What Are Generative AI, Large Language Models, and Foundation Models?

Recently, CSET was asked the question: What exactly are the differences between generative artificial intelligence (AI), large language models, and foundation models? These three terms suddenly seem to be everywhere and are often used interchangeably. This post provides a brief overview of what each term means, where they overlap, and how they differ.

To understand where these terms came from, it's helpful to know how AI research and development has changed over the last five or so years. AI is a very broad field encompassing research into many different types of problems, from ad targeting to weather prediction, autonomous vehicles to photo tagging, chess playing to speech recognition. While the field of AI research as a whole has always included work on many different topics in parallel, the seeming center of gravity involving the most exciting progress has shifted over the years.

Speaking loosely, one could say that in the [early 2010s](#) there was notable progress in image classification and speech recognition; in the [mid-2010s](#) the focus shifted to reinforcement learning (especially for games such as Go and StarCraft); and in the [late 2010s](#) and [early 2020s](#) there has been a boom in language and image generation. This chronological breakdown is very approximate, and any researcher would tell you that work on all of these areas—and many more—has been ongoing throughout that period and long before. The point is not to draw crisp boundaries, but to explain that terms like generative AI, large language models, and foundation models have emerged as attempts to point to a cluster of research directions and AI systems that have become especially noteworthy in recent years.

Generative AI is a broad term that can be used for any AI system whose primary function is to generate content. This is in contrast to AI systems that perform other functions, such as classifying data (e.g., assigning labels to images), grouping data (e.g., identifying customer segments with similar purchasing behavior), or choosing actions (e.g., steering an autonomous vehicle).

Typical examples of generative AI systems include image generators (such as Midjourney or Stable Diffusion), large language models (such as GPT-4, PaLM, or Claude), code generation tools (such as Copilot), or audio generation tools (such as VALL-E or resemble.ai). (Disclosure: I

serve in an uncompensated capacity on the non-profit board of directors for OpenAI, the company behind GPT-4.)

Using the term "generative AI" emphasizes the content-creating function of these systems. It is a relatively intuitive term that covers a range of types of AI that have progressed rapidly in recent years.

Large language models (LLMs) are a type of AI system that works with language. In the same way that an aeronautical engineer might use software to model an airplane wing, a researcher creating an LLM aims to model language, i.e., to create a simplified—but useful—digital representation. The "large" part of the term describes the trend towards training language models with more parameters.[1] A key finding of the past several years of language model research has been that using [more data and computational power](#) to train models with more parameters consistently results in better performance. Accordingly, cutting-edge language models trained today might have thousands or even millions of times as many parameters as language models trained ten years ago, hence the description "large."

Typical examples of LLMs include OpenAI's GPT-4, Google's PaLM, and Meta's LLaMA. There is some ambiguity about whether to refer to specific products (such as OpenAI's ChatGPT or Google's Bard) as LLMs themselves, or to say that they are powered by underlying LLMs.

As a term, LLM is the most specific of the three discussed here and is often used by AI practitioners to refer to systems that work with language. Nonetheless, it is still a somewhat vague descriptor. It is not entirely clear what should and shouldn't count as a language model—does a model trained on programming code count? What about one that primarily works with language, but can also take images as inputs? There is also no established consensus on what size of model should count as "large."

Foundation model is a term [popularized](#) by an institute at Stanford University. It refers to AI systems with broad capabilities that can be adapted to a range of different, more specific purposes. In other words, the original model provides a base (hence "foundation") on which other things can be built. This is in contrast to many other AI systems, which are specifically trained and then used for a particular purpose.

---

[1] Parameters are the numbers inside an AI model that determine how an input (e.g. a chunk of prompt text) is converted into an output (e.g. the next word after the prompt). The process of "training" an AI model consists in using mathematical optimization techniques to tweak the model's parameter values over and over again until the model is very good at converting inputs to outputs.

Typical examples of foundation models include many of the same systems listed as LLMs above. To illustrate what it means to build something more specific on top of a broader base, consider ChatGPT. For the original ChatGPT, an LLM called GPT-3.5 served as the foundation model. Simplifying somewhat, OpenAI used some chat-specific data to create a tweaked version of GPT-3.5 that was specialized to perform well in a chatbot setting, then built that into ChatGPT.

At present, "foundation model" is often used roughly synonymously with "large language model" because language models are currently the clearest example of systems with broad capabilities that can be adapted for specific purposes. The relevant distinction between the terms is that "large language models" specifically refers to language-focused systems, while "foundation model" is attempting to stake out a broader function-based concept, which could stretch to accommodate new types of systems in the future.[2]

—

This post aims to clarify what each of these three terms mean, how they overlap, and how they differ. It is important to note that at a technical level there are no bright lines that separate these terms from each other, or from other types of AI. Each term is an attempt to gesture towards a cluster of systems of interest, not a watertight category. There is also no clear hierarchy between the terms: for instance, generative AI could include LLMs or foundation models when these are used for generative use cases, but not when used in other ways.

Given how the AI landscape is evolving, the terms we use to describe these systems are sure to shift further in the future.[3] As is often the case with AI, anyone using these terms in a context where the precise definition matters should take care to delineate precisely what they intend—and don't intend—the term to cover.

---

[2] For instance, DeepMind's Gato model is designed to be able to perform a wide range of tasks that are not limited to language, including controlling a robot arm and playing video games. Gato could therefore be considered a foundation model, but not an LLM.

[3] Consider, for instance, "general-purpose AI," a term that is often used in similar ways to the three above and that has been a hot topic in discussions of the European Union's AI Act. This term aims to capture the fact that many of the systems in question have a wide range of possible use cases, while avoiding the baggage associated with the longer-standing idea of artificial general intelligence. Like the terms described above, general-purpose AI does not have clear or precise boundaries.