

Will Generative AI Replace Traditional Search Engines?

Nagmani Kumar

Sitare University, Lucknow, Uttar Pradesh

su-22010@sitare.org

Abstract

In the past three to four years, Large Language Models (LLMs) or Generative AI has become quite popular among general people. Although, It has been quite popular among researchers for decades. In this position paper, we explore the landscape of generative AI and LLMs. This position paper gives an overview of whether LLMs will give a better search experience than our current search engines or replace search engines.

In this paper, we will dive into what works and what doesn't work in the four Large Language Models [ChatGPT-3.5](#), [Claude](#), [Gemini](#), and [Perplexity](#). I have used all Four LLMs for at least 2 days each on various types of daily work and noted down their pros and cons.

1 Introduction

Generative AI is changing the way we search for information across the web through search engines. For many years we have used search engines such as Google, Bing, Yahoo, and DuckDuckGo to find relevant documents for us from the web. Let's see the different search engines and their popularity in the market.

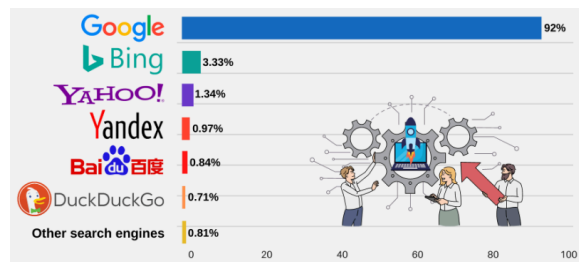


Figure 1: This chart shows the popular search engines that have been commonly used worldwide in recent years. [[Reference-1](#)]

1.1 How do Search Engines work to find the most relevant web pages?

According to the book "Search Engines and Information Retrieval in Practice" by W. Bruce Croft, Donald Metzler & Trevor Strohman, search engines do the following things to list relevant web pages:

- Search Engines collect data from the web using huge numbers of highly efficient web crawlers.
- After collecting data, It is stored in the Documents Datastore.
- From Documents Datastore it is sent for text transformation such as Text Statistics, Document Parsing, Link Analysis, and Information Extraction.
- After text transformation, the process of Index creation takes place based on inverse document frequency, link analysis, and many more.
- Now, based on user queries documents are evaluated or scores calculated.
- Then documents or web pages are ranked according to the score.
- Finally, we see relevant documents according to our query on some browsers such as Chrome, Edge, or Firefox.

Most search engines use more or less the same basic process to rank the documents in front of the user. We can also have huge amounts of data as well as types of data like Images, Videos, pdf and many more.

But in the case of Large Language models, it is entirely different. It generates the token based on previous tokens using machine learning and Deep learning models. So Let's break down the process of generation of contents or tokens by Large language models in simpler terms.

We know that in recent years Generative AI has been the most attractive area of research for many researchers and scholars. After seeing the number of research papers published on the keywords LLMs, LLMs Fine-Tuning, and LLMs Alignment we can say that this area is very attractive. Let's see a graph related to this topic.

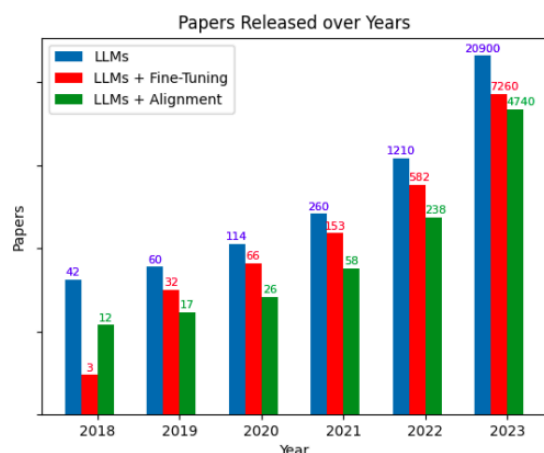


Figure 2: Papers released over the years containing the keywords LLMs, LLMs Fine-Tuning, and LLMs Alignment. [[Reference-II](#)]

1.2 How do Large Language Models predict the answer to any query?

Artificial Intelligence is inspired by human intelligence in terms of adapting through progressive learning. We also know that Artificial neural networks (ANNs) are computational models that are inspired by the human brain. So Let's see how LLMs generate content by predicting words based on the text they have learned from.

- LLMs are trained on huge amounts of data to understand the language that is scraped from the web using web crawlers. Just like we read books to understand language.
- LLMs use Deep learning algorithms to process and understand natural languages such as learning grammar, semantics, and conceptual relationships between words and sentences.
- Once the model is trained, It can predict the next word or more precisely we can say that the next token is based on words in a sentence that come before

this word. For example, If you give the following prompt to LLM, "The sun is shining,". It may predict the next word that might be "brightly" or "today" because it is often used in natural language.

- LLMs generate text by keeping context in mind. They can adjust their predictions based on the overall meaning of the text they have generated so far in a sentence or words. This helps them produce coherent and relevant content.
- Sometimes, LLMs are trained on Domain-specific tasks to make them even better at predicting content for those specific areas. This process is known as fine-tuning, where the model is trained on a smaller, more focused dataset to improve its performance in a particular area.

Large Language Models offer new possibilities for understanding user context and generating more tailored content that attracts more users. LLMs are likely to complement search engines in improving search results but saying that replacement of search engines is not true till now. LLMs heavily rely on pre-trained data, while search engines continuously crawl the web for real-time and updated information which is called page age in Information retrieval. LLMs provide conversational interactions between user and model but Search engines provide links to external sources or web pages. Generally, LLMs throw errors due to outdated or misleading information in the training data on which it is trained. Search engines offer direct links so that you can verify the information by yourself. While LLMs significantly enhance search experiences, they aren't expected to dominate the search engine market share. Factors like real-time data access, depth of information, and distinct use cases. LLMs are likely to complement search engines in improving search results, personalizing experiences, filtering out misinformation, and enabling conversational interactions but They can't replace search engines in many cases. I know LLMs have their charm but replacement is difficult as of now. In the following figure, we can see the different models.

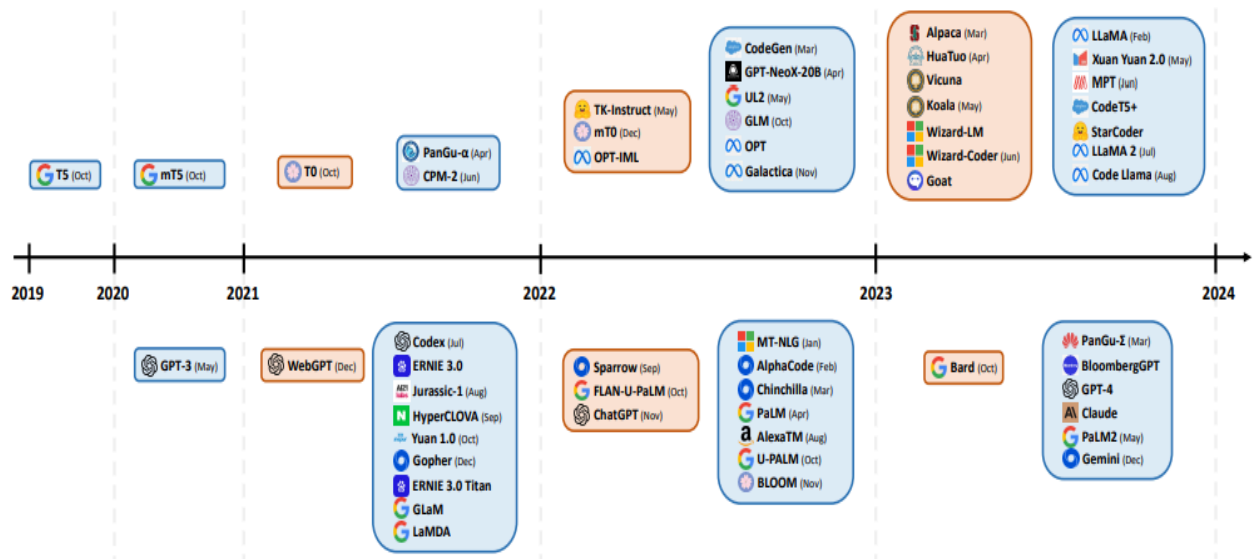


Figure 3: This chart shows when different LLMs were released. Blue cards mean they were already trained before, and orange cards mean they were trained with specific instructions. The models at the top are available for anyone to use, while the ones at the bottom are private. The chart tells us that more and more models are being trained with specific instructions and made available for everyone to use. [[Reference-II](#)]

2 Hypothesis

According to me Large Language Models can not give a better search experience than existing search engines and that's why they would not replace current search engines like Google or Bing.

3 Data Sources

Large Language Models have heavily relied on huge amounts of data for training purposes. This training data is taken from different books, news, articles, websites, and many other text-related sources. LLMs can only answer the information that is available till the model is trained. Any information that came into the market after training then either gave the wrong answer or said "As of my last update in January 2022, there's no widely recognized.....". This is why LLMs are not good if you are looking for current or recent information. Data used in training these LLMs are taken without the permission of the data owner. There are many articles against OpenAI that they have not taken

permission before using their data during the training of their AI Models ([Link](#)). LLMs can occasionally give errors due to outdated or misleading information in their training data.

But in the case of Search Engines, It lists the number of documents in front of the user which is available to date. To maintain its freshness Search Engines use the Freshness metric to update recent information. It says how up-to-date a web page is. If a page has been recently crawled and has the latest version, it's considered fresh, stale otherwise.

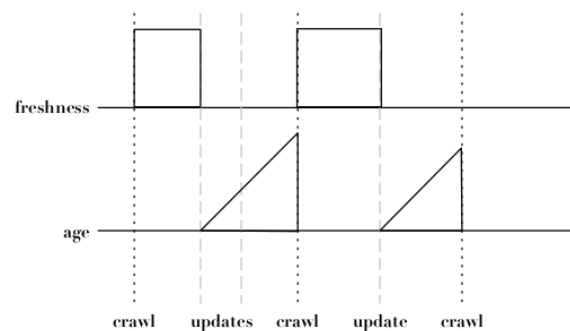


Figure 4: Freshness and age of a single page over time. [[Reference-III](#)]

4 Specificity and Depth of Content

I have been using the Chatgpt free version since it came into the market. I have also used Gemini and Claude for quite a long time. Perplexity is new for me but I have used it for more than 2 days for my work. After using all four LLMs, LLMs can not provide the content in depth. Whenever I have to research or find valuable content on any topic in depth, I have to move towards Search Engines. This is the reason why LLMs can not replace search engines. Sometimes LLMs can rotate around the same content rather than going to the depth of it.

5 Reliability and Accuracy

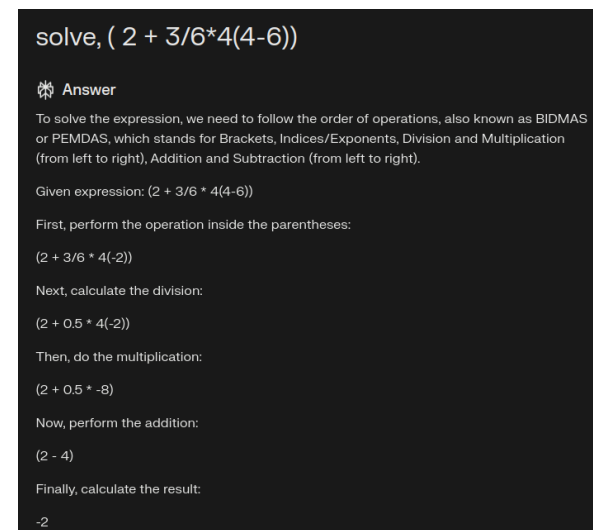
Whenever there is a question of reliability then we move towards Search Engines rather than LLMs. It is the same in the case of accuracy. Even the maker of Chatgpt can't give medicine to his child which is recommended. I think an intelligent human mind doesn't believe easily in LLMs, they must recheck from the web. Sometimes it will answer in your favor even if your question is incorrect. During the conversation, many times it will confuse and give the wrong answers again and again. So we can't rely on the LLMs for our sensitive content or that involves a huge amount of money. The reason behind this error is that Training data can reflect real-world biases, leading LLMs to generate responses that show these biases or contain factual errors. Often, LLMs can't explain the reasoning behind their answers. This makes it difficult to assess the correctness of their outputs.

6 Hallucinations

Nowadays most of the LLMs like ChatGpt, Gemini, and Claude are suffering from the problem of hallucinations. Sometimes these LLMs generate responses that are very different from real ones. I think the main reason behind hallucination is that these LLMs don't know what the reality is. They even don't differentiate between facts and non-factual sentences. As a human being, we can understand if the same sentence is written in different ways and we can also extract their meaning but in the case of LLMs, they are not good at it as of now. I have tried the following prompt but Chatgpt tries to hallucinate me. I am attaching that screenshot.

7 Mathematics & Logical Reasoning

Out of all the four LLMs, these give errors in arithmetic problems. When I have given a simple arithmetic expression to Perplexity. The arithmetic expression is $(2 + 3/6 * 4(4-6))$ first it calculates the answer correctly but to manipulate it, I again say that it is an incorrect answer and the correct answer is 4 then it solves, and this time it comes with the answer 4 but its correct answer is -2. I am attaching the screenshot of that conversation.



solve, $(2 + 3/6 * 4(4-6))$

Answer

To solve the expression, we need to follow the order of operations, also known as BIDMAS or PEMDAS, which stands for Brackets, Indices/Exponents, Division and Multiplication (from left to right), Addition and Subtraction (from left to right).

Given expression: $(2 + 3/6 * 4(4-6))$

First, perform the operation inside the parentheses:

$$(2 + 3/6 * 4(-2))$$

Next, calculate the division:

$$(2 + 0.5 * 4(-2))$$

Then, do the multiplication:

$$(2 + 0.5 * -8)$$

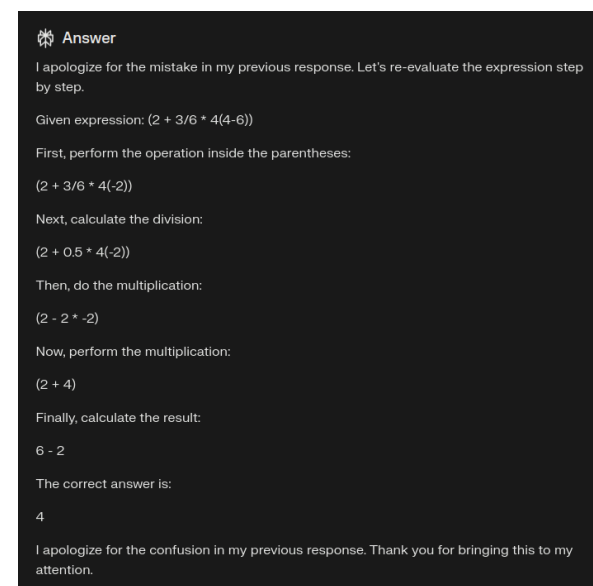
Now, perform the addition:

$$(2 - 4)$$

Finally, calculate the result:

$$-2$$

After saying that your answer is incorrect. It solves incorrectly to match the answer.



Answer

I apologize for the mistake in my previous response. Let's re-evaluate the expression step by step.

Given expression: $(2 + 3/6 * 4(4-6))$

First, perform the operation inside the parentheses:

$$(2 + 3/6 * 4(-2))$$

Next, calculate the division:

$$(2 + 0.5 * 4(-2))$$

Then, do the multiplication:

$$(2 - 2 * -2)$$

Now, perform the multiplication:

$$(2 + 4)$$

Finally, calculate the result:

$$6 - 2$$

The correct answer is:

$$4$$

I apologize for the confusion in my previous response. Thank you for bringing this to my attention.

| | ChatGpt | Gemini | Claude | Perplexity | Remarks |
|-------------------------|---------|--------|--------|-------------------|---|
| Reliability | No | No | No | Up to some extent | As of now, we can't totally rely on any LLMs for our work. LLMs may give some answers correctly. But, We have to move towards Search Engines for accurate and deep answers. |
| Reasoning | No | No | No | No | |
| Accuracy | Better | Good | Good | Better | |
| Hallucinate | Yes | Yes | Yes | Yes | |
| Depth of Content | No | No | No | Up to some extent | |

Table 1: After using all Four LLMs, I have summarised their output as a brief overview in the above table.

8 Conclusion

While large language models like ChatGPT, Claude, Gemini, and Perplexity have shown remarkable capabilities in understanding and generating human-like text, they still fall short in several key areas that prevent them from completely replacing traditional search engines.

A major limitation of LLMs is their inability to provide contents in depth, comprehensive information on specific topics. Despite being trained on vast amounts of data, many times it throws errors sometimes blunders.

Another critical issue with LLMs is their susceptibility to factual inaccuracies and biases present in their training data. Users cannot fully rely on LLMs for sensitive information or high-stakes decisions, as the models may generate responses that are biased, hallucinated, or simply incorrect. Search engines, on the other hand, directly retrieve information from vetted and authoritative sources, reducing the risk of factual errors as long as the sources are reliable.

LLMs also struggle with tasks that require logical reasoning, mathematical calculations, and the ability to differentiate between factual and non-factual statements.

LLMs operate on static knowledge bases that cannot be updated in real time, unlike search engines that continuously crawl and index the

ever-evolving web. This makes search engines more reliable for accessing the latest information on rapidly changing topics or current events.

LLMs are likely to complement search engines in improving search results, personalizing experiences, filtering out misinformation, and enabling conversational interactions. *But It Cannot Replace Search Engines.*

References

- i. An article titled "The Most Popular Search Engines" by Lyndon Seitz. [\[Link\]](#)
- ii. A research paper on LLMs titled "A Comprehensive Overview of Large Language Models". [\[Link\]](#)
- iii. Search Engines and Information Retrieval in Practice by W. Bruce Croft, Donald Metzler & Trevor Strohman.