

Large Language Models

Aleks Lillis

June 2025

Contents

1	Introduction	2
1.1	What are Large Language Models	2
1.2	Examples	3
1.3	Brief Statistics	4
2	LLMs Text Generation	6
2.1	Language	6
2.2	Meaning	6
2.3	Text Generation Process	7
2.3.1	The Power of the Prompt	7
3	Model Training	9
3.1	What is Training	9
3.2	How are Models Trained	9
3.3	Training Data and Bias	9
4	Training Example	10
4.1	Process Overview	10
4.2	Created Model Analysis	10
4.3	Alternate Model Analysis	10
4.4	Overall Comparison	10
5	Impacts of AI	11

Introduction

1.1 What are Large Language Models

In recent times artificial intelligence (AI) has been rapidly developing. With this large AI booms large language models (LLMs) have become a well known thing with many people mistaking them to be synonymous with AI as a whole. This is inaccurate because LLMs are a subset of AI which focus specifically on human language.

Large language models (LLMs) are a category of foundation models trained on immense amounts of data making them capable of understanding and generating natural language and other types of content to perform a wide range of tasks.

Above is the definition of LLMs by International Business Machines Corporation, more widely known as IBM. To put it simple, LLMs are models embedded into programs which are designed to not only understand human text but also respond by generating content. Generated content is based on the data that trains the model (which will be gone over more in depth in a later section) and the prompt that is provided.

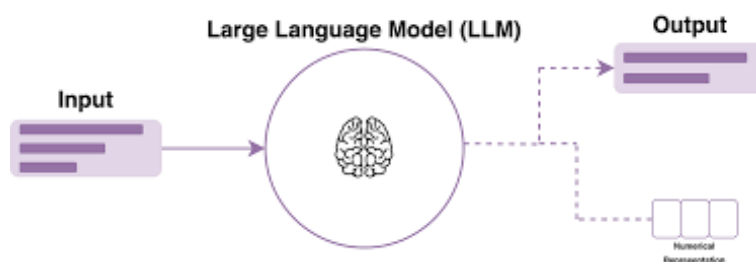


Figure 1.1: Simple LLM diagram [2]

To better understand what LLMs are it is very important to take into consideration what they aren't. LLMs primarily generate context based on patterns uncovered during training and fine-tuning phases of their development. These models can respond to questions by predicting what follows but they do not truly think, that is the job of reasoning models. Reasoning models simulate ac-

tual logic processes in order to solve problems. LLMs are more likely to come up with inaccurate answers because they are models predicting what the response should be.

1.2 Examples

LLMs are used in many different applications. What's nice about modern models is that they can be easily accessible through various mediums by: being built into applications, easily trained remotely from open-source access points, and just being incorporated through a remote server that can send over responses from specific prompts. Some specific usages of LLMs are as follow:

- Text Generation
- Content Summarization
- AI Assistance
- Code Generation
- Sentiment Analysis
- Language Translation
- Grading
- Teaching

The largest use of LLMs in modern day is the use of chat agents that are able to respond and generate context from given prompts but are also able to add prior information from the chat log and use it as context for future generation (while not guaranteed in every model this is typically an implemented feature). Very well-known chat agents are:

- ChatGPT

ChatGPT is an easily accessible model which has rapidly developed recently and is the face of AI chat bots.

- DeepSeek

DeepSeek was created by a Chinese AI startup and stirred up the world of AI by having similar capabilities at a much lower cost.

- Claude

Claude models are forms of conversational AI designed with the goal of being ethical and safe.

- Meta AI

Meta AI is a conversational generative AI that is popular mainly due to its integration to Meta platforms such as Facebook, Instagram, WhatsApp and more.

- Google Gemini

Google Gemini is Google's AI chatbot which is integrated into many of Google's different apps and it is widely popular due to its summarization of search queries on Google's search engine.

- Microsoft Copilot

Copilot is Microsoft's conversational AI model which is integrated into windows computers and other products that Microsoft has.

- Khan Academy Khanmigo

Khanmigo is an AI powered chatbot which focuses on helping students learn from a wide range of courses.



Figure 1.2: An image of the "friendly" icon of the khanmigo AI [3]

1.3 Brief Statistics

The chatbot market is a relatively new but very quickly growing market. According to the Chatbot Global Market Report 2025 the market was valued at \$8.27 billion in 2024 and is projected to have a growth to \$10.32 billion by the end of 2025. Seeing this compound annual growth rate of roughly 25%, the chatbot market is expected to experience an exponential growth to \$29.5 billion by 2029 [5].

The market is not the only place where AI numbers have seen a huge increase in recent times. AI usage has had a wild increase in the last few years globally. AI is very appealing in the work place due to it being able to very easily complete repetitive tasks that would otherwise take up a lot of time for someone to do it themselves. In a study done by the Melbourne Business School as of 2025 %58 of employees have reported intentionally (using specific tools to aid in task completion) using AI at work on a regular basis where applicable [7]. In a survey done in August 2024, about one out of three respondents reported using AI tools daily or at least a few times per week. One interesting data point was that AI usage was reported to be higher at home than at work but more frequent occurrence was more common at work [4]. This large-scale usage of AI tools has mixed impacts in our world and that will be mentioned in chapter 5.

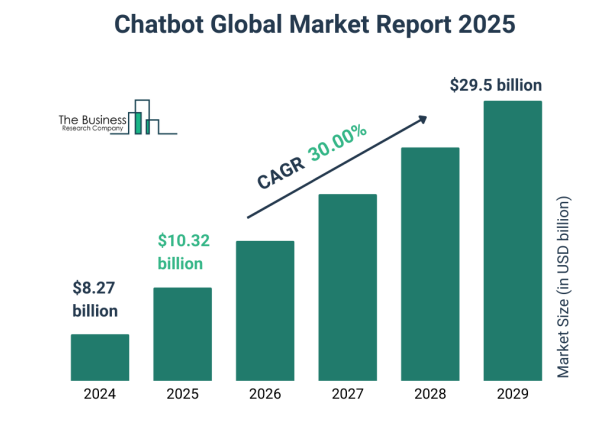


Figure 1.3: A graph depicting the chatbot market growth [2]

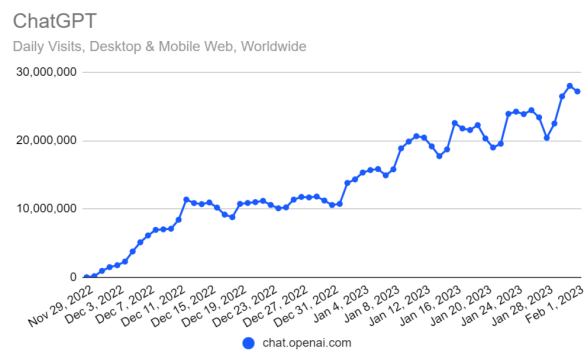


Figure 1.4: A graph showing the fast increase of AI usage [9]

LLMs Text Generation

2.1 Language

As mentioned earlier, the main purpose of LLMs is to generate text—but what, fundamentally, is language? Bernard Bloch and George L. Trager define it as “a system of arbitrary vocal symbols by means of which a social group cooperates.” This classic definition emphasizes spoken language and its social function. However, language has since broadened to include written, gestural, and even computational forms. At its core, language is a structured system of symbols—whether phonetic or textual—used to convey meaning, share knowledge, and build connections. Understanding this helps frame what LLMs are truly modeling: not intelligence in the human sense, but patterns in the way people communicate meaning through text.

2.2 Meaning



Figure 2.1: The famous sculpture of “The Thinker” which embodies a strong philosophical connection to meaning [6]

Does the text that LLMs produce actually convey any meaning and can invoke emotion? Lets start first by comparing this generated text to human language. In human language meaning is deeply tied to context and sentence intention. A sentence or statement isn't powerful because of the words that it uses but mostly because of what we mean by them. Large Language Models, however, do not have the ability to convey this level of meaning and have as high of a level of understanding of what is being conveyed. Instead, LLMs are able to generate text by using highly advanced probabilistic models based on statistical patterns from trained data to complete sentences.

2.3 Text Generation Process

How do Large Language Models actually generate text?

2.3.1 The Power of the Prompt

There are many different types of prompts that LLMs are capable of responding to.

- Summarization

LLMs are able to summarize text. This can be achieved by using advanced algorithms for generating a shorter version of a text that still keeps the main points. Summarization primarily reliant on **word frequency** (words that appear more often are likely to be more important), **sentence position** (some sentences are more important in conveying the point of a text such as thesis statements or conclusions), **specific names** (sentences which specify information like names, locations, etc. are more likely to be important), and much more [8]. As well as this, larger texts are typically broken down into smaller chunks of text in order to summarize each one into broader summaries of the entire text by keeping the most important information [1].

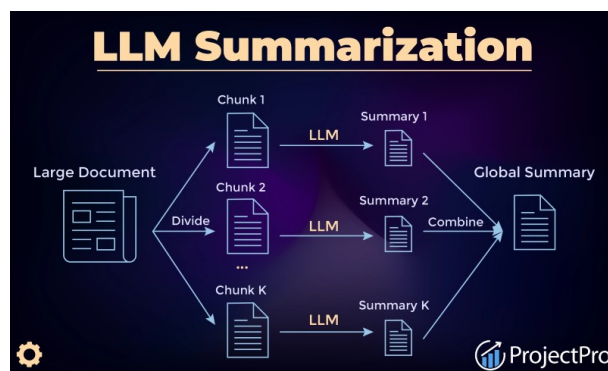


Figure 2.2: Simple diagram breaking down LLM summarization [1]

- Question Answering

AI models are able to answer questions which is one of the primary ways that LLMs are used. Well, LLMs don't actually recognize and know how to respond to a question by thinking. The way LLMs actually answer is by continuing text. The main way that this is able to be done is when the non-AI portion of the program transforms the original part into text that can be continued to provide an answer. This is because the model can finish and return a completed version of rephrased sentence which is the generated answer.

Model Training

3.1 What is Training

3.2 How are Models Trained

3.3 Training Data and Bias

Training Example

- 4.1 Process Overview
- 4.2 Created Model Analysis
- 4.3 Alternate Model Analysis
- 4.4 Overall Comparison

Impacts of AI

Bibliography

- [1] en. Jan. 2025. URL: <https://www.projectpro.io/article/llm-summarization/1082>.
- [2] URL: <https://images.app.goo.gl/KskyibfBQ4kLSzA67>.
- [3] URL: <https://support.khanacademy.org/hc/en-us/articles/13987846292621-Where-did-the-name-Khanmigo-come-from>.
- [4] Kseniia Burmagina. *Unveiling the Future: AI Usage Stats You Need to Know*. en-US. Jan. 2025. URL: <https://elfsight.com/blog/ai-usage-statistics/>.
- [5] *Chatbot Market Report 2025: Statistics and Key Trends by 2034*. en. URL: <https://www.thebusinessresearchcompany.com/report/chatbot-global-market-report>.
- [6] Wikipedia contributors. *The Thinker*. en. June 2025. URL: https://en.wikipedia.org/wiki/The_Thinker.
- [7] “Key findings on AI at work and in education₂₀₂₅”. In: (2025). URL: <https://www.mbs.edu/faculty-and-research/trust-and-ai/Key-findings-on-AI-at-work-and-in-education>.
- [8] Rajni Singh. “Large language models and text summarization: a powerful combination”. en. In: *Medium* (Nov. 2024). URL: <https://medium.com/@singhrajni/large-language-models-and-text-summarization-a-powerful-combination-6400e7643b70>.
- [9] Krunal Vaghasiya. *The Latest ChatGPT Statistics and User Trends (2022-2025)*. en-US. June 2024. URL: <https://wisernotify.com/blog/chatgpt-users/>.