CHALLENGES & OPPORTUNITIES IN ML

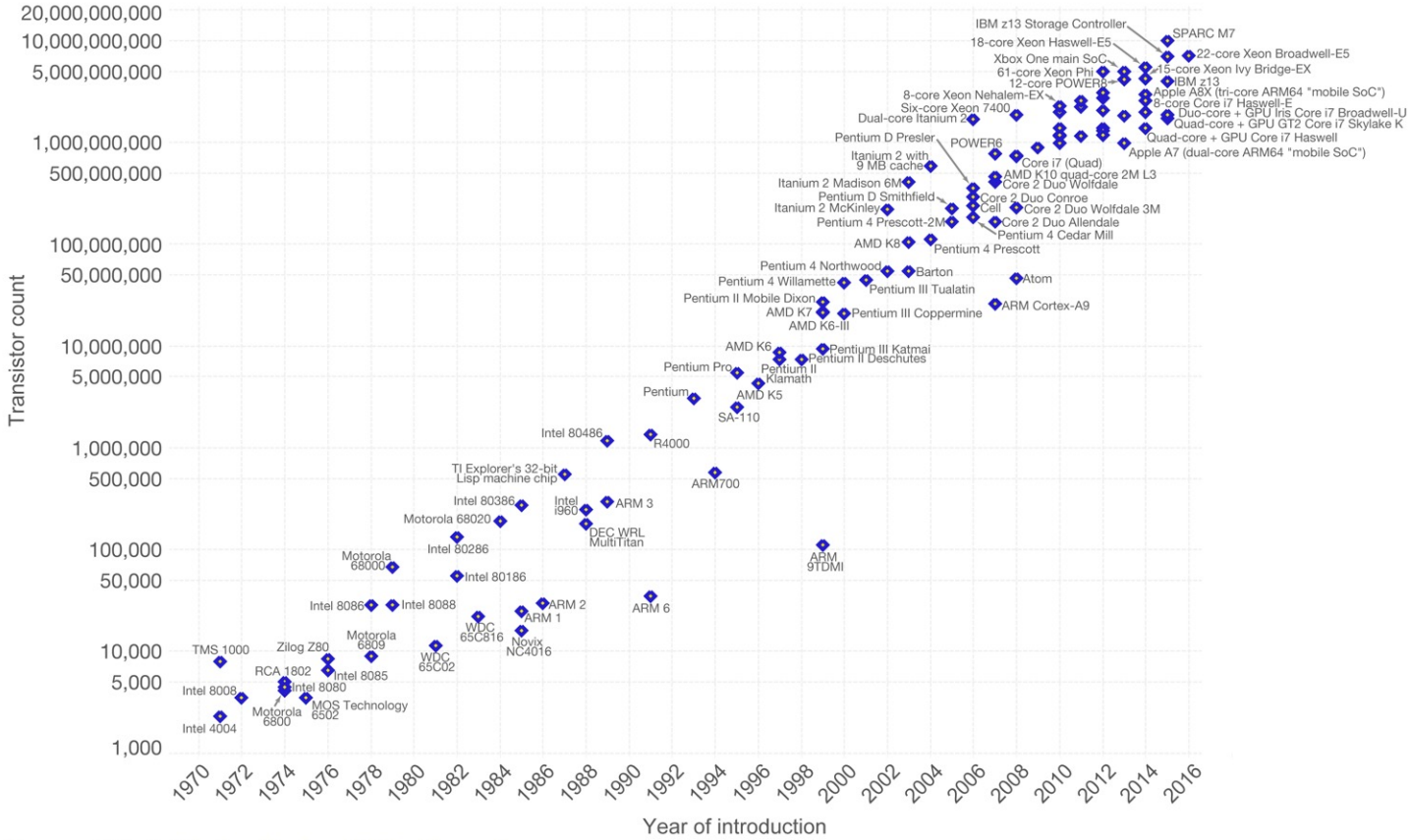
SAMET OYMAK


Machine Learning
is everywhere!
ВЫХОД В ГОРОД
ACCESS TO CITY
What can I help you with?

Moore's Law – The number of transistors on integrated circuit chips (1971-2016)

Our World in Data

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important as other aspects of technological progress – such as processing speed or the price of electronic products – are strongly linked to Moore's law.



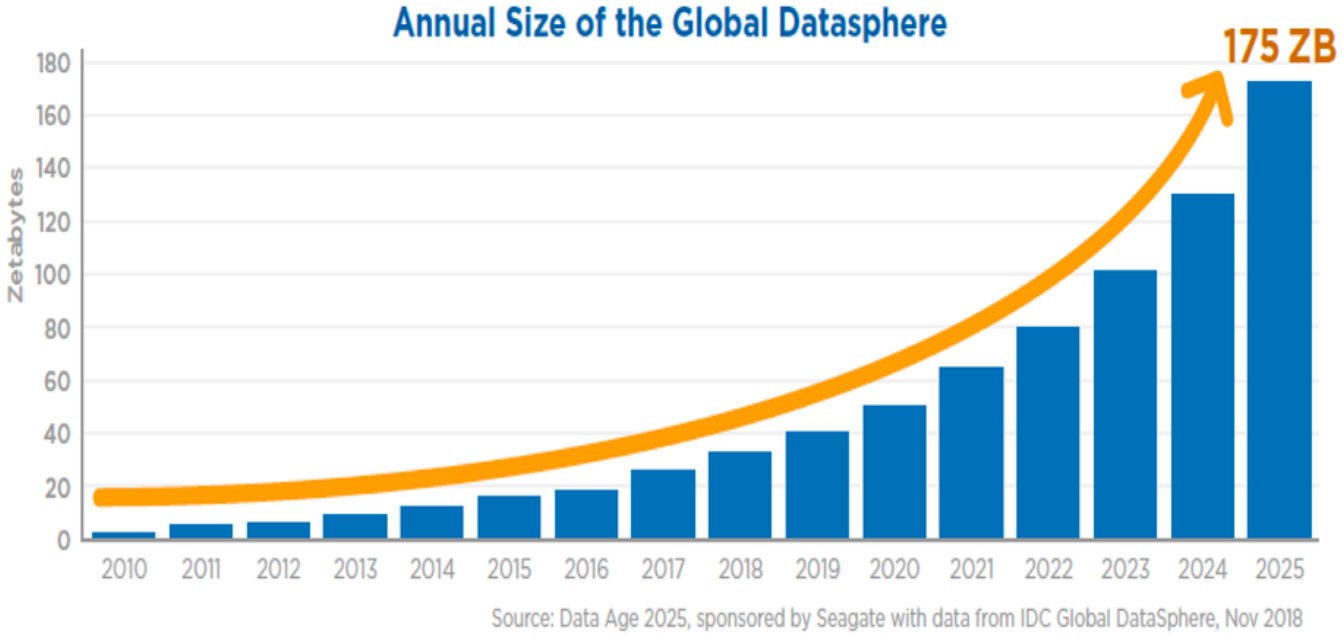
Data source: Wikipedia (https://en.wikipedia.org/wiki/Transistor_count)

The data visualization is available at OurWorldinData.org. There you find more visualizations and research on this topic.

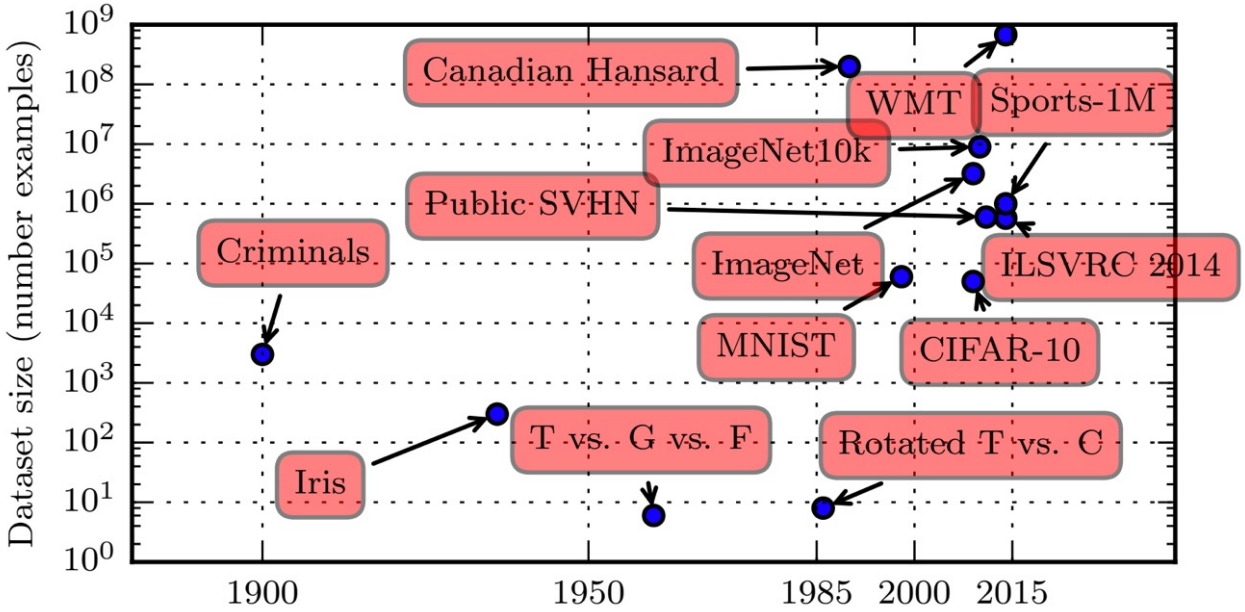
Rise of GPUs


GEFORCE GTX
GTX 1080

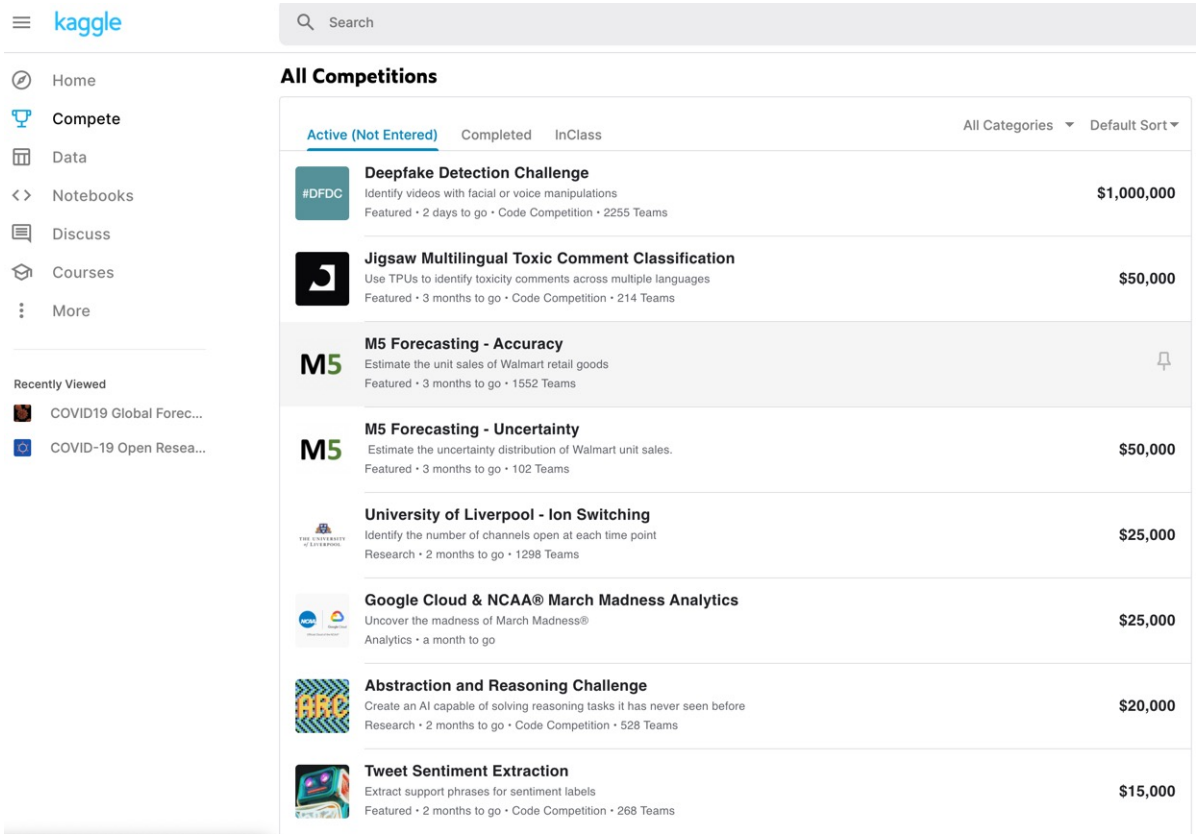

Market Summary > NVIDIA Corporation
269.42 USD
+243.42 (936.23%) ↑ past 5 years
Closed: Jan 14, 7:59 PM EST • Disclaimer
After hours 268.85 −0.57 (0.21%)
Follow
1D 5D 1M 6M YTD 1Y 5Y Max
400 25.37 USD Feb 24, 2017
300
200
100
0
2018 2019 2020 2021 2022
Open 263.00
High 271.97
Low 262.10
Mkt cap 673.55B
P/E ratio 83.17
Div yield 0.059%
CDP score B
52-wk high 346.47
52-wk low 115.66


Annual Size of the Global Datasphere
175 ZB
Zetabytes
180
160
140
120
100
80
60
40
20
0
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

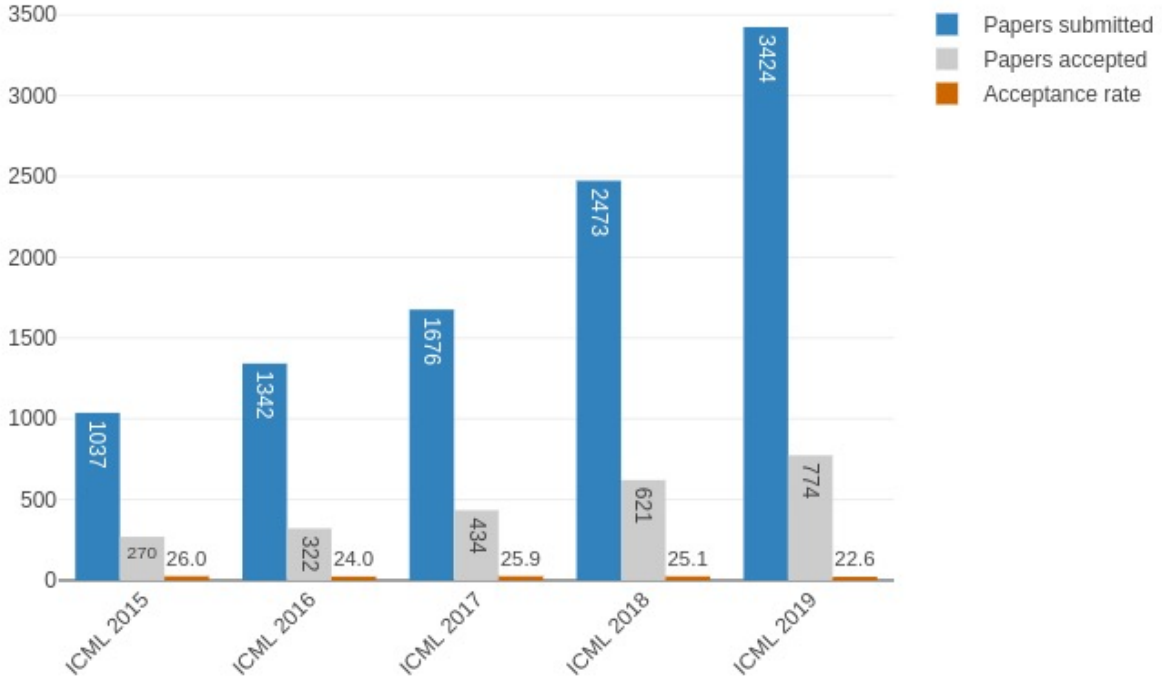
Growing ML Datasets



Resources for Quality Data



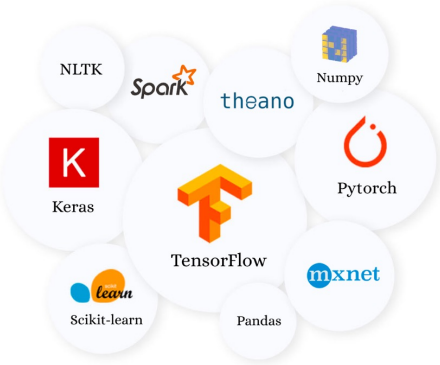
Academic Interest

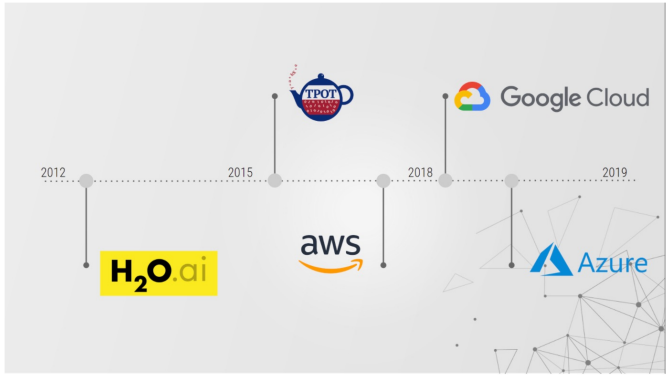


# of papers in International Conference on Machine Learning

Industry Interest

Growing ML/DL software





A brief history of AutoML platforms.

Research labs

Microsoft Research

Facebook AI Research

amazon | science

Google AI

many many more

Trends in ML

Deep learning as a rising ML paradigm

ML for healthcare

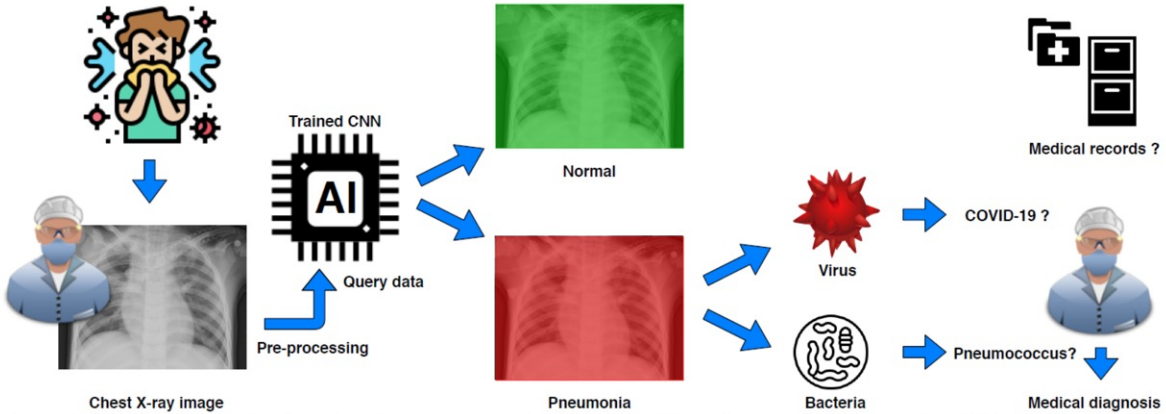
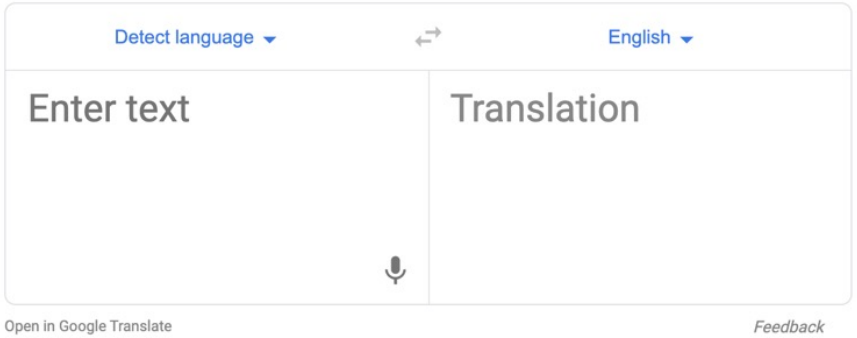
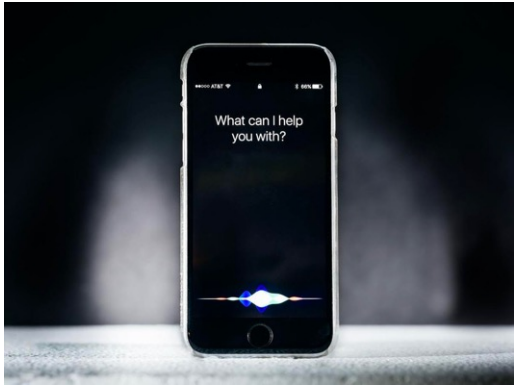


Fig. 1: Global workflow using deep learning for **automatic detection of infection towards supporting COVID-19 screening** from chest X-ray images. In a COVID-19 epidemic context, a detected viral pneumonia can particularly presume a COVID-19 infection.

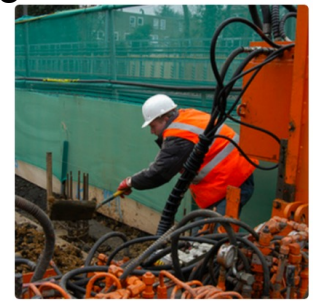
Natural Language Processing





Application: Image to Sentence

"man in black shirt is playing guitar."

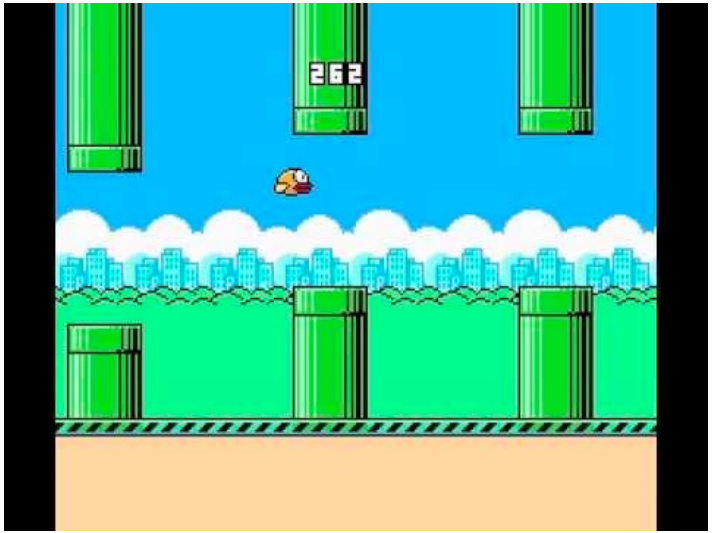


"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

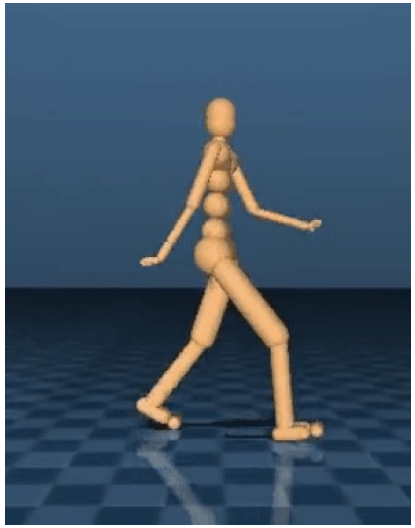
Deep Reinforcement Learning

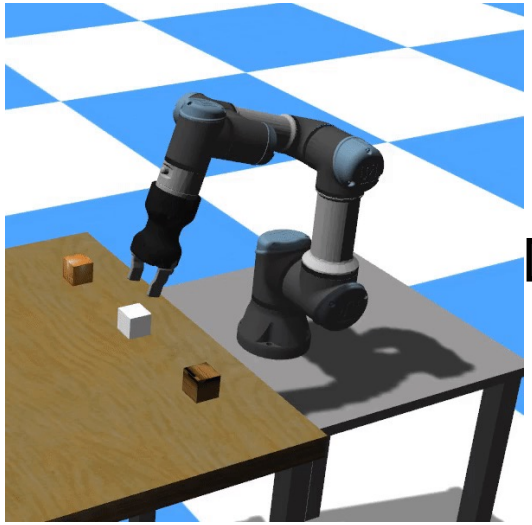
Q: How can we make optimal **decision** in competitive environments?

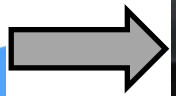


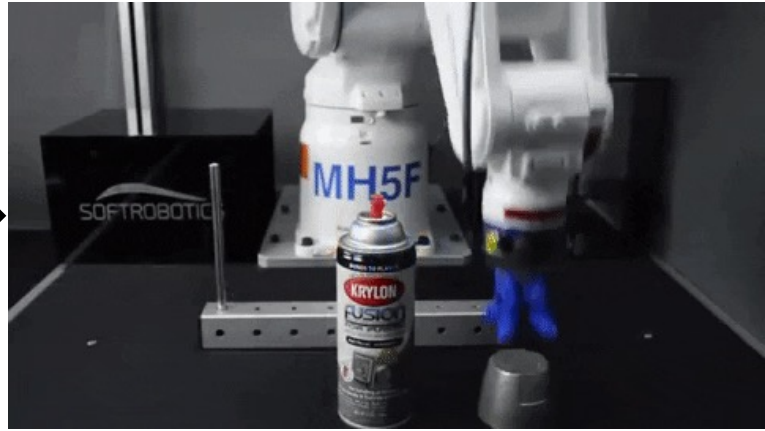
https://github.com/yenchenlin/DeepLearningFlappyBird

Simulation environments to real world









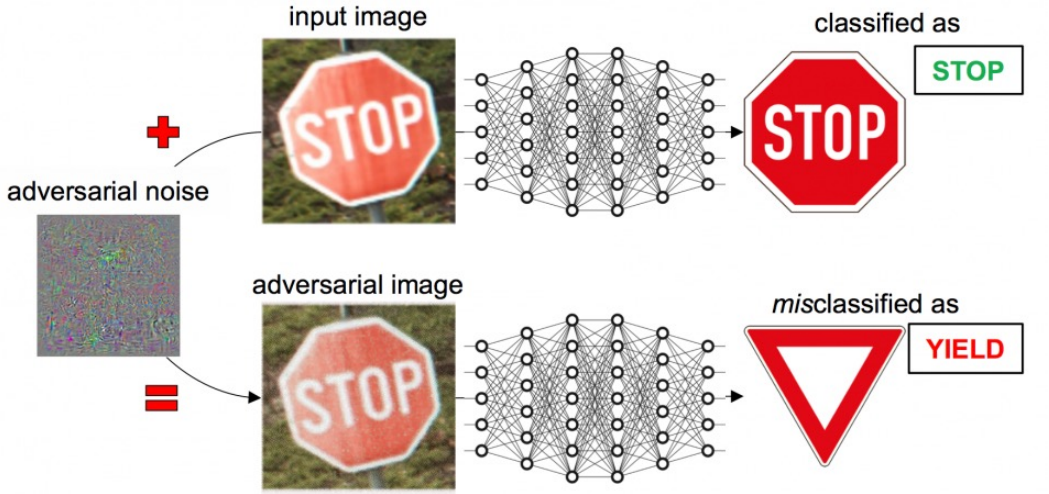
Simulation environments:

- OpenAI Gym
- MuJoCo

Challenges in ML

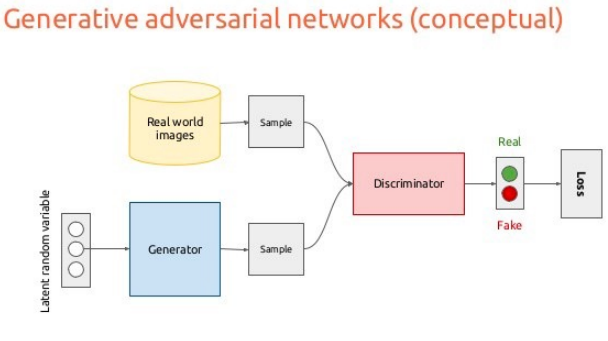
Adversarial Learning

- Can we mess with an ML model?

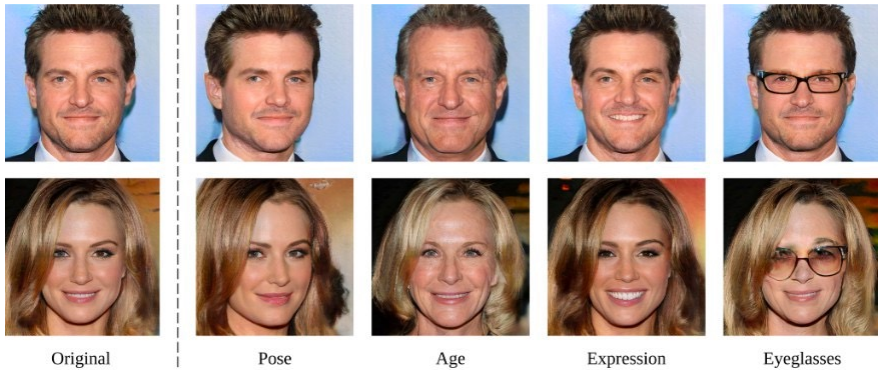


[Exercise: Digit recognition]

Generative Models



Q: Can we learn to generate fake content?



Deepfakes

All Competitions

Active (Not Entered) | Completed | InClass

All Categories | Default Sort

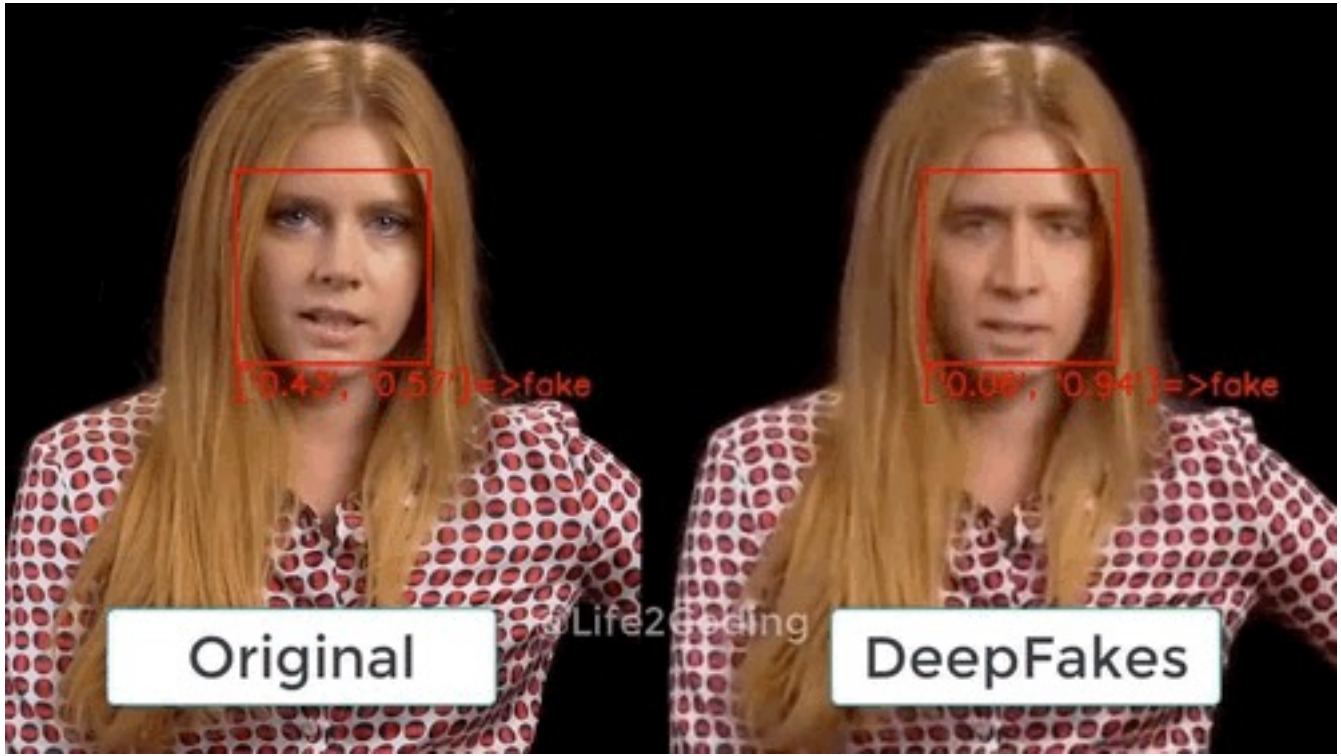
#DFDC

Deepfake Detection Challenge

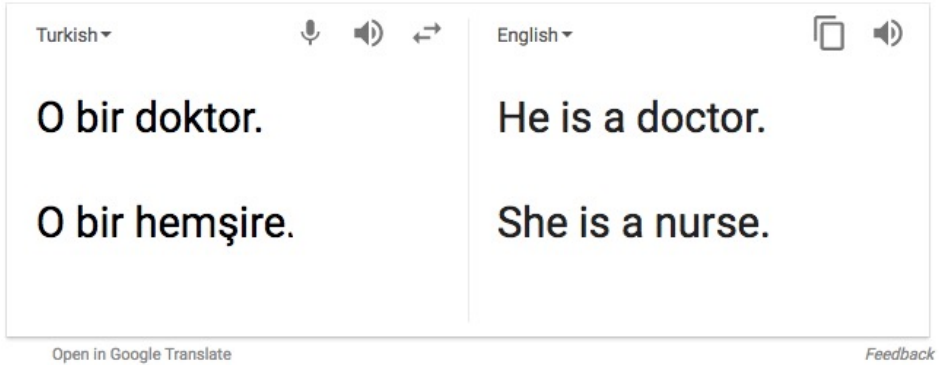
Identify videos with facial or voice manipulations

Featured • 2 days to go • Code Competition • 2255 Teams

$1,000,000



Algorithmic Bias



Amazon ditched AI recruiting tool that favored men for technical jobs

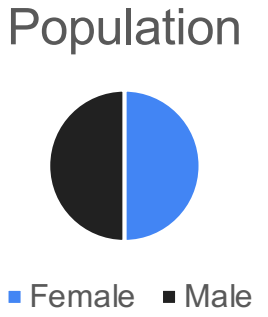
Specialists had been building computer programs since 2014 to review résumés in an effort to automate the search process

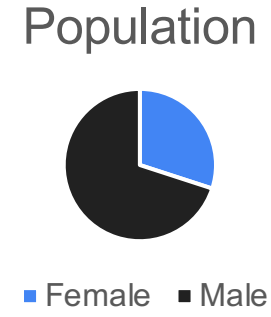
Amazon's automated hiring tool was found to be inadequate after penalizing the résumés of female candidates. Photograph: Brian Snyder/Reuters

[Exercise: Salary prediction on Adult dataset]

Algorithmic Bias

- Source of bias is often under-representation
- Female and Male population similar
- Example: Within high income earners, females might be under-represented
- ML algorithms might pick up such bias



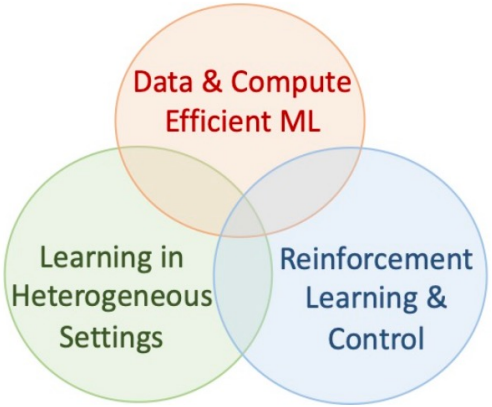


My research

On algorithmic bias: AutoML algorithms for fairness

AutoBalance: Optimized Loss Functions for Imbalanced Data

Label-Imbalanced and Group-Sensitive Classification under Overparameterization



Current focus: Foundations of learning & decision making via efficient models, algorithms, representations.

Current topics: heterogeneous data, reinforcement learning, autoML & model compression, deep learning theory.

Prospective PhDs: Research experience in ML and solid background on optimization/statistics is important.

Undergraduates: Basic Python knowledge is necessary. Experience with PyTorch or TensorFlow is a plus.

Send resume/interest to oymak@ece for research projects.

Thanks for listening