FUJITSU

Emerging Technologies

# ChatGPT for business

AI will not replace you.
A person using AI might.

**ChatGPT is an international news phenomena, and rightly so.**

**Every day, hundreds of new applications based on Large Language Model are created. But what should we be excited about using it for, and what should we be wary of?**

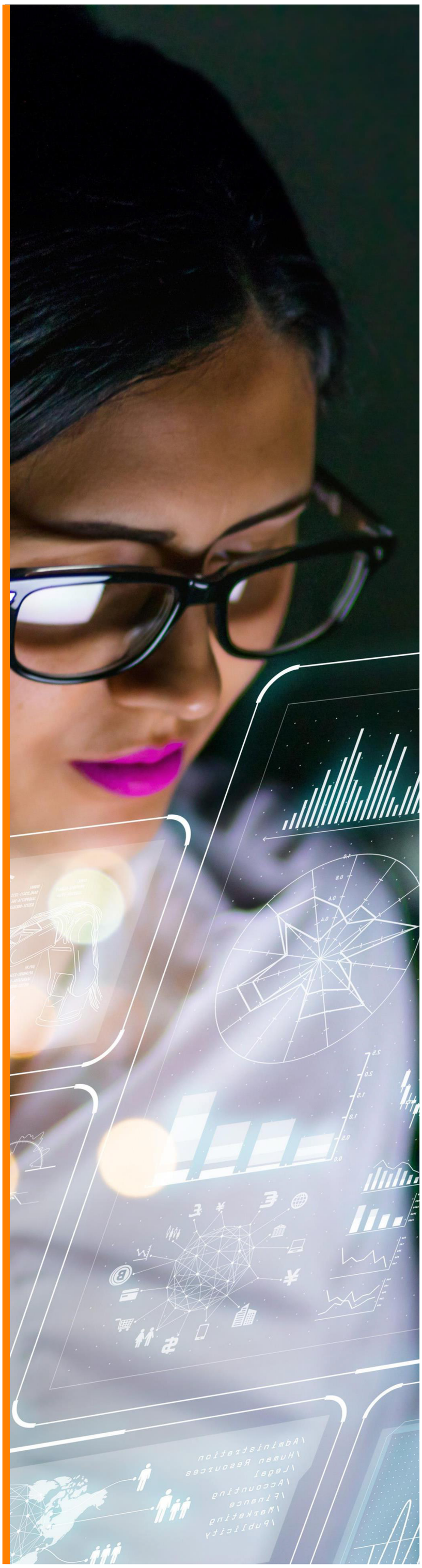
While this article talks about ChatGPT, what follows is also broadly true for all of the following comparator systems:

- **Playground** (OpenAI)
- **Lex**
- **Bard** (Google)
- **BingChat**
- **Claude** (AnthropicAI)
- **NeevaAI**
- **YouChat** (You.com)
- **Perplexity Chat** (Perplexity.ai)
- **Playground** (Cohere.AI)
- **Poe** (Quora)

## What is a Large Language Model (LLM)?

**ChatGPT is based on the 'Generative Pretrained Transformer (GPT)' architecture.**

**Generative** means it creates a synthesised output or output options. It is not an 'if-then' system where a given input links directly to a preset output. It is predisposed to give you an answer – often compellingly and clearly articulated – without any uncertainty quantification. There are engineered safety guardrails to try to minimise harmful answers, but they are brittle and easy to 'jailbreak'. For instance, you may ask it to write hate speech and it will refuse, however asking it to write a story in which a fictional character engages in hate speech can produce that hate speech.



**Pretrained** means it has been trained on a massive amount of text data from the internet. Wikipedia accounts for only 3% of the training data[1]. That means its intelligence is based on a predominantly 'knowing' rather than 'thinking' approach - more on that later.
In ChatGPT's case, the label 'pretrained' also refers to reinforcement learning from human feedback. OpenAI employed lots of human users to rank outputs during its development. This allowed a reinforcement learning system to learn to reward ChatGPT for outputs similar to those that humans rank highly.
In other words, it learned "what good looked like" to people.

A **Transformer** is a deep learning model that adopts the mechanism of self-attention, differentially weighting the significance of each part of the input data. To put it another way, it reads an entire input, such as a sentence, and assigns importance and meaning to parts of that sentence. It is worthy of note that Transformers have limitations on the size of the input they cope with. A good analogy is that a Transformer could read a book, but only in bite sized pieces. But having interpreted the meaning of the first chapter from the relationships between text elements, that meaning becomes fixed. It could not then change the meaning it assigned if later in the book there was a twist that should cause you to reinterpret events in the story. If it read *Fight Club* or *And Then There Were None*, it wouldn't see the earlier chapters in a different light after the big plot reveal. This is one of the reasons that ChatGPT has fixed sizes for inputs and outputs and can remember across a limited amount of its conversational interactions.

The combined effect of these elements is a machine that:

- reads a prompt from a user and extracts what it perceives to be the meaning of that request;
- then predicts the most likely string of words, sentences, or code to satisfy that prompt;
- all judged against the patterns inherent in the very large corpus of data it was trained on.

1. All of Wikipedia only accounts for 3% of the training data of GPT-3. Since then, we have had GPT-3.5 and more recently GPT-4.

# Intelligence as Knowing vs Intelligence as Thinking

**Because it has access to a vast array of information, ChatGPT is very good at leveraging from existing knowledge. This is not the same as logical thinking from first principles.**

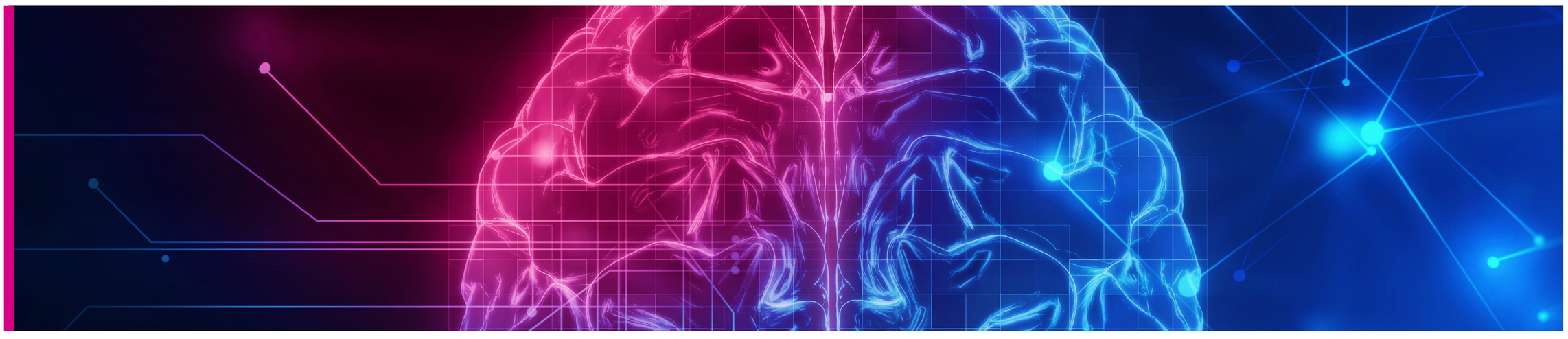
Humans have strong innate inductive priors; built-in assumptions, biases, and knowledge that shape their understanding of the world and guide their reasoning processes. These allow humans to make generalisations, learn from limited data, and reason about new situations based on their prior experiences. Critics argue that ChatGPT is just a 'stochastic parrot' incapable of anything like human intelligence because the LLM has no such tools and relies on probabilities derived from what was most common in its training data. But this risks underestimating how good outputs from ChatGPT can be.

Intelligence based on knowing can be far more advanced than something as basic as a lookup table, where every input has a simple matching output. It can allow for the solving of incomplete problems by inference.

It is also worth remembering that we humans expend a large amount of time, energy and money increasing the ‘knowing’ part of our own intelligence. The world of literature is full of people whose intelligence is not associated with reasoning from first principles, but from prodigious reading. Academic education has prescribed reading lists for almost every course. Artists study the styles and works of others to inspire, inform and shape their own works.

This is not the same as reasoning from first principles. For example, ChatGPT can do simple sums quite well. But this is because there are lots of examples of those simple sums in the slice of the internet that the system was trained on. It doesn’t have to ‘do maths’, it just remembers that “2 + 2 =” invariably gets followed the a “4”, exactly as “Mary went to the beach and built a ____” is more likely to be followed by “sandcastle” than “ice sculpture”. But LLMs are well known for getting complex maths questions wrong. There are simply fewer examples of complex maths questions, so the machine cannot find a 'likely' answer that happens to be correct.

Oddly enough LLMs are more likely to be able to write you software that can correctly answer complex maths questions, than to correctly answer the question.  This is because there are lots of examples of that sort of code in the training material. Later editions of LLMs have started to utilise 'plug ins'. The intent is that the LLM recognises the nature of the problem and farms the task out to a different type of software to get the right answer to incorporate into its reply. You can think of this as the LLM subcontracting part of the task to a better equipped program for that job. GPT-4 with a plug in to Wolfram -  software optimal for such maths problems - is a good for example.



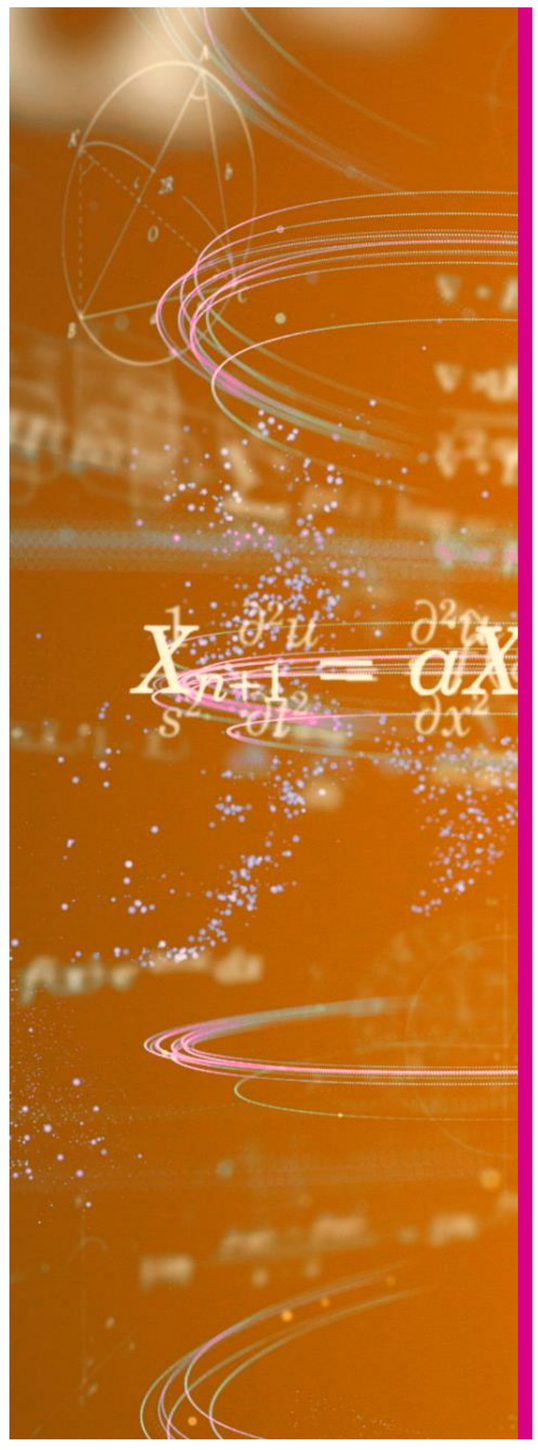
# What is ChatGPT bad at?

**Earlier in this article, we saw that ChatGPT can be very bad at basic maths. It also performs inconsistently at logic puzzles or logical thinking for very similar reasons. However, the ability of LLMs to perform such tasks jumps with each growth in scale. GPT-4 now masters some logic puzzles that defied GPT-3.5. We should expect future scaling to continue to improve that ability.**

ChatGPT is not grounded in real first-hand experience of the world. It has access to a huge amount of knowledge, but its relationship to the real world is mediated entirely through that data. In such an existence, there is no possibility of 'common sense' grounded on its own direct experiences. In addition to lacking ' grounding', the model will always tend towards reproducing any biases that exist in the training data, simply because the data and user interactions are the only representations it has of the world. It is notable that engineered rule-based guardrails have been added by OpenAI, for example, ChatGPT typically refuses to discuss views of religions.

ARC who partnered with both OpenAI and Anthropic as external safety testers indicated that *"models are error-prone, sometimes lacked important technical understanding, and easily became derailed. They were prone to hallucinations, were not fully effective at delegating large tasks between multiple copies and failed to tailor their plans to the details of their situation."*[2]

Of note in this list is *hallucinations*. This is where the model writes an incorrect but internally coherent answer. Because ChatGPT is very good at knowing "what a good answer looks like" and bad at identifying what it doesn't know or communicating uncertainty, its incorrect answers can appear compelling and wholly realistic to the naïve user. ChatGPT freely mingles fictions with facts which can make the problem even more challenging for the user. The question-and-answer coding site StackOverflow temporarily banned answers generated by ChatGPT because moderating the volume of misleading information submitted had become unmanageable.[3]



Copyright infringement and plagiarism are also risk areas. OpenAI 's terms and conditions indicate that users have free use of whatever output ChatGPT produces from prompts and OpenAI are not claiming any copyright fees or IP. However, ChatGPT frequently produces text likely to fail a plagiarism test because it comes too close to the content and structure of published texts that were part of its training corpus. For example, Tech news site CNET was recently caught posting AI-generated content that amounted to close paraphrasing of other journalists' original content.[4]

2. More information about the dangerous capability evaluations we did with GPT-4 and Claude. – Less Wrong
3. Why posting GPT and ChatGPT generated answers is not currently acceptable -Help Center -Stack Overflow
4. CNET's AI Journalist Appears to Have Committed Extensive Plagiarism (futurism.com)

## What is ChatGPT good at?

Lots of things! They are likely to have the same turbocharging effect on reading and writing that desktop calculators had on maths.

Every day, there are new API derived tools from LLMs. There is a list of features that LLMs are good at in the Blog for O365-Copilot, the GPT augmented Office 365 package from Microsoft. Note that for every O365 Copilot line listed below, you can already get a productivity tool that does something similar on the internet:

- Copilot gives you a first draft to edit and iterate on - saving hours in writing, sourcing, and editing time.
- Unlock productivity - 20% of our work that really matters, but 80% of our time is consumed with busywork… summarising long email threads to quickly drafting suggested replies, […] Copilot in PowerPoint helps you create beautiful presentations with a simple prompt, adding relevant content from a document you made last week or last year.
- It can summarise key discussion points [from a meeting] - including who said what and where people are aligned and where they disagree and suggest action items.
- Developers who use GitHub Copilot, 88% say they are more productive, 74% say that they can focus on more satisfying work, and 77% say it helps them spend less time searching for information or examples.
- The average person uses only a handful of commands - such as “animate a slide” or “insert a table” - from the thousands available across Microsoft 365. Now, all that rich functionality is unlocked using just natural language.

## What do we mean by bad or good?

Before we decide the list above or any list of pros and cons on LLMs is authoritative, you should always ask, “compared to what?” Can it write plays as well as Shakespeare? No. But then almost none of us can.

- **What are you comparing ChatGPT to?**
- **For what task? What are you trying to do?**
- **Does it matter whether the output is perfect, or good enough, or just cheap?**
- **What is the opportunity cost, or the actual monetary cost to do this? Compared to what alternative? What is the cost or saving of ‘not doing it’?**

# LLMs and business strategy

There has been an explosion of tools exploiting LLMs. The imperative to use them feels overwhelming. But there are a few factors to remember.

Where does your data go? Most LLMs have been released with the explicit statements that user interaction is invaluable for refining the tool for the future. The interactions with conversational agents will almost certainly form part of the next corpus of training data, which means that unless you know the system is in a secure segregated container - which can now be procured from many cloud or IT service providers - any input you make to the LLM should be treated if it were being logged onto the internet for future discovery. For example, Samsung banned LLM use for employees after engineers put proprietary code into GPT to fix it, risking leaking it.[5]

Microsoft, whose deal with OpenAI makes them heavily integrated described the layers of LLMs as follows:

| | |
|---|---|
| **Applications** | **Software programmes where the output of an AI model is put to work.**<br>• GitHub Copilot and the new Bing are examples |
| **API services** | **Application Program Interfaces (APIs) are endpoints that give our customers access to pre-trained models.**<br>• No need for customers to invest in significant infrastructure<br>• Customers can call multiple different APIs and access different models |
| **Pre-trained AI models** | **AI models that are already trained on a large amount of data and can be used to solve similar problems without starting from scratch.**<br>• The GPT series of models are pre-trained models |
| **Machine learning acceleration software** | **Helps to speed up the process of developing and deploying large AI models.**<br>• Optimises the performance and efficiency of ML models<br>• Simplifies the coding and debugging of ML models<br>• Helps to track experiments and manage workflows |
| **Infrastructure** | **Large AI models require advanced supercomputing infrastructure.**<br>• Clusters of specialised hardware, especially GPUs (Graphic Processing Units)<br>• High bandwidth networks<br>• Fast, low-latency connections |

Key points from Microsoft on understanding the GPT Technology Stack were that the costs are most heavily associated with the bottom two layers. Pre-training can be done for considerably less than the costs associated with the ML and Infrastructure layers but can still be expensive. The API and Applications layers are the cheapest, particularly the Applications layer. But this is also the least defensible layer from a business perspective. The ability to create near identical tools in different ways from the same foundation layers is difficult to protect against and cheap to emulate.

5. https://www.pcmag.com/news/samsung-bans-chatgpt-after-engineers-use-it-to-fix-proprietary-code

## Key takeaways

- **LLMs are Foundation Models.**
  A large variety of tools will continue to be built upon them.
- **LLMs derive what "intelligence" they have from the power of scale** – huge datasets and huge computation deriving probabilistically correct answers to prompts.
- **LLMs are not 'grounded' in the real world, only in what we have written about the world on the internet.** When that data is flawed the LLM will hallucinate convincing but wrong answers at us.
- **Copyright, security, and secrecy are all unresolved issues.**
  These are not trivial or easy questions for society.
- **LLMs can be fantastic aides now.** They have superb possibilities to assist in productivity and aide creativity. They are nowhere near ready to be part of safety critical systems.
- **Do not rely on ChatGPT outputs in making important decisions.** For important decisions check the answers LLMs give you against proven facts. You might think of this as similar to when you choose to accept the word of a  very persuasive but often unreliable friend occasionally prone to making things up to appear interesting.
  Or, a fact you found on a website where you know nothing about the site, the author, the publisher or the sources they used.

## Thinking about the future

Although this note contains important cautions and limitations about ChatGPT now, it is also important not to dismiss recent developments as either hype or the limit of what can be done. We are not talking down the profundity of the breakthroughs achieved in AI. The Alan Turing Institute described ChatGPT as the first tool that 'felt like AI' for the public. LLMs have been years in the making and are seem set to revolutionise the business world - and indeed the way government and society works.

Debate continues as to whether LLMs will get us to Artificial General Intelligence (AGI) – a machine that matches or exceeds human intelligence in all domains. But what is clear is that these technologies are advancing at a rate that will change how we understand the world.

**Contact**
Alan Brown, Director Neurosymbolic A.I.
alan.brown@fujitsu.com