

# CHALLENGES & OPPORTUNITIES IN ML

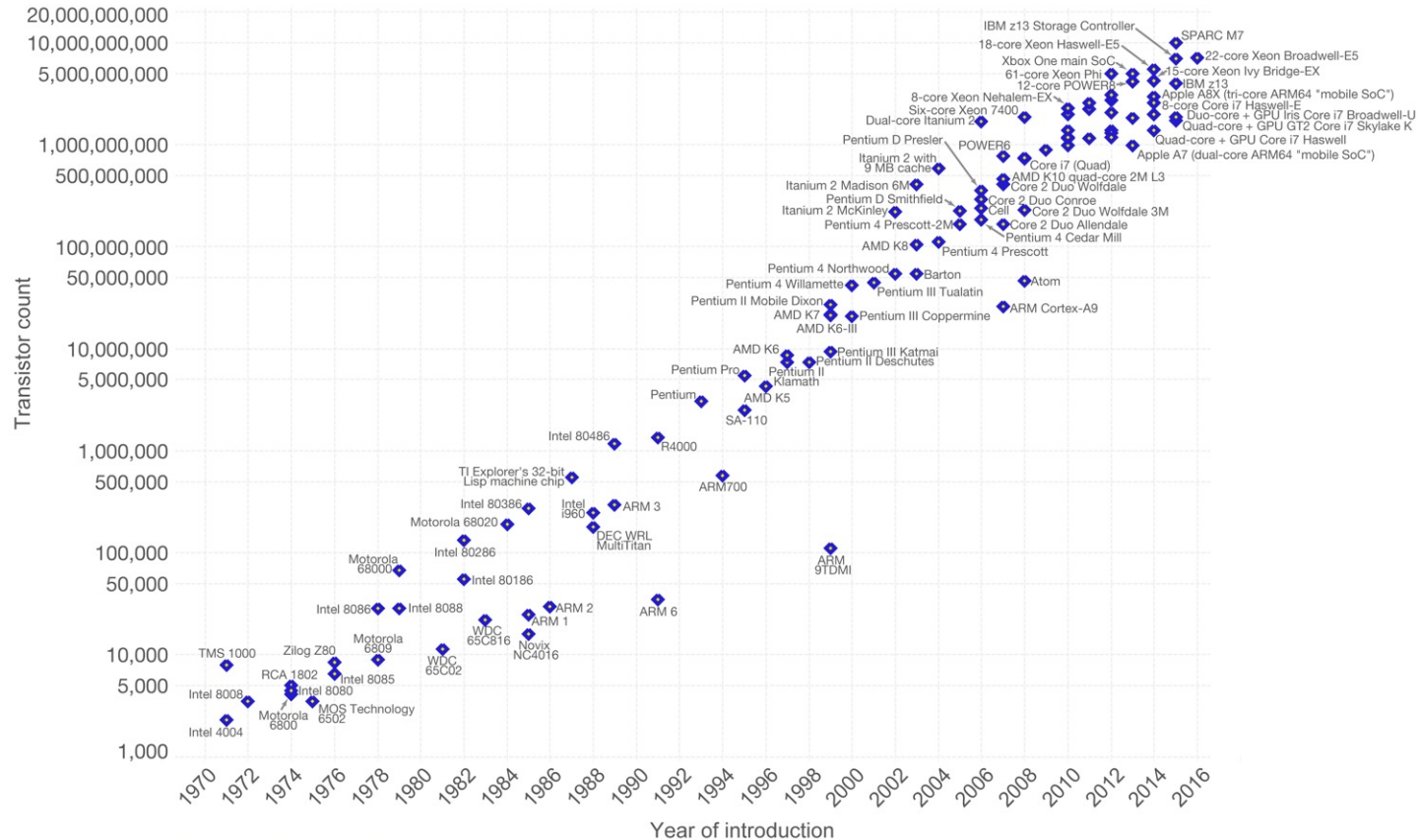
SAMET OYMAK





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

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](https://en.wikipedia.org/wiki/Transistor_count))

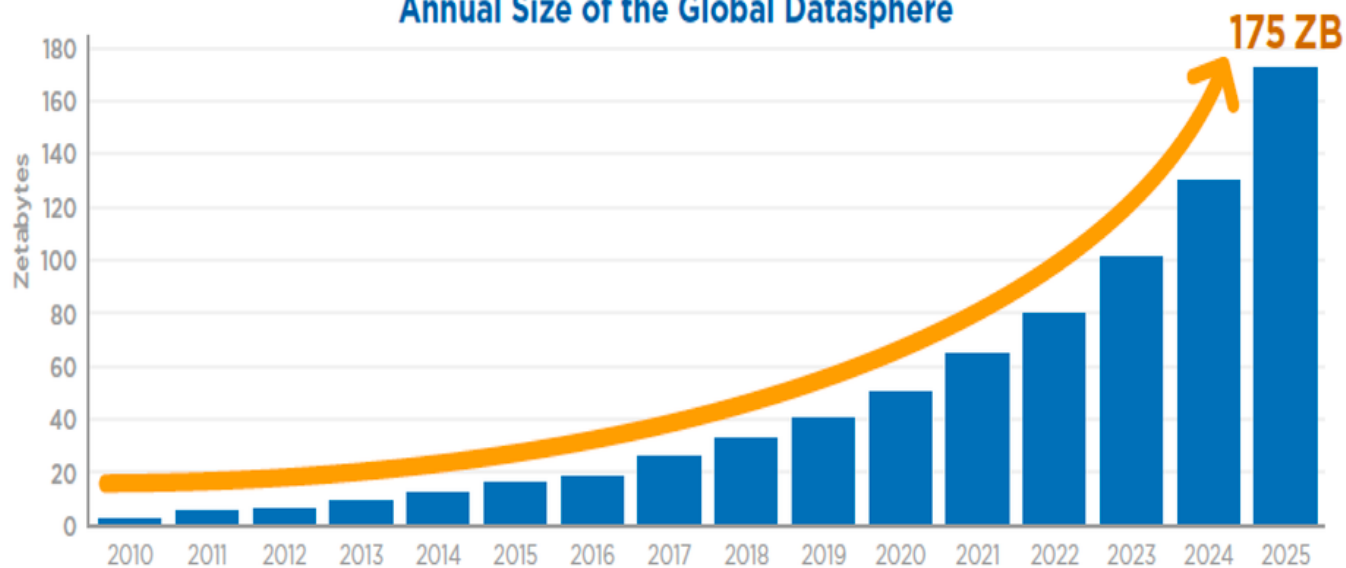
The data visualization is available at [OurWorldinData.org](https://ourworldindata.org). There you find more visualizations and research on this topic.

Licensed under CC-BY-SA by the author Max Roser.

# Rise of GPUs

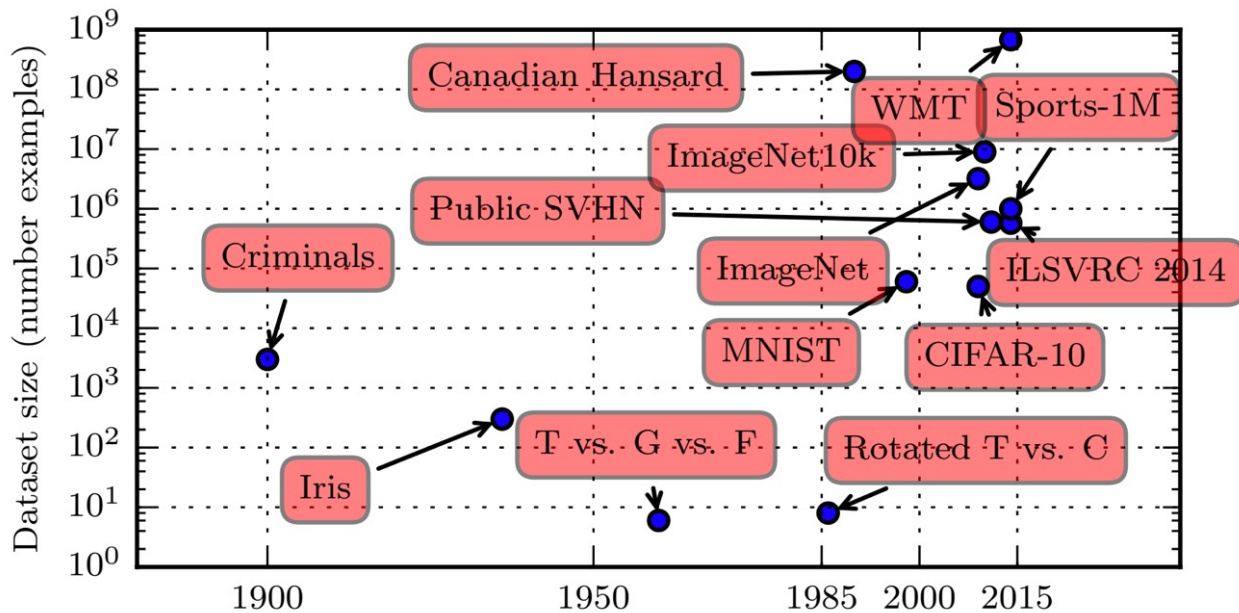


## Annual Size of the Global Datasphere






Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018


# Growing ML Datasets





# Resources for Quality Data


 Search


 Home


 Compete

 Data


 Notebooks


 Discuss

 Courses

 More

Recently Viewed

 COVID19 Global Forec...

 COVID-19 Open Resea...

## All Competitions

Active (Not Entered)CompletedInClass

All CategoriesDefault Sort

#DFDC

**Deepfake Detection Challenge**  
Identify videos with facial or voice manipulations  
Featured • 2 days to go • Code Competition • 2255 Teams

\$1,000,000



**Jigsaw Multilingual Toxic Comment Classification**  
Use TPUs to identify toxicity comments across multiple languages  
Featured • 3 months to go • Code Competition • 214 Teams

\$50,000

M5

**M5 Forecasting - Accuracy**  
Estimate the unit sales of Walmart retail goods  
Featured • 3 months to go • 1552 Teams



M5

**M5 Forecasting - Uncertainty**  
Estimate the uncertainty distribution of Walmart unit sales.  
Featured • 3 months to go • 102 Teams

\$50,000



**University of Liverpool - Ion Switching**  
Identify the number of channels open at each time point  
Research • 2 months to go • 1298 Teams

\$25,000



**Google Cloud & NCAA® March Madness Analytics**  
Uncover the madness of March Madness®  
Analytics • a month to go

\$25,000



**Abstraction and Reasoning Challenge**  
Create an AI capable of solving reasoning tasks it has never seen before  
Research • 2 months to go • Code Competition • 528 Teams

\$20,000

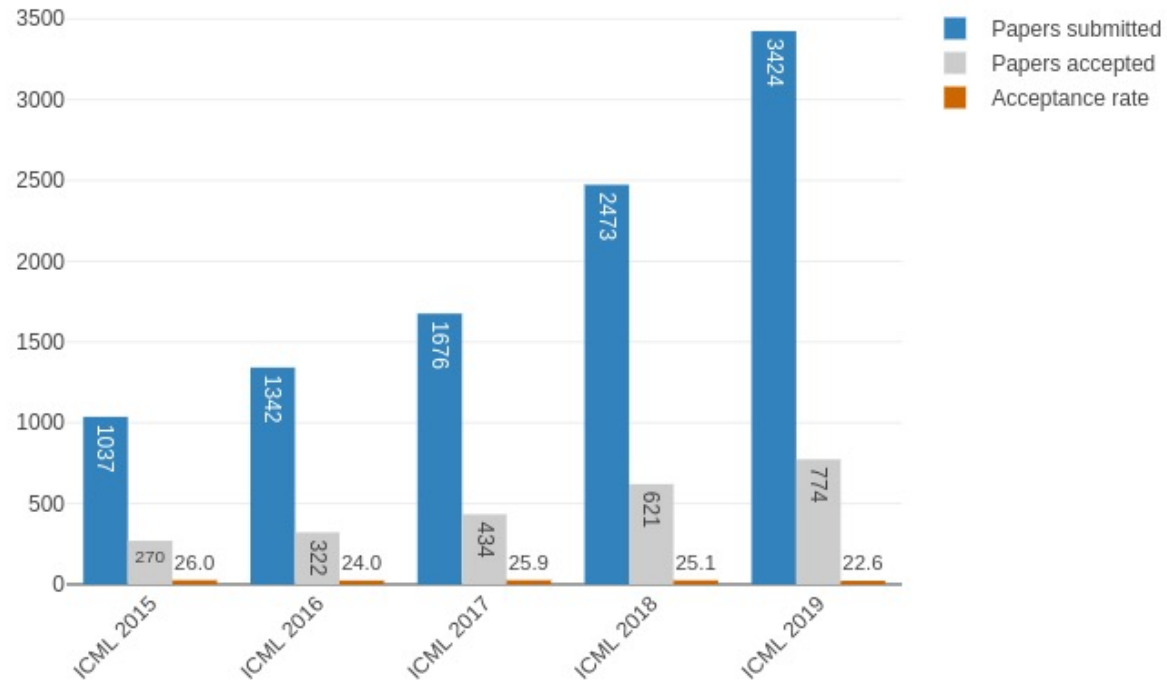


**Tweet Sentiment Extraction**  
Extract support phrases for sentiment labels  
Featured • 2 months to go • Code Competition • 268 Teams

\$15,000



## Academic Interest

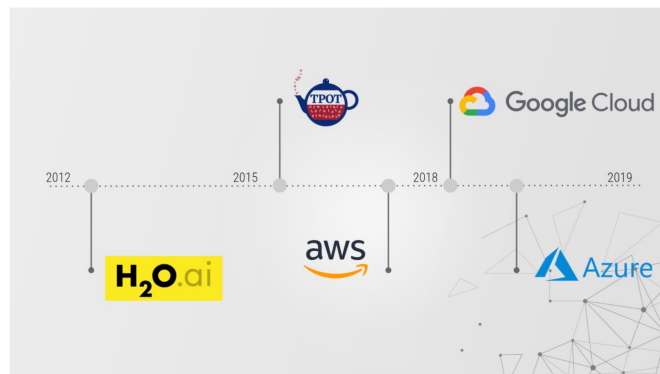


# of papers in International Conference on Machine Learning



# Industry Interest

## Growing ML/DL software



*A brief history of AutoML platforms.*

## Research labs



many many more

# Trends in ML

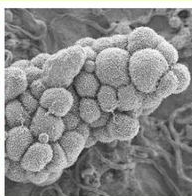
# Deep learning as a rising ML paradigm

## DEEP LEARNING EVERYWHERE



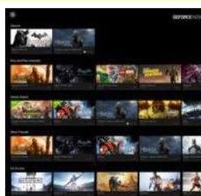
### INTERNET & CLOUD

Image Classification  
Speech Recognition  
Language Translation  
Language Processing  
Sentiment Analysis  
Recommendation



### MEDICINE & BIOLOGY

Cancer Cell Detection  
Diabetic Grading  
Drug Discovery



### MEDIA & ENTERTAINMENT

Video Captioning  
Video Search  
Real Time Translation



### SECURITY & DEFENSE

Face Detection  
Video Surveillance  
Satellite Imagery



### AUTONOMOUS MACHINES

Pedestrian Detection  
Lane Tracking  
Recognize Traffic Sign

# ML for healthcare

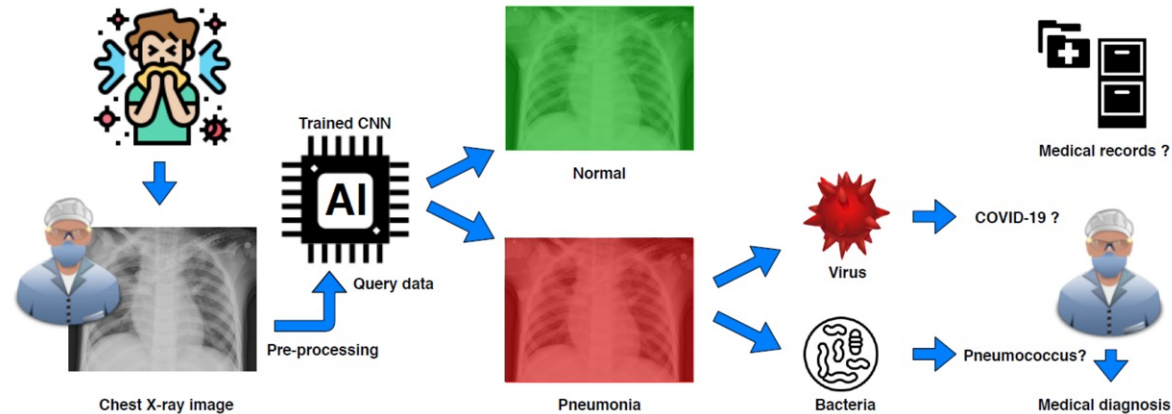

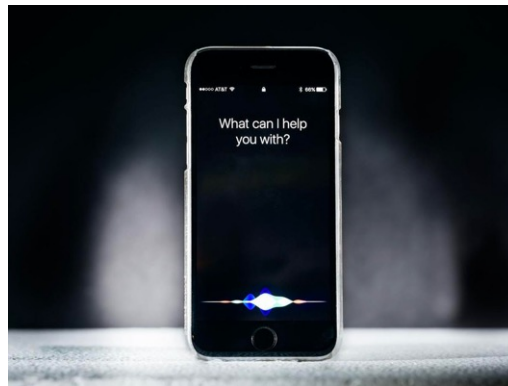


Fig. 1: Global workflow using deep learning for **automatic detection of infection towards supporting COVID-19 screening** from chest X-ray images. In a COVID-19 epidemic context, a detected viral pneumonia can particularly presume a COVID-19 infection.

# Natural Language Processing

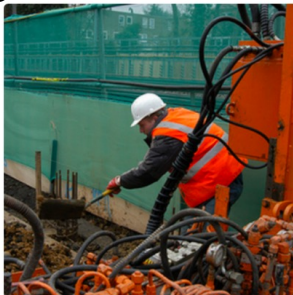
Detect language ▼	↔	English ▼
Enter text		Translation
		
<a href="#">Open in Google Translate</a>		<a href="#">Feedback</a>



## Application: Image to Sentence



"man in black shirt is playing guitar."



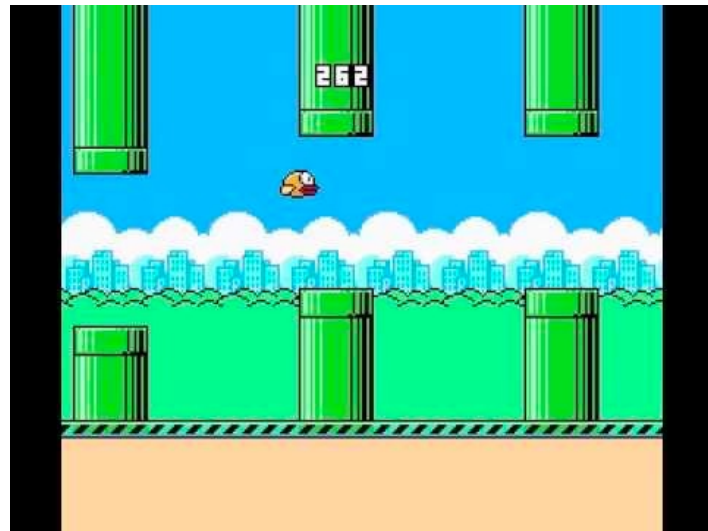
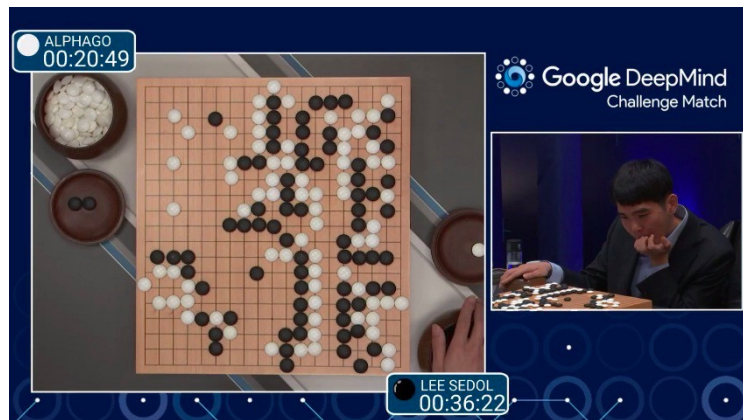
"construction worker in orange safety vest is working on road."



"two young girls are playing with lego toy."

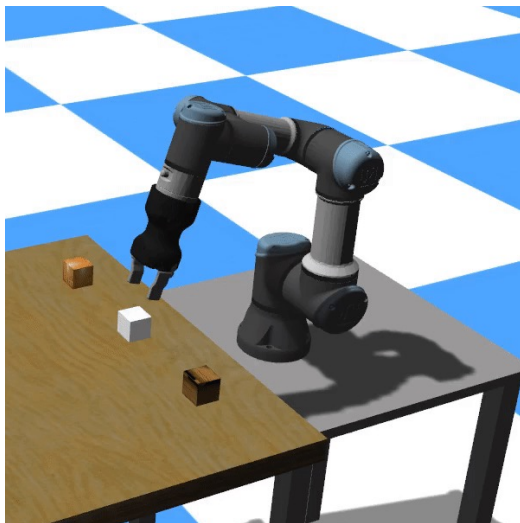
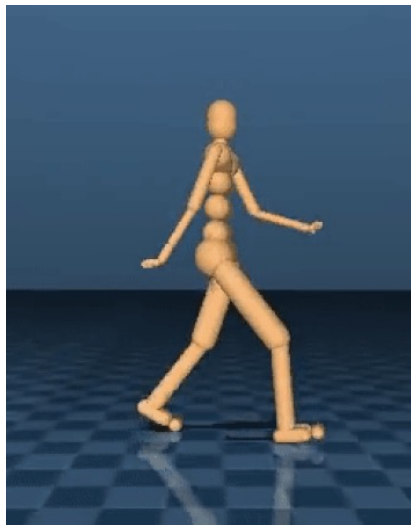
# Deep Reinforcement Learning

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



<https://github.com/yenchenlin/DeepLearningFlappyBird>

# Simulation environments to real world



Simulation environments:

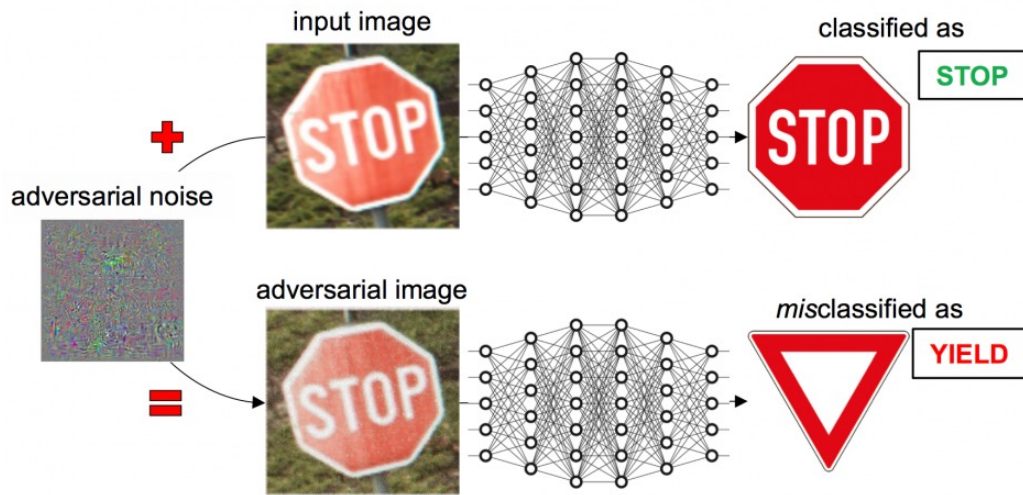
- OpenAI Gym
- MuJoCo



# Challenges in ML

# Adversarial Learning

- Can we mess with an ML model?

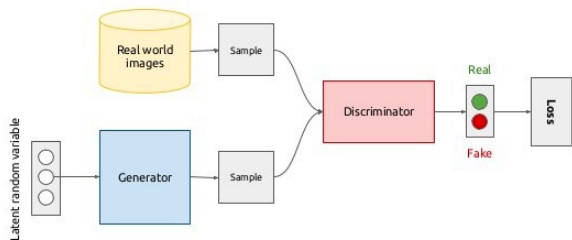


## Exercise: Digit recognition



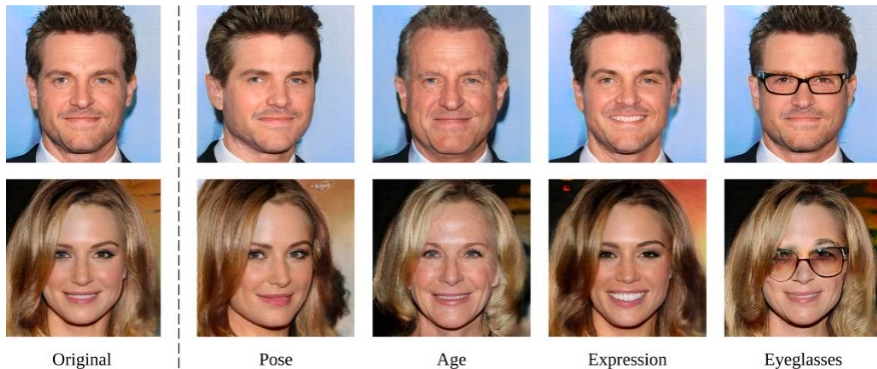
# Generative Models

## Generative adversarial networks (conceptual)



5

**Q:** Can we learn to generate fake content?



## Animating Faces

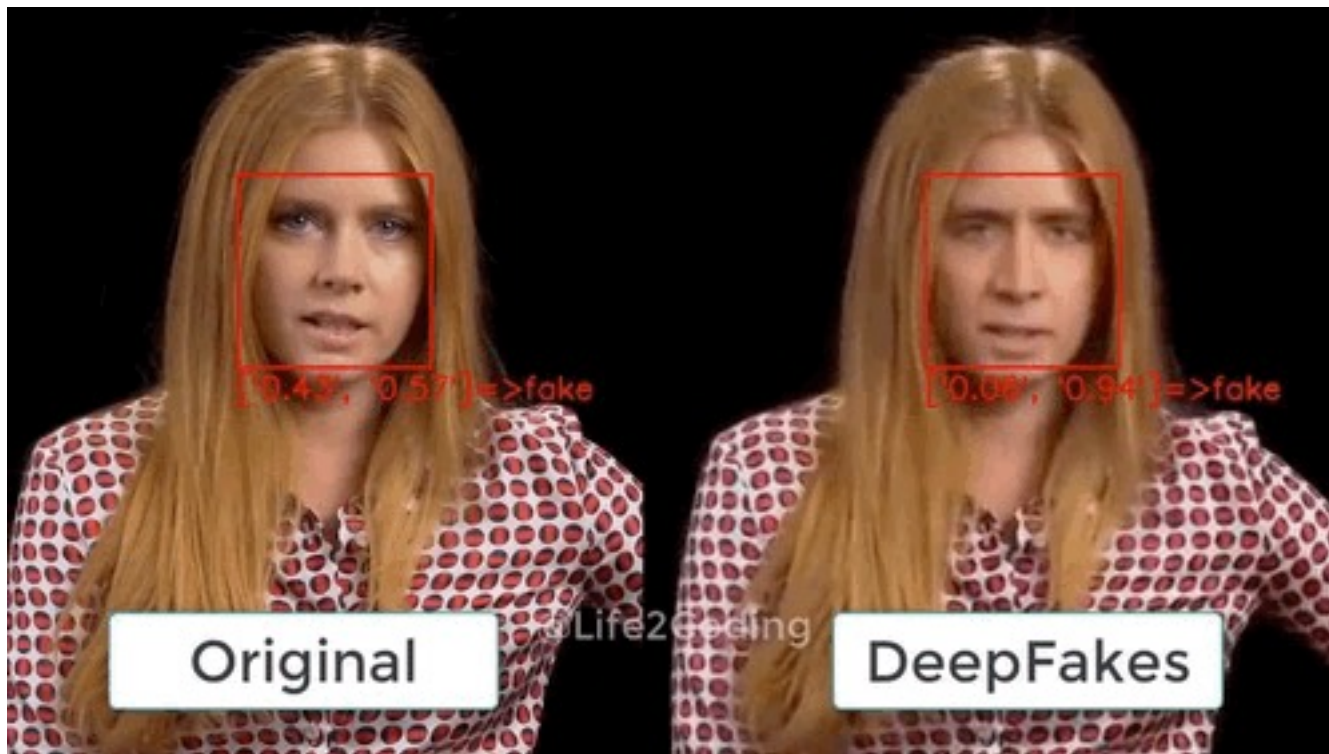
A single model animates all images given only a single source image



# Deepfakes

## All Competitions

All Competitions			All Categories ▾	Default Sort ▾
<a href="#">Active (Not Entered)</a> Completed InClass				
#DFDC	<b>Deepfake Detection Challenge</b> Identify videos with facial or voice manipulations Featured • 2 days to go • Code Competition • 2255 Teams	\$1,000,000		



# Algorithmic Bias

Turkish ▾



English ▾



O bir doktor.

He is a doctor.

O bir hemşire.

She is a nurse.

[Open in Google Translate](#)

[Feedback](#)

## 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**

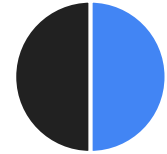


📷 Amazon's automated hiring tool was found to be inadequate after penalizing the résumés of female candidates. Photograph: Brian Snyder/Reuters

# Algorithmic Bias

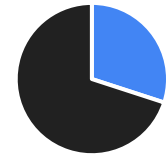
- Source of bias is often under-representation
- Female and Male population similar
- Example: Within high income earners, females might be under-represented
- ML algorithms might pick up such bias

Population



■ Female ■ Male

Population



■ Female ■ Male



# My research

On algorithmic bias: AutoML algorithms for fairness

AutoBalance: Optimized Loss Functions for Imbalanced Data

Label-Imbalanced and Group-Sensitive Classification under Overparameterization



Data & Compute  
Efficient ML

Learning in  
Heterogeneous  
Settings

Reinforcement  
Learning &  
Control

**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](mailto:oymak@ece) for research projects.

*Thanks for listening*