# TTIC 31230, Fundamentals of Deep Learning

David McAllester, Autumn 2020

## A Timeline of Deep Learning

## and A Course Overview

# Early History

**1943**: McCullock and Pitts introduced the linear threshold "neuron". Words in red are important — we will discuss them in detail in this class.

**1962**: Rosenblatt applies a "Hebbian" learning rule. Novikoff proved the perceptron convergence theorem.

**1969**: Minsky and Papert publish the book *Perceptrons*.

The Perceptrons book greatly discourages work in artificial neural networks. Symbolic methods dominate AI research through the 1970s.

## 80s Renaissance

**1980**: Fukushima introduces the neocognitron — a form of convolutional neural Network or CNN. CNNs created the deep revolution in 2012.

**1984**: Valiant defines PAC learnability and stimulates learning theory. Wins the Turing Award in 2010.

**1985**: Hinton and Sejnowski introduce the Boltzman machine

**1986**: Rummelhart, Hinton and Williams demonstrate empirical success with backpropagation (itself dating back to 1961).

## 90s and 00s: Research In the Shadows

**1997**: Schmidhuber et al. introduce LSTMs (a form of recurrent neural network or RNN).

**1998**: LeCunn draws attention to convolutional neural networks (CNNs) (LeNet).

**2003**: Bengio introduces neural language modeling.

## Current Era: 2012-13

**2012**: Alexnet dominates the Imagenet computer vision challenge.

Google speech recognition converts to deep learning.

Both developments come out of Hinton's group in Toronto.

**2013**: Refinement of AlexNet continues to dramatically improve computer vision.

## 2014

**2014**: Neural machine translation appears (Seq2Seq models).

Variational auto-encoders (VAEs) appear.

Generative Adversarial Networks (GANs) appear.

Graph neural networks appear (GNNs) revolutionizing the prediction of molecular properties.

Dramatic improvement in computer vision and speech recognition continues.

## 2015-16

**2015**: Google converts to neural machine translation leading to dramatic improvements.

Batch Normalization appears improving the performance of image classification.

Residual Connections appear. This makes yet another dramatic improvement in computer vision.

Diffusion Models are formulated which become important in 2021.

**2016**: Reinforcement Learning is used to develop Alphago which defeats Lee Sedol.

## 2017

**2017**: AlphaZero learns both go and chess at super-human levels in a mater of hours entirely form self-play and advances computer go far beyond human abilities.

Unsupervised machine translation is demonstrated.

Progressive GANs demonstrate high resolution realistic face generation.

The Transformer appears greatly improving language modeling.

## 2018

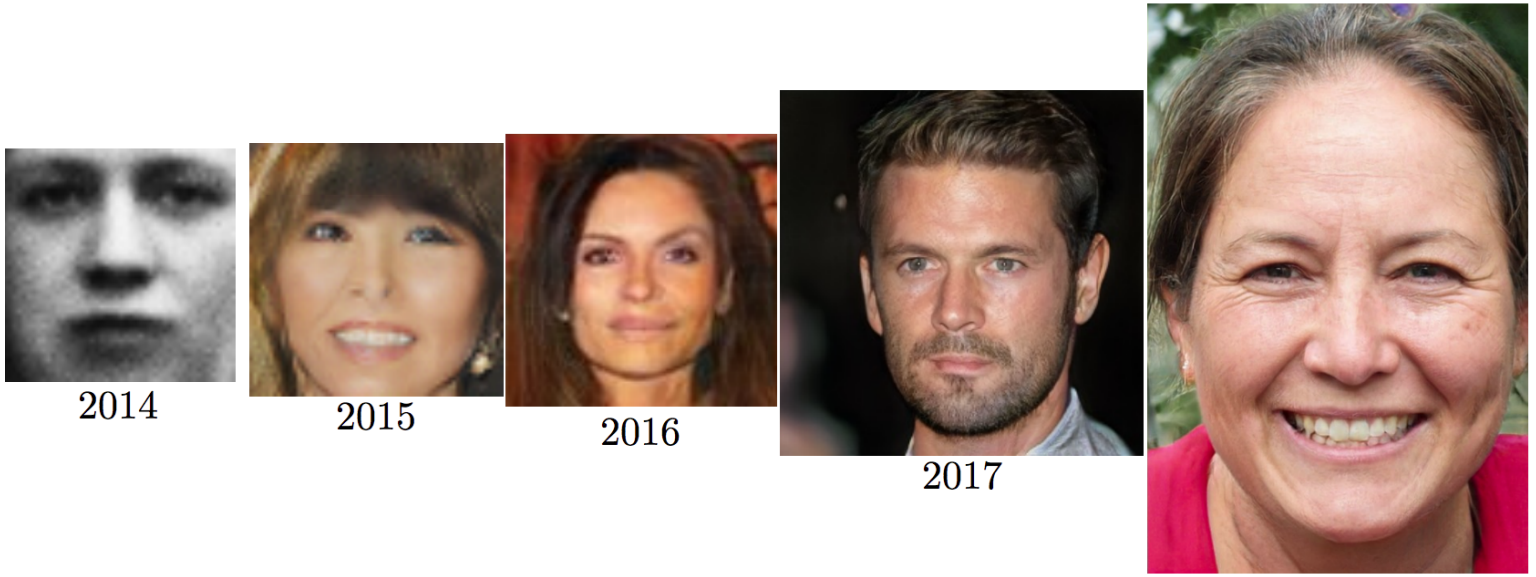
Unsupervised pre-training significantly improves a broad range of NLP tasks including question answering.

Contrastive learning is formulated which ultimately becomes the foundation of various systems.

AlphaFold revolutionizes protein structure prediction.

# Progress on GANS

## 4.5 years of progress on faces




(Goodfellow 2019)

ArXiv 1406.2661, 1511.06434, 1607.07536, 1710.10196, 1812.04948
Goodfellow, ICLR 2019 Invited Talk

# Progress on GANs


(Odena 2018)

# BigGANs, Brock et al., 2018

Figure 1: Class-conditional samples generated by our model.

# Variational Auto Encoders (VAEs, 2015)

[Alec Radford, 2015]

# 2019: Vector Quantized VAEs

VQ-VAE-2, Razavi et al. June, 2019

# VAEs in 2019

Figure 1: Class-conditional 256x256 image samples from a two-level model trained on ImageNet.

VQ-VAE-2, Razavi et al. June, 2019

# 2019: Natural Language Understanding

GLUE: General Language Understanding Evaluation

| Rank | Name | Model | URL | Score |
|---|---|---|---|---|
| 1 | T5 Team - Google | T5 | | 90.3 |
| 2 | ERNIE Team - Baidu | ERNIE | | 90.1 |
| 3 | Microsoft D365 AI & MSR AI & GATECH | MT-DNN-SMART | | 89.9 |
| 4 | 王玮 | ALICE v2 large ensemble (Alibaba DAMO NLP) | | 89.7 |
| 5 | Microsoft D365 AI & UMD | FreeLB-RoBERTa (ensemble) | | 88.4 |
| 6 | Junjie Yang | HIRE-RoBERTa | | 88.3 |
| 7 | Facebook AI | RoBERTa | | 88.1 |
| 8 | Microsoft D365 AI & MSR AI | MT-DNN-ensemble | | 87.6 |
| 9 | GLUE Human Baselines | GLUE Human Baselines | | 87.1 |

## June 2020: Wav2vec 2.0, Facebook

Trained on 53k hours of unlabeled audio (no text) they use contrastive learning to convert speech to a sequence of discrete quantized vectors they call "pseudo-text units".

By training on only one hour of human-transcribed audio, and using the Wav2vec transcription into pseudo-text, the outperform the previous state of the art in word error rate for 100 hours of human-transcribed text.

## February 2021: GLSM, Facebook

Generative Spoken Language Model (GSLM)

Using a form of VQ-VAE They then train a generative model of the sequences of pseudo-text units learned from unlabeled audio.

This model can continue speech from a speech prompt in much the same way that GPT-3 continues text from a text prompt.

Semantic and grammatical structure in a "unit language model" is recovered from speech alone.

## January 2021: CLIP, OpenAI

CLIP: Contrastive Language-Image Pre-training.

Trained on images and associated text (such as image captions or hypertext links to images) CLIP computes embeddings of text and embeddings of images ("co-embeddings") trained to capture the mutual information between the two.

This is done with contrastive learning.

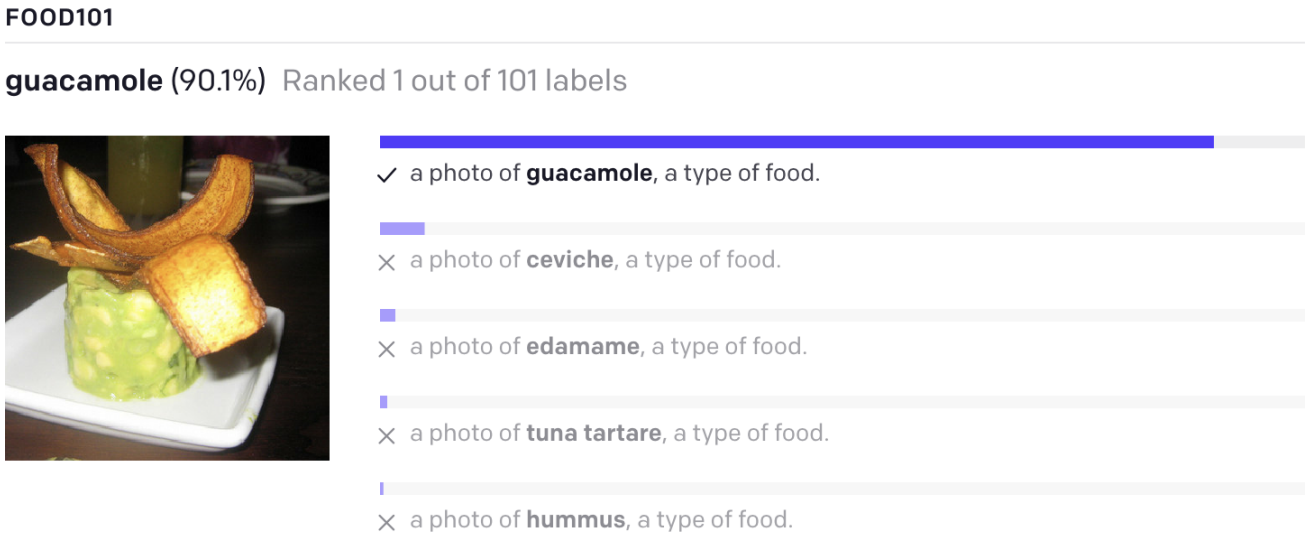
# CLIP

The model computes a probability of text given the co-embedding of the image.

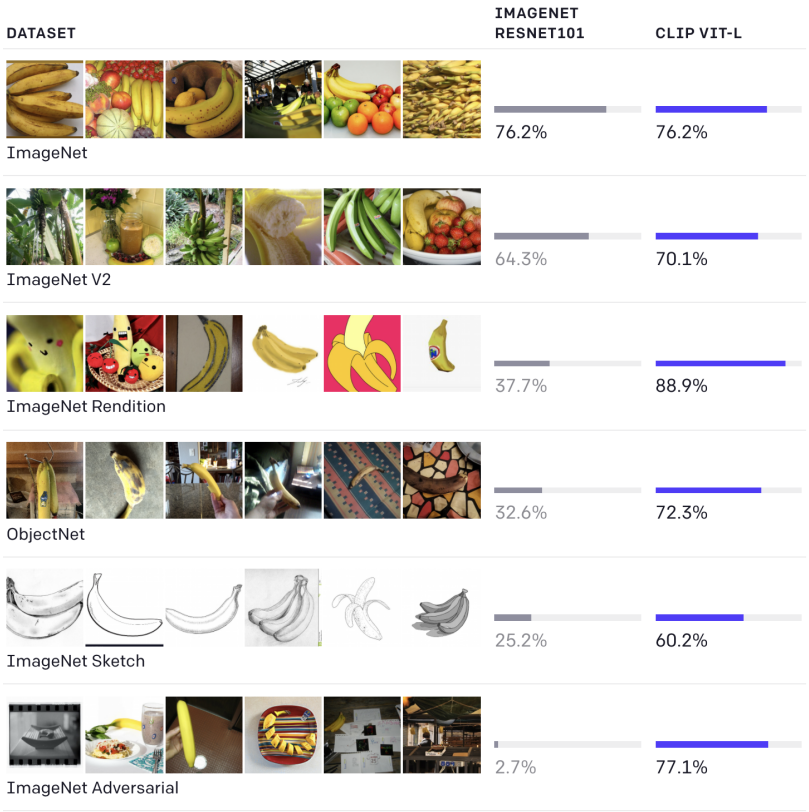
It is then used for zero-shot image classification on various datasets.

One can classify an image by comparing the probabilities that the model assigns to "prompts". There is a prompt for each class.

# Zero-Shot Image Classification



# Zero-Shot Image Classification




Although both models have the same accuracy on the ImageNet test set, CLIP's performance is much more representative of how it will fare on datasets that measure accuracy in different, non-ImageNet settings. For instance, ObjectNet checks a model's ability to recognize objects in many different poses and with many different backgrounds inside homes while ImageNet Rendition and ImageNet Sketch check a model's ability to recognize more abstract depictions of objects.

## January 2021: DALL·E, OpenAI

## April 2021: DALL·E-2

The name DALL·E is simply some kind of homage to the painter Dali and the Disney character WALL·E.

Both versions of DALL·E uses CLIP's co-embeddings of images and text.

Given text, DALL·E generates an image using a diffusion model.

# DALL·E-1 Zero-Shot Image Rendering from Language

TEXT PROMPT an illustration of a baby daikon radish in a tutu walking a dog

AI-GENERATED IMAGES

Edit prompt or view more images↓

TEXT PROMPT an armchair in the shape of an avocado. . . .

AI-GENERATED IMAGES

Edit prompt or view more images↓

# DALL·E-2

vibrant portrait painting of Salvador Dalí with a robotic half face

a shiba inu wearing a beret and black turtleneck

a close up of a handpalm with leaves growing from it

an espresso machine that makes coffee from human souls, artstation

panda mad scientist mixing sparkling chemicals, artstation

a corgi's head depicted as an explosion of a nebula

a dolphin in an astronaut suit on saturn, artstation

a propaganda poster depicting a cat dressed as french emperor napoleon holding a piece of cheese

a teddy bear on a skateboard in times square

## July 2021: Codex, OpenAI

Using an unsupersided pretrained language model they fine-tune on code, including comments, from public repositories.

Starting from an English prompt Codex continues with code — a form of automatic programming.

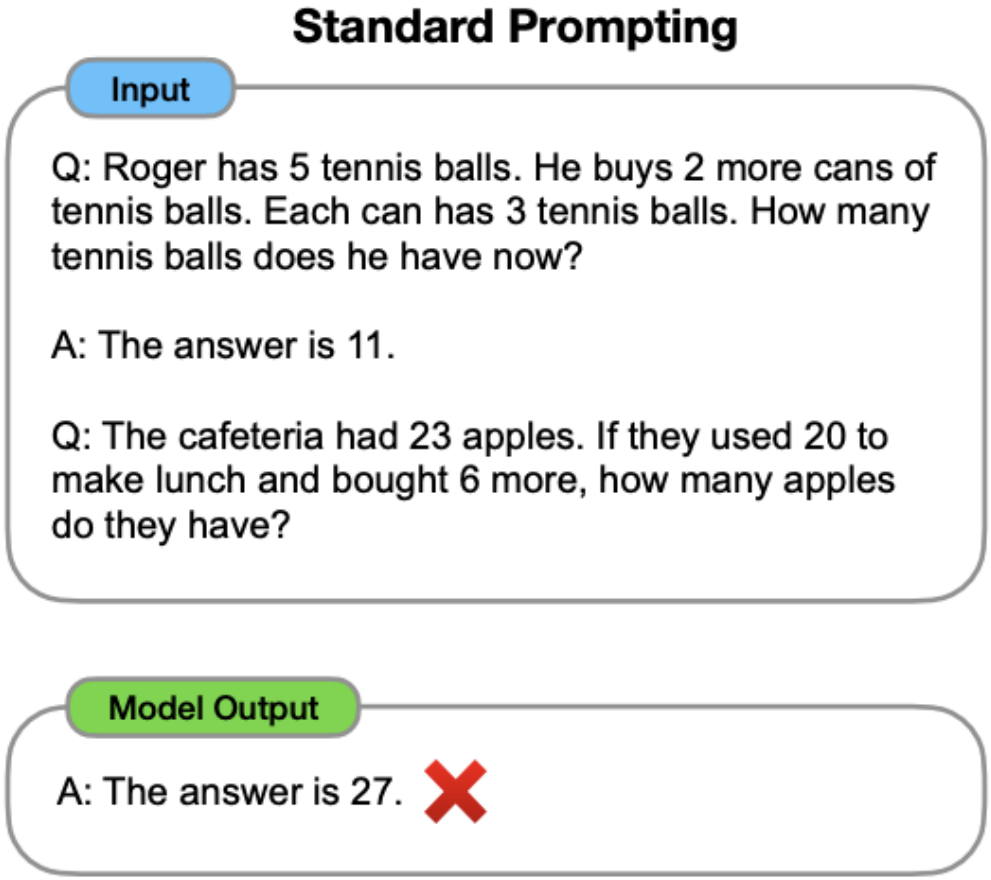
There is a published version (58 authors) and a production version that powers **GitHub Copilot**.

Copilot may supplant Stack Overflow for finding out how to do x in language y.

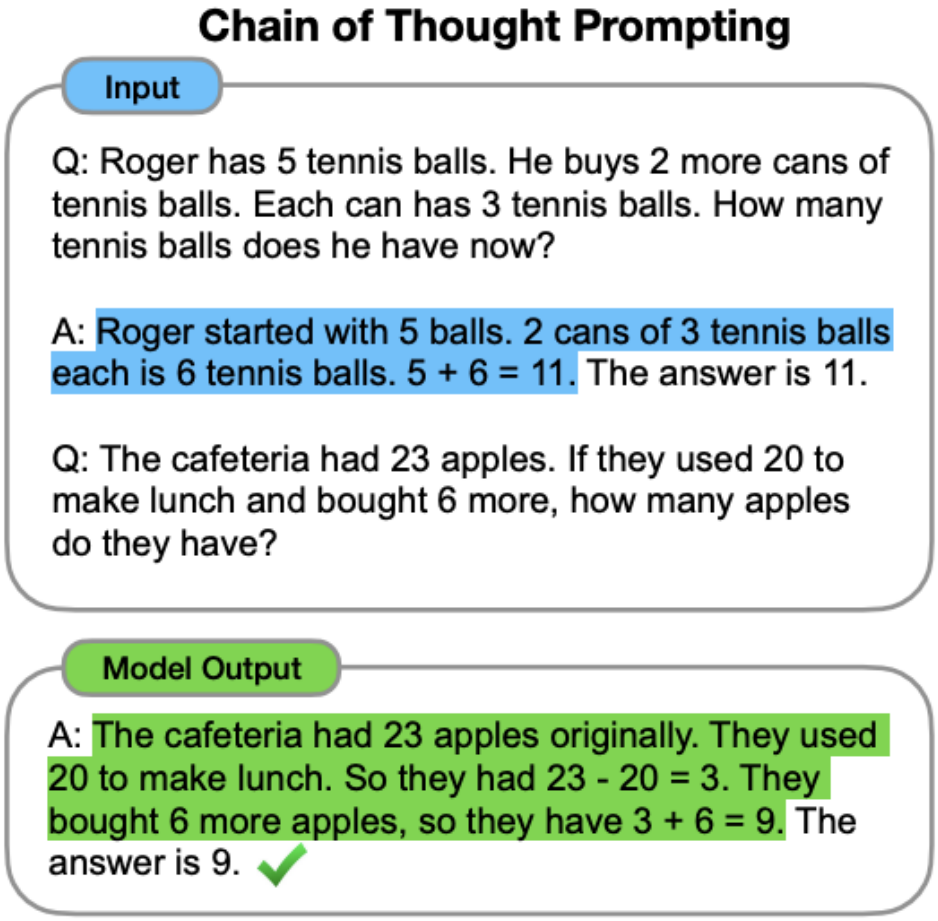
## January 2022: Chain of Thought Prompting

Give examples of "chains of thought" for few shot learning of reasoning steps.

# Naive Prompting



# Chain of Thought Prompting



## June 2022: Step by Step Prompting

It turns out that adding the simple instruction "take it step by step" elicits powerful chain of thought reasoning in GPT-3.

## September 2022: Humanoid Soccer, Deep Mind

Deep mind demonstrated two-on-two humanoid soccer in simulation (MuJoCo).

This a startling advance in the state of the art in humanoid control.

It also continues Deep Mind's effort in reinforcement learning.

## March 2023: GPT-4 is Unveiled

Dramatic improvement in conversational abilities and apparent understanding of common sense language meaning.

Training involves "instruction training" and "reinforcement learning with human feedback" (RLHF).

Most technical details are proprietary.

GPT-4 shows both a tendency to "hallucinate", presenting fiction as fact, and people find ways to "jailbreak" GPT4 into alternate and even malevolent personalities

Concerns over AI safety grows. Not everyone is concerned.

## September 2023

DALL·E-3 has just been announced.

ChatGPT continues to evolve. There is a new blog post from OpenAI today (Sept 26, 2023).

Significant advances in speech recognition and speech generation continue.

## Open Problems

**The MathZero Problem:** Since the development of AlphaZero people have been asking if the same thing could be done for the "game" of mathematics.

**The Autoformalization Problem:** This is the problem of automatically formalizing and verifying published proofs in mathematics possibly discovering legitimate gaps in the proofs where they exist.

**The Proof Assistant Problem:** This is the more modest goal of improving the level of automation in formal verification systems. The system most widely adopted by the mathematics community is LEAN based on dependent type theory.

## Application Advancements vs. Architecture Advancements

Advancements in the general principles of learning are having applications over very diverse applications.

When considering Moore's law of AI it seems worth distinguishing architectural advancements (new general learning methods) from new applications of established architectures or fine tuning of known methods.

This course will focus on general, architectural, ideas.

## Architectural Ideas

- Linear Threshold "neuron"
- Convolutional Neural Network or CNN
- Backpropagation
- Recurrent Neural Network or RNN
- Neural Language Modeling
- Variational Auto-Encoders (VAEs)
- Generative Adversarial Networks (GANs)
- Graph Neural Networks
- Normalization Layers
- Residual Connections
- Diffusion Models

- Reinforcement Learning
- The Transformer
- Unsupervised Pretraining
- Vector Quantization
- Contrastive Learning
- Prompting

# END