

[BACK TO RESOURCES](#)[PREVIOUS](#) | [NEXT](#)

January 25, 2023



Pre-training? Multimodal? Grounding?

Here are 12 terms to know about generative AI

12 Terms you need to know about generative AI



There is no question that generative AI is the hottest tech term of 2023. In fact, it sometimes seems that generative AI, and the impressive possibilities it represents, is all that the tech and VC community is talking about.

Unfortunately, unless you have a background in data science, it can be hard to grasp some of the technical terms that are being discussed as part of this generative AI revolution.

Don't worry – SambaNova has you covered – we've been building advanced generative AI models and products for years, and have organized a list of 12 of the most common terms to have you sounding like a generative AI expert in no time.

1. Foundation Models: 'Foundation model' as a term was popularized by the Center for Research on Foundation Models (CRFM) at Stanford University in the now famous paper, ["On the Opportunities and Risks of Foundation Models"](#). What makes foundation models so impressive is their ability to solve dozens of different tasks with human level accuracy, all with just a single model. They also have 'emergent' properties, which means they can learn new information very easily, often with just a very small amount of data. The implications for enterprises are significant: they can use a single foundation model to

replace thousands of different traditional machine learning models (and all the work that goes into training and maintaining them). So how do foundation models achieve this impressive combination of accuracy and versatility? The answer lies in their scale: they are trained on massive amounts of data, often referred to as 'web-scale' or 'internet-scale' data.

2. Generative AI: AI models that are able to generate new content, such as new language or images, after being trained on very large amounts of data, often referred to as 'web-scale' or 'internet-scale' (*similar to foundation models in #1 above*). These models are also known for being able to interact with and understand users through natural language interfaces 'prompts'. (*see #12 below*)

3. Large Language Models (LLMs): LLMs are a type of foundation model specifically focused on language. Because of their scale, they are able to generate an understanding of the context of language, enabling impressive versatility as well as the ability to generate new language content. (*see #1 and #2 above*)

So, what's the difference between #1, #2, and #3? After reading #1, #2, and #3 above, many of you may be asking, "So, what's the difference between these three terms?" The short answer is that in the news and the marketing of different AI companies, you may see these three terms used interchangeably. At SambaNova, we have a slightly more nuanced (and perhaps technical) definition. We view foundation models as the primary technology that will drive not only the next era of AI, but one of the fastest industrial revolutions the world has ever seen. Large language models are one type of foundation model. There are other types of foundation models, including those based on vision or multimodal (*see #5 below*) applications. Generative AI, with the ability to generate entirely new content while also enabling users to interact with foundation models using natural language, is the transformative capability that is bringing these technologies into the mainstream discussion and awareness.

4. Generative Pre-trained Transformer (GPT): GPT is an open source large language model used by a number of different AI product companies. Different companies have developed different versions of GPT models by pre-training them on different datasets and applying different training and fine-tuning techniques. GPT models are known as much for their impressive accuracy at solving complex language tasks, as they are for their large size and training complexity.

5. Multimodal model: A model that is able to understand more than one 'modality' of data, such as images or text, and associate the relationship between them. For example,

a multimodal model could analyze an image and generate a caption to describe what is happening in the image. Alternatively, some multimodal models work by generating completely new image content based on natural language input or guidance from a user. In both these cases, the multimodal model is able to connect an understanding of natural language with an understanding of what is happening in an image or video.

6. Pre-training: Pre-training refers to the first step in training a foundation model, enabling it to develop a general understanding of the features and representations of the data. Foundation models typically use unsupervised learning for pre-training. In very simple terms, this means the model is being trained on a large amount of unlabeled data to develop a fundamental understanding of the connections and context of the data. Pre-training usually involves a large percentage of the total training time for foundation models, often 80% or more. Once foundation models have been pre-trained they are able to solve a wide variety of different tasks, although their accuracy in solving these tasks may vary without additional fine-tuning (*see #7 below*).

7. Fine-tuning: Fine-tuning builds on the initial pre-training stage by adapting a model to perform specific tasks or objectives using a small amount of labeled data. Fine-tuning can improve how a model performs specific tasks, such as content generation, summarization, and sentiment analysis, even on tasks which the model has not been previously trained on. One specific type of fine-tuning that is particularly important for generative AI is 'instruction tuning', where a model is 'taught' to follow instructions by showing it examples that demonstrate what a 'good response' looks like. Fine-tuning often involves much less time and effort compared to pre-training, but requires labeled data to improve the accuracy of different tasks.

8. Few Shot Learning: The ability to guide a model on how to perform a certain task by providing it with examples. These examples are often provided via a 'prompt' interface, and can be provided in natural language instead of needing a structured dataset. (*see #12 below*) The fact that users can quickly and easily adapt a pre-trained model using natural language is a key part of what makes generative AI so exciting and impressive.

9. Domain Adaptation: Adapting a model to a specific subject matter. This could refer to an industry (such as banking), a highly specialized scientific discipline, or even topics or subject matter specific to an individual company or organization. Adapting a model in this way enables it to learn special terminology or terms and to develop a more complete and nuanced understanding of the subject matter. Domain adaptation is

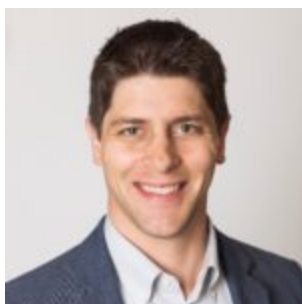
achieved by training a model using domain-specific data at either the pre-training or fine-tuning stage.

10. Model Parameters Count (“Parameters”): Model parameter count or “number of parameters” refers to the number of weights in all layers of a model. As a general rule, the more parameters a model has, the more capable and accurate it is assumed to be, although there are examples where [smaller models can achieve equal or even better accuracy than much larger models](#).

11. Grounding: A model’s ability to determine ‘factual information’ in the content it generates. One criticism of some generative AI models is that while the information they generate can be very convincing, in some cases it is not always factual or accurate. Grounding has implications not only related to the spread of mis-information, it is also extremely important for enterprises who need to ensure a high threshold for accuracy, particularly when interacting with customers via chatbots or similar interfaces.

12. Prompt: An interface where a user can interact with a generative AI model using natural language. Within these prompt interfaces, a user can make a specific request of a model such as “Write a paragraph summary of generative AI”, or “Draw a picture with horses running through a field with a forest in the background, during the spring”. In response, the model ‘generates’ a response to the user’s request. The unconstrained possibilities enabled by natural language interaction between user and model have generated some of the most impressive and unexpected results from generative AI.

What other terms are you seeing related to generative AI? Connect with me on [LinkedIn](#) and let me know what other terms we should cover in our next blog post.



Erik Goelz

[BACK TO RESOURCES](#)

[PREVIOUS](#) | [NEXT](#)



[PRODUCTS](#)

[RESOURCES](#)



Headquarters:
2200 Geng Road, Unit 100
Palo Alto, CA 94303
(650) 263-1153

[Terms & Conditions](#)
[Privacy Policy](#)

Copyright © 2023 SambaNova
Systems, Inc. All rights reserved.

[INDUSTRIES](#)

[ABOUT](#)

[CONTACT](#)

Customers turn to SambaNova to quickly deploy state-of-the-art AI capabilities to meet the demands of the AI-enabled world. Our purpose-built enterprise-scale AI platform is the technology backbone for the next generation of AI computing. We enable customers to unlock the valuable business insights trapped in their data. Our flagship offering, SambaNova Suite, overcomes the limitations of legacy technology to power the large complex foundation models that enable customers to discover new services and revenue streams, and boost operational efficiency. Headquartered in Palo Alto, California, SambaNova Systems was founded in 2017 by industry luminaries, and hardware and software design experts from Sun/Oracle and Stanford University. Investors include SoftBank Vision Fund 2, funds and accounts managed by BlackRock, Intel Capital, GV, Walden International, Temasek, GIC, Redline Capital, Atlantic Bridge Ventures, Celesta, and several others.



[PRODUCTS](#)

[RESOURCES](#)

[INDUSTRIES](#)

[ABOUT](#)

[CONTACT](#)

Customers turn to SambaNova to quickly deploy state-of-the-art AI capabilities to meet the demands of the AI-enabled world. Our purpose-built enterprise-scale AI platform is the technology backbone for the next generation of AI computing. We enable customers to unlock the valuable business insights trapped in their data. Our flagship offering, SambaNova Suite, overcomes the limitations of legacy technology to power the large complex foundation models that enable customers to discover new services and revenue streams, and boost operational efficiency. Headquartered in Palo Alto, California, SambaNova Systems was founded in 2017 by industry luminaries, and hardware and software design experts from Sun/Oracle and Stanford University. Investors include SoftBank Vision Fund 2, funds and accounts managed by BlackRock, Intel Capital, GV, Walden International, Temasek, GIC, Redline Capital, Atlantic Bridge Ventures, Celesta, and several others.

Headquarters:
2200 Geng Road, Unit 100
Palo Alto, CA 94303
(650) 263-1153

[Terms & Conditions](#)
[Privacy Policy](#)

Copyright © 2023 SambaNova
Systems, Inc. All rights reserved.





[PRODUCTS](#)

[RESOURCES](#)

[INDUSTRIES](#)

[ABOUT](#)

[CONTACT](#)

Customers turn to SambaNova to quickly deploy state-of-the-art AI capabilities to meet the demands of the AI-enabled world. Our purpose-built enterprise-scale AI platform is the technology backbone for the next generation of AI computing. We enable customers to unlock the valuable business insights trapped in their data. Our flagship offering, SambaNova Suite, overcomes the limitations of legacy technology to power the large complex foundation models that enable customers to discover new services and revenue streams, and boost operational efficiency. Headquartered in Palo Alto, California, SambaNova Systems was founded in 2017 by industry luminaries, and hardware and software design experts from Sun/Oracle and Stanford University. Investors include SoftBank Vision Fund 2, funds and accounts managed by BlackRock, Intel Capital, GV, Walden International, Temasek, GIC, Redline Capital, Atlantic Bridge Ventures, Celesta, and several others.

Headquarters:
2200 Geng Road, Unit 100
Palo Alto, CA 94303
(650) 263-1153

[Terms & Conditions](#)
[Privacy Policy](#)

Copyright © 2023 SambaNova
Systems, Inc. All rights reserved.