

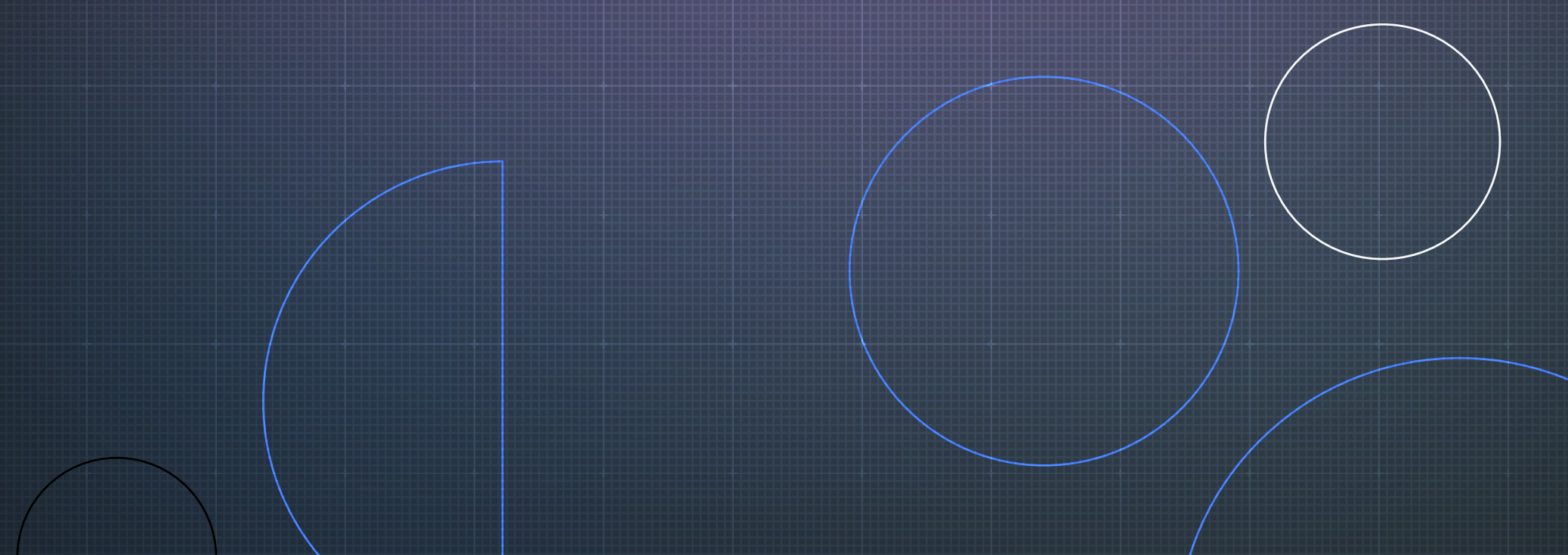


A BAYTREE GUIDE

# Large Language Models, A Guide for Digital Publishers

Welcome to this introductory guide on the application of AI, specifically Large Language Models (LLMs), in the field of digital publishing.

[hellobaytree.com](https://hellobaytree.com)



<b>3</b>	<b>Introduction</b>
<b>4</b>	<b>What are LLMs?</b>
<b>5</b>	<b>What does current adoption look like?</b>
<b>7</b>	<b>Key Takeaways</b>
<b>8</b>	<b>How can and how should publishers be working with LLMs?</b>
<b>15</b>	<b>Limitations and considerations</b>
<b>21</b>	<b>Current players</b>
<b>26</b>	<b>About Baytree</b>

# Introduction

Welcome to this guide on the application of AI, specifically Large Language Models (LLMs), in the field of digital publishing.

In this guide, we will explore how LLMs can aid content generation and enhance various aspects of the publishing workflow. We'll also discuss the key considerations and limitations associated with using LLMs in this context.



# What is an LLMs?

LLMs are a type of AI technology, with General Purpose Transformers (GPTs) being one notable example developed by OpenAI, such as GPT-3.5 and GPT-4. These models have the capability to process and generate human language-based content on a large scale, making them extremely interesting tools for content creation and augmentation.

Almost every industry is currently buzzing with the potential uplift in productivity that LLMs can offer. At Baytree, we see huge opportunity for LLMs to improve the output of anyone involved in digital publishing, with potential for the technology to replace some of the more time-consuming aspects of publishing and support editors and journalists in their day-to-day tasks.

# What does current adoption look like?

We set out to speak to publishers about their thoughts on LLMs and to understand current adoption rates around the technology.

This was a qualitative process involving informal chats and email conversations with our network in the publishing industry. We spoke to 16 publishing executives, all of whom were directly responsible for editorial output, either as a journalist, contributing writer, editor or owner.



We spoke to publishers about their thoughts on LLMs and current usage of the technology.

- 14% operate worldwide, 54% in Europe and the rest have either regional or national focus.
- 27% had an annual turnover in 2022 of more than £10 million. 11% of more than £5 million, 28% £500 to £1 million, 27% had a turnover of less than £500,000.
- They included single brand to multi brand publishers.

27%

were currently using LLMs in a limited ad hoc fashion without any guidelines or defined process in place for their application.

67%

were not using LLMs but intend to do so in the future.

19%

will not use LLMs based on an artistic belief that they devalue their editorial output and or pose a threat to their publication from a regulatory or compliance standpoint.

### How they are are using LLMs

- CRM content tagging and management (42%)
- Article generation (53%)
- Email and newsletters content creation (37%)
- Audio and video management (31%)

# Key Takeaways

- **Only one publisher has put in place formal documentation, guidelines, and processes for working with this technology.**
- **One message we continually heard back is that publishers are not clear on the ‘right’ way to be using LLMs and were unsure if they were taking full advantage of the technology.**
- **There was also a general feeling that there was not much information available to them that provided clear guidance on how best to approach the technology.**

# How can and how should publishers be working with LLMs?

Let's delve into the current specific use cases of LLMs in digital publishing and understand how they can transform the way publishers create and present content to their audience.

Depending on the application, it's important to note that the output of an LLM should be seen as providing "inspiration". Like Watson's insights to Sherlock Holmes, the results of an LLM should be viewed as an aide to spark further investigation. They will require editorial oversight.



## Changing Style and Tone of Articles

LLMs can be utilized to transform or enhance the style and tone of articles to make them more accessible and engaging. For instance, a dry and technical article about a scientific breakthrough can be rewritten using LLMs to captivate a wider audience by incorporating relatable language and real-world examples. This ability to reshape articles allows publishers to broaden their readership and foster greater interest in their content.

### EXAMPLE

A scientific journal has an article about a new research study that is important but written in a dry and technical language. They want to make it more accessible and interesting to a wider audience. By using an LLM, they can rewrite the article in a way that is more engaging and understandable. For instance, the LLM could transform the article into a story that includes real-world examples and relatable language. This would make the article more accessible to a wider audience and increase engagement with the content.

## Making it sound authentic

The base outputs ChatGPT produces are often very generic and dry.

If you want to give the LLMs output a unique voice, one of the fastest ways to do that is to specify a writing style or name a famous author in your LLMs prompt.

Say you work for the The Times, you could prompt that you want the article to be written in the style of the The Times. This can be taken a stage further and you can ask ChatGPT to mimic the style of a specific journalist e.g., “Write up the article in the style of the British historian and Times contributing journalist Max Hastings’.

## I want it to sound like me?

By showing an LLMs some previous content that you have written it will be able to mirror your syntax and style e.g. show ChatGPT some of your previous articles and ask it to ensure that the output it provides is the same in tone and style.

The more text you feed ChatGPT, the better it will be able to copy your style.

### **NOTE**

LLMs can also be used to check syntax, grammar and spelling.

## Headline and Metadata Generation

Creating catchy headlines and metadata is crucial for attracting readers' attention. LLMs can assist publishers by generating headlines of various lengths that can be seamlessly integrated into different formats e.g. print, website, mobile site and socials. This capability simplifies the task of creating multiple versions of headlines and ensures that they are both informative and interesting. By employing LLMs, publishers can enhance audience engagement and optimize the impact of their content.

### EXAMPLE

A lifestyle website wants to create a feature article about healthy eating habits. They want to include headlines of various lengths to fit different slots on their webpage. By using an LLM, they can generate attention-grabbing headlines that are both informative and engaging. For example, the LLM could generate a long headline like "10 Essential Healthy Eating Habits for a Better Lifestyle," a medium headline like "Improve Your Health with These 10 Eating Habits," and a short headline like "10 Habits for a Healthier You."

## Topic Research

Journalists often face the challenge of comprehending unfamiliar topics within tight deadlines. LLMs can function as “industry experts” by providing insights and answering questions in a conversational style. By leveraging LLMs for topic research and understanding, journalists can produce more informative and engaging content that resonates with their audience. This capability enables publishers to deliver timely and well-informed articles even on complex subjects.

### EXAMPLE

A journalist is covering a breaking news story about a new healthcare policy. They want to gain a better understanding of the topic so that they can produce informative and engaging content. By using an LLM, they can ask questions and receive answers in a conversational style, gaining insights into the policy and its implications. This would allow the journalist to produce more informed and engaging content that resonates with their audience.

## Statistical Analysis

With the advent of advanced LLMs like GPT-4, statistical analysis becomes a viable application. Equipped with code plugins, LLMs can generate graphs, charts, and other statistical analysis tools from datasets like Excel sheets. They can detect trends, spot anomalies, and present data in visually appealing and understandable formats. This approach to statistical analysis allows publishers to enhance the presentation and impact of their data-driven content.

### EXAMPLE

A financial website wants to present data about the stock market in a more visually appealing and understandable way. By using an LLM equipped with a code plugin, they can create graphs, charts, and other statistical analysis tools from Excel sheets. The LLM can also detect trends and spot anomalies, which can save time and effort for data engineers or individuals with a strong understanding of statistics and Excel. This would allow the website to produce more informative and engaging content that resonates with their audience.

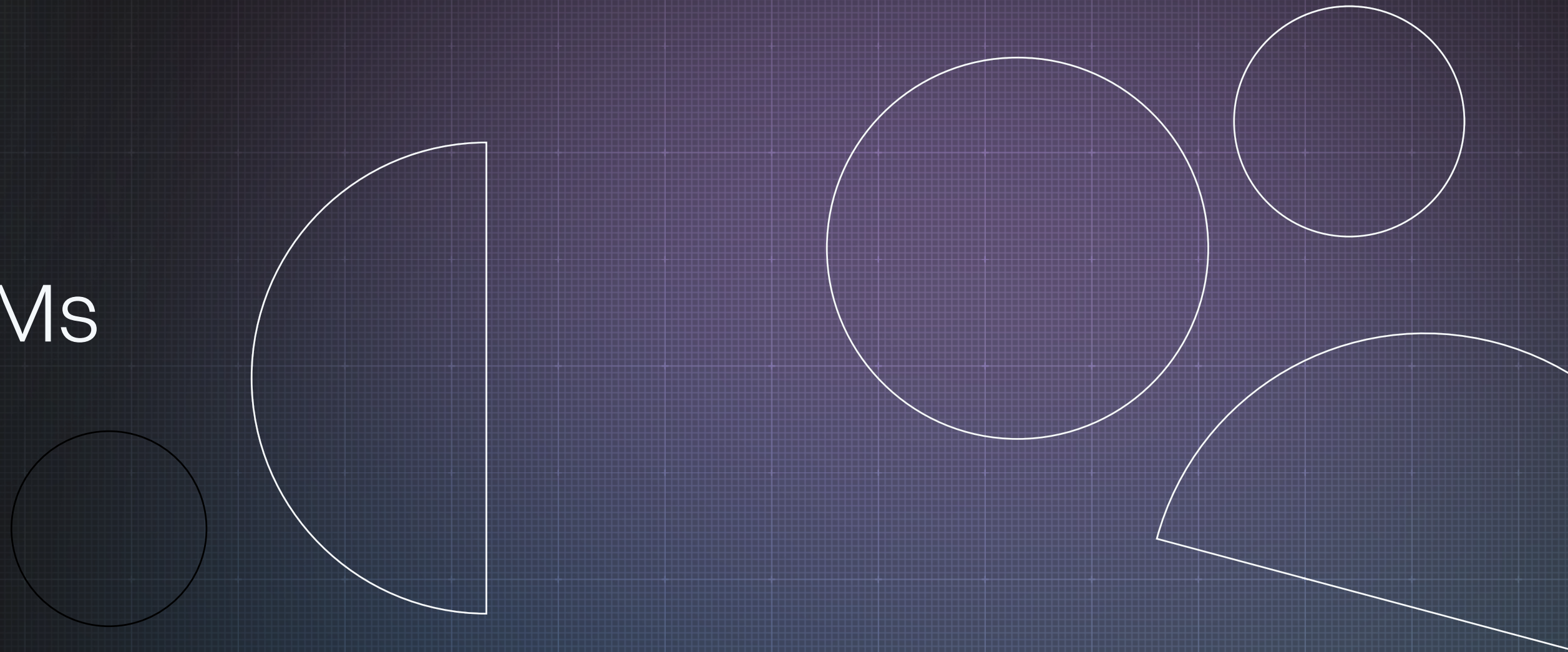
## Use of LLMs to Suggest Next Best Articles (recirculation)

LLMs can be utilized to suggest the next best articles for readers to read next. This involves using an LLM to analyze the content of an article and provide recommendations for other articles that are related or relevant. This can be particularly useful for publishers who want to increase engagement with their content and keep readers on their website for longer. By using LLMs to suggest the next best articles, publishers can create a more personalized experience for their readers and increase the likelihood of return visits. Moreover, with the advent of modern headless CMS technologies, this application is expected to take off massively over the next year.

### EXAMPLE

A news website wants to keep readers on their website for longer by suggesting the next best articles to read. By using an LLM to analyze the content of an article, they can provide recommendations for other articles that are related or relevant. For example, if a reader is reading an article about a new technology product launch, the LLM could suggest related articles about the company or the technology itself. This would create a more personalized experience for the reader and increase the likelihood of return visits.

# Limitations to adoption of LMMs



## Legal and Regulatory Limitations: Copyright Issues with LLM Outputs

The use of large language models (LLMs) by publishers raises several legal and regulatory issues, particularly in the area of copyright. When an LLM generates content, it relies on its training data, which typically includes copyrighted works. Consequently, the output of an LLM may inadvertently include text or elements that are substantially similar to or derived from copyrighted material. This poses a risk of copyright infringement claims against publishers who use LLM-generated content.

Moreover, the ownership of LLM-generated content is a contentious issue. It remains unclear whether the copyright of LLM outputs belongs to the creator of the LLM, the user who prompts the LLM, or if the content is even eligible for copyright protection. The current legal framework does not provide clear guidance on this matter, leaving publishers in a precarious position. This is an important discussion to have with your legal team.

There are various legal cases underway at the moment which should start to provide further clarity on this matter.



## Data Privacy Concerns and Solutions

Another significant concern for publishers using LLMs is the data privacy of the information they send to hosted LLM providers like OpenAI. When publishers use hosted LLM services, they often transmit sensitive or personally identifiable information (PII) to the LLM provider. This data transfer exposes publishers to potential data breaches, unauthorized access, or misuse of the information by the LLM provider or third parties. Moreover, this data sharing may violate data privacy regulations, such as the General Data Protection Regulation (GDPR) or the California Consumer Privacy Act (CCPA), resulting in fines and reputational damage.

One solution to mitigate data privacy risks is using open-source LLMs running locally on publishers' own infrastructure. By hosting and processing the data within their own systems, publishers can maintain control over the data and minimize the risk of unauthorized access or misuse. Furthermore, this approach allows publishers to better comply with data privacy regulations, as they can implement appropriate security measures and data processing agreements tailored to their specific needs.

## Running LLMs locally is not without its challenges

The computational resources required to host and operate large-scale LLMs can be significant, leading to increased costs and technical complexities. Additionally, maintaining and updating locally hosted models may require ongoing investments in hardware and expertise.

In conclusion, publishers need to carefully consider the legal and regulatory implications of using LLMs for content generation. While there are potential solutions to address these concerns, such as hosting LLMs locally, these approaches may also introduce new challenges and costs. It is crucial for publishers to stay informed on the evolving legal and regulatory landscape surrounding LLMs and to adopt best practices that minimize risks and ensure compliance.



## Hallucinations

One of the challenges of using LLMs is the potential for "hallucinations" - generating bad data or suggestions that are not accurate or relevant. This can occur when the LLM is trained on biased or incomplete data, or when the input data is noisy or ambiguous. When this happens, it can lead to incorrect conclusions or recommendations, which can have serious consequences.

To address this issue, workflows need to be put in place to ensure that the output of LLMs is validated and verified before it is acted upon. This could involve having human experts review the output and make corrections where necessary, or using other machine learning algorithms to verify the output. Additionally, there may be a need to retrain the LLM on better data or to use different techniques to reduce the risk of hallucinations.

### Old workflow

write article

check copy

publish

### Suggested new workflow

input facts +  
prompt

refine

fact check

publish

## Context Size

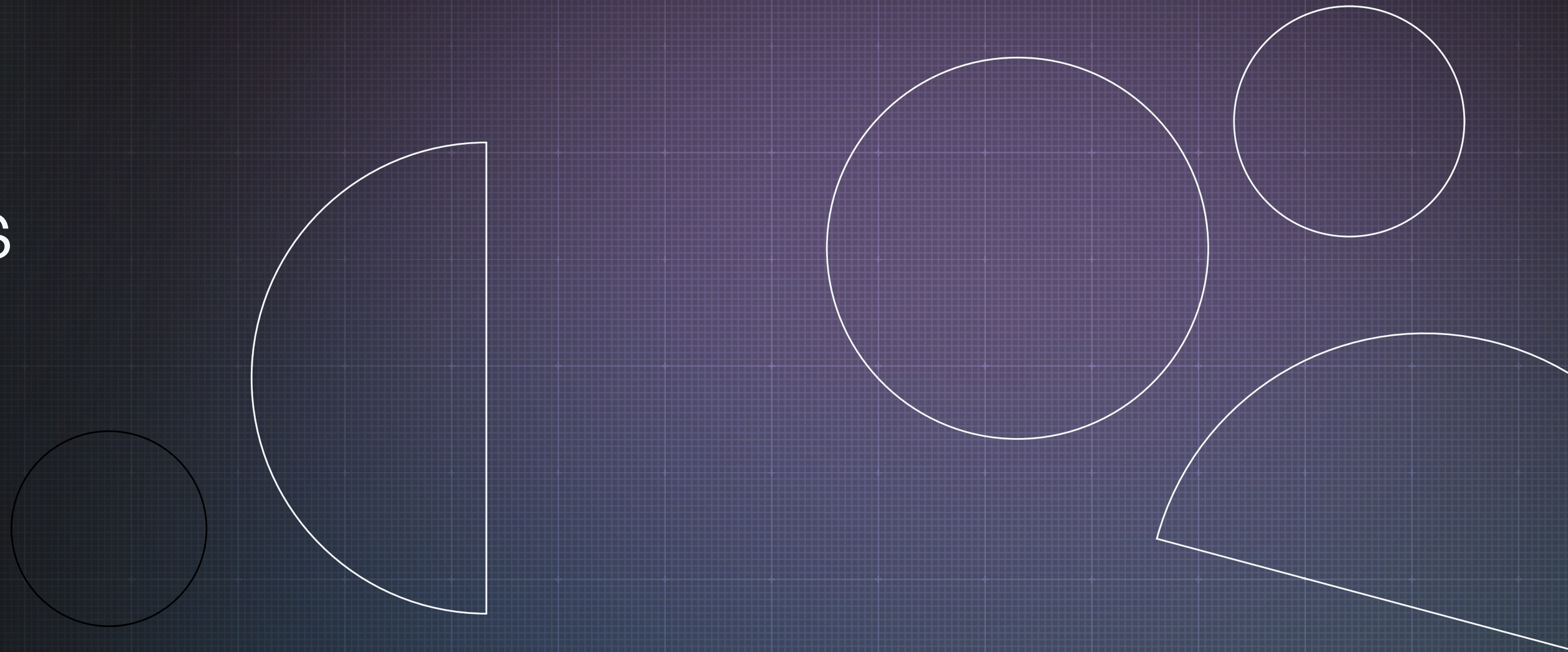
One of the key limitations of LLMs is the size of the context that can be used for generating text. For example, ChatGPT currently allows 2k tokens, which is equivalent to around 1,500 words. However, newer models like GPT4 can handle up to 32k tokens, while other models can handle up to 1 million tokens.

This has significant implications for the use of LLMs in various contexts. For example, an LLM with a larger context size could be used to analyze and summarize large amounts of legal cases or SEC filings in one go. This could save time and effort compared to doing them one by one. Similarly, an LLM with a larger context size could be used to generate longer and more detailed articles from press releases or other sources.

However, there are also limitations to using LLMs with larger context sizes. For example, they may require more computing power and memory, making them more expensive to run. Additionally, they may be more prone to generating hallucinations or other errors, particularly if the input data is noisy or ambiguous. As such, there is a need to balance the benefits of using larger context sizes with the potential risks and limitations.

GPT3.5	2k tokens
GPT4	32k tokens
Other models	up to 1 million tokens

# Current players



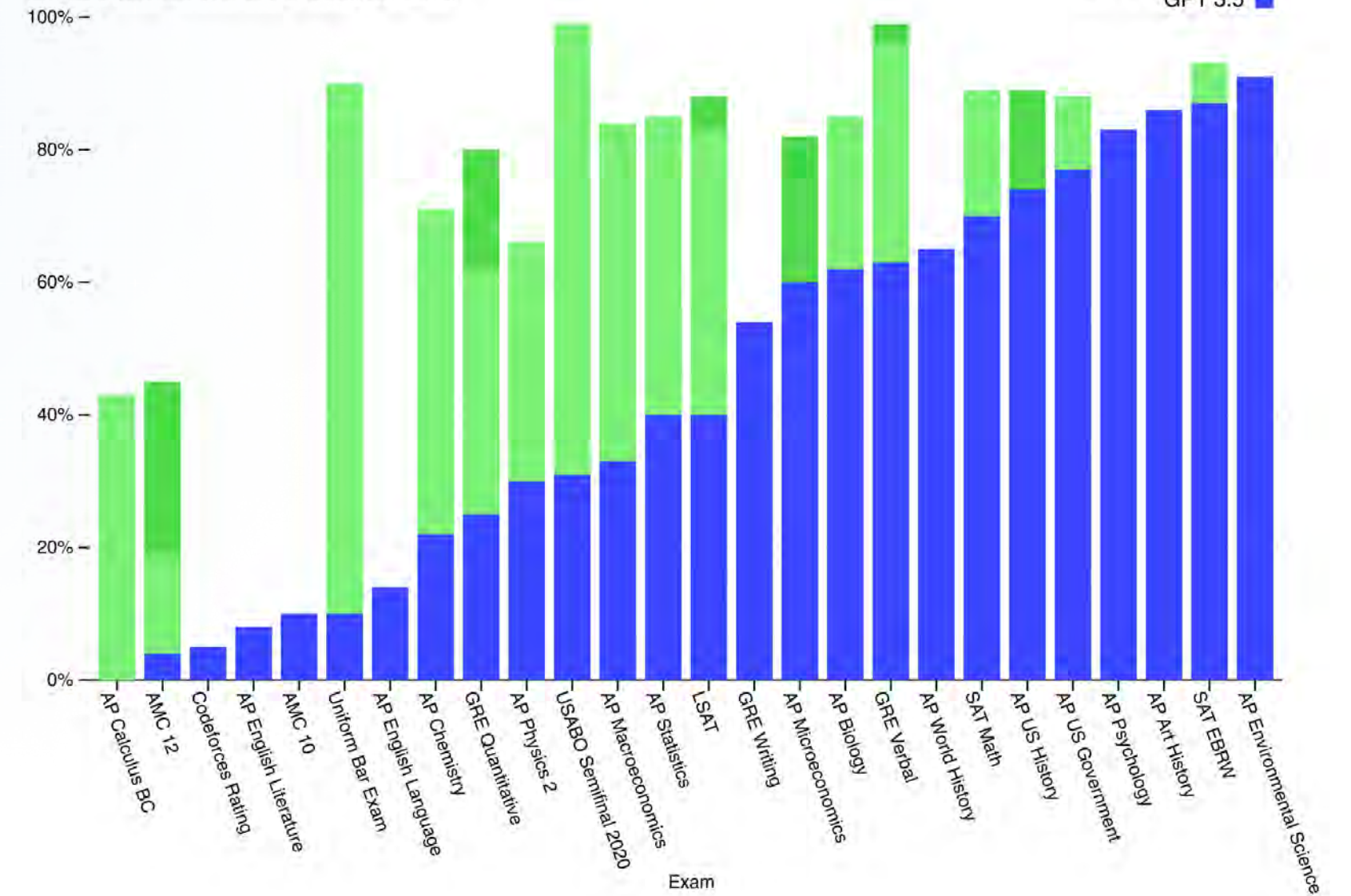
## OpenAI

Open AI is by far the biggest player in the current LLMs market. It offers two services (one paid and one free) GPT3.5 and GPT4. Both are similar in performance for a lot of text based tasks; but GPT4 is much better at complex reasoning and technical/scientific “specialist” knowledge.

You can see from the side-by-side comparison that GPT3.5 performs well on many humanity based tests, but GPT4 is vastly better at technical fields (legal, economics, STEM). It’s also better at verbal reasoning. GTP4 can output 4k tokens (3,000 words) which makes it good for use in the publishing industry.

Exam results (ordered by GPT 3.5 performance)

Estimated percentile lower bound (among test takers)



## Google

Bard and PaLM2 are Google's latest LLMs. The current perception is that their output is roughly equal to the quality of GPT3.5. Both tools have limited public availability, especially to use in an automated way. It's also important to note that their token input and output is limited to 8k and 1k respectively, which means they will struggle on the creation of longer form articles. This market position may change soon though as Google looks to throw their vast resources at improving their offering.

## Anthropic

A new startup, their main LLM is called Claude2, which offers token sizes up to 1million. Limited availability at this time.

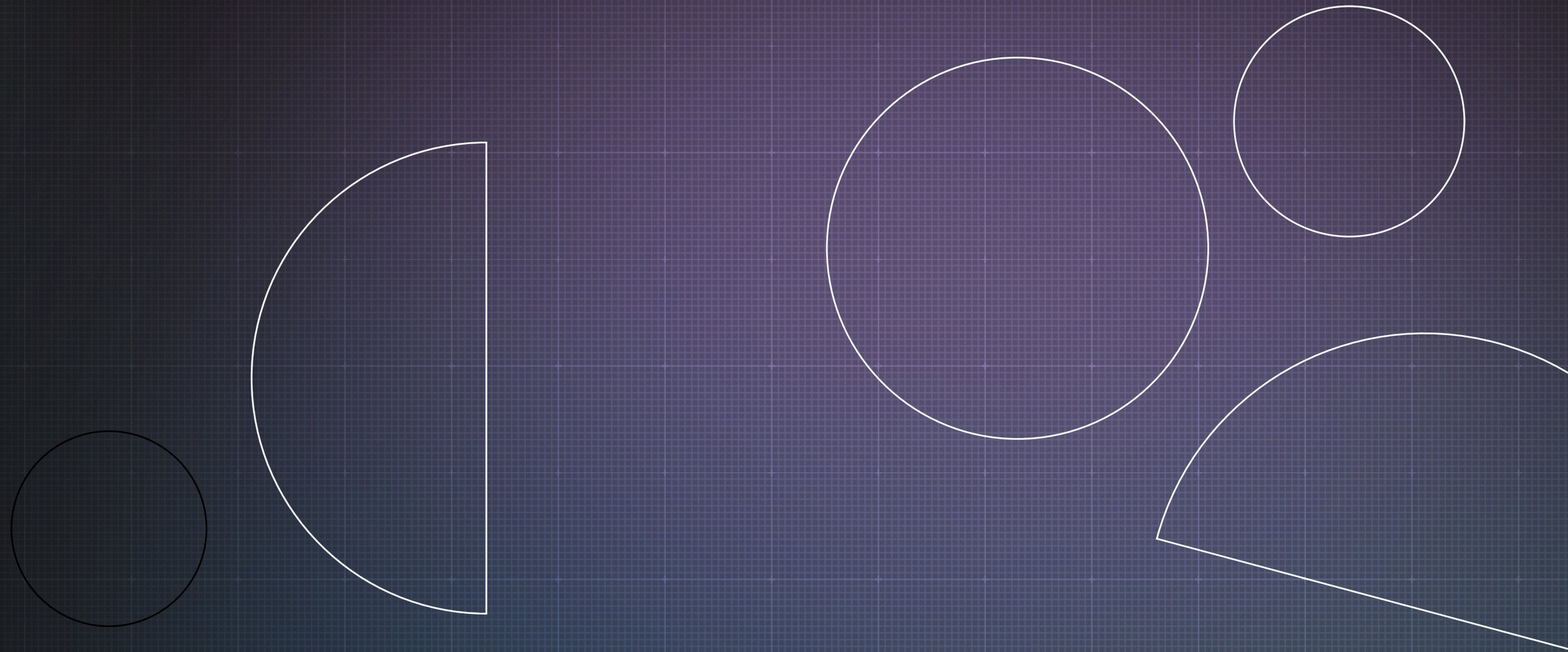
## Open source LLMs

We are also seeing a new trend towards open source LLMs which you can run on your own computer/server. This solves the problem of sending potentially sensitive data to one of the other private providers. This market has seen rapid growth and huge improvements in some of the products available, most notably with the Falcon model performing close to GPT3.5.

We expect the open source market to continue to develop rapidly, especially in industries with sensitive data.

The key thing to note is that currently it is hard to setup and run an open source LLMs without technical knowledge and it requires extremely expensive computer hardware to run for the larger models. We expect this to change as they become more efficient.

# Key takeaways





- You can use generative AI now in your organisation and a 1/3rd of the organisations we surveyed are already employing it.
- There is no reason that you can't be using generative AI at least as a 'topic expert' in your organisation. Think of this as an extra research assistant that your staff can go to when an unfamiliar topic crops up and they need quick info on it.
- OpenAI is the best provider in the market right now and has the lowest barriers to entry to use.
- Ensure you are ready to leverage the power of LLMs by using a headless CMS that will allow you to "slot" in AI features as and when you need to. Monolithic legacy CMSs are extremely badly positioned for this trend; as they require your vendor to implement support for all this and are not as modular.
- Big opportunities are coming, with increased context windows allowing much more data to be analysed at once, and GPT4 allows interesting statistical analysis. Expect this to continue to improve.
- Open source LLMs potentially solve a lot of regulatory issues. If this is holding you back from using generative AI in your organisation, it is imperative that your technical leadership team are on top of these developments as the performance of the open source solutions have recently reached a tipping point of usability.



# About Baytree

Today's digital publishers and news organisations need a partner that understands their unique market and their reach and revenue focused business models. Navigating the complexity of customer and business demands in this environment is no small feat. Publishers must adapt to the ever-evolving needs of customers while creating real growth. You need the right digital strategy and implementation to achieve this.

At Baytree, we bring together product managers, designers, data scientists, architects, and engineers so that you can harness this opportunity and solve your biggest challenges.

Want our help or simply fancy a chat about the content of this guide?

Email Tom at [toma@hellobaytree.com](mailto:toma@hellobaytree.com)