# The Technology of ChatGPT and its Impact in Education

Harry Nelson
*School of Electronics and Computer Science*
*University of Southampton*
*hjn2g19@soton.ac.uk*

## Abstract

*In November 2022 OpenAI released a product called ChatGPT, a chatbot built on their GPT-3.5 model, which has led to a massive increase in the use of generative AI both privately and commercially. While ChatGPT is very capable at generating (mostly) accurate, natural-looking written content in a fraction of the time it would take a human, there are many limitations of the model that need consideration, like "hallucinations" of facts and the presence of bias due to the data used and the training of the model.*

*The use of conversational chatbots in education has been researched prior to the release of ChatGPT, and the power of newly released chatbots improves upon those applications significantly. However the educational sector will need to teach students and staff on the proper use of chatbot technology to prevent misuse and over-reliance, and assessments will likely need to be significantly restructured to compensate for the accessibility of these writing tools.*

## 1. Introduction

With the release of ChatGPT [1] in November 2022, the accessibility to powerful chatbots to the general public increased substantially, leading to increased knowledge of the uses of such programs - for example, many students using ChatGPT to help write their homework [2], people using ChatGPT for therapy [3], or using the chatbot as a research aid for summarising papers or explaining difficult concepts [4]. Since then, the commercial availability of the API behind ChatGPT has led to many companies integrating these Large Language Model (LLM)-based chatbots into their products, such as the "New Bing" [5] from Microsoft overhauling a popular search engine, further normalising the use of this technology in day-to-day life.

While the use of chatbots in education has previously been explored [6] [7], the power of the new models may open up new avenues for benefits to students and education staff, but also introduce new difficulties to areas like student accessibility or academic integrity that will need to be considered or mitigated. This paper aims to investigate the capabilities and limitations of LLM-based chatbots like ChatGPT, and explore their potential positive and negative impact on the educational sector.

## 2. ChatGPT Background

### 2.1. Introduction to Chatbots

A chatbot is a program that simulates talking to a human, designed to accept commands from users and respond to those commands, all in a conversational style using natural language. Early chatbots like ELIZA [8], a chatbot designed to simulate talking to a psychologist, used specific rules to generate their responses based on keywords that appear in the "prompt". Until recently, chatbots commonly used this technique in customer support roles for businesses, helping a user find answers to their queries without needing a human assistant to guide them [9] [10]. Other chatbots rely on a collection of text (a corpus) that is used to construct their response, either by retrieving specific information from the corpus or, like ChatGPT, using information from the corpus to generate a response using deep learning techniques.

Recent approaches to chatbot technology involve using deep learning techniques to create a language representation that is learned from a very large corpus [12]. This is generally achieved using an "encoder-decoder" model (shown in Figure 1), where the program needs to "encode" the input sequence of text into a numerical representation of the contextual information (known as a "hidden state" or "context vector"), and then "decode" that information to produce an

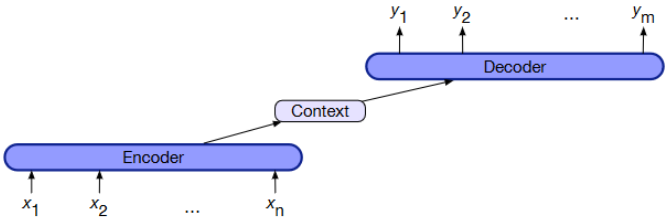


Figure 1. An example of the structure of an encoder-decoder model. [11]

output sequence token by token [13]. Encoder-decoder models are used for a variety of NLP tasks, such as response generation in chatbots, text summarisation and machine translation for translating between languages.

## 2.2. Language Encoding

In these encoder-decoder models, the input sequence of text first needs to be "embedded" into a numerical vector representation of the words that the model can process. An example of a method to create these vector representations is "Bag-of-words", where the frequency of each word in the document (e.g. sentence or paragraph) is counted, across all of the input documents, creating a vector where each row represents a document and each column represents a word in the vocabulary across all documents. Representing the words in this way can give an idea of what words appear alongside each other, and in what documents, which the model can use to help determine its response to future input. The bag-of-words approach can be improved using n-grams, where instead of the frequencies of single words being represented in the vector, it uses a sequence of $n$ words (for example, in a 2-gram or bigram the text "the quick brown fox jumped..." would be split into "(the, quick), (quick, brown), (brown, fox), ..."). This allows the relative positions of the words to be learnt as well as the frequencies.

These word embeddings can be improved by adding context learnt from pre-trained word embeddings. "Continuous Bag-of-Words" and "Continuous Skip-Gram" [14] were two new architectures for computing vector representations of large datasets. Compared to Bag-of-words, which is a technique for creating embeddings that only involves counting the frequencies of words in the vocabulary, CBoW and CSG are neural networks themselves. CBoW is a fully-connected neural network that tries to predict a word when given the surrounding words, and CSG tries to predict the surrounding words when given a word. A visualisation is shown in Figure 2.

The vectors produced by these models contain semantic information about the words and their relationships to each other, which can be extracted by using simple vector arithmetic, and using these word embeddings to pre-train natural language models had benefits to models' performance in NLP tasks [15]. Training the models on these large unlabelled word vectors allowed them to learn the semantic information about the words, and then could be fine-tuned using labelled data curated for the specific problem, such as sentiment analysis.

## 2.3. Attention and Transformers

The models used for NLP tasks (including the encoders and decoders of encoder-decoder models) were usually Recurrent Neural Networks (RNN) or Long Short-Term Memory (LSTM - a variant of RNN) models [15] [16]. The main difference between basic Feedforward Neural Networks (FFNN) and RNNs is that in FFNNs the "signals" only go one-way and it takes a fixed-size window of the input sequence as input to produce an output, whereas an RNN only takes a single token in the sequence as input, and the input to the hidden layers in the RNN is augmented based on the weights from the hidden layers for previous elements of the sequence. A downside of the way previous weights influence the current state in an RNN is that as you progress along the sequence, the impact of weights further from the current input gets lower and lower in comparison to the closer weights. This can be mitigated using LSTMs, which are similar to RNNs but pass on a hidden state at each step, managed by a series of "forget gates" which can selectively retain or discard context from previous elements.

Attention was another popular addition to the RNN architectures for encoder-decoder models to prevent performance from deteriorating with greater input lengths, which involves encoding the input sentence into a series of vectors and choosing a subset of those vectors dynamically as it generates and decodes the output [17]. The output of the attention mechanism is a vector containing a mapping of a query to key/value

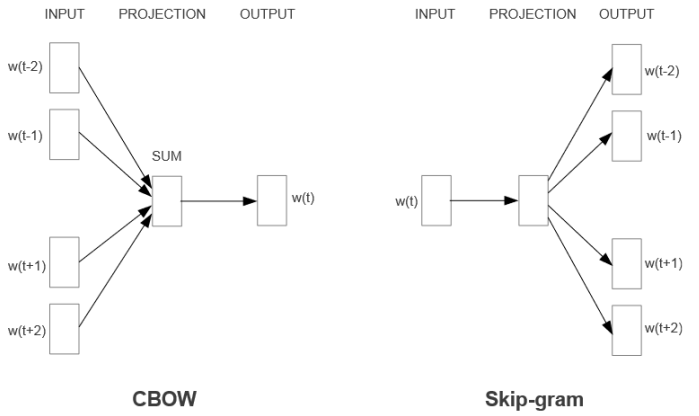


Figure 2. The proposed architectures used by Word2Vec [14]. *w(t)* refers to the word at time-step t.

pairs, weighted by relevance according to some compatibility function (e.g. cosine distance). The attention mechanism is usually used between the encoder and decoder stages, supplying additional information about the relevance of the current word to previous words to the decoder as it generates output. Without this attention mechanism, the model would need to compress all the information of a source sentence into a fixed-length vector to achieve the same effect.

In 2017, a novel networked architecture called the Transformer was released to out-perform the state-of-the-art solutions for sequence modelling problems at the time (shown in Figure 3) [18]. These existing architectures suffered from being constrained to sequential computations, as each word is processed once the hidden states of the previous word have been computed. The previous word is also weighted much more heavily than the words before it, and so the influence of words further backdrops unless steps are taken to retain them at a large memory cost. While there have been improvements in the computational efficiency of these architectures [19] [20] the fundamental limitation of sequential processing remains.

The transformer model processes each element of the input sequence in parallel, using "multi-headed attention" to supply the contextual information to different parts of the input at the same time, providing significantly improved computational performance as well as greater quality output. The transformer model forgoes the recurrent neural network and relies entirely on the attention mechanism to construct the dependencies used in the model. The benefits of multi-headed attention in this model include the lower computational complexity per layer, the larger amount of computation that can be parallelised, and the lower "path length" between dependencies in the network.

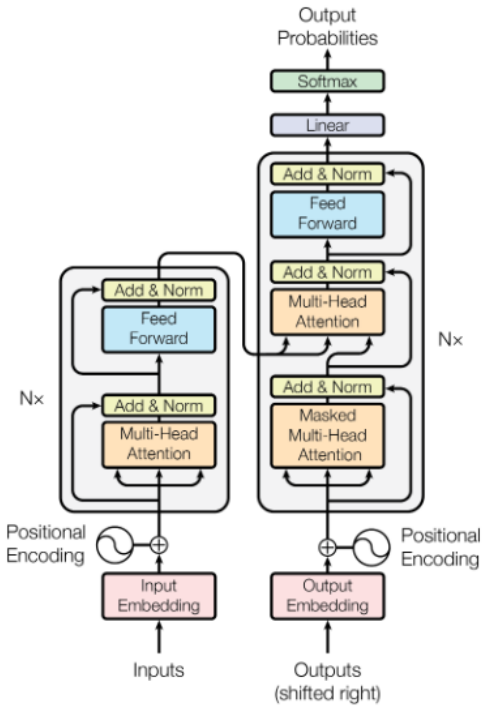


Figure 3. Transformer Architecture from *Attention is All You Need* [18]

## 2.4. Models Building on Transformers

Some examples of popular Language Models that made use of the Transformer architecture following the publishing of *Attention is All You Need* [18] were OpenAI's GPT [21] and Google's BERT [22].

GPT aimed to take a semi-supervised approach to language understanding tasks using unsupervised pre-training and supervised fine-tuning, with the aim of creating a model that required little adaptation to a wide range of NLP tasks based on a Transformer Decoder model [23]. As labelled data is expensive to produce, and the datasets tend to be much smaller, they hypothesised that giving the model greater "world knowledge" through unsupervised training would allow it to learn a lot of relationships between words naturally, and their model did in fact improve upon the state-of-the-art for 9 of their 12 experiments using this expanded knowledge. OpenAI trained the initial model on the BookCorpus dataset [24], an unlabeled dataset of unique books across a range of genres with long stretches of text, and then fine-tuned the model for a few different NLP tasks using similar hyperparameters but on a new, smaller, labelled dataset.

BERT, similarly to GPT, uses a semi-supervised approach, with the model pre-trained on a large unlabelled dataset and fine-tuned on a task-specific dataset afterwards. Unlike GPT though, BERT used a Transformer model almost identical to the one initially outlined in *Attention is All You Need*. Devlin *et al.* argued that the performance of unidirectional models (left-to-right or right-to-left) was lacking due to the inability to attend to tokens on both sides of the input sequence, as context from words later in the sequence could be important. They propose a "masked language model" pre-training objective to train a bidirectional transformer and use a "next sentence prediction" task to help train text-pair representations. The Masked LM task involves masking random tokens from the input sequence and predicting those, and the next sentence prediction involves being given a pair of sentences and predicting if sentence A is followed by sentence B. This allowed the model to understand the relationship between sentences as well as just the relationship of words in a context. Similar to GPT before it, BERT was able to perform competitively with state-of-the-art solutions for a variety of NLP tasks thanks to the pre-training and minimal fine-tuning.

GPT-2 attempted to further demonstrate the power of general pre-training [25] using a new WebText dataset to train a model and evaluate it on various NLP tasks without any fine-tuning. The WebText dataset is a dataset made by OpenAI with an emphasis on "document quality", by crawling pages that had been linked to on the social media platform Reddit with a positive "upvote" count (meaning that they were endorsed by humans). Using a similar model architecture to GPT but with the pre-training on this dataset, they were able to "zero-shot" (no demonstrations, just instructions in natural language) to state-of-the-art performance on 7/8 tested NLP tasks, demonstrating the power of unsupervised learning even without supervised fine-tuning.

## 2.5. GPT-3 and 4, and ChatGPT

As the capacity of Transformer language models increased to 17 billion parameters [26], OpenAI hypothesised that in-context learning capabilities would improve with the increasing dataset size. Their paper, *Language Models are Few-Shot Learners* [27], focused on few-shot learning, one-shot learning and zero-shot learning, meaning the model is given a few, one and no demonstrations of the task at inference time, and isn't fine-tuned on a task-specific dataset. The model is given examples (or none for zero-shot) and then a natural language description of the problem.

Several models were trained, with a range from 125 million to 175 billion parameters, with the last of those being the one referred to as GPT-3. The dataset used includes a (processed to improve document quality) Common Crawl dataset [28], an expanded version of the WebText dataset, English-language Wikipedia and the BookCorpus dataset. These models were evaluated on over two dozen NLP datasets and some novel tasks OpenAI created, testing the few/one/zero-shot learning. They found that it achieved promising results in the zero and one-shot settings, and competitive with the state-of-the-art results in the few-shot setting.

While GPT-3 excels at NLP tasks, it doesn't meet the social requirements of a product released to the public, as the content it generates can contain bias and inappropriate stereotypes present in the dataset, and when intentionally prompted can produce harmful and toxic responses. To try and align GPT-3 with users, OpenAI created InstructGPT [29], fine-tuning the model using reinforcement learning from human feedback to make GPT-3 follow written instructions more appropriately. Using data manually labelled by their team, and demonstrations submitted by users to their API, they train the supervised-learning baselines. They then collect human-labelled comparisons between outputs from the models on a larger set of API prompts. They also train a reward model focused on predicting which model output the labellers would prefer. This reward model is then used to fine-tune the supervised learning. This method significantly improved the quality of the outputs of the model with improvements in truthfulness, but there was not much improvement in the frequency of toxic content or bias of the dataset. This model and other models trained on text and code from before Q4 2021 are called

"GPT-3.5", with variants more optimised for chat or code-completion tasks, and were available through the OpenAI API [30].

ChatGPT was then created and trained using a different RLHF method [1], where human AI trainers would provide both sides of the conversation and this dataset was mixed with the InstructGPT dataset. The reward model used was based on conversations that the AI trainers had with the chatbot, with the AI trainers manually ranking different outputs to the same random prompt. This was then used to fine-tune the model and the process was repeated several times. The resulting model launched publicly as a chatbot that users could prompt to generate text on a given subject and in a given format, as well as interact with conversationally.

Six months after ChatGPT was made publicly available, OpenAI announced GPT-4, the next in their series of LLMs, using similar training to GPT3.5 but with an altered dataset to try and help reduce the potential for misuse [31]. They used internally trained classifiers and lexicon-based approaches to filter out documents containing erotic content and removed it from the pre-training set, and then used RLHF methods identical to InstructGPT [29]. They then used their models as tools to help steer the model towards an appropriate level of "alignment" using supervised learning, with a GPT-4-based classifier. It would be given the output of a previous training prompt and would classify whether the model had responded appropriately (for example, if it had refused to answer a dangerous prompt or not), and the output of this would be given back to the original model to tune the responses further.

The model is also used in other ways to aid the tuning process. OpenAI used GPT-4 to rewrite prompts requesting disallowed content into prompts as close as possible to the original but without requesting disallowed content, to ensure that the model doesn't refuse those. They also were able to use GPT-4 to iteratively generate prompts, have it identify "hallucinations" in its output (a confident response that is incorrect), and rewrite the prompts without those hallucinations.

This iteration of GPT focuses heavily on the safety challenges of the model and mitigation strategies, but it also introduced new capabilities to the model. GPT-4 can take images as input and is able to describe the contents, and through the use of MATH and GSM-8K datasets being mixed into the training set OpenAI were able to increase its capacity to do mathematical reasoning [31].

### 2.6. ChatGPT Limitations

OpenAI have critically evaluated the limitations of ChatGPT in their article on its release [1], stating that ChatGPT can often write plausible-sounding but incorrect or nonsensical answers when prompted, due to its reluctance to reject a prompt that hasn't been classed as inappropriate. It was found by those testing the chatbot's ability to answer more complex questions [32] that it would often hallucinate explanations for concepts it was unfamiliar with, or invent academic references that look plausible and even have a DOI number but don't exist.

Studies also found that despite the pre-training dataset containing more text than a human will see in their lifetime, the models still consistently underperformed when generating content about a specific domain or when performing reasoning about these domain-specific concepts [33] [34]. It was also found that it often had the potential to make up facts or not give explanations in enough detail as it lacked "understanding" of the content that it was writing about [35], meaning that any text generated by it will still need human validation before it can be used reliably.

The model is also not yet fully aligned (doesn't entirely work how the creators intended it to) [29] [1] - it is still capable of producing "toxic or biased outputs" or "sexual and violent content", even without explicit prompting. Steps have been taken to try and reduce this, but this mitigation strategy relies on the one training and providing the model, meaning that a language model generated and provided by a less ethical company using similar training methods could forgo these mitigations entirely. Even with all of the restrictions provided by OpenAI trying to limit the chances of illegal or immoral output, there are communities of users trying to "jailbreak" ChatGPT through prompt engineering [36].

Due to the data that the model has been trained on (a large number of web pages), even though it has been curated to some degree, there is still a heavy bias that can lead to the model "generating stereotyped or prejudiced content" [27]. OpenAI were able to detect noticeable bias in the outputs regarding gender, race and religion, although that was specifically in an experimental setup with prompts about those topics.

## 3. Chatbots and LLMs in Education

### 3.1. Potential Uses of Chatbots in Education

As well as recent research on the concept of using chatbots in an educational setting, the topic has been

proposed and experimented with prior to the release of ChatGPT. The main examples identified and explained in the following section are:

1) Playing a character for students to interact with.
2) Providing support to students on the content being taught.
3) Providing support during software training for students or educators.
4) "Helpdesk" chatbots providing administrative support.
5) Brainstorming, Idea Generation and Cognitive offloading.
6) Generating content for educators (e.g. lesson plans).
7) Help students and educators work around a language barrier.

One use of chatbots prototyped prior to the release of ChatGPT is where the chatbot takes the role of a character for the student to interact with, allowing them to practice specific conversational skills or assessing them on their knowledge interactively, for example having to convince a "customer" to not change suppliers [37]. The example chatbot would identify keywords in the messages sent by the user and would respond with pre-determined answers based on those messages. Students found it helpful for learning, and it was able to guide the student through the interaction by prompting them helpfully if they were stuck too. However, this chatbot was identifiable as a bot rather than a real customer, which students felt decreased its effectiveness for conversational practice, and any particularly difficult questions it was unable to answer would have to be passed along to a teacher, which couldn't happen automatically.

Chatbots have also been experimented with for supporting educators and students through the training and use of educational applications [37]. The chatbot would answer questions about the software and try to train the users to resolve technical issues themselves based on common problems that the developers knew were encountered, and failing that it would be able to pass the users along to further support if needed for more complex issues.

Another use of chatbots that has been proposed is using the technology to teach basic content to students [38]. The learner would ask specific questions and receive personal guidance without taking up the time of a teacher, who would have multiple students to attend to. When the capabilities of GPT-3 were tested, it was found that it was good at speaking like a teacher but scored worse in terms of how much the students felt they learnt or were helped [39].

Another application of chatbots in an educational setting is the idea of a "helpdesk" chatbot [7] [38], which is able to conversationally get responses to frequently asked questions, answer meta-information questions about a course (e.g. timetabling, study tips) and obtain additional learning resources without needing to contact a teacher. This removes the time-of-day restrictions on the questions as well and generally allows students to get answers to "mundane" questions without occupying the teacher's time. The "adaptability and frequent feedback" of chatbots as learning support led to higher student engagement and allowed the students to feel the social benefits of communicating when searching for information [32]. With the release of GPT-4 and its ability to understand image input, it has introduced the capacity for greater capabilities as a helpdesk chatbot, allowing for potential explanations of images. For example, if the chatbot could look at and "understand" a timetable or a map, it would be able to provide students advice based purely on the image without being frequently provided up-to-date information about the site, or it could be able to explain mathematical concepts to a student based off of an image of the problem.

It has been found that ChatGPT provides significant value as a brainstorming and idea-generation tool that a student can then take and develop [34] [40], as the conversational style can help students discuss their ideas and get suggestions in a natural way. It can also provide feedback on the text that students have written, checking both the contents of the writing and the accuracy in grammar or punctuation [41]. This idea of "cognitive offloading" is similar to other grammar and spellchecking tools in the past [42], preventing students from wasting time on things that aren't the learning objectives, and allowing them to focus on developing their content in the time they are given.

Another proposed use of ChatGPT's ability to generate text is creating content for educators, for example generating quizzes based on a topic, or lesson planning [40]. It would also be able to tailor content for specific students in a fraction of the time it would take a human educator.

Finally, ChatGPT's multilingual capabilities have led to the proposal that it be used to aid educators in overcoming a language barrier when teaching students in a language other than their first language and for teaching languages in general [40]. ChatGPT would be able to provide deeper explanations of concepts in a student's first language and could give dynamic conversational practice in the language of choice.

### 3.2. Discussion on Challenges in Education

Due to the capabilities of ChatGPT and other Large Language Models in generating large bodies of text on prompting, many learning facilities globally have found students using ChatGPT to assist and even entirely write homework or assessments for them [43], leading to academic institutions having to consider whether the way that they test students is appropriate.

One of the major considerations with the generative power of ChatGPT is with regard to its ability to complete standard academic assessments, such as long-form questions or essays. As most assessments aren't designed with access to this technology in mind, several academic institutions have made statements that the use of ChatGPT for assessments isn't allowed [44], and instances where a teacher has detected a student using ChatGPT to aid with or write their homework are often being treated as a breach of academic integrity [45] [46]. The QAA, an independent organisation for quality assurance across Higher Education, issued a paper for higher education providers that noted the potential benefits of LLMs to education briefly but only discussed the challenges it provided to academic integrity, and potential mitigations [47].

Some ways that studies suggest working around the existence of ChatGPT is by bringing the focus back to in-person examinations or adding oral/video elements that are harder for AI to generate [48], as invigilators will be able to ensure that they are doing the work on their own, and more exams have already been moved to in-person after the COVID-19 Pandemic. Other suggestions take after OpenAI's stance that it will be "necessary for students to learn how to navigate a world where tools like ChatGPT are commonplace" [49]. It is suggested that assessments should be reworked so that the learning objectives focus on creative or critical thinking, to try and test understanding rather than skills like memorisation, allowing for the tools to be used without invalidating the assessment [50]. OpenAI and other studies also emphasise the need to communicate ahead of time any policies regarding the use of these generative text tools so that it isn't a "grey area", and students know for sure whether what they are doing is going to be considered cheating [49] [41] [42].

Despite plagiarism-checker tools not being able to detect AI-written text (and it is debated whether AI-written text is plagiarism), there are a few examples of tools capable of detecting whether a passage of text has been generated which are in development [51] [52]. They all note that much like standard plagiarism detection tools, they shouldn't be believed fully and be used more as a prompt for investigation. OpenAI's own detector only labelled 26% of AI-written text as AI-written and 9% of human-written text was labelled as AI-written [53].

Another concern with generative text tools is the ability for students to use it to pass courses that they otherwise wouldn't be able to, as well as just using it as a tool for convenience. Studies have researched ChatGPT's ability to complete assessments, finding that it is capable of answering the exam questions on medicine and law well enough to obtain a passing grade with minimal intervention [34] [54]. OpenAI also evaluated GPT-4 on an array of academic exams to try and measure its performance, including exams like the Uniform Bar Exam, AP calculus and Leetcode challenges and was able to obtain not just competitive but high-scoring grades in over half of them [31]. There are numerous guides online targeted at students who provide advice and tutorials on prompt engineering in order to write essays more effectively [55] [56]. Alongside ChatGPT there are also other AI-based writing services with a greater focus on writing specifically for academic assignments [57] [58], with both free and paid avenues boasting thousands of users.

While GPT-4 has focused on creating safeguards against intentional misuse and accidental misinformation, the tendency for bias and hallucinations would be an issue for the introduction of AI-based aid in education [31] [40]. If the information being given to students and teachers isn't guaranteed to be accurate then it will still need to be verified by human experts before it would be reasonable to use it to teach. AI models also have a tendency to be biased, with GPT-4 mentioning that their fine-tuning team is made up of people with specific educational and professional backgrounds who tend to come from English-speaking and Western countries. Experimentation with GPT-3 also found bias in the responses about specific groups, like gender or race [27].

Biased information and misinformation is still generated frequently by GPT-4 after fine-tuning [31]. For models trained for a specific educational institution (for example, a helpdesk chatbot) there would likely be a considerable cost to maintaining the chatbot and updating it, as well as training the teachers to be able to use it effectively [40] [37] [59]. Any content produced by chatbots for use in teaching would still need to be vetted by humans before use to ensure that it was accurate.

Some other considerations that require work before integrating AI into education would be ensuring accessibility to all students, due to the varying levels of quality of AI-assistant models out there, some of which

cost money to access [40]. If assessments are written assuming that students have access to state-of-the-art AI generative tools, it will be unfair to those that can't access them.

### 3.3. Future of Chatbots

Since the release of ChatGPT, the number of publicly available AI tools utilising GPT-4 has rapidly increased. OpenAI have partnered with several companies to integrate GPT-4 into their products, such as Duolingo [60] using it to improve their language-learning service, and Stripe [61] using the model to enhance its documentation and improve user experience. OpenAI have also partnered with Microsoft, and GPT-4 has been used in the newly released "New Bing", a chatbot with access to the internet [5]. Statistics aggregator ThinkImpact reports that over 500 companies across the technology, education and business sectors make use of the OpenAI platform as of 2023 [62]. Competitors have also been rushing to release their own models to the market, such as Google's Bard chatbot based on the LaMDA family of LLMs [63], which caused massive damage to Google's stock after it made a factual error in its first public demonstration [64].

The speed at which new tools are coming out is potentially damaging to the AI landscape, with the potential for competitors to be lax on safety standards due to racing dynamics. OpenAI have documented a concern on the impact that releasing GPT-4 would have on the AI research and development ecosystem [31], and have referred to their charter [65] that commits to ceasing work on their own artificial general intelligence if another project looks closer to completion than theirs and working with the other project.

From the discussion around chatbots, it is evident that there is a large advantage to using chatbot technology in different areas of education, providing support for learners in learning activities and helping educators in training, and producing materials and assessments. The primary challenge will be handling the disruption from the technology and making use of it effectively, rather than seeing it as a threat. As the models and tools making use of them are actively being developed, it is likely that limitations like an inability to generate accurate references or the frequency of hallucinations will be fixed. GPT-4 has increased mathematical ability over GPT-3.5 [31], and ChatGPT now has a plugin framework available [66], advertising new plugins focused on interacting with services on the web.

## 4. Conclusion

ChatGPT's proficiency at text generation has the capacity to greatly improve upon existing applications of chatbots in education as well as finding new roles within educational organisations, helping to facilitate learning as well as supporting them with administrative content. Academic organisations should take the lead in making use of this technology for the benefit of their learners, and educate people properly on the limitations of the technology while still exploring its potential. The main difficulty with adopting new tools will be disentangling the capabilities of the AI from the capabilities of the student. ChatGPT and other tools are accessible enough that it is almost impossible that they won't be used regardless of guidance given, so it would be more effective for students to know how to use them effectively and for assessments and exams to be restructured to test the students given the existence of these tools, rather than prohibiting their use and having to rely on detection tools to catch them.

## References


[1] Introducing ChatGPT, November 2022. https://openai.com/blog/chatgpt [Online; accessed 14. Mar. 2023].

[2] Chris Westfall. Educators Battle Plagiarism As 89% Of Students Admit To Using OpenAI's ChatGPT For Homework, January 2023. https://www.forbes.com/sites/chriswestfall/2023/01/28/educators-battle-plagiarism-as-89-of-students-admit-to-using-open-ais-chatgpt-for-homework [Online; accessed 7. May 2023].

[3] Lance Eliot. People Are Eagerly Consulting Generative AI ChatGPT For Mental Health Advice, Stressing Out AI Ethics And AI Law. *Forbes*, January 2023. https://www.forbes.com/sites/lanceeliot/2023/01/01/people-are-eagerly-consulting-generative-ai-chatgpt-for-mental-health-advice-stressing-out-ai-ethics-and-ai-law.

[4] Use Cases of GPT-3 ChatBot - Make Your Life Easier, March 2023. [Online; accessed 3. May 2023].

[5] Introducing the new Bing, April 2023. https://www.bing.com/new [Online; accessed 17. Apr. 2023].

[6] Sebastian Wollny, Jan Schneider, Daniele Di Mitri, Joshua Weidlich, Marc Rittberger, and Hendrik Drachsler. Are we there yet? - a systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence*, 4, 2021.

[7] Nitirajsingh Sandu and Ergun Gide. Adoption of ai-chatbots to enhance student learning experience in higher education in india. In *2019 18th International Conference on Information Technology Based Higher Education and Training (ITHET)*, pages 1–5, 2019.

[8] Joseph Weizenbaum. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966.

[9] Jagdish Singh, Minnu Helen Joesph, and Khurshid Begum Abdul Jabbar. Rule-based chabot for student enquiries. In *Journal of Physics: Conference Series*, volume 1228, page 012060. IOP Publishing, 2019.

[10] GD Souza. Chatbot for organizational faq's. *no. May*, pages 5591–5594, 2019.

[11] Speech and Language Processing (3rd Edition), April 2023. https://web.stanford.edu/~jurafsky/slp3 [Online; accessed 23. Apr. 2023].

[12] Eleni Adamopoulou and Lefteris Moussiades. An overview of chatbot technology. In Ilias Maglogiannis, Lazaros Iliadis, and Elias Pimenidis, editors, *Artificial Intelligence Applications and Innovations*, pages 373–383, Cham, 2020. Springer International Publishing.

[13] Ziang Xie. Neural text generation: A practical guide. *CoRR*, abs/1711.09534, 2017.

[14] Tomas Mikolov, Kai Chen, Greg S. Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.

[15] Andrew M Dai and Quoc V Le. Semi-supervised sequence learning. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

[16] Bryan McCann, James Bradbury, Caiming Xiong, and Richard Socher. Learned in translation: Contextualized word vectors. *CoRR*, abs/1708.00107, 2017.

[17] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473, 2016.

[18] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[19] Oleksii Kuchaiev and Boris Ginsburg. Factorization tricks for LSTM networks. *CoRR*, abs/1703.10722, 2017.

[20] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc V. Le, Geoffrey E. Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *CoRR*, abs/1701.06538, 2017.

[21] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. Improving language understanding by generative pre-training. *OpenAI Blog*, 2018.

[22] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[23] Peter J. Liu, Mohammad Saleh, Etienne Pot, Ben Goodrich, Ryan Sepassi, Lukasz Kaiser, and Noam Shazeer. Generating wikipedia by summarizing long sequences. *CoRR*, abs/1801.10198, 2018.

[24] Yukun Zhu, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[25] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

[26] Alexis Hagen. Turing-NLG: A 17-billion-parameter language model by Microsoft - Microsoft Research. *Microsoft Research*, February 2020.

[27] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc., 2020.

[28] Common Crawl, March 2023. https://commoncr

awl.org [Online; accessed 14. Mar. 2023].

[29] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback, 2022.

[30] OpenAI. Openai api. https://platform.openai.com/docs/model-index-for-researchers [Online; accessed 14. Mar. 2023].

[31] OpenAI. Gpt-4 technical report, 2023.

[32] Junaid Qadir. Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education. *Qatar University*, December 2022.

[33] Tal Linzen. How can we accelerate progress towards human-like linguistic generalization? In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5210–5217, Online, July 2020. Association for Computational Linguistics.

[34] Jonathan H Choi, Kristin E Hickman, Amy Monahan, and Daniel Schwarcz. Chatgpt goes to law school. *Minnesota Legal Studies Research Paper No. 23-03*, 2023.

[35] Yu Chen, Scott Jensen, Leslie J. Albert, Sambhav Gupta, and Terri Lee. Artificial Intelligence (AI) Student Assistants in the Classroom: Designing Chatbots to Support Student Success. *Inf. Syst. Front.*, 25(1):161–182, June 2022.

[36] ChatGPT-Dan-Jailbreak.md, March 2023. https://gist.github.com/coolaj86/6f4f7b30129b0251f61fa7baaa881516 [Online; accessed 14. Mar. 2023].

[37] Shanshan Yang and Chris Evans. Opportunities and challenges in using ai chatbots in higher education. In *Proceedings of the 2019 3rd International Conference on Education and E-Learning*, ICEEL 2019, page 79–83, New York, NY, USA, 2020. Association for Computing Machinery.

[38] Rehan Ahmed Khan, Masood Jawaid, Aymen Rehan Khan, and Madiha Sajjad. Chatgpt - reshaping medical education and clinical management. *Pakistan Journal of Medical Sciences*, 39(2), Feb. 2023.

[39] Anaïs Tack and Chris Piech. The ai teacher test: Measuring the pedagogical ability of blender and gpt-3 in educational dialogues, 2022.

[40] Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, and et al. Chatgpt for good? on opportunities and challenges of large language models for education, Jan 2023.

[41] Mike Perkins. Academic integrity considerations of ai large language models in the post-pandemic era: Chatgpt and beyond. *Journal of University Teaching & Learning Practice*, 20(2):07, 2023.

[42] Phillip Dawson. Cognitive Offloading and Assessment. *Deakin University*, Jan 2020.

[43] Vishwam Sankaran. Cheating by students using ChatGPT is already on the rise, surveys suggest. *Independent*, February 2023.

[44] Jonathan Holmes. Universities warn against using ChatGPT for assignments. *BBC News*, February 2023. https://www.bbc.co.uk/news/uk-england-bristol-64785020 [Online; accessed 17. Mar. 2023].

[45] Cooper Worth. UI student cheats the system with AI program ChatGPT, March 2023. https://dailyiowan.com/2023/02/19/university-of-iowa-student-cheats-the-system-with-ai-program-chatgpt [Online; accessed 17. Mar. 2023].

[46] Thomas Germain. A Student Used ChatGPT to Cheat in an AI Ethics Class. *Gizmodo*, February 2023. https://gizmodo.com/ai-chatgpt-ethics-class-essay-cheating-bing-google-bard-1850129519 [Online; accessed 17. Mar. 2023].

[47] The rise of artificial intelligence software and potential risks for academic integrity: A qaa briefing paper for higher education providers, January 2023. https://www.qaa.ac.uk/news-events/news/qaa-briefs-members-on-artificial-intelligence-threat-to-academic-integrity [Online; accessed 17. Apr. 2023].

[48] Teo Susnjak. Chatgpt: The end of online exam integrity?, 2022.

[49] OpenAI API, March 2023. https://platform.openai.com/docs/chatgpt-education [Online; accessed 17. Mar. 2023].

[50] Xiaoming Zhai. Chatgpt user experience: Implications for education. *Available at SSRN 4312418*, 2022.

[51] AI Writing | AI Tools, March 2023. https://www.turnitin.com/solutions/ai-writing [Online; accessed 17. Mar. 2023].

[52] Mohammad Khalil and Erkan Er. Will chatgpt get you caught? rethinking of plagiarism detection, 2023.

[53] New AI classifier for indicating AI-written text, March 2023. https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text [Online; accessed 17. Mar. 2023].

[54] Tiffany H Kung, Morgan Cheatham, Arielle

Medenilla, Czarina Sillos, Lorie De Leon, Camille Elepaño, Maria Madriaga, Rimel Aggabao, Giezel Diaz-Candido, James Maningo, et al. Performance of chatgpt on usmle: Potential for ai-assisted medical education using large language models. *PLOS Digital Health*, 2(2):e0000198, 2023.

[55] Pragati Gupta. How to use ChatGPT to write an essay. *Writesonic Blog - Making Content Your Superpower*, March 2023. https://writesonic.com/blog/how-to-use-chatgpt-to-write-essay [Online; accessed 17. Mar. 2023].

[56] GripRoom, March 2023. https://www.griproom.com/fun/how-to-write-better-prompts-for-chat-gpt, [Online; accessed 17. Mar. 2023].

[57] Free Essay Writing Tool | AI Essay Writer, March 2023. https://www.the-good-ai.com [Online; accessed 17. Mar. 2023].

[58] Jasper Commercial, December 2022. https://www.jasper.ai [Online; accessed 17. Mar. 2023].

[59] Ismail Celik. Towards intelligent-tpack: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (ai)-based tools into education. *Computers in Human Behavior*, 138:107468, 2023.

[60] Duolingo Team. Duolingo Max Uses OpenAI's GPT-4 For New Learning Features. *Duolingo Blog*, March 2023.

[61] Stripe and OpenAI collaborate to monetize OpenAI's flagship products and enhance Stripe with GPT-4, April 2023. {https://stripe.com/gb/newsroom/news/stripe-and-openai} [Online; accessed 17. Apr. 2023].

[62] OpenAI Statistics, February 2023. [Online; accessed 5. May 2023].

[63] Bard, April 2023. https://bard.google.com/?hl=en [Online; accessed 17. Apr. 2023].

[64] James Vincent. Google's AI chatbot Bard makes factual error in first demo. *Verge*, February 2023.

[65] OpenAI Charter, April 2023. https://openai.com/charter [Online; accessed 17. Apr. 2023].

[66] ChatGPT plugins, April 2023. https://openai.com/blog/chatgpt-plugins [Online; accessed 17. Apr. 2023].