AI (including ChatGPT tech) Basics For Non-techies

By Shankar Saikia (cofounder, Lexim.ai)

ChatGPT led to a huge increase in media coverage of and interest in ***artificial intelligence (AI)***. While there are many sources of technical education in AI (which I will roughly translate as those that require knowledge of Python programming, linear algebra and statistics) what's in short supply is a basic overview for those with nontechnical backgrounds. This article provides an introduction to the core ideas and concepts in AI, with the objective of enabling a non-technical reader to better understand how AI works. AI has its own vocabulary, with words borrowed from general grammar but with meaning very specific to AI. Let's start with an explanation of the term AI itself. AI is the capability of computers to perform tasks that are normally associated with human capabilities, such as recognizing images, understanding natural language (i.e., words spoken or written in a form that people can comprehend, such as in English), responding in natural language etc.

Imagine that you are on vacation and you see a historical monument such as the one below:

[1]

You want to know the name of this structure and so on your mobile phone you open your favorite search application and do an image search. Within seconds the search software gives

[1]

you the name of that monument (which in this case is Deoksugung Palace in Seoul). The act of producing the result is referred to as ***inference*.** You can probably guess that an inference is the response to a request, whether it is a request to identify an image, or, in the case of ChatGPT, a request to write a poem !

A follow-on question would be: what produces the inference ? The answer is that there is a ***model*** (some may use the phrase **prediction model** or **trained model**) that produces the inference. There are different model types or architectures. Here is a fairly simplistic but chronologically correct list of model architectures:

Expert System: this type of model was introduced in the mid 60s and first commercially used in the early 80s. Expert systems are often referred to as being rule-based. The rules are relatively simple and the answers do not have any ambiguity (i.e., the answer does not have the concept of what statisticians refer to as variability or stochasticity). We say that the model is for situations where the answer is **deterministic** (i.e., the question produces the same exact result every time). For example: the answer to the question “what is the capital of the United States?” has only one answer, and hence can be answered by an expert system model. In the image recognition example above, what if the picture only showed the roof of the building, and what if the roof resembled more than one building? In this situation the answer has to be guessed and hence an expert system would not be the right type of prediction model.

Machine Learning: Machine Learning models are used where the answer (i.e., inference) involves some guesswork. This type of model is for situations where there is potential for variability in the answer. The model has to first learn from data in order to be able to produce answers. Machine learning models are used for tasks like classifying different types of data (e.g., grouping items by color) or predicting values (e.g., housing prices). Spam detection is the first commonly known application of a machine learning model. As you can imagine the definition of spam is fairly ambiguous: while some examples of spam are easy to identify, others may not be. Because of the guesswork involved, spam detection lends itself to machine learning models.

Deep Learning: Deep Learning models are a type of machine learning that breaks a problem (e.g., image recognition, or speech recognition) into smaller subtasks or steps, and solves each step sequentially. The result of one step (e.g. identifying whether the structure is a building or a bridge) becomes an input to the next task (e.g., determining location of the building). These models, when depicted in a diagram, resemble nerves connected in biological species and hence the term **neural network** is used to refer to deep learning models. In the image recognition example above a neural network known as a Convolutional Neural Network (CNN) was used to produce the result.

Transformers: The transformer is the newest type of AI model and was first introduced in 2017. ChatGPT is based on a transformer model. ChatGPT is the commercial name given to the model by the software vendor (i.e., OpenAI, the company), whereas transformer is the model type or architecture. The transformer model is useful for tasks that involve natural languages, such as being able to ask ChatGPT a question in a sentence or phrase. Another term used to

refer to the type of transformer model used by ChatGPT is **large language model** (**LLM**). A LLM is one that is trained on a **large volume of data** and has a **large number of parameters**. (I explain both these concepts (**data** and **parameter**) below). ChatGPT is an example of **Generative AI**. Generative AI is the ability of an AI program to create entirely new content (e.g., a letter to a customer, or a poem !). Yet another term related to ChatGPT is that the model is a **foundational** model. Foundational means that the model is trained to answer general questions, and can be **fine-tuned** (explained below) for more specific uses such as to answer healthcare-related questions.

At this point you are probably asking: how does the model produce the answer, or maybe: what has to happen to enable the model to produce the answer?. The answer is that the model has to be **trained** before it can produce answers. During training the model is fed data (e.g., images and information about the image such as location, approximate height, width etc). The model learns from the data (hence the term learning used in machine learning, deep learning etc.) and the final result of training is the ability to produce answers such as the name of an object in an image. In the image recognition example the information that you provide the model (by pointing the phone's camera at the object) is used by the model to produce the name of the image .

Another way I explain training is to consider a situation where a model has to predict the price of houses. (This paragraph touches on basic algebra (e.g., a simple equation with one variable), not linear algebra which involves matrices, vectors etc. :)). You are given the following **data**: the price of each house and the number of rooms in each house. Here, we refer to each type of data as a **feature**. For the sake of explaining the concept of training, imagine that the price of a house only depends on the number of rooms. The objective of the exercise is to enable the model to predict the price of a house when the model is given the number of rooms. The training process has to derive a relationship between the **target** (i.e, what you are trying to predict) and the input feature (i.e., data, which in this case is the number of rooms in each house). The data used for training includes the price of each house and the number of rooms. Let's say that the trained model yields the following relationship: **y = 3x + 6**, where y is the price and x is the number of rooms. In this example the number **3** is referred to as a **parameter** (sometimes also called a **weight**). After a model is trained one refers to the model as being **trained** or **pre-trained**.

Fine tuning is another term related to training and one that is especially relevant to LLMs**.** Fine tuning is the process of taking a pre-trained model and modifying it for more specific tasks. For example, taking the model used for ChatGPT and fine-tuning it for answering questions related to healthcare. In general pre-training a LLM is very expensive whereas fine-tuning is relatively less expensive, and hence the value of fine-tuning.

As mentioned above, during training data is fed into the model. If prior to training you have identified the target variable (i.e., what the model has to predict) then we say that the training is a case of **supervised learning.** The training data is supervised by the target value. A different case is **unsupervised learning** where the target variable is not specified prior to training (e.g., you have marbles in a jar, each with a different color. In this situation there is no specific target task.). In the image recognition example above, the objective of the model is to identify the

name of the image and so it is an example of supervised learning. In the context of ChatGPT training the model is an example of **un**supervised learning (i.e., the target variable is not identified prior to training).

The last term that I would like to refer to is **optimization (**or **model optimization). Model optimization** is a part of **model training.** There are different ways to train a model and each way has a set of steps. One way of training is an iterative process known as optimization.

Let's review the terminology* that we have learned so far:

Inference

Model

Model Type or Architecture

Machine Learning (ML)

Deep Learning (DL)

***Transformer**

***Large Language Model (LLM)**

***Foundational Model**

***Generative AI**

Training

Pre-training

Fine-tuning

Data

Feature

Parameter

Supervised Learning

Unsupervised Learning

Optimization

*** = associated with ChatGPT**

In the context of ChatGPT when you type a request the result is an example of AI **inference**. The **model,** a **large language model (LLM),** is based on the **transformer** architecture. **Machine learning** (ML) and **deep learning**(DL) are other types of **model types** or **architectures**. In order to produce results the model has to be **trained** or **pre-trained**. In the case of the model underlying ChatGPT, because there is no specific target objective prior to training, the training is an example of **unsupervised learning.** ChatGPT is an example of a **Generative AI** solution. The model underlying ChatGPT is a **foundational** model and is a large language model (LLM)**.** A LLM uses a large volume of **data** and has a large number of **parameters** (one version of the model had 175 billion parameters !) Question: what would have to happen to enable a foundational model to answer questions specific to a domain like healthcare ? If your answer includes the term **fine tuning** you are correct.

Can you explain the image recognition example at the beginning of this writeup in the context of the AI terms mentioned above ?

Now that you have an overview of AI concepts like training, inference, parameters, models etc., you are better prepared to understand AI terminology and to ask and answer questions such as:

- What data is the model trained on?
- How often is the model trained ?
- Is it a LLM ?
- What kinds of generative AI capabilities does it have ?
- Is it a case of supervised or unsupervised learning ?

These terms are a subset of a larger vocabulary related to AI and there are several nuances even within these concepts. For example, in addition to supervised and unsupervised learning, there are others in the continuum such as semi-supervised, self-supervised etc. Even within the context of ChatGPT there are related concepts such as different types of transformer models for different tasks like summarizing text, translating languages etc. Stay tuned for further write ups on explanations of additional AI concepts.