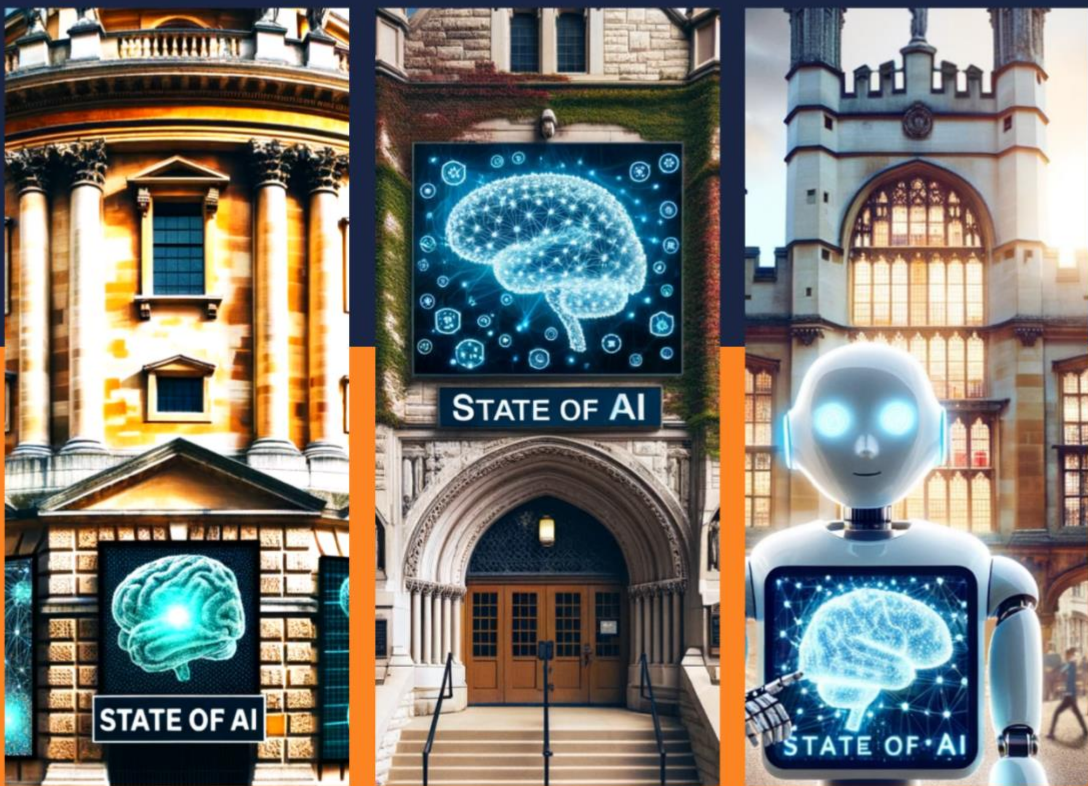
Centre for Teaching and Learning

UNIVERSITY OF OXFORD

Beyond ChatGPT



State of AI in Academic Practice

October 2023

Beyond ChatGPT: The state of generative AI in academic practice for autumn 2023

It has only been 10 months since ChatGPT was announced. The initial response was so overwhelming that the CTL published an overview of four early lessons from ChatGPT only two months later at the end of January 2023. Even two months in, there were many lessons to learn.

Since those early lessons, we have seen a number of developments that have given us more clarity but also opened up even more questions. For many people, generative AI has become synonymous with ChatGPT. While ChatGPT is still an important driver of many developments, the field has grown beyond a single tool.

This report seeks to provide an overview of the key developments in generative Artificial Intelligence (AI) relevant to academic practice that extend beyond ChatGPT.

Generative AI is a fast-moving field with new consequential developments being announced regularly. This report reflects announcements made until 11 October 2023.

Five key lessons

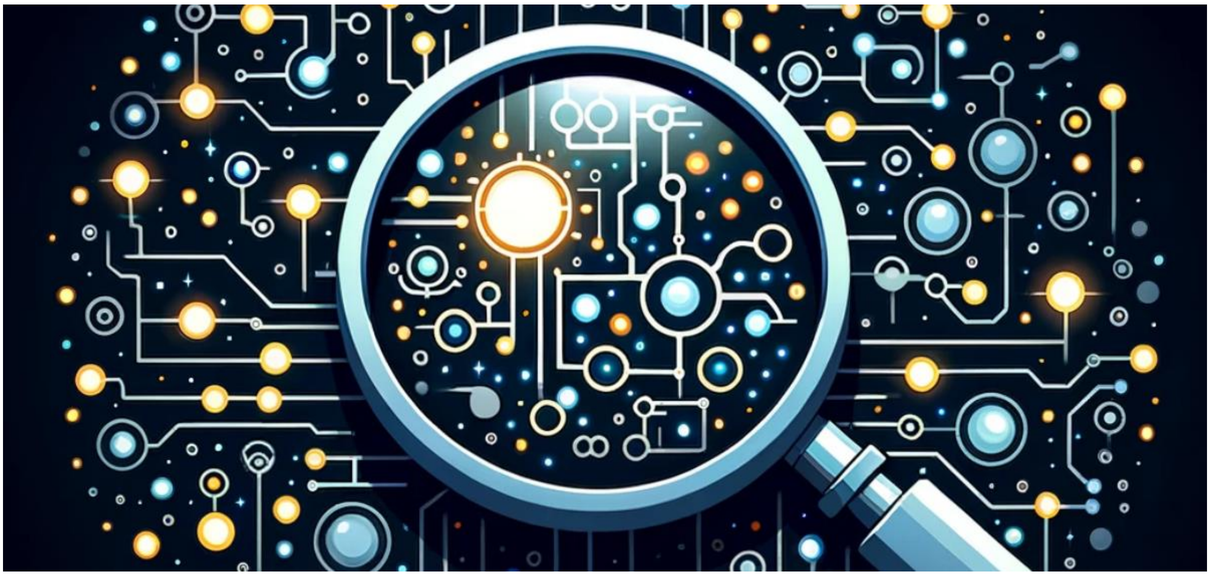


Image generated by DALL-E

This report tries to be as comprehensive as possible. However, many of the key lessons can be summarised in five points:

1. Generative AI is here to stay and it already makes a real difference to many of its users. Despite being new, people are already using it to make real and meaningful contributions to assist with their work. Generative AI capabilities overlap with the tasks performed as part of academic work to a degree unmatched by any previous technology. Most higher education institutions and the sector as a whole have already recognised this and are advising for incorporation and against blanket bans.
2. There is more to generative AI than ChatGPT. There are now at least four credible tools that can and are being used: ChatGPT, Claude, Bard, Bing, as well as many more apps that are built on the models from these companies. Although in the university context, most attention is devoted to models that generate text, we are also seeing great advances in image generation and audio generation.
3. The fundamental capabilities of generative AI are still being explored, but we now know that it can be used even more widely than initially expected. The changes in scope come partly from technological advancements of the underlying generative models, partly from more effective implementation of existing models and partly from increased knowledge of how to use them.
4. Hallucination, producing plausible but non-existent facts, is still one of the defining features of generative AI and it has proved resistant to efforts at mitigation. Despite often producing accurate and factual responses to prompts, generative AI can easily switch to ‘hallucination’

without any indication it has done so. Despite improvements in technology and ongoing efforts at reducing hallucination, the output of generative AI remains unpredictable and needs to be regarded as a first draft to be checked or a hypothesis to be confirmed.

5. Because of its probabilistic nature, generative AI produces different outputs to the same prompts every time. This makes it different from traditional software and goes against the many expectations users bring to it. A single use does not give a good indication of its capabilities because its performance differs radically on tasks that appear very similar to people used to traditional software. Even knowledge about how generative AI works is not sufficient to be able to predict its behaviour. As a result, generative AI capabilities have to be discovered empirically and over time.

Developments in AI since February 2023

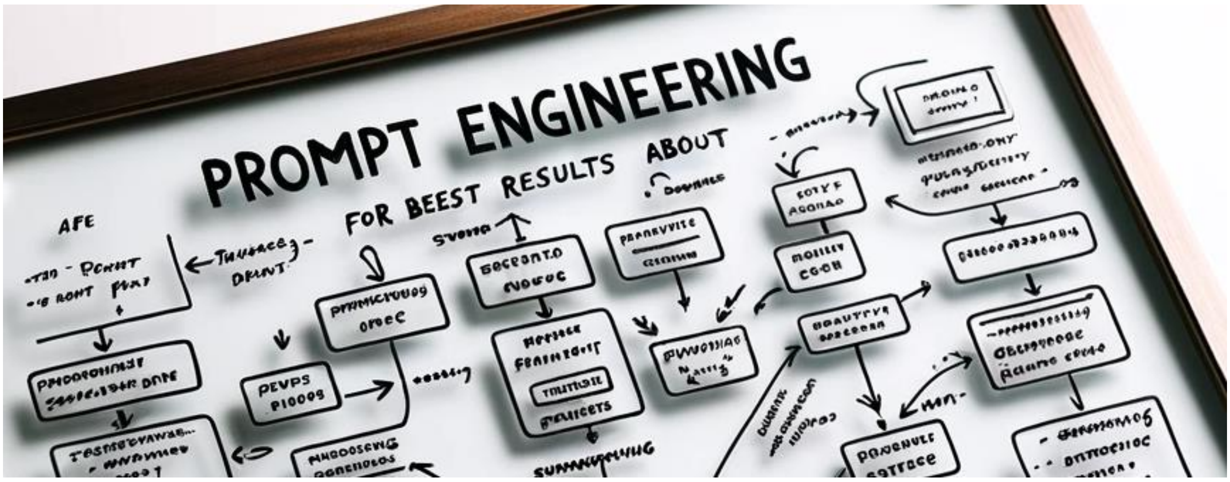


Image generated by DALL-E 3

There have been many developments since we published Four lessons from ChatGPT: Challenges and opportunities for educators in late January 2023. We can divide these into four broad categories:

1. Developments in knowledge about how to best use AI
2. Developments in practical applications
3. Developments in AI technology
4. Developments in in Higher Education sector response

1. Developments in knowledge about how to best use AI

When OpenAI released ChatGPT at the end of November 2022, they positioned it not as a product but as a ‘research preview’ with the intention to discover different potential uses and ‘strengths and weaknesses’. However, within a few weeks, it became clear that the scope of potential uses of ChatGPT far exceeded what OpenAI could imagine.

Since then, we have collectively learned many lessons about what ChatGPT and other generative AI tools based on Large Language Models can and cannot be used for.

The lessons fall broadly into three categories:

1. Discovering potential uses
2. Formulating prompts
3. Managing hallucinations and other limitations.

1.1 Discovering potential uses

The biggest overall lesson is that generative AI can have many creative uses beyond simply generating text in response to prompts in a chat interface.

Some of the creative uses that have emerged include:

- Extracting information from text (for instance, finding people mentioned in text)
- Presenting information in tables and/or structured lists
- Asking it to generate and/or correct computer code in a variety of languages
- Generating multiple choice questions about text
- Translating between languages and/or genres
- Supporting idea generation through outlines
- Explaining and analysing abstract concepts including metaphors
- Answering questions about well-known facts (including the contents of some books)
- Writing and translating poetry in different styles
- Interpreting images (only recent multimodal models).

However, any of these uses come with limitations that are not always immediately obvious. Generative AI is radically different from other tools and users cannot rely on existing mental models and intuitions about it.

That is why much of the effort in expanding our knowledge about generative AI focused on developing a better understanding of where Large Language Models (LLMs) can fail as well as where they succeed. Given the dynamic nature of this research, it is impossible to reference a single systematic review.

1.2 Formulating prompts (prompt engineering)

One of the early lessons of generative AI that preceded ChatGPT was that its output depends on how it is prompted. Formulating the prompts for best results became known as 'prompt engineering'. One of the advances of ChatGPT was that it needed much less prompt engineering than previous models.

Unlike previous interfaces, anybody who could type into a ChatGPT prompt was able to get a plausible answer. However, it soon became clear that the phrasing of the prompt still matters in ways that will not be obvious to the user. And because ChatGPT generates a different answer for the exact same prompt, simply using it will not provide a user with sufficient feedback about which phrasing works better.

Many of the things that users of computers learned to pay attention to, such as exact phrasing, punctuation or spelling accuracy, do not have a large impact on the quality of output from ChatGPT and other generative AI tools.

What matters more is directing the model into the appropriate 'conceptual space'. The key lesson learned was that Large Language Models (LLMs) are highly contextual and the more context they receive, the better the outcome. Some examples of techniques that take advantage of this feature are:

1. **Chain of thought:** Asking the model to 'think about the problem step by step' will often increase accuracy on problem solving and reasoning tasks.
2. **Personas and context:** Telling the model to answer as 'an expert in field X who is good at explaining things to Y' often produces more relevant and accurate answers.
3. **Examples:** Giving the model an example to follow (both in style and structure). The output is better even if the model is first asked to generate the example.
4. **Self-correction:** Asking the model to check its own answer will often (but not always) find problems and leads to self-correction.

Many of these techniques were identified through systematic evaluation and are not something an individual user is likely to discover for themselves. Guides are now available that collate the key insights. For example, Prompt Engineering Guide and Learn Prompting. There are also several online courses focusing on prompt engineering, such as Prompt Engineering for ChatGPT.

1.3 Managing hallucinations and other limitations

There are many limitations of Large Language Models that can be divided into five categories:

1. Hallucinations and accuracy issues
2. Probabilistic nature of LLMs
3. Lack of internal logic
4. Context length
5. Knowledge cut off and retrieval augmented generation.

1.3.1 Hallucinations and accuracy issues

The original announcement of ChatGPT included the warning: 'ChatGPT sometimes writes plausible-sounding but incorrect or nonsensical answers.' This has come to be known as 'hallucination'.

Large Language Models (LLMs) are trained to generate text that is likely to occur. As a result, they will often generate plausible but entirely fictional facts such as:

1. Non-existent links
2. Numbers (particularly larger numbers)
3. Non-existent titles of books or papers
4. Made up biographical details about known people
5. Facts about what is in a text it is presented that are not there
6. Information about their internal processes that do not exist.

Unfortunately, these 'plausible but incorrect' statements will often be seamlessly embedded within perfectly factual statements.

Hallucinations have proved to be extremely resilient to technical improvements and none of the new and updated models were able to eliminate them.

Even implementing traditional algorithmic techniques to discover factual errors only yields limited results. For example, Google's Bard's 'implicit code execution' feature referenced above still misses many issues. Equally, ChatGPT Plus can access WolframAlpha through a plugin, but it can still 'hallucinate' facts based on the input it receives.

Using some of the prompt engineering techniques above will reduce hallucinations. This has proved very useful to products who preformulate prompts for their users at scale. But any individual prompt may still generate incorrect information.

1.3.2 Probabilistic nature of LLMs

Large Language Models (LLMs) work by predicting the next token (word or part of a word). In order for LLMs to work at all, the token chosen has to be somewhat randomised. If the most likely token were to be chosen at every step, the output would be circular and unusable.

As a consequence, identical prompts result in different outputs (sometimes significantly different). This makes it much more difficult for users to discover reliable prompting techniques and very difficult to give precise instructions for others to follow.

It is also easy for first-time or even frequent users to be left with inaccurate impressions of what the tool can and cannot do (see section on “Limits to AI use in practice: Jagged frontier” for more details).

1.3.3 Lack of internal logic or access to internal data

Large Language Models (LLMs) work by predicting the next token (word or part of word) in a step-by-step fashion. Yet, their output often gives the impression of structured reasoning and information retrieval. This is made worse by models hallucinating in response to queries about how they arrived at their result. This often confuses users as to what is actually happening.

The question of whether LLMs can reason has been subject of a vigorous debate over the last few years. The limits of their ability to reason are also an active research subject for the AI community.

What is not in question, however, is whether the models can produce outputs that would require reasoning by humans. It is often helpful to interact with the model as if it was able to reason and had access to databases, but it is equally necessary to evaluate the outputs with great caution.

Some of the lessons important for users to keep in mind are:

1. LLMs **do not have access to a database of information** or their own training data despite generating factual texts.
2. LLMs **cannot reliably compute complicated calculations** directly despite being able to write computer code that can.
3. LLMs **do not search and copy information** from a text even if they produce a summary with individual facts (in fact, they may randomly hallucinate one item among many correct ones).
4. LLMs have **no ability to self-correct** without being prompted to do so.
5. LLMs can often generate text with factual information but **cannot complete related tasks** that require that information unless specifically prompted to use it.
6. LLMs produce seemingly perfectly grammatical text in English and to variable extent in many other languages, yet they produce very **unreliable metalanguage** (such as incorrectly labelling parts of speech across languages they can produce text in).
7. LLM **performance in one topic or area may not translate** into seemingly closely related tasks.

Because of these subtle discrepancies between the surface output and internal processes, it is easy for users to overly rely on LLMs even in situations where it is no longer warranted.

1.3.4 Context length

The final limitation of LLMs that users have to contend with is the length of the input that tools such as ChatGPT, Claude or Bard can use to generate their responses.

As an LLM is generating individual tokens (words or parts of words), it uses all the preceding context of the entire chat to compute their probabilities. This includes all the tokens the tool itself generated. There is a limit in how many tokens each LLM can take into account. This is called the **context window**.

The context window of most LLMs available to users is **quite short (4,000-8,000 tokens)**. This translates to about **3,000-6,000 words for English text** but it could be much less for other languages and computer code. (Languages using non-Latin characters may use up to 15 times as many tokens for the same text.)

As the interaction proceeds, **it is possible to exceed the context window**. This can happen very quickly if the user pastes in a long text to the prompt. Since every word generated is also a part of the context, the LLM may soon lose access to the early parts of the text and it may start hallucinating facts. No mainstream tools give any indication that a context window has been exceeded. Some tools based on LLMs get around this limitation by summarising text outside of the context window and injecting it back into the chat. However, this is mostly not in any way visible to the user.

These are the published context windows of the three main chatbots:

1. **ChatGPT (free using GPT 3.5):** 4,000 tokens (longer context windows available via API)
2. **ChatGPT Plus (paid using GPT-4):** 8,000 tokens (longer context windows available via API)
3. **Claude:** 100,000 tokens
4. **Bard and Bing:** Do not publish their context windows.

Anthropic's Claude stands out with its extremely large context window which makes it the only tool suitable to summarise the text of an entire academic paper.

Note: A longer context window reduces the likelihood of hallucination about the text, but it does not eliminate it. The impacts of increasing the context window length are a subject of continuing research as well as innovations in the industry.

1.3.5 Knowledge cut off and retrieval augmented generation

Most Large Language Models also publish a “knowledge cut off” or the last date covered by the training data. This means that the model will not be able to reflect any events or developments after that date. Most models have been fine-tuned to make this clear to users who ask questions about current events.

ChatGPT: On release, OpenAI’s ChatGPT only included training data up to June 2021 and latest models now cover data up to September 2021. OpenAI’s models are also used by Microsoft.

Claude: Anthropic’s latest model Claude 2 only specifies training data “until early 2023”.

Bard: Google does not publish details about the implementation of Bard but specifies that the underlying PaLM 2 model used training data up to “mid-2021”.

Note: It is important to reiterate that despite the commonly used term “knowledge cut off”, the LLM does not actually have direct access to its training data and may still hallucinate facts (see above).

To facilitate access to knowledge outside of the model’s training set, a method called Retrieval Augmented Generation (RAG) is commonly deployed. Using this technique, a platform uses aspects of the LLM to match the user’s query to a pre-processed document or a part(s) of the document. It then sends the query accompanied by the matched text back to the LLM as a prompt.

This is approach is used by products that claim to allow individuals or organisations to “ask questions about their own documents”. A version of this technique is also used by chatbots such as Bing Chat with search capabilities.

It is sometimes erroneously reported that connecting an LLM to search removes the “knowledge cut off” but in fact, the underlying model remains the same.

Also, while RAG can in some cases be an effective way to reduce hallucinations it does not eliminate them. For example, even cursory testing revealed that Microsoft’s Bing Chat can report facts not on the web pages it finds. Equally, Google Bard’s integration with Documents in Google Drive has shown examples of hallucinated content.

2. Developments in practical applications

In addition to advances in knowledge, we are seeing developments in actual uses of ChatGPT and other tools in industry and in academic practice. We are seeing an increasing number of reports of academics using Large Language Models (LLMs) in both their teaching and research. We have also seen the first systematic evaluations of the suitability of LLMs to specific disciplines and uses within those disciplines.

Impact of LLMs in practice

Aside from informal personal accounts, there have now been numerous studies that show that knowledge workers (including students) in a number of areas have seen great productivity increases.

Studies that investigated different areas of practice have found similar outcomes in fields as diverse as consulting, customer support, business document writing, and coding.

There are some lessons that seem to hold across a number of these studies.

1. Using generative AI almost uniformly increases the productivity of knowledge workers in a variety of contexts.
2. Using generative AI tools particularly narrows the gap between high performing and low-performing workers.
3. Expert performers are often not seeing the same level of benefit and may occasionally see a performance decrease.
4. In some instances, using AI uncritically may lower quality.

AI performance on benchmarks and standardised assessments

One of the more consequential developments has been the high performance of the new models on standard examinations across many disciplines including law, architecture and medicine. Until recently, these results excluded non-textual questions, but the new multimodal models made available in September 2023 are showing great promise in image recognition and may extend AI performance further.

We have also seen some initial investigations into AI performance in specific academic disciplines. This systematic investigation revealed high performance in a variety of data science tasks. A similar analysis revealed AI outperforming average students on a variety of legal tasks, while also finding performance decline in top students.

It is important to note that performance on exam questions does not necessarily imply that LLMs can perform independently or replace professionals in the respective fields. However, they have implications for examinations in various context as well as broader academic practice outside purely instructional contexts. For instance, a recent study showed a high level of overlap between peer reviews conducted by AI and human peer reviewers (the variance of AI from human peer reviewers was equivalent to the variance between peer reviewers themselves).

These and other developments led QAA to release an advice paper on Reconsidering assessment for the ChatGPT era. In it they propose reviewing assessment strategies and suggesting that 'Three desirable outcomes of reviewing assessment strategies could be:

1. **Reducing the volume of assessment** by removing items that are susceptible to misuse of Generative Artificial Intelligence tools to generate unauthorised outputs and repurposing the time available for other pedagogical activities.
2. Promoting a shift towards **greater use of synoptic assessments** that test programme level outcomes by requiring students to synthesise knowledge from different parts of the programme. Some of these may permit or incorporate the use of Generative Artificial Intelligence tools.
3. **Developing a range of authentic assessments** in which students are asked to use and apply their knowledge and competencies in real-life, often workplace related, settings. Ideally authentic assessments should have a synoptic element.'

[Emphasis added]

Identifying AI-generated content

So far, the ability to detect AI-generated content has not kept pace with developments in generative AI models. AI detectors may have some success, they do not offer nearly the level of reliability required in academic settings. Particularly, the rate of false positives is alarmingly high.

This led OpenAI to discontinue their AI classifier due to 'low level of accuracy' in July 2023. Equally, TurnitIn's AI detector was shown not to be sufficiently reliable in systematic evaluations.

OpenAI and other LLM providers are actively exploring this area with the most promising being a text-fingerprinting method that would leave detectable traces in AI-generated text. This is described in more detail in the JISC report on AI in tertiary education. However, no announcements in this direction have been made by any of the major model providers. But even these methods will be inherently probabilistic and cannot be expected to be 100% reliable in the way that pure text matching is. Two recent papers showed that both text and image watermarking have inherent limits that may prove resistant to technical developments.

Limits to AI use in practice: Jagged frontier

Despite the clear promise of generative AI in a variety of disciplines, we have yet to see actual and systematic impact across an industry. One of the reasons for this is that while generative AI is a great tool for increasing personal productivity, many of the inherent limitations discussed above make it less suitable for deployment in environments where uniformity is important.

However, even as an individual productivity tool, it is not clear how much generative AI can change the landscape across the board. This is due to what a recent paper called the 'jagged technological frontier'. The authors offer the following elaboration:

> 'We suggest that the capabilities of AI create a "jagged technological frontier" where some tasks are easily done by AI, while others, though seemingly similar in difficulty level, are outside the current capability of AI.'

Because high-performing individuals were not able to correctly anticipate whether similar tasks were within the scope of AI capabilities, their performance on those outside the frontier suffered.

This has significant implications within academic practice and perhaps the most significant lesson of AI use: It is easy to be 'lulled' into thinking 'AI can do this' and to stop paying attention to the real limitations of AI. This should be foremost in any guidance for students and academics.

3. Developments in AI technology

We have seen significant improvements in what AI can do and what can be done with AI since the introduction of ChatGPT. These improvements have happened on three levels:

1. Improvements in the underlying Large Language Models
2. Improvements in how the models are implemented as chatbots
3. Explosion in AI-based tools and features.

1. Improvements in Large Language Models

Why Large Language Models matter

Large Language Models (LLMs) are what powers products like ChatGPT. They are incredibly expensive to produce and they are only produced by large companies (or sometimes by large open source projects).

The companies that produce the most impactful LLMs at the moment are OpenAI (with investment from Microsoft), Google, Anthropic (with investment from Amazon), and Meta. All companies other than Meta also offer a chatbot based on their model. Google, OpenAI and Anthropic also offer access to their models to other companies who build AI-based products based on them.

Provider	LLM	Products
OpenAI	GPT-3.5 GPT-4	ChatGPT (also used by Microsoft in Bing Chat)
Google	PALM	Bard
Anthropic	Claude 2	Claude
Meta	Llama 2	Open source model used in research Also powers chatbots in Meta products

Improvements in LLMs since ChatGPT

All currently used LLMs have seen significant improvements that greatly expand their reach. Most relevant to academic contexts are:

1. GPT-4 – Introduced in March 2023 with greatly enhanced capabilities across all benchmarks and introduced multimodality (ability to integrate sound and images). It is available to paying customers subscribed to ChatGPT Plus or to users of Bing Chat for free.
2. Claude 2 – Introduced in July 2023 with enhanced capabilities including the ability to directly process text of 100,000 tokens (about 75,000 English words – equivalent to a short monograph)
3. Llama 2 – Powers Meta's chatbots in WhatsApp but also available to researchers and commercial vendors to integrate in other products.

Another prominent model is PaLM 2 used by Google and rumoured to be soon succeeded by a more powerful Gemini model.

Code interpretation and data analysis

Both OpenAI and Google released features that leverage the ability of LLMs to write code to add features to their tools.

Google announced that it is now adding an 'implicit code execution' feature where its Bard chatbot will write code in the background to help it answer mathematical queries with greater accuracy.

OpenAI introduced a feature for ChatGPT Plus users initially called Code Interpreter and later renamed to Advanced Data Analysis that makes this capability more explicit. For instance, it is possible to upload a structured data set and ChatGPT Plus will write code to analyse and visualise it.

Both of these features are still subject to the fundamental limitations of LLMs discussed above.

Multimodality

The most potentially impactful change in LLM capabilities is multimodality. This means the ability of LLMs to interpret images as part of the user interaction.

OpenAI announced GPT-4 as a multimodal model in spring 2023, but the multimodal features were delayed until autumn. As of early October 2023, users of ChatGPT Plus can now ask the model to interpret images and on mobile apps, they can have free flowing voice conversations.

As of writing, this feature has only been widely available for a very brief time for any systematic evaluations to become available, but early reports show impressive results. Some reported use cases have included replicating a screenshot of user-interface in code, interpreting complex images with text, explaining context required to understand the meaning of visual humour, etc.

This feature is now also available for free to users of Bing Chat.

Google Bard also offer image interpretation and has shown great promise at simple graph interpretation and text extraction. It can also often interpret the meaning in cartoons. In early informal tests, the outcomes are slightly behind ChatGPT Plus but nevertheless can be used for practical applications.

Note: While the improvement in image interpretation is impressive and has clear benefits to users, it is still limited in how well it can interpret fine details, most notably spatial configuration of objects and subtle differences in lines. Hallucination also still remains a problem in that the model can report features not present in the image.

Improvements in image generation models

Large Language Models are not the only generative models used in AI. The past year has also seen a huge increase in the capabilities and availability of image generation models (most using the diffusion methods). These models are distinct from LLMs and their development has proceeded largely independently.

DALL-E 2 was announced prior to ChatGPT and ushered in a revolution of models and products. The most prominent of these were the Open Source model Stable Diffusion (used in a number of products) and Midjourney which have both had a number of releases so far this year (Midjourney has introduced versions 4 and 5).

At the same time, image generation has become integrated into products such as Adobe Photoshop or the graphic design tool Canva.

In September 2023, OpenAI announced DALL-E 3 which will be integrated into ChatGPT Plus, meaning it will be possible to ask ChatGPT to generate images as part of a conversation with the chatbot. Microsoft who integrate OpenAI's models into Bing Chat have now made DALL-E 3 available for free in their Image Creator service.

Apart from general quality, perhaps the most noticeable improvement to DALL-E 3 is its capacity to generate text as part of images which will greatly expand its usefulness in many contexts.

2. Improvements in implementation

All four major generative AI chatbots (ChatGPT, Bard, Bing, Claude) have received significant feature updates making them much more usable for practical applications in ways that extend model improvements. For example:

1. Chat history makes it easier to review and build on previous work (this was initially missing from Bing Chat and Google Bard)
2. Chat sharing and collaboration makes it possible to share example chats and for others to continue them (now possible with all tools except Claude)
3. Integrations with other tools enables input from other sources of information not included in the model's capabilities (available as extensions in Google Bard for free and in ChatGPT Plus as plugins).

For example, Google's Bard now integrates with Google Workspace, YouTube and Collab notebooks. Bard also has the ability to invoke an implicit code interpreter to increase to accuracy of numerical responses.

Perhaps the most significant update was the introduction of Advanced Data Analysis (previously Code Interpreter) by OpenAI (available to subscribers of ChatGPT Plus). This allows users to upload files with data and ask the language model to suggest ways of analysing it. The model uses the same implicit code generation as Google but can apply it to large data sets.

3. Explosion in AI-based tools and features

The last 10 months have also witnessed an explosion of AI-based tools many of which are aimed at students and educators.

AI-first products

Sites like TheresanAIforThat.com list thousands of generative AI products (over 8,600 as at 10 October 2023). Many of these are aimed directly at researchers, educators and students. Almost all of these products are built on one of the LLMs from the big providers, although many may use additional custom models for different tasks.

Many of these products simply replicate the functionality available through the chat interface in Claude, Bard or ChatGPT but add specific features to format the output. They may also use a mixture of models for different tasks based on their own empirical testing. Some products implement more complex data and prompt workflows that would not be possible simply in the chatbot's primary interface.

Further growth of these is supported by developments in generative AI workflow frameworks such as LangChain or LlamaIndex and supported by a growth in professional development opportunities in what has been called AI engineering.

Some illustrative examples of these applications include:

- **Elicit** uses a LLM to extract information from multiple papers and presents comparisons in tabular form.
- **Consensus** searches the scientific literature and summarises aggregate results in response to user questions.
- **AudioPen** is a mobile app that takes transcribes free flowing notes and summarises them into actionable items.
- **Scispace** finds graphs and formulas in scientific papers and interprets them.
- **Teachermatic** offers a simple user interface to generate text from various tasks commonly performed by teachers.

All of these products have the same strengths and limitations as the underlying language models, and at least some of their functionality is often available to users of the primary chatbots such as Bard, Claude or ChatGPT.

AI features in established products

Although slower to appear, even more impactful is the release of many generative AI features into established products. This is likely to contribute to the blurring of boundaries between content created by AI and content created manually using traditional approaches. Some notable examples include:

1. **Google** announced generative AI features in their Workspace Suite of products and it is already available to some users of Google Docs and in Gmail.
2. **Microsoft** announced Copilot as a feature in its Office suite available initially to enterprise users only, but with roll out to consumers in the future. Copilot is now also available as preview in the latest version of Windows 11 to users in the US and UK.

3. **Meta** announced integration of generative AI bots in popular messaging app WhatsApp and AI image editing in Instagram.
4. **Notion**, the popular note taking and project management app has integrated generative AI with features including the population of tables.
5. **Grammarly** has introduced a text generation feature in addition to its traditional grammar and spelling checker. While many of these are only available in limited contexts to paying customers, we can expect further growth and expanding availability in this area.
6. Both Adobe and Canva who had released a first version of image generation features into their products earlier in the year have now announced more advanced versions.

4. Developments in Higher Education sector response

As noted in Four lessons from ChatGPT: Challenges and opportunities for educators in late January 2023, the initial response to the introduction of ChatGPT has been mostly measured and cautiously optimistic. From the start, many (if by no means all) academics advocated for incorporating generative AI tools into the educational context, including assessment, in a way that minimises their weaknesses and takes advantage of their strong points.

Responses by sector bodies

Since then, this attitude has been reflected in reports and statements by various sectoral bodies. Most comprehensive of these have been:

1. Russell Group principles on the use of generative AI tools in education
2. Reconsidering assessment for the ChatGPT era: New QAA advice published
3. Artificial intelligence (AI) in tertiary education (report by JISC National Centre for AI) AI Maturity Framework
4. Artificial Intelligence in Education: Report by University of Warwick with contributions from University of Oxford.

The five principles published by the Russell Group in June 2023 are particularly representative of the general consensus in the sector:

1. Universities will support students and staff to become AI-literate.
2. Staff should be equipped to support students to use generative AI tools effectively and appropriately in their learning experience.
3. Universities will adapt teaching and assessment to incorporate the ethical use of generative AI and support equal access.
4. Universities will ensure academic rigour and integrity is upheld.

5. Universities will work collaboratively to share best practice as the technology and its application in education evolves.

There is a clear call for developing best practices to augment what can be done using generative AI while maintaining academic integrity. However, there is not universal agreement yet on what those best practices will be. But we are seeing some institutions adopt these principles formally, for instance, King's College London have used them to frame their guidance.

The University of Warwick report on the work of cross-institutional teams (including Oxford) investigating the impact of AI on education presents some early suggestions:

- Embrace the potential of AI to innovate educational practices and create impactful learning experiences for students.
- Question assumptions about what AI can or cannot do and critically examine AI-generated feedback.
- Approach AI tools as powerful aids that augment human expertise, not replace it.
- Reflect on preconceptions about the learning process and how AI challenges existing beliefs.
- Maintain a balance and continually evaluate the effectiveness of AI tools.

The JISC National Centre for AI offers a perspective on how institutions can approach the task of integrating AI into their institutional practices through their AI Maturity Model:

Stage of AI adoption	Characteristics
Approaching and understanding	• Interested in AI • Understanding how it has impacted or is transforming other sectors.
Experimentation and pilots	• Experimentation and pilots within existing processes • Data culture to support AI emerging • AI ethics processes established.
Operational	• AI used for one or more processes across an organisation, for example, chatbots for a specific purpose or adaptive learning systems.
Embedded	• AI embedded in strategy • Data maturity allows AI to be considered for all new systems and processes.

Transformational	• AI models and systems' effectiveness monitored as part of business as usual • AI has transformed the learning and teaching experience • The tutor is free from all routine admin tasks to focus on supporting students • The student has a fully personalised learning experience.

(Table extracted from an image in the JISC report using Google Bard.)

Variability in institutional guidance

We are seeing more and more institutions releasing guidance to both staff and students on the best uses of AI. Some illustrative examples from across the globe include:

1. Guidance for students by the University of Sydney: Supporting students to use AI responsibly and productively
2. Guidance for instructors and students published by Harvard Business Publishing: Student Use Cases for AI
3. Guidance for students by University College London: Engaging with AI in your education and assessment
4. Guidance for students, staff and departments at King's College London: King's guidance on generative AI for teaching, assessment and feedback
5. Guidance for students and staff on academic integrity from University of British Columbia: ChatGPT Q&A - Academic Integrity at UBC
6. Guidance on AI for researchers by Brown University: Generative AI as a Research Tool.

What all of these examples have in common are:

1. Acknowledgement of legitimate and often beneficial use cases of generative AI in academic contexts
2. Focus on maintaining academic integrity.

However, they differ in their approach and scope across a number of dimensions:

1. **Format of the guidance**: We are seeing policy documents, position papers, courses, FAQs.
2. **Location of the guidance:** Each institution may have multiple sources of guidance. Some are provided by a centre for teaching and learning, others by libraries, and yet others by examination bodies. None of the illustrative examples above represent the only place where the institution provides guidance on AI.

3. **Target audience:** Some guidance is addressed directly to students, some to instructors suggesting how they guide students, some to researchers and other to departments.
4. **Level of detail:** The guidance also varies in how much depth it offers users. Some give specific suggestions for uses, others even include prompts and mention specific tools, others only answer more general questions, and provide definitions.

Given how new and rapidly developing this field is, it is not surprising that there is also not a perfect agreement on terminology, definitions or even how to address some knotted issues arising from integrating AI into academic practice. Some of the unsettled issues are:

1. **How to acknowledge the use of AI tools:** Some guides prefer detailed acknowledgement, others traditional citation, some simply describe the issue.
2. **How to refer to AI:** Some guides always refer to generative AI (some shortening it to genAI), others focus on specific tools most often ChatGPT. Some use the term hallucinations consistently; others simply list possible issues.
3. **How to describe how generative AI works and its limitations:** Some guides go into some level of technical detail, using some of the technical terminology and definitions. Others focus more on practical uses sometimes even supporting these with examples.

Individual perceptions and responses

Beneath this variability in institutional responses, lies an even greater diversity of approaches and attitudes by individual academics and students. A large number of academics have not personally engaged with generative AI. A survey by de Greuter's Insights conducted between 15 June and 26 July found that from among 748 academics from 82 countries, 39% had not yet used ChatGPT at all and only 14% use it at least weekly for academic work.

Given the global reach of the respondents, unsurprisingly, the most common use case for ChatGPT was translation (47%) and "correcting text" (52%). Other popular uses were "searching for meanings and definitions" (41%) and "clarify/simplify complex concepts" (39%). ChatGPT was used for writing text by 36% and writing code by 25% of the 289 respondents who reported using it.

Respondents in the survey also expressed a range of concerns about ChatGPT that ranged from misuse by students to reliability of the output. While some scholars expressed complete scepticism of the technology, others reported already using it in their teaching.

Conclusions and future prospects

Image generated by DALL-E 3

There is no doubt that generative AI and Large Language Models will continue to play an increasingly significant role in all academic contexts. Their quality and reliability are likely to increase even if the limits on their performance are not currently known.

The trend we are seeing of more AI tools and more AI-features in existing tools is also set to continue. But as with any new field, we are also certain that many tools will disappear as companies creating them fail or merge.

Many of these developments will be straightforwardly positive for increasing access and reducing effort for many repetitive tasks. Others will present challenges for many of the academic tasks that have been so far outside the possibility of automation.

This will mean having to think more deeply about how to ensure academic integrity in the face of ubiquitous content generation tools. We will have to continue to search for a balance between developing students' and our own skills to critically engage with academic content and preparing us for a world in which AI will play an increasing role in generating that content.

We are at the start of an exciting and challenging journey. This is not the first time the world of education and academic practice has faced a considerable challenge from technology, but the speed of adoption of generative AI is in many ways unprecedented.

Unlike many previous technologies, AI has fulfilled its early promise and there are very few voices advocating banning or ignoring it entirely. But its deeper long-term impact on Higher Education and academia is still uncertain. We hope that this overview of the current state can provide a starting point for exploration.

Appendices

Key AI terms

The following is a summary of key terms used in this report.

Key AI terms	Explanation
Generative AI	A type of artificial intelligence that can generate and understand text. This is different from other types of ‘analytical’ artificial intelligence, such as that used to optimise shopping recommendations or label images.
Large Language Model (LLM)	Underlying technology that gives generative artificial intelligence tools its capabilities. ChatGPT is an example of a product based on a Large Language Model (called GPT). Most generative AI products use one of a few LLMs.
Prompt	The text that is sent to an LLM to generate a response. This could be a question or a document with a question. Newly, some systems allow document attachments or images as part of the prompt.
Prompt engineering	The craft of formulating prompts in such a way that the LLM responds in an appropriate way. This can include phrasing prompts in a certain way or including templates. Prompt engineering is used both by users of chatbots such as ChatGPT, but also makers of products based on LLMs to generate the appropriate response.
Context	Everything the LLM uses to generate more text. This includes the prompt or any text the LLM has already generated as part of its response up to that point.
Context window	The maximum length of the context that the LLM can use to generate its response. Most LLMs have relatively short context windows (about 4,000-8,000 tokens with the exception of Anthropic’s Claude that has 100,000). The context window is reset in chatbots by starting a new chat. LLM-based AI products may breakdown documents into smaller chunks to get around the context window limitations.

Tokens	The basic units of text that an LLM works with. Tokens can be words, punctuation marks, or even parts of words. Any prompt sent to the LLM is first broken down into tokens and the LLM generates its response one token at a time based on the tokens that came before. Similar length text will use radically different amounts of tokens in different languages.
Hallucination	Commonly accepted term for instances where the LLM generates plausibly looking but false information. This can include numbers, references, links, biographical details, non-existent quotes. There is no way to differentiate between hallucinated facts and those that are not hallucinated. Hallucination is usually not a result of incorrect information in the training data, it is rather the result of the probabilistic nature of LLMs. More advanced models, prompt engineering techniques and implicit code execution can be used to reduce hallucination, but so far there has been no solution to eliminate it entirely.
Knowledge cut off	Many tools report a knowledge cut off which refers to last update of the training data set. This means that questions about events since that cut off date will not be reflected. Some tools, such as Bard, Bing, ChatGPT Plus can pull in information using search. However, this data is simply summarised using the model and not reflected in the model training itself. The use depends on the tools search capacity and ability summarise the information accurately.

Main AI tool comparison

This table compares the four main chatbots as of October 2023.

Feature	ChatGPT	ChatGPT Plus	Bing Chat	Claude	Bard
Company	OpenAI	OpenAI	Microsoft	Anthropic	Google
Pricing	Free	Paid	Free	Free (with paid tier)	Free
Model Version	GPT-3.5	GPT-4	GPT-3 and 4	Claude 2	PaLM 2
Context Window	4k	8k	Unknown	100k	Unknown
Mobile App	Yes	Yes	Yes	Browser only	Browser only
Multilingual	Yes	Yes	Yes	Yes	40 languages
Search Integration	No	Yes	Yes	No	Yes
Voice Interaction	No	In app	Limited	No	Limited
Image Generation	No	Yes	Yes	No	No
Image Interpretation	No	Yes	Yes	No	Yes
Plugins	No	Yes	No	No	Yes
File Uploads	No	Yes	Image only	Text only	Image only
Code Generation	Yes	Yes	Yes	Limited	Yes
Chat Sharing	Yes	Yes	Yes	No	Yes
Other features	Custom instructions	Custom instructions	Built into MS Edge and Windows 11 Modes Style options	None	Response checking Docs / sheets exports Draft suggestions

Notes on individual chatbots

Here are additional notes on the individual tools.

ChatGPT

- The free version of ChatGPT can be used successfully for most text-based tasks. It is a good starting place for learning about generative AI functionalities, but it is also sufficient in most situations.
- The free version uses the original model with which ChatGPT was first released.
- However, it is worth exploring free and paid alternatives for certain tasks.

ChatGPT Plus

- The paid version of ChatGPT (currently US$20 per month) is the most advanced generative AI tool available. It uses the most advanced LLM currently available – GPT-4 (with usage limits).
- Its most advanced feature is Advanced Data Analysis that can generate code to interpret and visualise data sets.
- It also includes image generation and interpretation.

Bing Chat (Microsoft)

- Bing Chat is the best way to access the most advanced LLM (GPT-4) for free. It works best in Microsoft Edge (free browser).
- Bing Chat offers three modes (Creative, Normal, Precise). Creative mode uses GPT-4.
- Bing Chat integrates results of Bing search into its responses but these are still subject to hallucination.
- Despite using GPT-4, Bing often gives shorter and more limited responses than ChatGPT.

Claude (Anthropic)

- Claude is the only tool that can interpret very large documents (up to about 70,000 words of English text) without breaking them up into parts.
- It allows uploads of PDFs, Word documents and text files.
- It does not have a mobile app but it works fully in mobile browsers (including file uploads).
- Claude is not best at computer code and it has no ability to interpret images.
- Anthropic recently introduced a paid tier, but as of October 2023, it is not differentiated by features, only usage limits.

Bard (Google)

- Bard generates three draft responses to choose from. This makes it very useful for learning about the variability of responses.
- Bard is very useful for free image interpretation on par with ChatGPT.
- Bard has a feature to check responses for quality.
- It can also be used to generate code and continue working on it later.

Note: Bard uses PaLM 2 model which is inferior to GPT-4, but Google is widely expected to upgrade to its latest Gemini model soon.

Other generative AI tools

In addition to ChatGPT, Claude, Bing Chat and Bard covered in detail above, the following tools were mentioned in the report.

Tool	Short Description
Adobe Firefly	Adobe's image generation tool (free with limits).
Audiopen.ai	Mobile app that transcribes user's voice notes and creates an AI generated summary.
Consensus.app	Users can ask general questions to receive a summary of the research consensus.
Elicit.org	Uses LLMs to extract data from multiple academic papers.
LangChain	Development framework for building systems with Large Language Models.
Microsoft Designer	Free AI design tool by Microsoft, used to create layout based on text instructions with the ability to generate images using DALL-E.
Midjourney	Independent image generation tool (paid).
Scispace.io	One of many tools that lets users ask questions about PDFs. Also offers interpretation of graphs and tables.
Teachermatic.com	Uses LLMs to generate drafts of commonly used teaching materials such as learning outcomes or lesson plans.
There's An AI For That	Online resource that compiles a list of generative AI tools across a number of categories.

Other developments outside the scope of this report

Whilst this report is an attempt to provide a comprehensive overview of the state of generative AI as at October 2023, it has been impossible to encompass all aspects of AI debates.

Perhaps the most significant omissions are:

1. Discussions of existential risk around AI developments
2. Discussions about AI ethics and bias outside of questions of academic integrity
3. Developments in speech generation and voice cloning
4. Developments in personality-based chatbots such as Character.ai.
5. Developments in AI not driven by Large Language and Image Generation models such as robotics and self-driving automobiles
6. Developments in hardware for generative AI
7. Developments in tools for AI implementation and AI engineering (see The State of AI for August 2023: 6 trends from the last 3 months and 6 actions you can take today for more details)
8. Developments in open source models (see The State of AI for August 2023: 6 trends from the last 3 months and 6 actions you can take today for more details).

Example: AI generated overview of developments and conclusions

As an example of a practical application, this overview was fully generated by Claude.ai, a generative tool from Anthropic that can take longer texts as input.

Using the full text of the report and the prompt, 'make an outline of the key points by section.'

1. Developments in practical applications
    - Studies show AI increases productivity in many domains
    - AI can outperform humans on some academic tasks
    - Uneven capabilities revealed in practice ('jagged frontier')
    - Reliable AI detection still a challenge.
2. Developments in AI technology
    - Improvements in large language models expand capabilities
    - Better implementation makes models more usable
    - Explosion of new AI-based tools and features.
3. Developments in knowledge about AI use
    - Discovering creative potential applications of AI
    - Importance of prompt engineering to guide AI
    - Managing inherent limitations like hallucination
4. Developments in Higher Education sectoral response
    - Most institutions recognise potential of AI and advise incorporation, not bans
    - Sector bodies released principles and frameworks for ethical AI use
    - Key is developing best practices to augment human capabilities
5. Conclusions and future prospects
    - AI adoption will continue rapidly, long-term impacts uncertain
    - Developing best practices for ethical AI use remains critical
    - Important to balance developing critical thinking and AI literacy

References and key links

This is a categorised list of the links and references mentioned in this report.

Sector reports and policy statements

Abbas, Ramlah, and Alexandra Hinz. 2023. 'Cautious but Curious: AI Adoption Trends Among Scholars'. https://blog.degruyter.com/chatgpt-in-academia-how-scholars-integrate-artificial-intelligence-into-their-daily-work/.

Centre for Teaching and Learning. 2023. 'Four Lessons from ChatGPT: Challenges and Opportunities for Educators'. https://www.ctl.ox.ac.uk/article/four-lessons-from-chatgpt-challenges-and-opportunities-for-educators.

Fischer, Isabel et al. 2023. 'Transforming Higher Education: How We Can Harness AI in Teaching and Assessments and Uphold Academic Rigour and Integrity'. https://warwick.ac.uk/fac/cross_fac/academy/activities/learningcircles/future-of-learning.

JISC. 2023. 'AI in Tertiary Education: A Summary of the Current State of Play – Third Edition'. https://beta.jisc.ac.uk/reports/artificial-intelligence-in-tertiary-education.

King's College London. 2023. 'King's Guidance on Generative AI for Teaching, Assessment and Feedback'. King's College London. https://www.kcl.ac.uk/about/strategy/learning-and-teaching/ai-guidance.

QAA. 2023. 'Maintaining Quality and Standards in the ChatGPT Era: QAA Advice on the Opportunities and Challenges Posed by Generative Artificial Intelligence'. https://www.qaa.ac.uk/news-events/news/reconsidering-assessment-for-the-chatgpt-era-new-qaa-advice-published.

Shepperd, Paddy. 2023. 'Navigating the Terms and Conditions of Generative AI'. *National Centre for AI* (blog). 27 September 2023. https://nationalcentreforai.jiscinvolve.org/wp/2023/09/27/navigating-the-terms-and-conditions-of-generative-ai/.

The Russell Group. 2023. 'Russell Group Principles on the Use of Generative AI Tools in Education'. https://russellgroup.ac.uk/news/new-principles-on-use-of-ai-in-education/.

Webb, Michael. 2023. 'AI Detection - Latest Recommendations'. *National Centre for AI* (blog). 18 September 2023. https://nationalcentreforai.jiscinvolve.org/wp/2023/09/18/ai-detection-latest-recommendations/.

Product announcements and manufacturer guidance

Adobe. 2023. ‘Adobe Unveils Firefly, a Family of New Creative Generative AI’. 21 March 2023. https://news.adobe.com/news/news-details/2023/Adobe-Unveils-Firefly-a-Family-of-new-Creative-Generative-AI/default.aspx.

Anthropic. 2023a. ‘Introducing 100K Context Windows’. Anthropic. 2023. https://www.anthropic.com/index/100k-context-windows.

———. 2023b. ‘Model Card and Evaluations for Claude Models’. https://www-files.anthropic.com/production/images/Model-Card-Claude-2.pdf.

———. 2023c. ‘Claude 2’. Anthropic. Accessed 15 October 2023. https://www.anthropic.com/index/claude-2.

———. n.d. ‘How Long Do You Store Personal Data? | Anthropic Help Center’. 2023. https://support.anthropic.com/en/articles/7996866-how-long-do-you-store-personal-data.

Canva. 2023. ‘Turn Imagination into Reality with AI Image Generation Apps in Canva’. Newsroom - Latest Canva News Announcements, Brand Guidelines and Media Kit. 4 October 2023. https://www.canva.com/newsroom/news/text-to-image-ai-image-generator/.

Google. 2023a. ‘Announcing New Generative AI Experiences in Google Workspace’. Google Workspace Blog. 2023. https://workspace.google.com/blog/product-announcements/generative-ai.

———. 2023b. ‘PaLM 2 Technical Report’. https://ai.google/static/documents/palm2techreport.pdf.

———. 2023c. ‘What Can Bard Do and Other Frequently Asked Questions – Bard’. 2023. https://bard.google.com/faq.

———. 2023d. ‘Bard Is Getting Better at Logic and Reasoning’. Google. 7 June 2023. https://blog.google/technology/ai/bard-improved-reasoning-google-sheets-export/.

———. n.d. ‘Models | PaLM API’. Generative AI for Developers. 3 October 2023. https://developers.generativeai.google/models/language.

Grammarly. 2023. ‘Ushering in a New Era of Communication Assistance With GrammarlyGO’. https://www.grammarly.com/blog/grammarlygo-augmented-intelligence/.

Microsoft. 2023a. ‘Introducing the New Bing’. 2023. https://www.bing.com/new.

———. 2023b. ‘Introducing Microsoft 365 Copilot – Your Copilot for Work’. The Official Microsoft Blog. 16 March 2023. https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/.

Meta. 2023. ‘Introducing New AI Experiences Across Our Family of Apps and Devices’. *Meta* (blog). 27 September 2023. https://about.fb.com/news/2023/09/introducing-ai-powered-assistants-characters-and-creative-tools/.

Notion. 2023. ‘Explore Notion AI: Augment Your Writing & Creativity Now’. Notion. 22 February 2023. https://www.notion.so/blog/notion-ai-is-here-for-everyone

OpenAI. 2022a. ‘DALL·E 2’. https://openai.com/dall-e-2.

———. 2022b. ‘Introducing ChatGPT’. https://openai.com/blog/chatgpt.

———. 2023a. ‘ChatGPT Plugins’. https://openai.com/blog/chatgpt-plugins.

———. 2023b. ‘How Your Data Is Used to Improve Model Performance | OpenAI Help Center’. https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance.

———. 2023c. ‘New AI Classifier for Indicating AI-Written Text’. https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text.

———. n.d. ‘OpenAI Platform’. Accessed 15 October 2023. https://platform.openai.com.

Victor, Jon. 2023. ‘Google Nears Release of Gemini AI to Challenge OpenAI’. *The Information*, 2023. https://www.theinformation.com/articles/google-nears-release-of-gemini-ai-to-rival-openai.

Wolfram, Stephen. 2023. ‘ChatGPT Gets Its “Wolfram Superpowers”!’ 23 March 2023. https://writings.stephenwolfram.com/2023/03/chatgpt-gets-its-wolfram-superpowers/.

Academic papers, reports, and opinion pieces

Arredondo, Pablo. 2023. ‘GPT-4 Passes the Bar Exam: What That Means for Artificial Intelligence Tools in the Legal Profession’. *Stanford Law School* (blog). 19 April 2023. https://law.stanford.edu/2023/04/19/gpt-4-passes-the-bar-exam-what-that-means-for-artificial-intelligence-tools-in-the-legal-industry/.

Brynjolfsson, Erik, Danielle Li, and Lindsey R. Raymond. 2023. ‘Generative AI at Work’. Working Paper. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w31161.

Choi, Jonathan H., and Daniel Schwarcz. 2023. ‘AI Assistance in Legal Analysis: An Empirical Study’. SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.4539836.

Cowen, Tyler, and Alexander T. Tabarrok. 2023. ‘How to Learn and Teach Economics with Large Language Models, Including GPT’. SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.4391863.

Dell’Acqua, Fabrizio, et al. 2023. ‘Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality’. SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.4573321.

Dhuliawala, Shehzaad, et al. 2023. ‘Chain-of-Verification Reduces Hallucination in Large Language Models’. arXiv. http://arxiv.org/abs/2309.11495.

GDELT Project. 2023. ‘LLM Infinite Loops & Failure Modes: The Current State Of LLM Entity Extraction – The GDELT Project’. 9 August 2023. https://blog.gdeltproject.org/llm-infinite-loops-failure-modes-the-current-state-of-llm-entity-extraction/.

Liang, Weixin, et al. 2023. ‘Can Large Language Models Provide Useful Feedback on Research Papers? A Large-Scale Empirical Analysis’. arXiv. https://doi.org/10.48550/arXiv.2310.01783.

Linkov, Denys. 2023. ‘Why Is GPT-3 15.77x More Expensive for Certain Languages?’ *Medium* (blog). 10 January 2023. https://denyslinkov.medium.com/why-is-gpt-3-15-77x-more-expensive-for-certain-languages-2b19a4adc4bc.

Liu, Nelson F., et al. 2023. ‘Lost in the Middle: How Language Models Use Long Contexts’. arXiv. https://doi.org/10.48550/arXiv.2307.03172.

Lukeš, Dominik. 2023. ‘The State of AI for August 2023: 6 Trends from the Last 3 Months and 6 Actions You Can Take Today’. 2023. https://www.linkedin.com/pulse/state-ai-august-2023-6-trends-from-last-3-months-actions-lukes/.

Martineau, Kim. 2021. ‘What Is Retrieval-Augmented Generation?’ IBM Research Blog. 9 February 2021. https://research.ibm.com/blog/retrieval-augmented-generation-RAG.

Mollick, Ethan R., and Lilach Mollick. 2023. ‘Using AI to Implement Effective Teaching Strategies in Classrooms: Five Strategies, Including Prompts’.

SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.4391243.

Nielsen, Jakob. 2023. 'ChatGPT Lifts Business Professionals' Productivity and Improves Work Quality'. Nielsen Norman Group. 31 May 2023. https://www.nngroup.com/articles/chatgpt-productivity/.

Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. 2023. 'The Impact of AI on Developer Productivity: Evidence from GitHub Copilot'. arXiv. http://arxiv.org/abs/2302.06590.

Pittenger, William. 2023. 'Leveraging ChatGPT for Advanced Data Analysis: A Deep Dive with Synthea's COVID-19 Data'. 2023. https://www.linkedin.com/pulse/leveraging-chatgpt-advanced-data-analysis-deep-dive-william/.

Potter, Brian. 2023. 'Could ChatGPT Become an Architect?' 2023. https://www.construction-physics.com/p/could-chatgpt-become-an-architect.

Salem, Lori, Stephanie Fiore, Stephen Kelly, and Benjamin Brock. 2023. 'Evaluating the Effectiveness of Turnitin's AI Writing Indicator Model'.

Strong, Eric et al.. 2023. 'Chatbot vs Medical Student Performance on Free-Response Clinical Reasoning Examinations'. *JAMA Internal Medicine* 183 (9): 1028–30. https://doi.org/10.1001/jamainternmed.2023.2909.

swyx. 2023. 'The Rise of the AI Engineer'. https://www.latent.space/p/ai-engineer.

Tu, Xinming, James Zou, Weijie J. Su, and Linjun Zhang. 2023. 'What Should Data Science Education Do with Large Language Models?' arXiv. http://arxiv.org/abs/2307.02792.

Weber-Wulff, Debora, et al. 2023. 'Testing of Detection Tools for AI-Generated Text'. arXiv. https://doi.org/10.48550/arXiv.2306.15666.

Wolfram, Stephen. 2023. 'What Is ChatGPT Doing … and Why Does It Work?' 14 February 2023. https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/.

Usage guides and sources of further information

- Student Use Cases for AI | Harvard Business Publishing Education
- Have you asked your bot to make you a table? ChatGPT as a cognitive scaffolding engine (linkedin.com)
- How (not) to learn about AI with metaphors: And how to use ChatGPT as a metaphor generation assistant (linkedin.com)

- Prompt Engineering Guide | Prompt Engineering Guide (promptingguide.ai)
- Learn Prompting: Your Guide to Communicating with AI
- Prompt Engineering for ChatGPT | Coursera
- There's An AI For That - The Biggest AI Aggregator (theresanaiforthat.com)

Acknowledgements

Contributions

The original draft of this document was produced by Dominik Lukeš, with contributions from Dr Xavier Laurent, Dr Jane Pritchard, Professor Rhona Sharpe, Dr Chloe Walker (all from the Centre for Teaching and Learning, University of Oxford).

Other work

Parts of this report are based on The State of AI for August 2023: 6 trends from the last 3 months and 6 actions you can take today published on LinkedIn by Dominik Lukeš.

Use of AI in this report

No AI tools were used for generating the initial draft.

Claude was used to:

- Provide suggestions for missing areas which resulted in adding several sections
- Generating a bullet point summary (see Appendices)
- Extract key terms used in the document to create a list of definitions.

ChatGPT Plus was used to:

- Extract links from the draft document and categorise them – this is now reflected in the appendix
- Convert a narrative about terms into a table which was then adapted in the list of key terms in the appendix.

ChatGPT Plus with DALL-E 3 was used to generate images throughout the report including those included in the cover.

The cover was generated using Microsoft Designer with manual edits.

Google Bard was used to extract text from an infographic published by JISC on AI maturity.