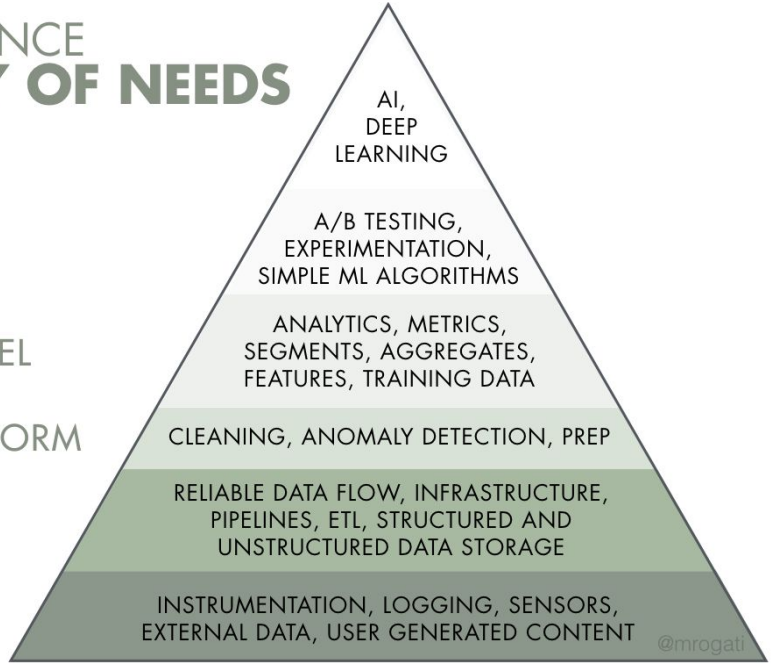
- Data Scientist vs ML Researcher vs ML Engineer

ML Jobs	ML Researcher	Data Scientist	ML Engineer
Purpose	create **papers**	create **insights**	create **products**
How	new algos, academic datasets	existing algos, real datasets	existing algos, real datasets

ML is **teamwork**

There's a role for **you**

THE DATA SCIENCE
HIERARCHY OF NEEDS



LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

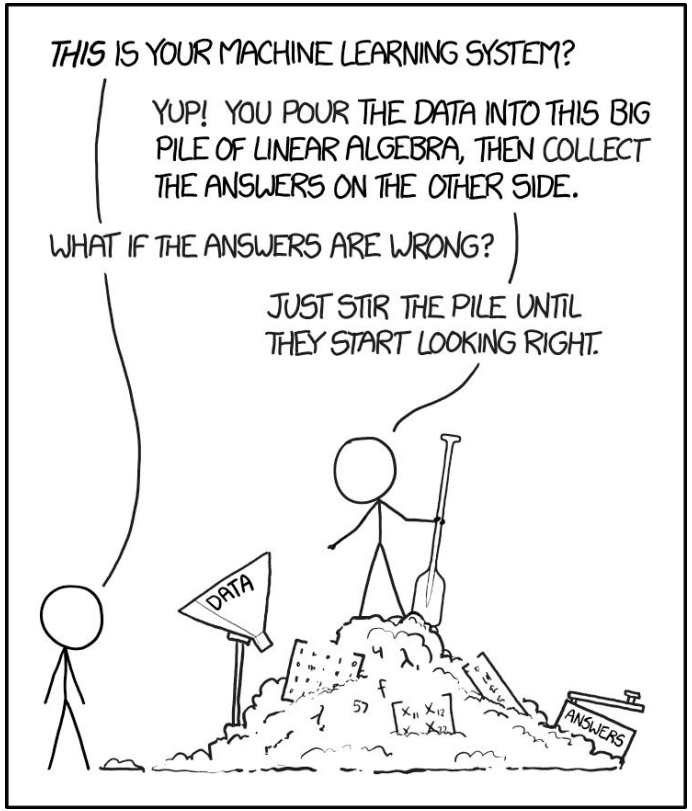
ML is **teamwork**

ML is **experimental**

Experiments are **exciting**

- ML accuracies are **unpredictable** and must be **tested**
- Models are developed **empirically** and **incrementally**

ML is **experimental**



xkcd 30

ML is **experimental**

- A new software paradigm: **software 2.0**

Software Engineer

1. Assess the problem
2. Design a solution
3. Implement the solution

ML Engineer

1. Assess the problem
2. Design experiments to find solution
3. Implement experiments
4. If found a solution, design solution
5. Implement the solution

ML is experimental

ML is **imperfect**

- Systems that can’t be debugged need **better testing**
- Systems that can’t be predicted need **better ethics**

ML is imperfect

Bugs without ML:
"My app froze"

Bugs with ML:
"My app told me to kill myself" [1]

[1] Nabla - GPT-3

ML is **imperfect**

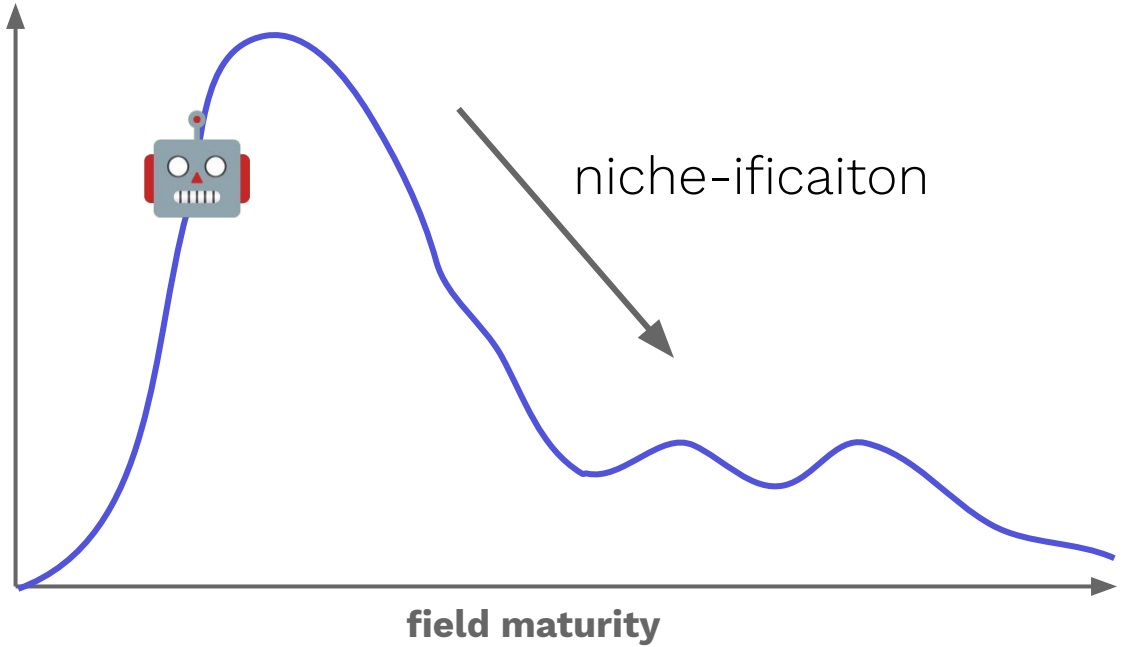
The field is still **growing**

- New roles for new challenges:
  ML QA, **MLOps**, **MLSecOps**, **AI Ethics**, ...

ML is creative

- Building effective ML models still requires **innovation**

Innovation is **challenging**



ML is **creative**

ML is creative

- The place of humans in ML engineering is still **evolving**

ML is creative

Gwern - On GPT-3
Andrew NG - Data Centric

ML is
open

- open source **research**
- open source **data**
- open source **code**
- open source **models**

3439

ML papers released last month on arxiv

78

ML papers released *today* on arxiv

ML is open

ML development is **fast**

Transformers

40M$

40k ⭐

ML is open

Tech Crunch - Huggingface raises 40M

ML is open

- The ML field has grown as a **collaborative** effort
- The AI revolution couldn’t have happened without an **open culture**

Contributing is
rewarding

ML is
open

When a GitHub repo you have contributed reaches 1k stars

ML is **fast**

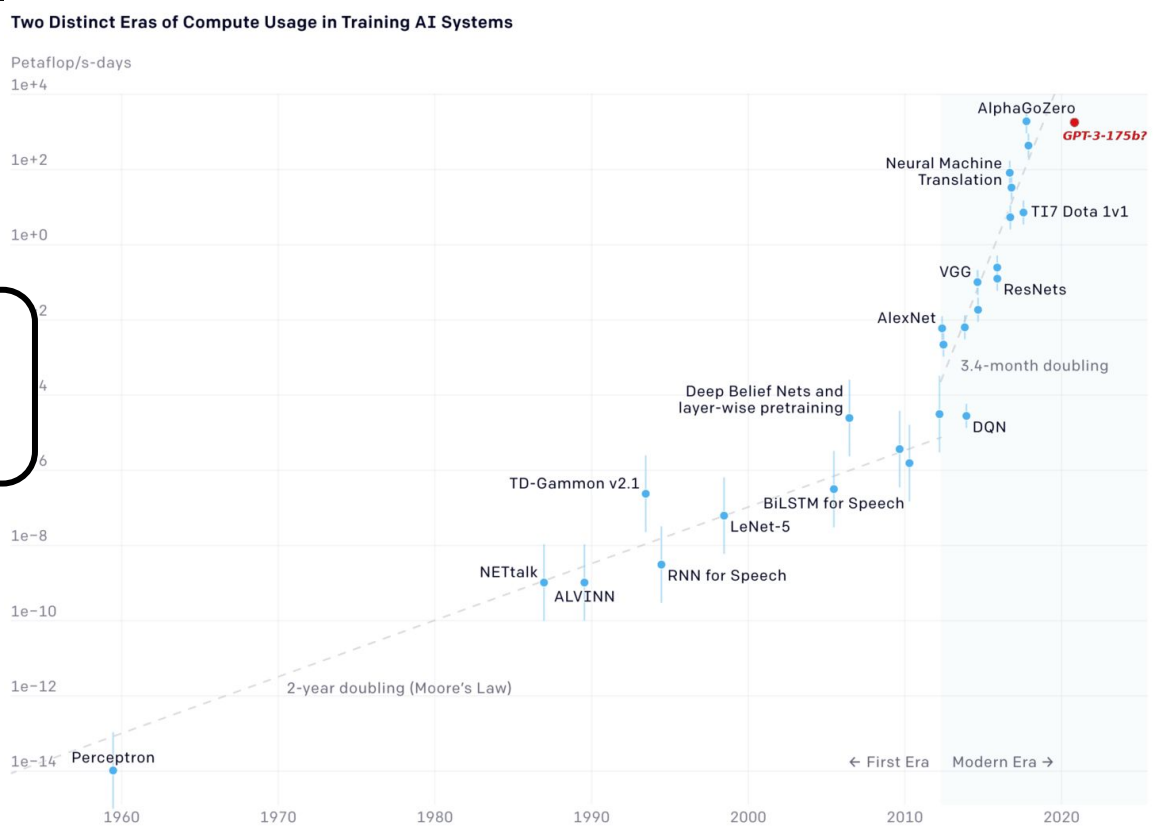
- Like **blazingly** fast 🔥

Image Classification SOTA

ML is fast

Working with cutting-edge tech is **cool**

ML is **fast**



ML is **fast**

ML feels like a
level playing field

- Fast learning skills > deep expertise

ML is impactful

- ML has power for both **good** and **bad**

ML is **impactful**

The Ethical Questions that Haunt Facial Recognition Research - Nature

Greg Baker/AFP/Getty

Having an impact is **motivating**

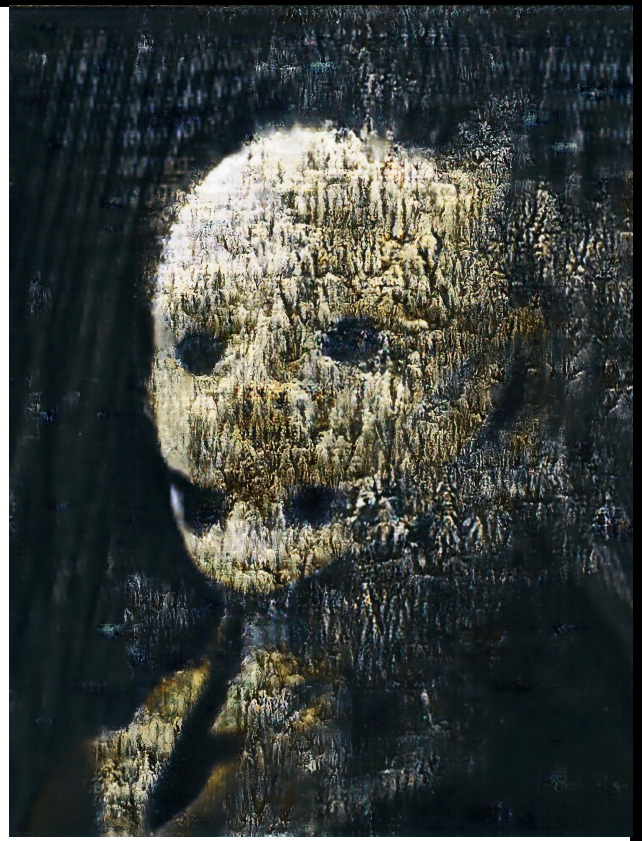
AI for Good Global Summit

ML is **impactful**

ML is impactful

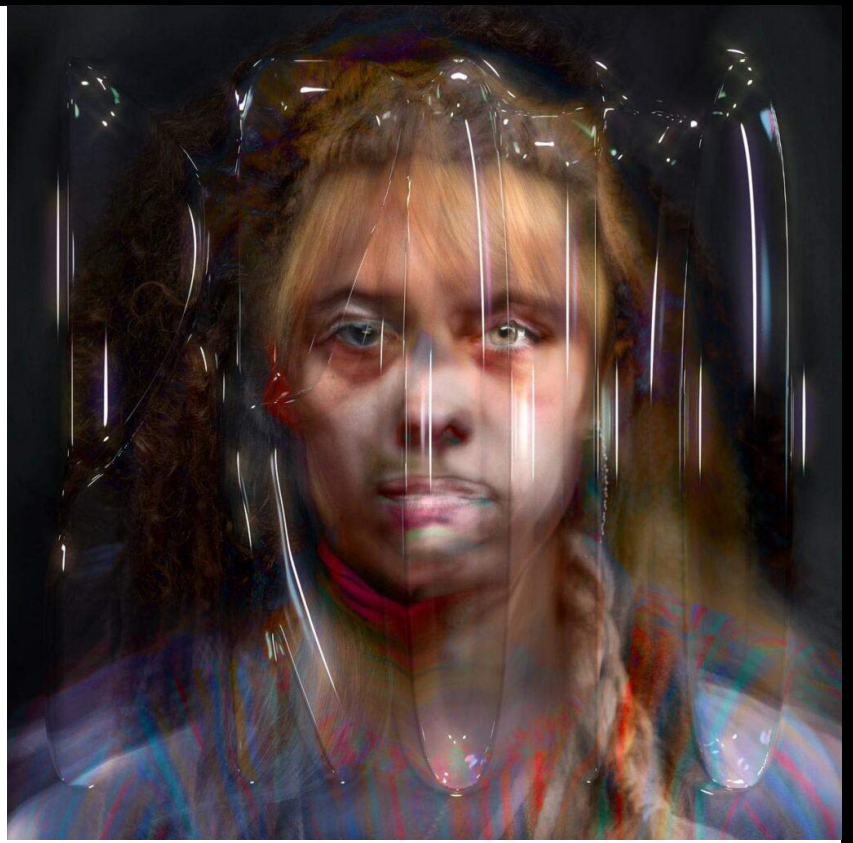
- ML is changing **automation** beyond regular software engineering

ML is **impactful**



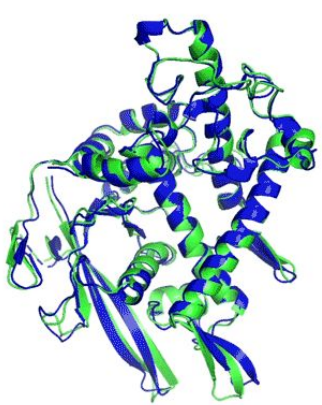
Mario Klingemann aka Quasimondo - Neural Decay

ML is **impactful**

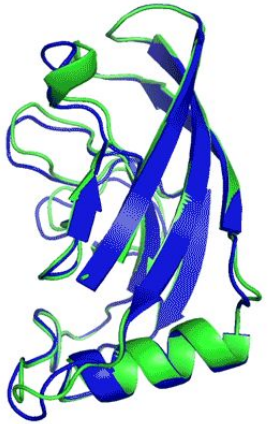


ML is enlightening

- ML models can **level-up** Science



T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)



T1049 / 6y4f
93.3 GDT
(adhesin tip)

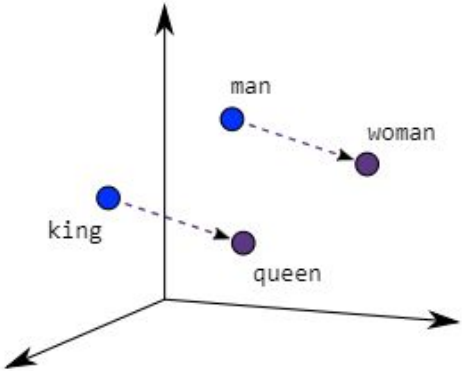
ML is **enlightening**

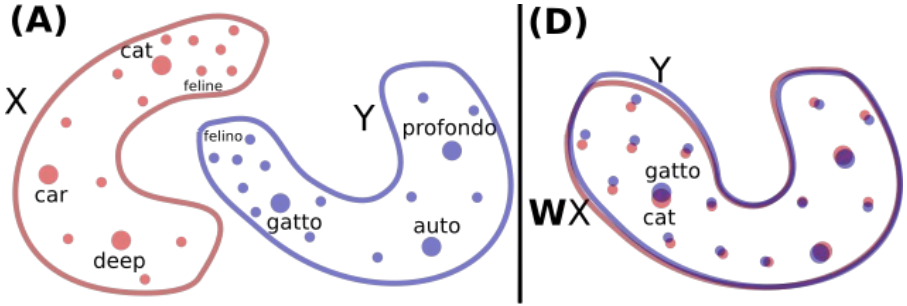
Experimental result
Computational prediction

ML is enlightening

- ML models can **teach us** about Science

Pushing the limits of knowledge is **inspiring**





ML is **enlightening**

Facebook - MUSE
Google - NLP Embeddings

ML is

cross-functional
teamwork
experimental
imperfect
creative
open
fast
impactful
enlightening

6-week live training

A hands-on journey from raw data to useful insights

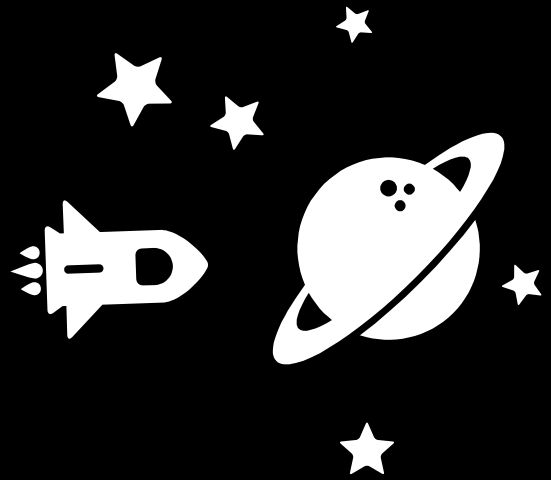
Training period: May 24 - July 2 2021

6-week training to become Data Scientist, practice and gain strong foundations in machine learning.

Learn more

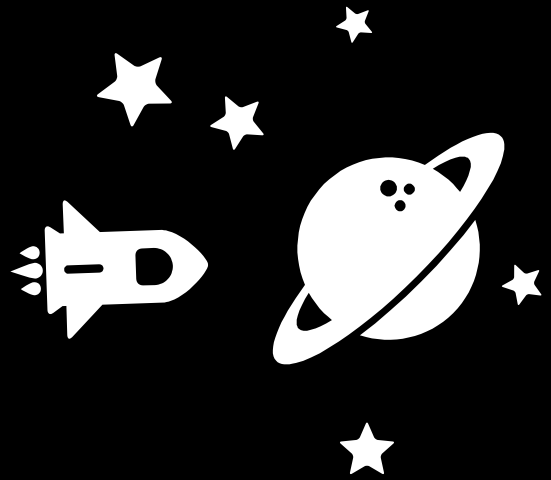
Q&A

(also we're hiring 🦐)



camille.vanhoffelen@gmail.com
https://www.linkedin.com/in/camillevanhoffelen/

List of explainability and bias in AI frameworks
Have my photos been used to build face recognition AI?
Discussion of bias in technology



Bonus Refs