What is AI?

An introduction to AI for a non-technical audience.

What is AI?

Artificial intelligence is broadly defined as creating machines that can act like humans. A classic example is the Turing Test [TUR-1950]. Can a machine hold human-like conversations with another human? Other examples of AI:

- Identifying objects like pedestrians in an image
- Transcribing speech to text
- Driving a car
- Generating images and long-form text

How is AI different from traditional software?

Traditional software is created with just four basic building blocks:

- The ability to do math.
- The ability to remember information (e.g. save a file to your computer).
- The ability to send electrical signals to physical devices like a monitor or receive them from a keyboard. These are called inputs and outputs.
- The ability to describe which mathematical, memory, or output operations to do in the form of simple if-then-else statements. The computer science term for this is control flow.

This is great for things like adding up a column of numbers in Excel or typing a document in word and saving it. You expect Excel to be 100% accurate with its arithmetic, to do exactly what you ask of it, and to save its information perfectly.

But those basic instructions don't work very well for things like understanding images, driving a car, or generating long-form text. The early inventors of computer science in the 1940s recognized this and basically defined AI as anything that traditional software couldn't do.

How is AI implemented?

There are two competing approaches on how to implement AI. The first is symbolic systems. The symbolic systems approach believes that a system of simple, handcrafted symbols and rules could collectively become rich enough to produce AI behavior [MIN-1988]. For example, if a system is programmed with the parent child rules and then fed symbols that Alice is the parent of Bob, then the system can deduce that Bob is the child of Alice. This field draws broadly from linguistics and logic. The pinnacle of this approach is arguable IBM's Watson, which dazzled the public in 2011 by beating human players in the game show Jeopardy! Unfortunately, the symbolic systems approach

never delivered any commercially viable AI, and was basically eclipsed by the second approach, Machine Learning (ML).

The second approach is machine learning (ML). ML rejects the idea that symbols and rules should be crafted by humans, and instead believe that machines should learn automatically from large datasets.

ML to this day boils down to guessing and checking values against a sample of data called training data, and then hoping that these values generalize to the real world. This approach draws heavily on statistics, scientific computing, linear algebra, and calculus. In fact, at the heart of most types of ML is the simple linear regression.

How does machine learning relate to deep learning?

ML itself can be divided into two broad approaches: deep learning and everything else.

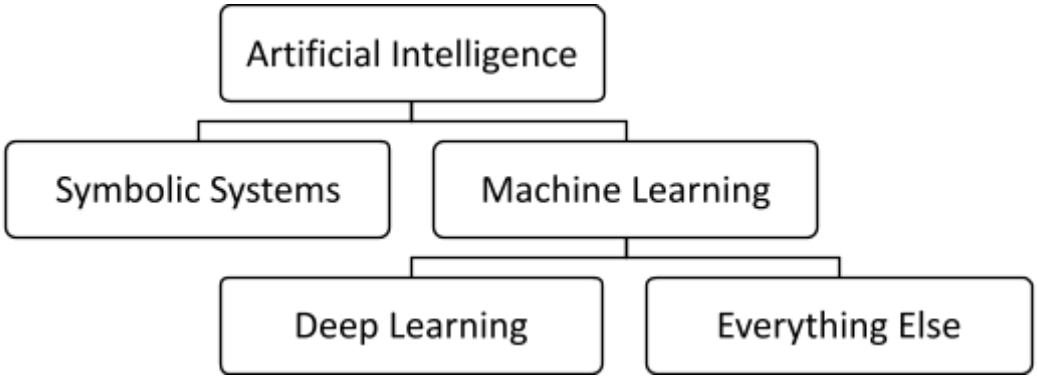


Figure 1: Hierarchy of AI.

Deep learning uses neural networks to analyze unstructured data. It is deep learning which has fueled the major advances in AI in the last decade.

TERMINOLOGY

Unstructured data refers to images, audio, and text files including emails and documents and social media posts. Individual data points, like a pixel value, are generally uninterpretable – it is the synthesis of hundreds or thousands of these datapoints that form the image of a cat or singer's unique voice. Humans can easily process this data. These are tasks that machines only recently started solving with deep learning.

Neural Networks are a method to take input data, either structured or unstructured, and output an estimated value (regression) or predict a class (classification). Although the term conjures up images of brains and neuron and synapses, neural networks are in

fact quite simple. For example, a model to predict the presence of diabetes could simply sum up fat intake and number of cigarettes smoked and apply a simple function (called an activation) so that as the sum gets higher, the prediction of diabetes increases towards 100%, but never exceeds it. This simple model is called logistic regression, or, a single neuron neural network. A second neuron could be placed alongside the first neuron to predict say heart disease. Stacking layers of these would be called a deep neural network, with the intermediate layers called hidden layers.

In the everything else category, we find many powerful tools that generally focus on structured data. Clustering methods like k-means clustering attempt to find natural groupings of objects. Decision trees are the ML approach of choice for financial data like credit scoring. Common commercial packages in this space include SAS, Databricks, Apache Spark, LightGBM, and XGBoost. Even linear regression and logistic regression can create powerful models for structured data.

TERMINOLOGY

Structured data is the kind of data we are used to working with in spreadsheets and databases: numerical values like age and categories like zipcode. Individual data points are easy to understand for humans and machines alike.

How did deep learning breakthrough?

The two main factors that allowed deep learning to breakthrough in 2012 were large volumes of data and the use of graphical processing unit (GPUs).

TERMINOLOGY

Graphical processing units were built to render computer graphics. A typical computer screen will have about 1-3 million pixels. Each pixel is rendered independently. GPUs are by design highly parallelized so that they can render these pixels in parallel. For example, in a single instruction, a GPU could add a vector of 1000 numbers to another vector of 1000 numbers. The equivalent operation on a CPU would require 1000 instructions, one for each pair of numbers.

It's no surprise that the consumer internet giants like Google and Microsoft are leaders in deep learning: they were the only companies that controlled sufficiently large volumes of data. In many cases they in fact pioneered the management of these large volumes of data via tools like Hadoop and Cassandra, and the acquisition of these large datasets via consumer applications like YouTube [Le-2011].

And the key insight that the intimidating sounding “neural network” is really just a series of linear algebra operations, just like computer graphics, led to the use of the tool of choice for computer graphics, GPUs. Nvidia released CUDA in 2006, which allows programmers to access the raw computing power of GPUs directly [NVI-2006]. In 2009, the first research applying GPUs to training neural networks was published [RAI-2009]. Indeed, Nvidia now produces GPU chips that don’t do graphics at all, but instead are focused on AI. Google has released similar chips branded Tensor Processing Units (TPUs). Tesla Motors produces its own GPUs as well.

TERMINOLOGY

Tensors are groups of numbers. A scalar is a type of tensor, namely, a single number, like 9. A vector (also a type of scalar) is a set of numbers arranged along a line, like

1, 2, 3

Since a line is a one dimensional object, a vector is called a rank-one tensor.

A matrix is a set of numbers arranged along a 2D grid, like

4 5
6 7

This could be useful for example for storing the pixel values of a black and white photo along the x and y axes. Since a grid is a two dimensional object, a matrix is called a rank-two tensor.

A tensor is a set of numbers arranged along a dot, line, grid, cube or hypercube. A color photo with different channels for red, green, and blue would need a tensor of rank three, corresponding to width, height, and color channel. A color video would require a tensor of rank four, corresponding to width, height, color channel, and frame. An autonomous vehicle with multiple cameras might use a tensor of rank 5, corresponding to width, height, color channel, frame, and camera.

Leading up to 2012, neural networks were out of the mainstream in the 1990s and early 2000s. A core of group of researchers sought funding from the Canadian Institute for Advanced Research (CIFAR) to continue research into the field and decided to rebrand the entire field “Deep Learning”. Those three researchers, Geoffrey Hinton, Yann LeCun, and Yoshia Bengio, are now household names in the field of ML.

TERMINOLOGY

Backpropagation is the method deep learning uses to inform and improve its guesses for its internal parameters. During training, an ML model makes a prediction. The prediction is compared to ground truth to calculate the error (which for historical reasons is called a loss). The error is backpropagated through the layers of the deep learning model to update all internal parameters of the model to become slightly more accurate. This can be visualized as a marble in a bowl. The coordinates of the marble are the parameters, and the perfect parameters are the lowest point in the bowl. As long as the marble follows the slope of the bowl via gravity, it will eventually find the lowest point in the bowl.

In 2012, Alex Krizhevsky finally put all the pieces together in a groundbreaking demonstration that marks the beginning of the commercial explosion of deep learning. Krizhevsky successfully trained a deep convolutional neural network called AlexNet to smash the previous record for accuracy for image classification [KRI-2012]. His company was acquired by Google soon thereafter.

TERMINOLOGY

Convolutional Neural Networks, also called **CNNs** and **ConvNets**, are neural networks optimized for analyzing images. CNNs are built on convolutions. Convolutions are shallow, mini-neural networks trained to match specific patterns, such as vertical edges or the color red. As with other forms of ML, these convolutions are learned via training on training data, and not handcrafted by humans.

Computer Vision (CV) refers to the task of analyzing image.

In recent years, **Transformers** have started to replace Convolutional Neural Networks when analyzing images. Transformers were originally used for language tasks, like translating languages. this is why Transformers are called transformers - they originally transformed a sentence from one language to another. But the same transformer architecture has been successfully applied to Computer Vision tasks.

Natural Language Processing (NLP) refers to the task analyzing writing.

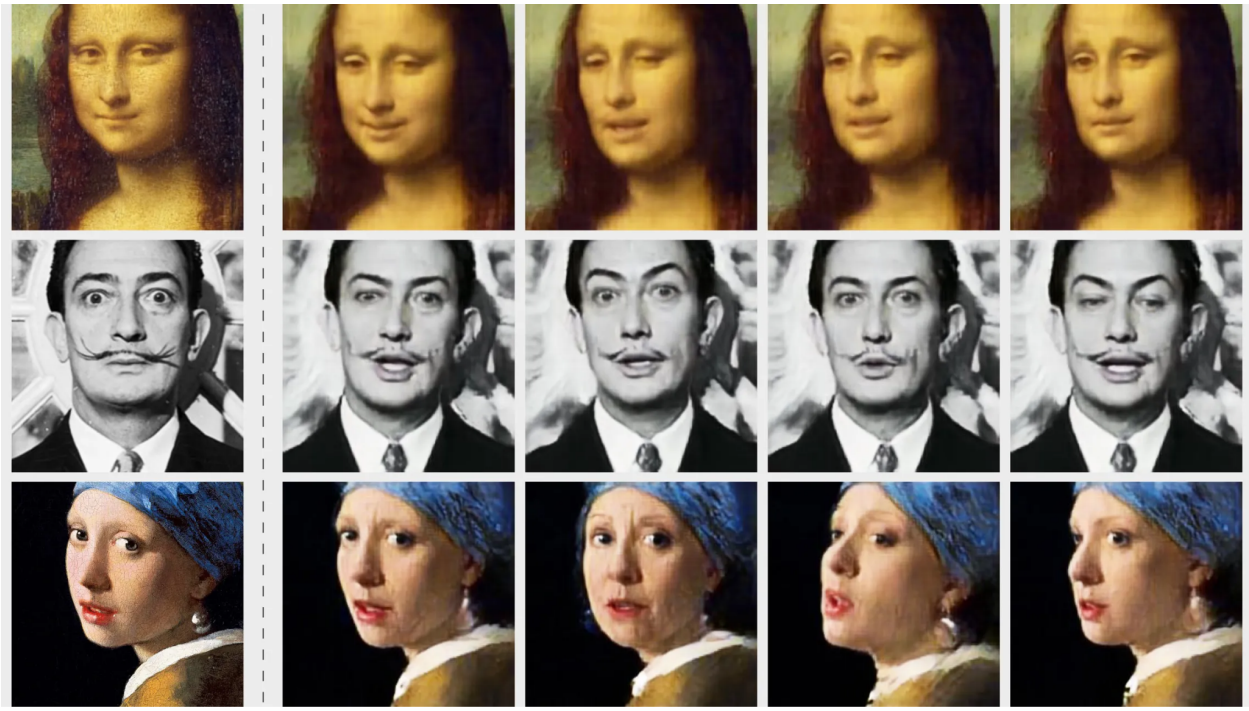
AlexNet was improved by a series of innovations, including VGG, Resnet by Microsoft, GoogLeNet (aka Inception) by Google, and others. Today, these innovations have been packaged as mature APIs that require zero understanding of the underlying AI, such as Google Cloud Vision API, AWS Rekognition, and Microsoft Azure Computer Vision.

Generative AI in Imagery

The domain Computer Vision traditionally focused on analysis: what objects are in an image, and where are those objects. From an image with one million pixels, a CV model might predict just 1 attribute, such as whether the image of a machine part is defective or not.

But if a CV model can make a prediction on a single attribute… could it be tasked to make a prediction on all one million pixels of a brand new, synthetically created image?

In 2013 and 2014, the first models to do just that were released, technically known as Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN) but mostly known by the product name Deep Fakes.



[Examples of Deep Fakes]

As of 2023, the state of the art in image to image generation still remains techniques based on VAEs and GANs. In fact, a VAE is at the heart of Stable Diffusion, a model that can generate images from a text prompt. And GANs may be at the heart of current techniques to de-age actors (e.g. Robert De Niro in The Irishman, Mark Hamill in The Mandalorian).

Natural Language Processing

Unlike CV, NLP has emphasized generative use cases more than analytical ones from the start. For example, one of the most basic NLP tasks is translation between languages, which generates a new sentence in a different language.

An example of an analytical NLP task is sentiment analysis, e.g., based on a movie review, would the reviewer give a thumbs up or thumbs down.

A major innovation in deep learning that cracked NLP problems was the move away from Recurrent Neural Networks towards Transformers. It's beyond the scope of this document to discuss the technical differences, but the key insight is that RNNs analyze one word at a time, whereas Transformers can look at all the words in a sentence at the same time.

Generative AI

The two AI techniques that have captured the public's imagination are

- ChatGPT by Open AI
- Image generation from a text prompt such as Dalle*E 2, stable Diffusion, and Midjourney (prompt2image).

ChatGPT is the latest model from a long line of NLP models OpenAI. It's built on top of GPT3 and more recently GPT4, which are the underlying NLP models. What ChatGPT did was to add a conversation interface, which lets a human user almost imagine that there might be another person on the other end of the conversation. The first set of use cases for ChatGPT are writing assistants, and ideas for writing heavy industries like legal, creative, medicine, and finance have been imagined.

prompt2image had its groundbreaking success with Dall*E 2, also by OpenAI, and MidJourney and Stable Diffusion were fast followers. Initial use cases for prompt2image have focused on creative applications, e.g. the realm of Photoshop. Missing functionality that is quickly being filled by new entrants include:

- Better generation of images of text
- faces (notice that astronauts do not have faces, making them an easier demo)
- complex 3D geometry
- video
- realistic imagery

[commonly used image to prompte Dall*E 2.]

References

[TUR-1950]: Computing Machinery and Intelligence
Alan Turing
https://link.springer.com/chapter/10.1007/978-1-4020-6710-5_3

[MIN-1988]: The Society of Mind
Marvin Minsky

http://aurellem.org/society-of-mind/

[LE-2011]: Building high-level features using large scale unsupervised learning
Quoc V. Le, Marc'Aurelio Ranzato, Rajat Monga, Matthieu Devin, Kai Chen, Greg S. Corrado, Jeff Dean, Andrew Y. Ng
https://arxiv.org/abs/1112.6209

[NVI-2006]: NVIDIA Unveils CUDA™-The GPU Computing Revolution Begins
https://www.nvidia.com/object/IO_37226.html

[RUM-1986]: Learning representation by back-propagating errors
David E. Rumelhart, Geoffrey E. Hinton, Ronald J. Williams
https://www.iro.umontreal.ca/~vincentp/ift3395/lectures/backprop_old.pdf

[HOC-1997]: Long Short-term Memory
Sepp Hochreiter, Jürgen Schmidhuber
https://www.researchgate.net/publication/13853244_Long_Short-term_Memory

[BEN-1998]: Object Recognition with Gradient Based Learning
Yann LeCun, Patrick Haffner, Leon Bottou, Yoshua Bengio
http://yann.lecun.com/exdb/publis/pdf/lecun-99.pdf

[RAI-2009]: Large-scale Deep Unsupervised Learning using Graphics Processors
Rajat Raina Anand Madhavan Andrew Y. Ng
http://robotics.stanford.edu/~ang/papers/icml09-LargeScaleUnsupervisedDeepLearningGPU.pdf

[GLO-2010]: Understanding the difficulty of training deep feedforward neural networks
Xavier Glorot, Yoshua Bengio
http://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf

[GLO-2011]: Deep sparse rectifier neural networks
Xavier Glorot, Antoine Bordes, Yoshua Bengio
http://proceedings.mlr.press/v15/glorot11a/glorot11a.pdf

[KRI-2012]: Imagenet classification with deep convolutional neural networks
A Krizhevsky, I Sutskever, GE Hinton
http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networ

[SCH-2015]: Critique of paper by "deep learning conspiracy"
Jürgen Schmidhuber
http://people.idsia.ch/~juergen/deep-learning-conspiracy.html

[BOL-2016]: Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings
Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, Adam Kalai
https://arxiv.org/abs/1607.06520

[TEN-2019]: Experimental security research of Telsa autopilot
Tencent Keen Security Lab
https://keenlab.tencent.com/en/whitepapers/Experimental_Security_Research_of_Tesla_Autopilot.pdf

[SZA-2014]: Intriguing properties of neural networks
Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, Rob Fergus
https://arxiv.org/abs/1312.6199?context=cs

[BRE-2019]: Approximating CNNs with bag-of-local features models works surprisingly well on ImageNet
Wieland Brendel and Matthias Bethge
https://openreview.net/pdf?id=SkfMWhAqYQ

[BRO-2018]: Adversarial Patch
Tom B. Brown, Dandelion Mané, Aurko Roy, Martín Abadi, Justin Gilmer
https://arxiv.org/pdf/1712.09665.pdf

[CAR-2017]: Towards Evaluating the Robustness of Neural Networks
Nicholas Carlini, David Wagner
https://arxiv.org/pdf/1608.04644.pdf

[EYK-2018]: Robust Physical-World Attacks on Deep Learning Visual Classification
Kevin Eykholt, Ivan Evtimov, Earlence Fernandes, Bo Li, Amir Rahmati, Chaowei Xiao, Atul Prakash, Tadayoshi Kohno, Dawn Song
https://arxiv.org/pdf/1707.08945.pdf

[IJG-2014]: Generative Adversarial Networks
Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio
https://arxiv.org/abs/1406.2661

[LIN-2017]: Focal loss for dense object detection
Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollar
https://research.fb.com/publications/focal-loss-for-dense-object-detection/

[PAP-2016]: Transferability in Machine Learning: from Phenomena to Black-Box Attacks using Adversarial Samples
Nicolas Papernot, Patrick McDaniel, Ian Goodfellow

https://arxiv.org/abs/1605.07277

[KAN-2016]: Evasion and Hardening of Tree Ensemble Classifiers
Alex Kantchelian J. D. Tygar Anthony D. Joseph
http://proceedings.mlr.press/v48/kantchelian16.pdf

[CHE-2019]: Query-Efficient Hard-label Black-box Attack:An Optimization-based Approach
Minhao Cheng, Thong Le, Pin-Yu Chen, Jinfeng Yi, Huan Zhang, Cho-Jui Hsieh
https://arxiv.org/abs/1807.04457

[CHN-2019]: Robust Decision Trees Against Adversarial Examples
Hongge Chen, Huan Zhang, Duane Boning, Cho-Jui Hsieh
http://proceedings.mlr.press/v97/chen19m/chen19m.pdf