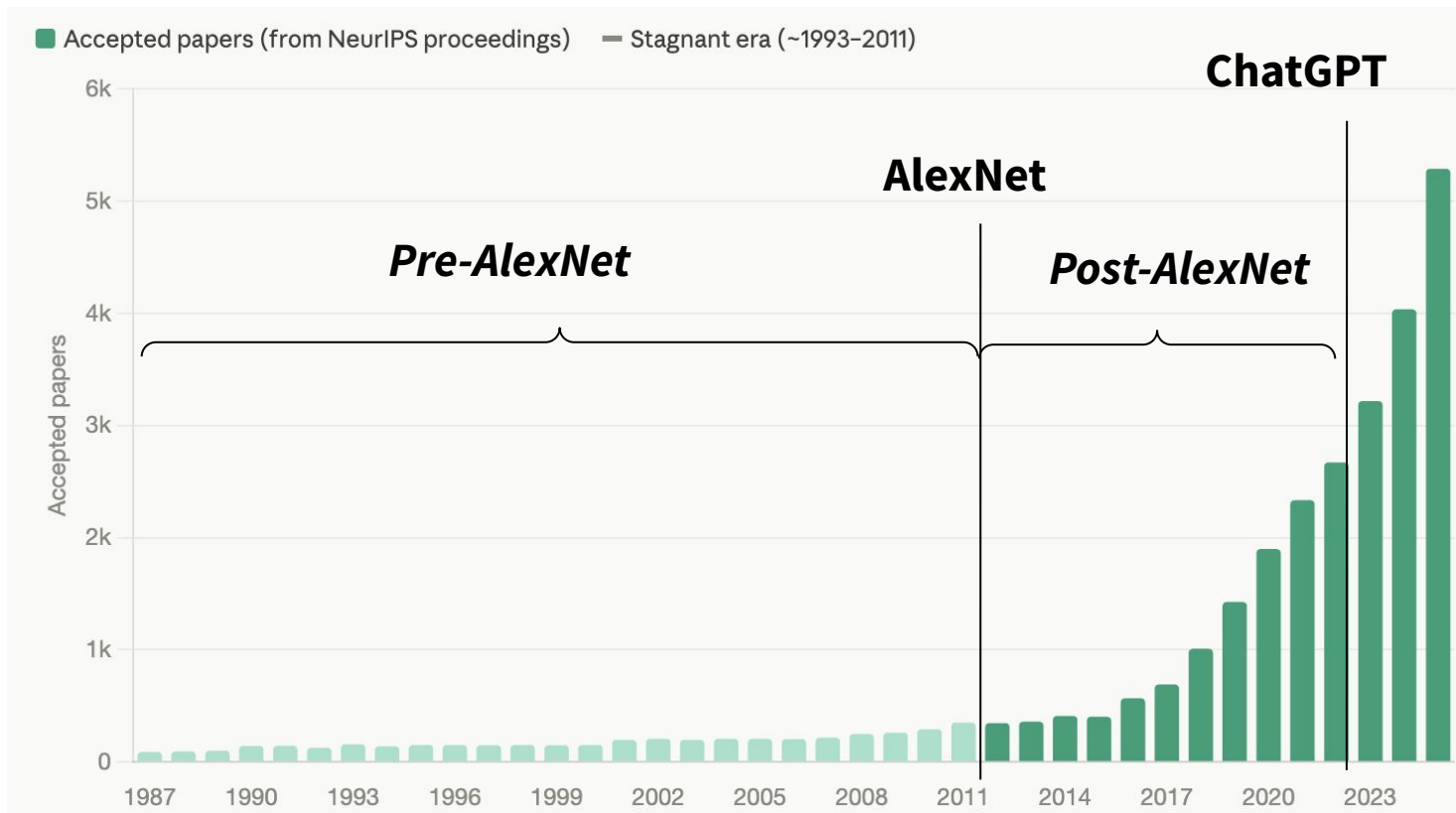


# Who's who of deep learning

CSE 493G1, Spring 2026

Materials prepared by Ben Newman

# Three phases



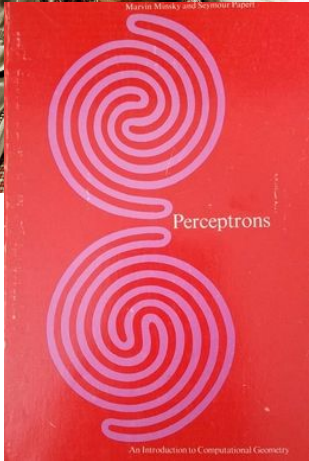
***In the beginning...***

# Marvin Minsky

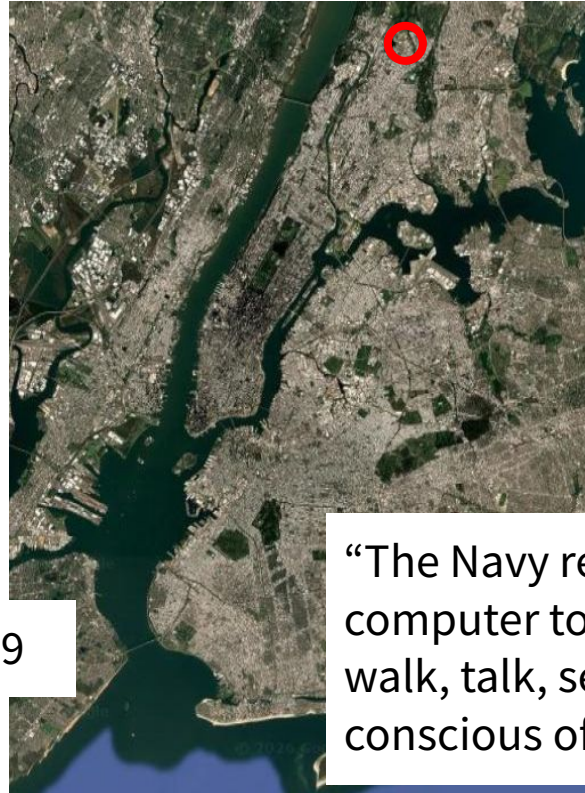
(1927-2016)



**"Symbolist"**

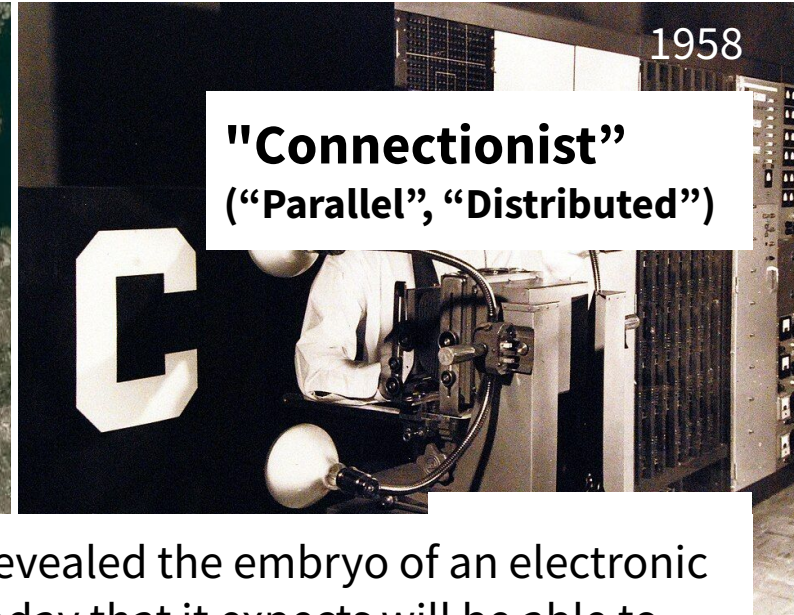


1969



# Frank Rosenblatt

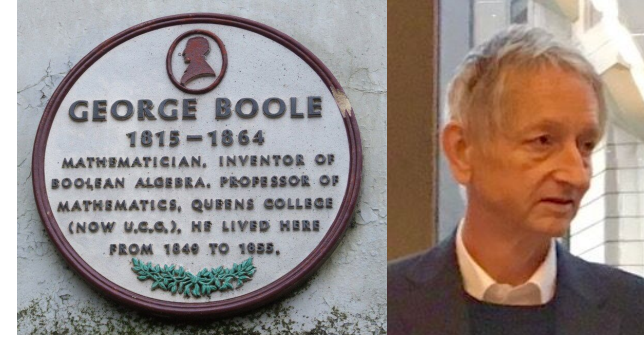
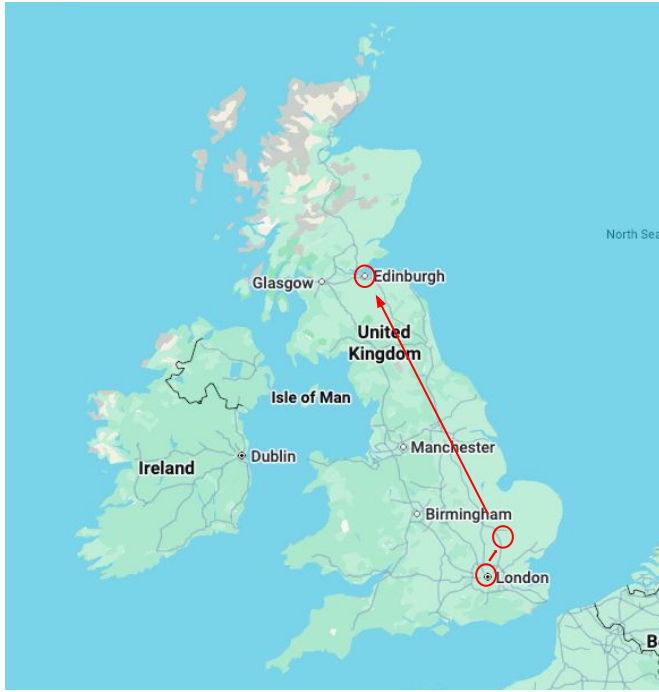
(1928-1971)



**"Connectionist"**  
(**"Parallel", "Distributed"**)

“The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.” - New York Times

# Geoffrey Hinton



1947-

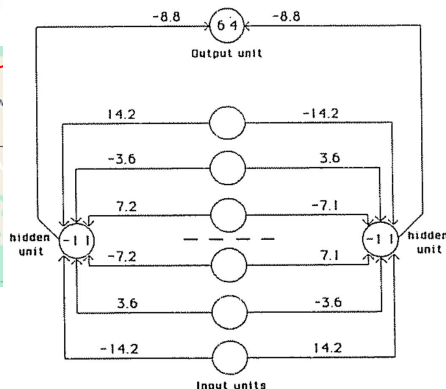
- Attended Cambridge in **Experimental Psychology**
- Super interested in how the brain works
- **Connectionist**

# Geoffrey Hinton

## Learning representations by back-propagating errors

1986

David E. Rumelhart\*, Geoffrey E. Hinton†  
& Ronald J. Williams\*



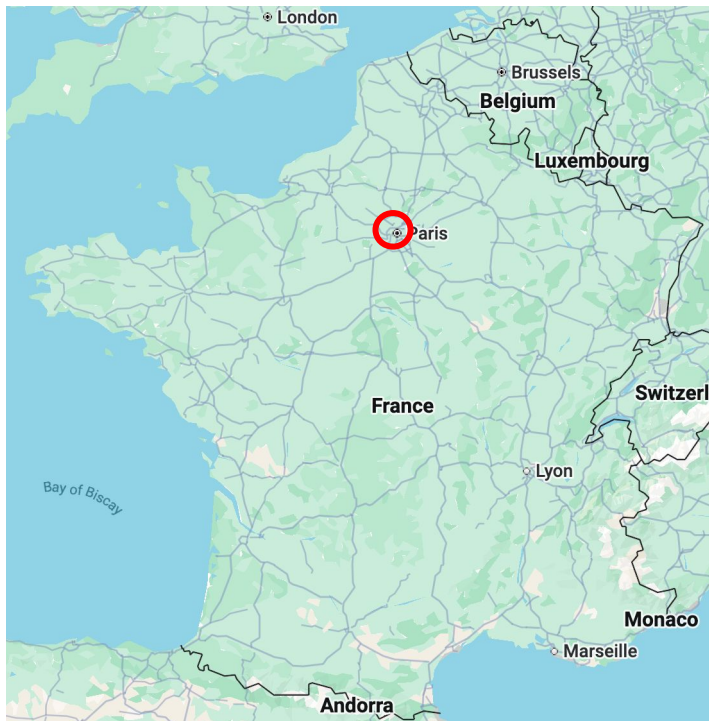
By Rolf Kickuth - Own work, CC BY-SA 4.0, [link](#)

Post-Doc with David Rumelhart in  
the PDP group at UCSD (1978)

Then prof at CMU

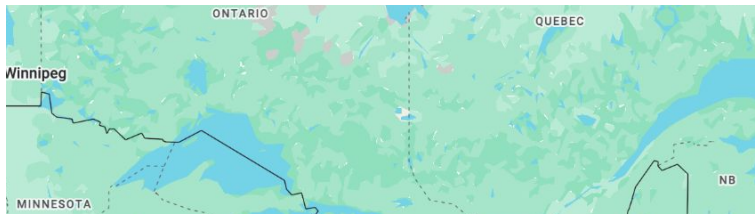
And then prof at U. Toronto

# Yann Lecun



- Born in France
- Undergrad in electrical engineering
- PhD from Sorbonne Université (1987)

# Yann Lecun



## Backpropagation Applied to Handwritten Zip Code Recognition

Y. LeCun  
B. Boser  
J. S. Denker  
D. Henderson  
R. E. Howard  
W. Hubbard  
L. D. Jackel

AT&T Bell Laboratories, Holmdel, NJ 07733 USA



1989

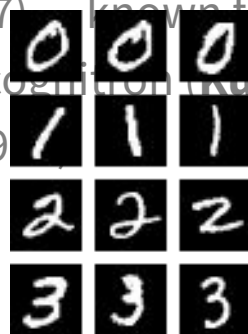
“LeNet”

PostDoc in Toronto w/ Hinton →

AT&T Bell labs (1987) known for CNNs

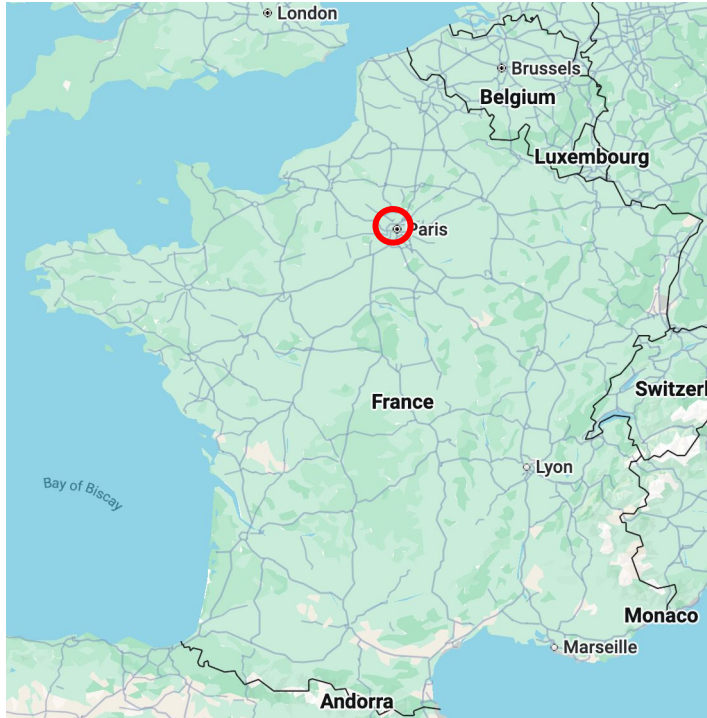
- based on Neocognitron **nihiko**

Fukushima, 1980



MNIST (1992)

# Yoshua Bengio



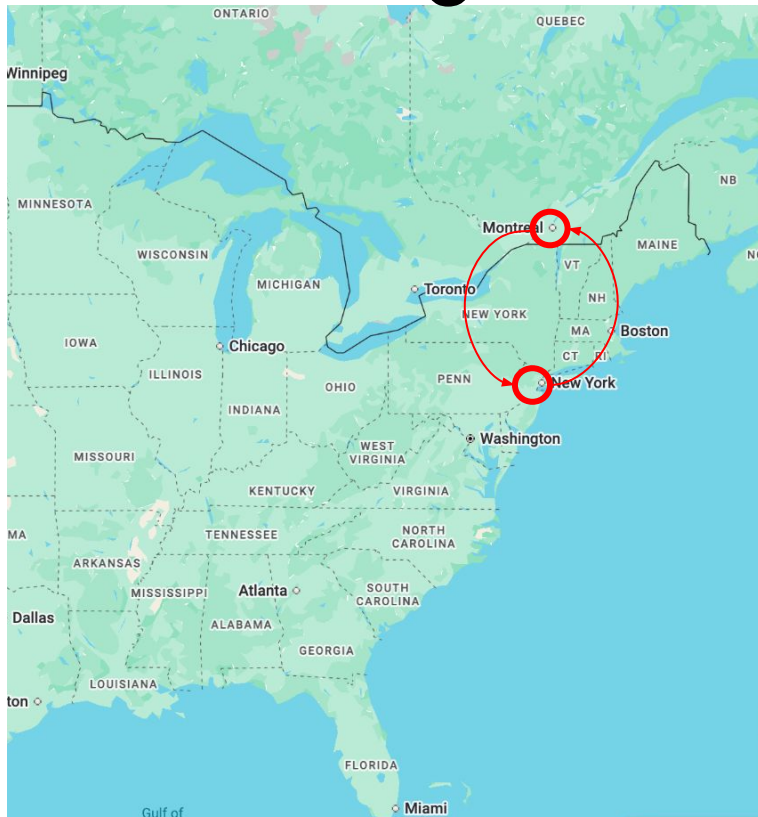
Born in France and moved to Quebec as a child

Brother is **Samy Bengio**



[Source](#) - [License](#)

# Yoshua Bengio



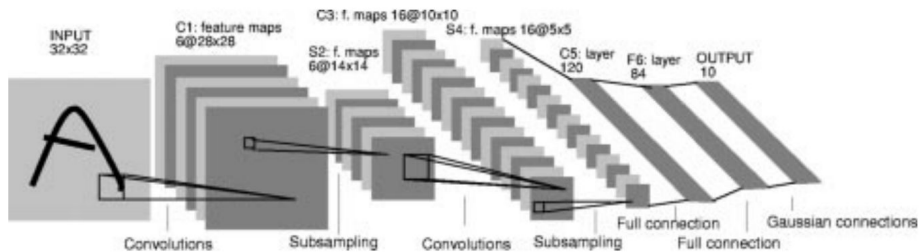
Born in France and moved to Quebec as a child  
Worked with Yann Lecun at Bell Labs in 1993 on using Conv Nets for OCR on checks

Prof at Université de Montréal and founded Montreal Institute for Learning Algorithms (MILA)

# Gradient-Based Learning Applied to Document Recognition



YANN LECUN, MEMBER, IEEE, IÉON BOTTOU, YOSHUA BENGIO, AND PATRICK HAFFNER



1998

“LeNet-5”

1998 LeCun and Bengio published a work on OCR

But, AT&T Bell Labs was completely dissolving at this point, so LeCun moved to NYU in 2003

# Jürgen Schmidhuber



Raised and PhD in Munich

Co-created the LSTM in 1997

**Long Short-Term Memory**

**1997**

**Sepp Hochreiter**

*Fakultät für Informatik, Technische Universität München, 80290 München, Germany*

**Jürgen Schmidhuber**

*IDSIA, Corso Elvezia 36, 6900 Lugano, Switzerland*

**But the late 90s/early 2000s was an AI winter...**

**So where was the interest?**

# The Internet!

Google

YAHOO!

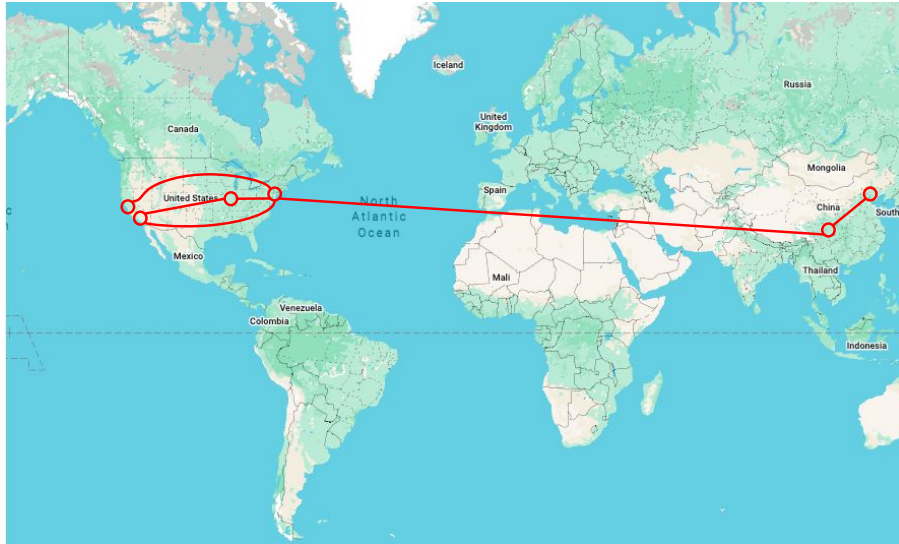


amazon.com  
EARTH'S BIGGEST BOOKSTORE

flickr

facebook

# Fei-Fei Li



- Born in Beijing, and grew up in Chengdu, before moving to New Jersey during high school and Princeton for undergrad
- PhD at CalTech and then professorships at Princeton, before moving to Stanford

# Fei-Fei Li

2009

**ImageNet: A Large-Scale Hierarchical Image Database**

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li and Li Fei-Fei  
Dept. of Computer Science, Princeton University, USA



Image Classification Datasets in 2007

Led by Jia Deng, her group collected:

Pascal VOC



20 object classes

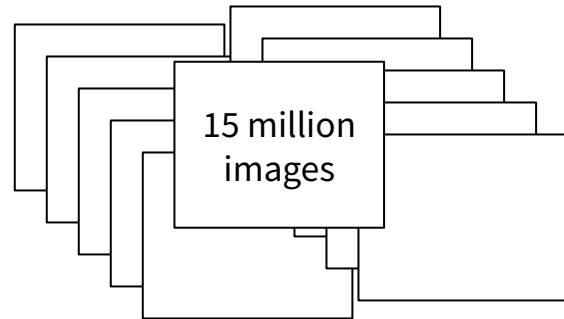
Find Images



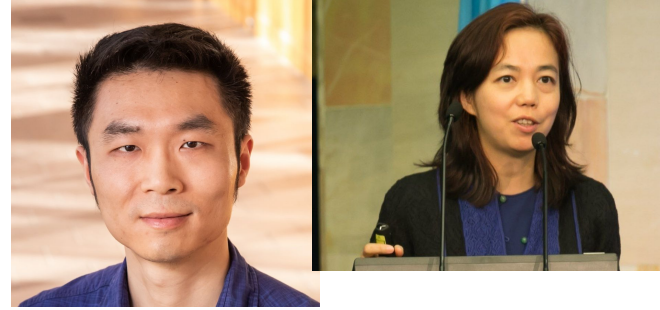
Label Them



IMAGENET



22k object classes



Some of Fei-Fei Li's other students to know about:

Andrej Karpathy (graduated 2015)



Timnit Gebru (graduated 2017)



Ranjay Krishna (graduated 2021 )





[Source](#)



[Source](#)



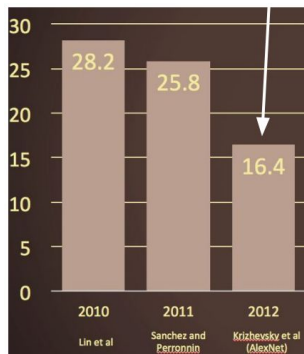
IMAGENET

14,197,122 images, 21841 synsets indexed

[Home](#) [Download](#) [Challenges](#) [About](#)

Not logged in. [Login](#) | [Signup](#)

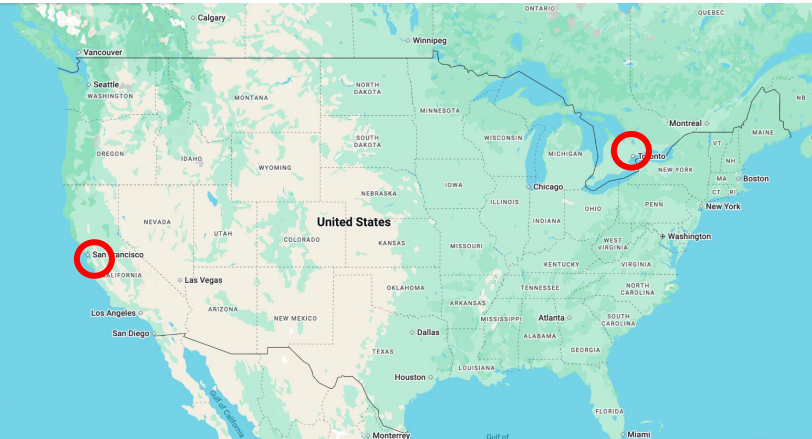
## ImageNet Large Scale Visual Recognition Challenge (ILSVRC)



# Jitendra Malik



How can I convince you that  
Neural Nets are the future?



Real tasks like object detection,  
Like Pascal VOC

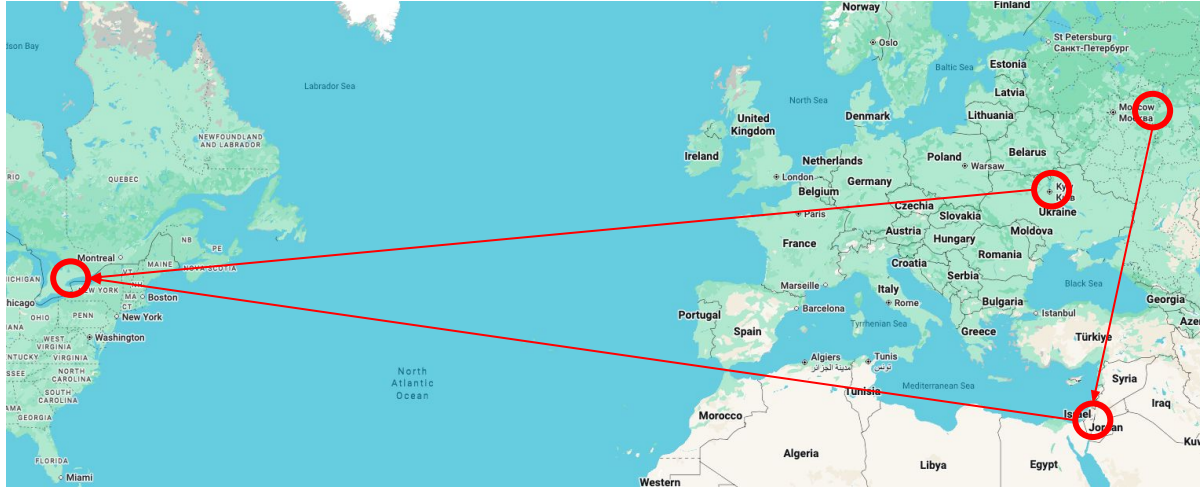
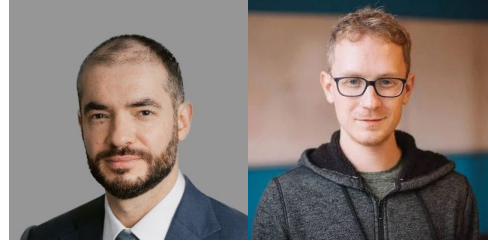


It's too small



Have you heard of ImageNet?

# Ilya Sutskever & Alex Krizhevsky



IMAGENET

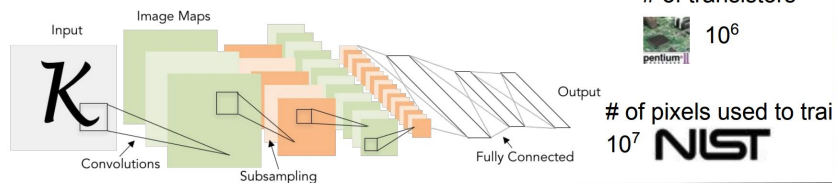
Ilya: Math prodigy; wanted to work on ImageNet

Alex: skilled at getting neural nets to work on GPUs

# Ilya Sutskever & Alex Krizhevsky



1998 LeCun et al.



2012 Krizhevsky et al.

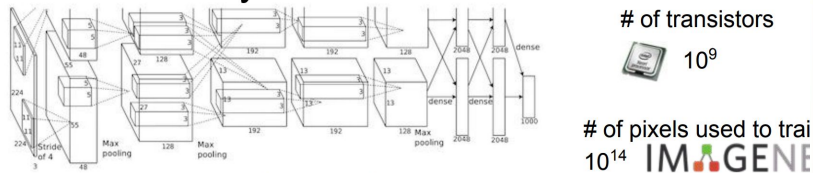
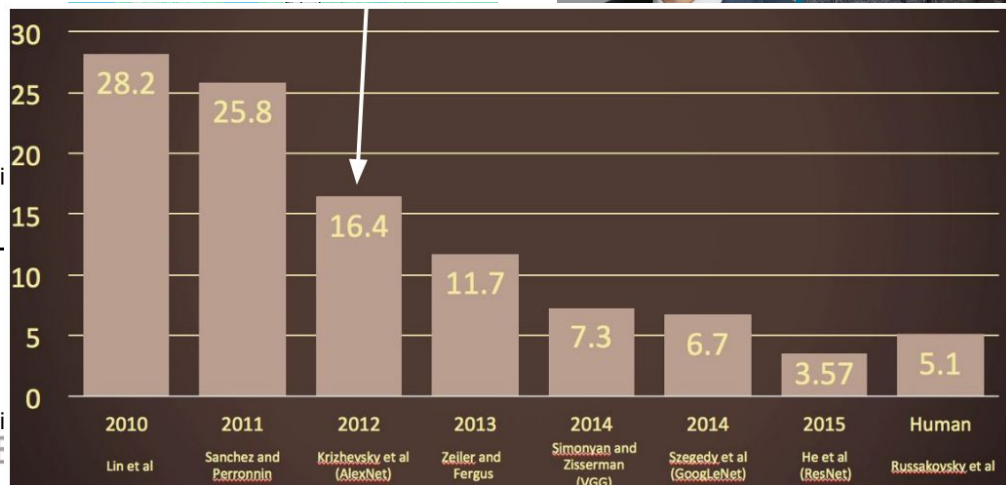


Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.



October 2012 — Alex Krizhevsky presented their work at the 3rd ImageNet competition

December 2012 — Geoff Hinton, Ilya Sutskever and Alex Krizhevsky formed a company that was sold to Google for \$44 million

# Reactions to these Improvements

1. Increased corporate interest in deep learning
2. Improvements to original architecture
3. More study of bias, fairness and ethics of AI systems
4. AGI and AI Safety
5. Expansion to other applications

# Andrew Ng & Jeff Dean - Start (Google Brain



## Jeff Dean *Facts*

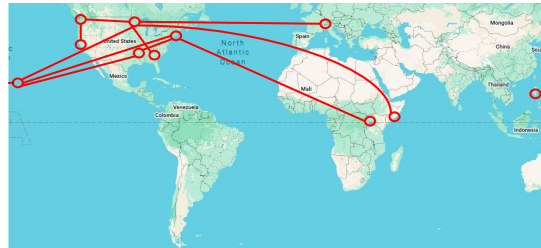
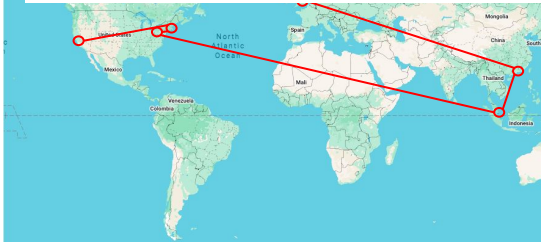


“Jeff Dean once failed a Turing test when he correctly identified the 203rd Fibonacci number in less than a second.”

“The speed of light in a vacuum used to be about 35 mph. Then Jeff Dean spent a weekend optimizing physics.”

“Jeff Dean's IDE doesn't do code analysis, it does code appreciation.”

[Source](#)



# Andrew Ng & Jeff Dean - Start (Google Brain



## Google Brain (2011)

“Jeff Dean once identified the 203rd Fibonacci number in less than a second.”

“The sp  
a week

Building High-level Features  
Using Large Scale Unsupervised Learning

Dean spent

“Jeff De

Quoc V. Le  
Marc'Aurelio Ranzato  
Rajat Monga  
Matthieu Devin  
Kai Chen  
Greg S. Corrado  
Jeff Dean  
Andrew Y. Ng



QUOCLE@CS.STANFORD.EDU  
RANZATO@GOOGLE.COM  
RAJATMONGA@GOOGLE.COM  
MDEVIN@GOOGLE.COM  
KAICHEN@GOOGLE.COM  
GCCRADO@GOOGLE.COM  
JEFF@GOOGLE.COM  
ANG@CS.STANFORD.EDU

[Source](#)



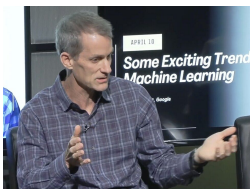
# 1. Corporate Era of Deep Learning Research

Google  
Brain

Baidu 百度

facebook

Microsoft



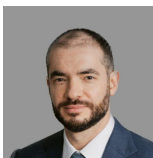
(2011)



(2014)



(2013)



(2012)

## 2. Improvements to AlexNet

(2014)

(2012)

Ross Girshick

**Rich feature hierarchies for accurate object detection and semantic segmentation**

Tech report (v5)

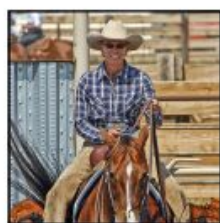
Ross Girshick Jeff Donahue Trevor Darrell Jitendra Malik  
UC Berkeley

Christian Szegedy

Kaiming He

Ian Goodfellow

### **R-CNN: *Regions with CNN features***

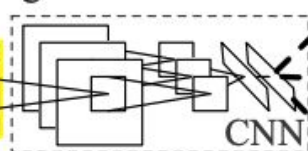
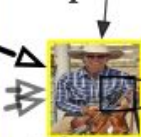


**1. Input image**



**2. Extract region proposals (~2k)**

warped region



CNN

**3. Compute CNN features**

aeroplane? no.

⋮

person? yes.

⋮

tvmonitor? no.

**4. Classify regions**

## 2. Improvements to AlexNet (2012)

“GoogLeNet”

Ross Girshick

Christian Szegedy

Google

Kaiming He

Ian Goodfellow

### Going deeper with convolutions (2014)

Christian Szegedy

Google Inc.

Wei Liu

University of North Carolina, Chapel Hill

Yangqing Jia

Google Inc.

Pierre Sermanet

Google Inc.

Scott Reed

University of Michigan

Dragomir Anguelov

Google Inc.

Dumitru Erhan

Google Inc.

Vincent Vanhoucke

Google Inc.

Andrew Rabinovich

Google Inc.

### Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift (2015)

Sergey Ioffe

Google Inc., [sioffe@google.com](mailto:sioffe@google.com)

Christian Szegedy

Google Inc., [szegedy@google.com](mailto:szegedy@google.com)

## 2. Improvements to AlexNet (2012)

“ResNet”

Ross Girshick

Christian Szegedy

Kaiming He

Microsoft → **facebook**

Ian Goodfellow

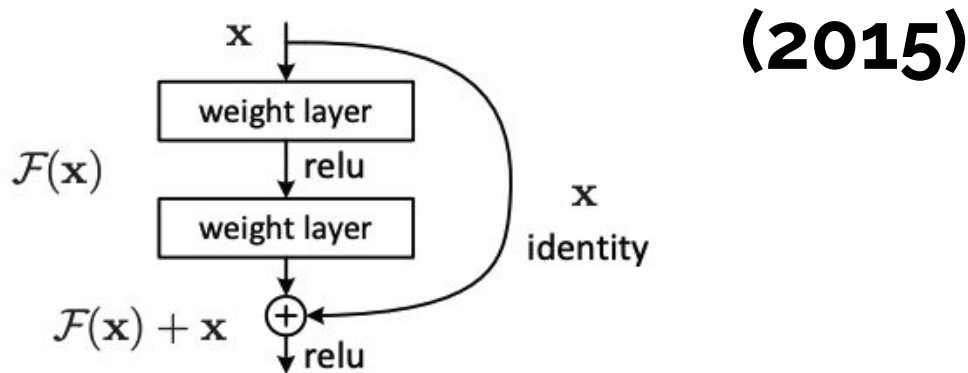
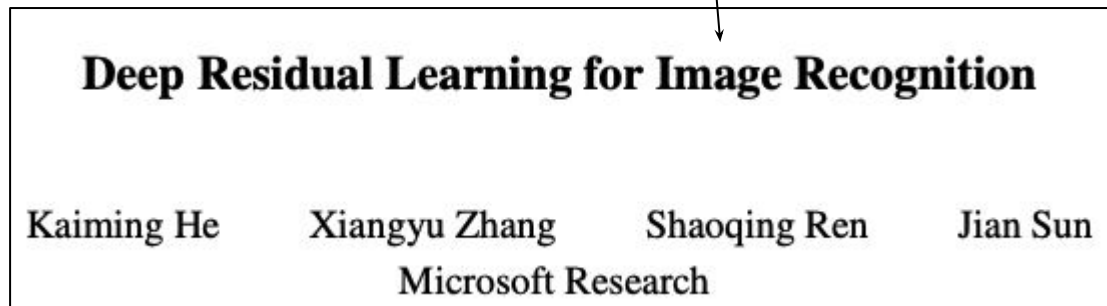


Figure 2. Residual learning: a building block.

## 2. Improvements to AlexNet (2012)

Ross Girshick

---

# Generative Adversarial Nets (2014)

---

Christian Szegedy

**Ian J. Goodfellow, Jean Pouget-Abadie\*, Mehdi Mirza, Bing Xu, David Warde-Farley,  
Sherjil Ozair†, Aaron Courville, Yoshua Bengio‡**  
Département d'informatique et de recherche opérationnelle  
Université de Montréal  
Montréal, QC H3C 3J7

Kaiming He

**Ian Goodfellow**

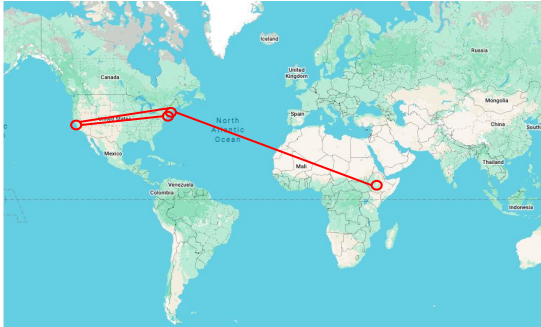




# 3. Fairness, Bias, and Ethics

Timnit Gebru

- Electrical Engineer from Stanford
- CS PhD advised by Fei-Fei Li
- Biased datasets → biased/harmful predictions
- Datasheets for Datasets and Model Cards



## Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification\*


**Joy Buolamwini**  
MIT Media Lab 75 Amherst St. Cambridge, MA 02139

JOYAB@MIT.EDU

**Timnit Gebru**  
Microsoft Research 641 Avenue of the Americas, New York, NY 10011

TIMNIT.GEBRU@MICROSOFT.COM

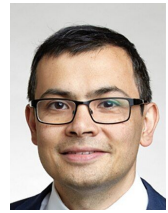
## Datasheets for Datasets

TIMNIT GEBRU, Black in AI  
 JAMIE MORGENSTERN, University of Washington  
 BRIANA VECCHIONE, Cornell University  
 JENNIFER WORTMAN VAUGHAN, Microsoft Research  
 HANNA WALLACH, Microsoft Research  
 HAL DAUMÉ III, Microsoft Research; University of Maryland  
 KATE CRAWFORD, Microsoft Research

## Model Cards for Model Reporting

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, Timnit Gebru

## 4. AGI & AI Safety



[Source](#) by John Sears - [License](#)



- Co-founded in 2010 by Demis Hassabis
- Played Atari Games (2013)
- Bought by Google in 2014
- Alpha Go (2016)
- Alpha Fold (2020)



- Founded in 2015 as a non-profit
- Ilya Sutskever, Andrej Karpathy and others were founding scientists

# *Techniques filter from vision to NLP*

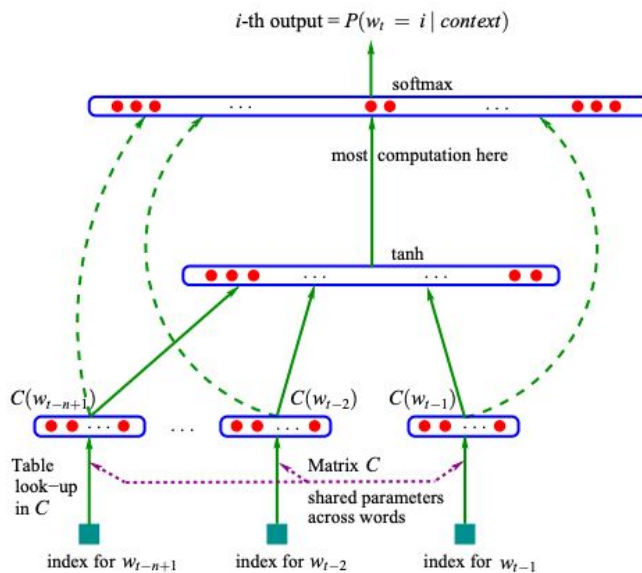
**How did NLP work at this point?**

# Yoshua Bengio



## A Neural Probabilistic Language Model 2003

Yoshua Bengio  
Réjean Ducharme  
Pascal Vincent  
Christian Jauvin



# How to generate word vectors?

Ronan Collobert & Jason Weston (2011)

## Natural Language Processing (Almost) from Scratch

**Ronan Collobert\***

**Jason Weston†**

**Léon Bottou‡**

**Michael Karlen**

**Koray Kavukcuoglu§**

**Pavel Kuksa¶**

RONAN@COLLOBERT.COM

JWESTON@GOOGLE.COM

LEON@BOTTOU.ORG

MICHAEL.KARLEN@GMAIL.COM

KORAY@CS.NYU.EDU

PKUKSA@CS.RUTGERS.EDU

# How to generate word vectors?

Ronan Collobert & Jason Weston (2011)

Tomas Mikolov & Jeff Dean (2012)

Jeffrey Pennington, Richard Socher and Cr

---

## Efficient Estimation of Word Representations in Vector Space

---

**Tomas Mikolov**

Google Inc., Mountain View, CA  
tmikolov@google.com

**Kai Chen**

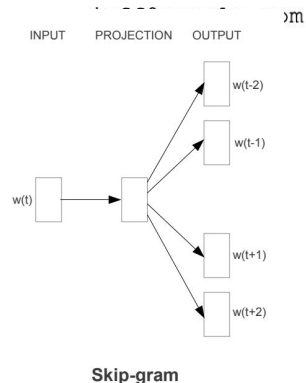
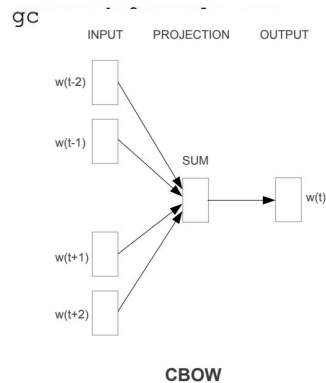
Google Inc., Mountain View, CA  
kaichen@google.com

**Greg Corrado**

Google Inc., Mountain View, CA

**Jeffrey Dean**

Google Inc., Mountain View, CA



# How to generate word vectors?

Ronan Collobert & Jason Weston (2011)

Tomas Mikolov & Jeff Dean (2012)



[Source](#) by Xuthoria [License](#)

**Jeffrey Pennington, Richard Socher and Chris Manning (2013)**

## **GloVe: Global Vectors for Word Representation**

**Jeffrey Pennington, Richard Socher, Christopher D. Manning**  
Computer Science Department, Stanford University, Stanford, CA 94305

# Word vectors → Sequences

Sequence-to-sequence models (2014)

---

## Sequence to Sequence Learning with Neural Networks

---

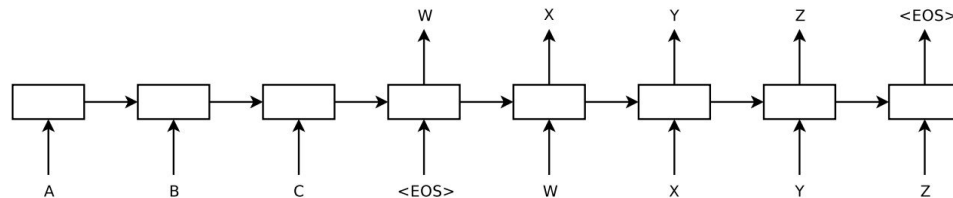
Attention (2015)

**Ilya Sutskever**  
Google  
ilyasu@google.com

**Oriol Vinyals**  
Google  
vinyals@google.com

**Quoc V. Le**  
Google  
qvl@google.com

Attention is All You Need (2017)



# Word vectors → Sequences

Sequence-to-sequence models (2014)

**Attention (2015)**

Attention is All You Need (2017)

NEURAL MACHINE TRANSLATION  
BY JOINTLY LEARNING TO ALIGN AND TRANSLATE

**Dzmitry Bahdanau**  
Jacobs University Bremen, Germany

**KyungHyun Cho**   **Yoshua Bengio\***  
Université de Montréal

# Word vectors → Sequences

Sequence-to-sequence models (2014)

Attention (2015)

**Attention is All You Need (2017)**

---

## Attention Is All You Need

---

<b>Ashish Vaswani*</b> Google Brain avaswani@google.com	<b>Noam Shazeer*</b> Google Brain noam@google.com	<b>Niki Parmar*</b> Google Research nikip@google.com	<b>Jakob Uszkoreit*</b> Google Research usz@google.com
<b>Llion Jones*</b> Google Research llion@google.com	<b>Aidan N. Gomez* †</b> University of Toronto aidan@cs.toronto.edu	<b>Łukasz Kaiser*</b> Google Brain lukaszkaizer@google.com	
<b>Illia Polosukhin* ‡</b> illia.polosukhin@gmail.com			

# Modes of Finetuning

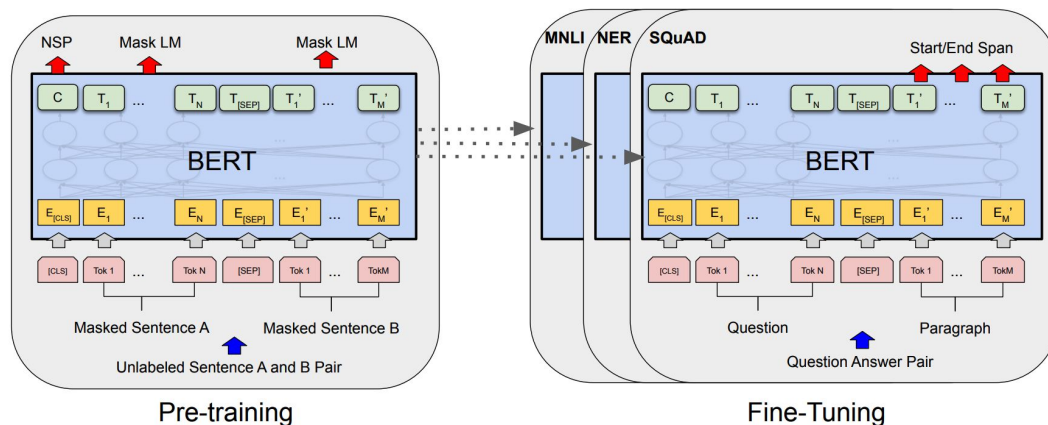
BERT - Kristina  
Toutanova

2018

**BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**

**Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova**  
Google AI Language

GPT, GPT-2 - Alec Radford



# Modes of Finetuning

BERT - Kristina Toutanova

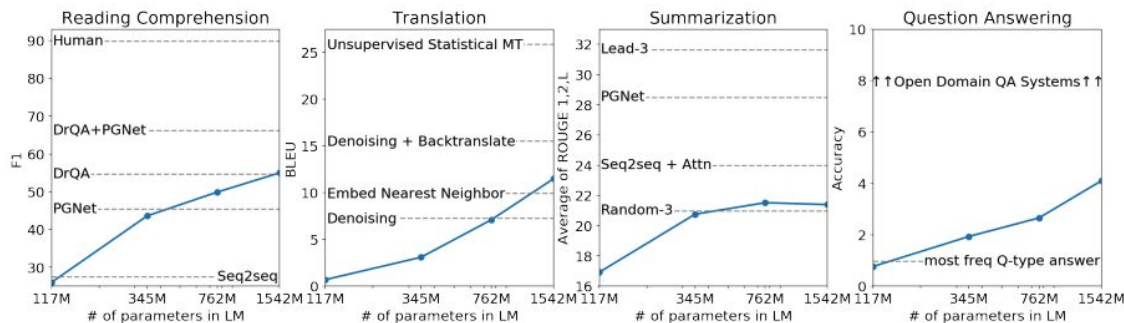
2019

## Language Models are Unsupervised Multitask Learners

Alec Radford<sup>\*1</sup> Jeffrey Wu<sup>\*1</sup> Rewon Child<sup>1</sup> David Luan<sup>1</sup> Dario Amodei<sup>\*\*1</sup> Ilya Sutskever<sup>\*\*1</sup>

## GPT-2 - Alec Radford

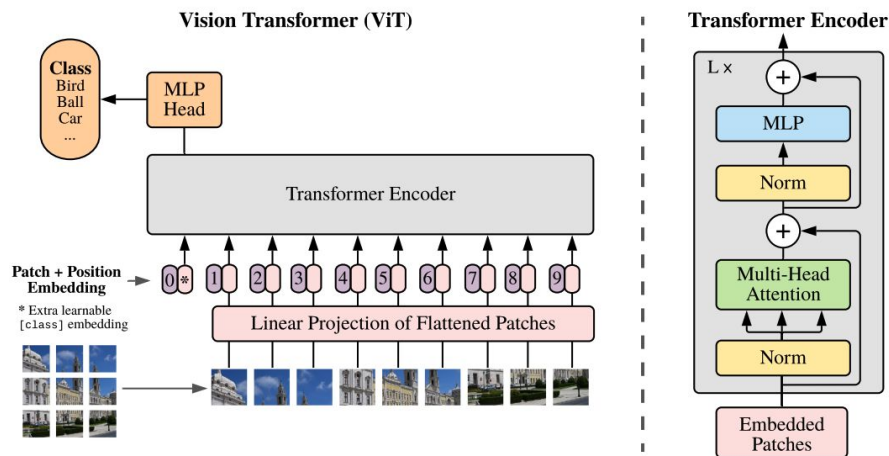
We test GPT-2's ability to perform summarization on the CNN and Daily Mail dataset (Nallapati et al., 2016). To induce summarization behavior we add the text `TL;DR:` after the article and generate 100 tokens with Top- $k$  random sam-



# Transformers for Computer Vision (2021)

## AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE

Alexey Dosovitskiy<sup>\*,†</sup>, Lucas Beyer<sup>\*</sup>, Alexander Kolesnikov<sup>\*</sup>, Dirk Weissenborn<sup>\*</sup>,  
Xiaohua Zhai<sup>\*</sup>, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer,  
Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby<sup>\*,†</sup>



# *ChatGPT*

December 2022: OpenAI released ChatGPT

April 2022: Deepmind and Google Brain merged into *Google Deepmind*

Google DeepMind



2020



2023

International AI Safety Report



Meta



2025

Advanced Machine Intelligence

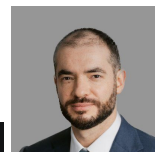
OpenAI

2021

ANTHROPIC

THINKING MACHINES

Safe Superintelligence Inc.



2024

## Turing Award (2018)



## Nobel Prize (2024)

Physics

Chemistry



***Thanks!***

***(And sorry to all those I left out!)***

# Sources

- *Genius Makers*, 2021 by Cade Metz
- *The Worlds I See* by Dr. Fei-Fei Li
- *Empire of AI* by Karen Hao
- [Perceptrons](#), 1969 by Marvin Minsky and Seymour Papert