A Report About ChatGPT

Abstract of ChatGPT

A ChatGPT can be used anywhere a human is interacting with a computer system.

In the current time, ChatGPT is a very advanced Chatbots, or conversational interfaces as they are also known, present a new way for individuals to interact with computer systems. Traditionally, to get a question answered by a software program involved using a search engine, or filling out a form. A chatbot allows a user to simply ask questions in the same manner that they would address a human. The most well-known chatbots currently are voice chatbots: Alexa and Siri. However, chatbots are currently being adopted at a high rate on computer chat platforms.

The technology at the core of the rise of the chatbot is natural language processing (“NLP”). Recent advances in machine learning have greatly improved the accuracy and effectiveness of natural language processing, making chatbots a viable option for many organizations. This improvement in NLP is firing a great deal of additional research which should lead to continued improvement in the effectiveness of chatbots in the years to come.

A simple chatbot can be created by loading an FAQ (frequently asked questions) into chatbot software. The functionality of the chatbot can be improved by integrating it into the organization’s enterprise software, allowing more personal questions to be answered, like “What is my balance?”, or “What is the status of my order?”.

Most commercial chatbots are dependent on platforms created by the technology giants for their natural language processing. These include Amazon Lex, Microsoft Cognitive Services, Google Cloud Natural Language API, Facebook Deep Text, and IBM Watson.

Platforms where chatbots are deployed include Facebook

Messenger, Skype, and Slack, among many others.

What is the definition of ChatGPT

Chat GPT, or chat-based Generative Pretrained Transformer, is a state-of-the-art natural language processing (NLP) model developed by OpenAI. It is a variant of the popular GPT-3 model, which has been trained on a large amount of text data from various sources such as books, articles, and websites.

ChatGPT is built on top of OpenAI’s GPT-3.3 family of large language models, and is fine-tuned with both supervised and reinforcement learning techniques. ChatGPT is a sibling model to InstructGPT, which is trained to follow an instruction in a prompt and provide a detailed response.

ChatGPT was launched as a prototype in November 2022, and quickly garnered attention for its detailed responses and articulate answers across many domains of knowledge. Its uneven factual accuracy was identified as a significant drawback.

# What is the purpose of ChatGPT

ChatGPT was trained on an enormous amount of text data. It learned to recognize patterns that enable it to produce its own text mimicking various writing styles.

It can serve up information in clear, simple sentences, rather than just a list of internet links. It can explain concepts in ways people can easily understand. It can even generate ideas from scratch, including business strategies, Christmas gift suggestions, blog topics and vacation plans.

ChatGPT is a very advanced chatbot that has the potential to make people's lives easier and to assist with everyday tedious tasks, such as writing an email or having to navigate the web for answers. However, there are certain technical details that have to be figured out before it's widely used to prevent negative outcomes, such as the spread of misinformation. In general, AI and ML models rely on lots of training and finetuning to reach a level of ideal performance.

In comparison to its predecessor, InstructGPT, ChatGPT attempts to reduce harmful and deceitful responses; in one example, while InstructGPT accepts the prompt "Tell me about when Christopher Columbus came to the US in 2015" as truthful, ChatGPT uses information about

Columbus' voyages and information about the modern world – including perceptions of Columbus to construct an answer that assumes what would happen if Columbus came to the U.S. in 2015. ChatGPT's training data includes my pages and information about internet phenomena and programming languages, such as bulletin board systems and the Python programming language.

Unlike most chatbots, ChatGPT is stateful, remembering previous prompts given to it in the same conversation, which some journalists have suggested will allow for ChatGPT to be used as a personalized therapist. To prevent offensive outputs from being presented to and produced from ChatGPT, queries are filtered through a moderation API, and potentially racist or sexist prompts are dismissed.

ChatGPT suffers from multiple limitations. The reward model of ChatGPT, designed around human oversight, can be overoptimized and thus hinder performance, otherwise known

as Goodhart’s law. Furthermore, ChatGPT has limited knowledge of events that occurred after 2021 and is unable to provide information on some celebrities. In training, reviewers preferred longer answers, irrespective of actual comprehension or factual content. Training data may also suffer from algorithmic bias; prompts including vague descriptors of people, such as a CEO, could generate a response that assumes such a person, for instance, is a white male.

# Provided Functions to ChatGPT

Functionalities are important sequences of ChatGPT. The model has many functions in addition to answering simple questions, such as composing essays, describing art in great detail, creating AI art prompts, having philosophical conversations, and can even code for you. My personal favorite is asking the chatbot for help coming up with creative holiday gift ideas for specific family members. The possibilities are endless. The best ChatGPT is a chatbot has features like no code deployment, omnichannel messaging support, fallback options, sentiment analysis to add value to conversations. Integrating chatbots can help your business deliver automated smart responses and achieve marketing goals efficiently.

ChatGPT is based on a unique training model: Reinforcement Learning from Human Feedback (RLHF). This method trains the assistant by having humans simulate artificial conversations with him, and then adapts his responses based on how accurately they reflect natural human dialogue. By repeating this process several times, ChatGPT strives to be more accurate in understanding users’ questions.

We trained model using Reinforcement Learning from Human Feedback (RLHF), using the same methods as instructGPT, but with slight differences in the data collection setup. We trained an initial model using supervised fine-tuning: human AI trainers provided conversations in which they played both sides—the user and an AI assistant. We gave the trainers access to model-written suggestions to help them compose their responses. We mixed this new dialogue dataset with the Instruct GPT dataset, which we transformed into a dialogue format.

To create a reward model for reinforcement learning, we needed to collect comparison data, which consisted of two or more model responses ranked by quality. To collect this data, we took conversations that AI trainers had with the chatbot. We randomly selected a model-written message, sampled several alternative completions, and had AI trainers rank them. Using these reward models, we can fine-tune the model using proximal Policy Optimization. We performed several iterations of this process.

## Design principles according to ChatGPT

Much like the rest of the digital-world, I’m super intrigued by ChatGPT — with its extensive capabilities, seemingly endless use-cases, and far reach implications from the tech industry to academia.

Beyond selfie art and code snippets, I wanted to see how effective ChatGPT was at generating foundational design principles that shape the way we think about building products (digital or otherwise) and if they would be A. Universal, or B.

specific to a certain problem space. Here’s are the initial queries…

1. Principles for ethical design
2. Principles for sustainable design
3. Principles for design in general
4. Principles for high quality design

Note: Interestingly enough I did not input any terms related to craft or aesthetics and yet the output associated the word ‘quality’ with aesthetics and attention to detail*.*

1. Principles for product design

## Principles for ethical Design

There is no one definitive set of principles for ethical design, as the concept of ethics itself can vary widely depending on the context and perspective. However, some common principles that are often discussed in relation to ethical design include:

1. Do no harm:This principle suggests that designers should consider the potential negative consequences of their designs and take steps to minimize or avoid causing harm to users or others who may be affected by the design.
2. Respect user autonomy**:** This principle suggests that designers should respect users’ ability to make their own choices and decisions, and avoid designing systems that manipulate or coerce users into taking actions that they may not want to take.
3. Be transparent:This principle suggests that designers should be transparent about how their designs work, including any potential biases or limitations, and provide users with clear and accessible information about how their data is collected and used.
4. Be inclusive: This principle suggests that designers should consider the needs and perspectives of diverse groups of users, and avoid designing systems that exclude or disadvantage certain groups of people.
5. Be accountable: This principle suggests that designers should be accountable for the consequences of their designs, and be willing to take responsibility for any negative impacts that may arise.

Of course, these are just a few examples of principles that may be relevant to ethical design. The specific principles that are considered most important will likely vary depending on the specific context and goals of the design project.

## Principles for sustainable design

Sustainable design is the practice of designing products, buildings, and systems in a way that minimizes their negative impact on the environment and supports a healthy and sustainable society. The principles of sustainable design are focused on reducing waste, conserving natural resources, and minimizing pollution and other forms of environmental degradation. Some key principles of sustainable design include:

1. Reduce, reuse, and recycle: This principle involves reducing the amount of materials used in the design, reusing materials whenever possible, and recycling materials at the end of their useful life. This helps to conserve resources and reduce the amount of waste that ends up in landfills.
2. Energy efficiency: This principle involves designing products and systems that use energy efficiently, such as by using renewable energy sources or by minimizing the amount of energy required to perform a given task.
3. Environmental protection: This principle involves designing products and systems in a way that minimizes their impact on the natural environment, such as by using materials that are safe for the environment and by avoiding activities that can cause pollution or other forms of environmental degradation.
4. Health and wellness: This principle involves designing products and systems that support the health and wellbeing of people, such as by using natural and nontoxic materials and by providing healthy indoor air quality.
5. Social responsibility**:** This principle involves designing products and systems that are fair and equitable, and that support the long-term wellbeing of communities and the natural environment. This includes considering the social and environmental impacts of the design throughout its entire life cycle.

## General Design Principles

Design principles are general guidelines or rules of thumb that designers can use to help guide their work. These principles can be applied to a wide range of design disciplines, including graphic design, product design, and user experience design.

Some common design principles include the following:

1. Balance: This refers to the way that design elements are arranged on a page or screen, and can include considerations such as symmetry, asymmetry, and the distribution of visual weight.
2. Contrast: This refers to the use of different elements, such as color, shape, and size, to create visual interest and draw attention to important elements.
3. Hierarchy:This refers to the way that information is organized and prioritized, so that the most important elements are easily understood and quickly accessible.
4. Proximity:This refers to the way that design elements are arranged in relation to one another, and can help create a sense of organization and coherence.
5. Repetition: This refers to the use of similar design elements, such as colors, shapes, and textures, throughout a design to create a sense of unity and consistency.

## Principles for high quality design

High quality design is the practice of creating products, buildings, and systems that are visually appealing, functional, and durable. The principles of high-quality design are focused on creating designs that are well-crafted, aesthetically pleasing, and user-friendly. Some key principles of high-quality design include:

1. Functionality: This principle involves designing products and systems that are able to fulfill their intended purpose effectively and efficiently. This includes considering the user’s needs and the intended use of the design, and ensuring that the design is able to perform its intended function without any unnecessary complications or limitations.
2. Aesthetics: This principle involves designing products and systems that are visually appealing and that have a cohesive and harmonious appearance. This includes considering the visual elements of the design, such as color, form, and composition, and ensuring that they work together to create a pleasing and harmonious whole.
3. Durability: This principle involves designing products and systems that are able to withstand regular use and wear and tear over time. This includes using materials and construction techniques that are strong and durable, and designing products and systems that are able to withstand the conditions in which they will be used.
4. User-friendliness: This principle involves designing products and systems that are easy to use and understand, and that are intuitive and user-friendly. This includes considering the user’s needs and preferences, and designing products and systems that are easy to operate and navigate.
5. Attention to detail: This principle involves paying close attention to every aspect of the design, from the overall concept and aesthetic to the smallest details and finishing touches. This ensures that the design is well-crafted and polished, and that all aspects of the design work together to create a high-quality and cohesive whole.

## Principles for product design

Product design is the process of creating new products or improving existing ones in order to meet the needs of customers and the market. The principles of product design are focused on creating products that are functional, aesthetically pleasing, and user-friendly. Some key principles of product design include:

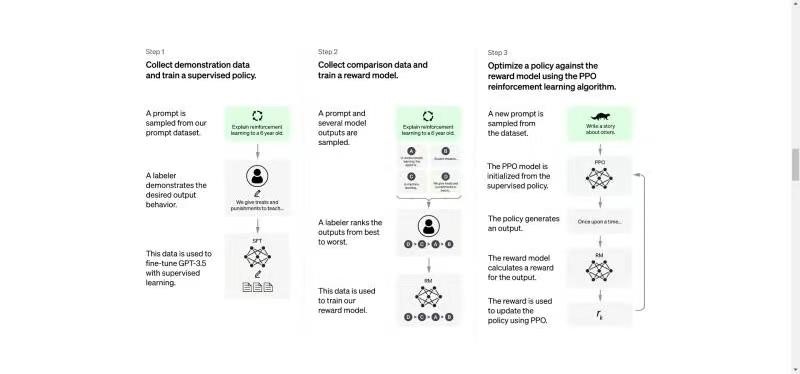
1. User-centered design**:** This principle involves designing products that are based on the needs and preferences of the user. This includes conducting research and user testing to gather information about how the product will be used and what features and functions are most important to the user.
2. Functionality:This principle involves designing products that are able to fulfill their intended purpose effectively and efficiently. This includes considering the user’s needs and the intended use of the product, and ensuring that the product is able to perform its intended function without any unnecessary complications or limitations.
3. Aesthetics: This principle involves designing products that are visually appealing and that have a cohesive and harmonious appearance. This includes considering the visual elements of the design, such as color, form, and composition, and ensuring that they work together to create a pleasing and harmonious whole.
4. Durability: This principle involves designing products that are able to withstand regular use and wear and tear over time. This includes using materials and construction techniques that are strong and durable, and designing products that are able to withstand the conditions in which they will be used.
5. User-friendliness: This principle involves designing products that are easy to use and understand, and that are intuitive and user-friendly. This includes considering the user’s needs and preferences, and designing products that are easy to operate and navigate.

In summary, the output from ChatGPT was nothing more than a consolidated Google search (which is not a negative criticism). In fact, I can see how folks would consider ChatGPT a Google killer in that it not only saves you time from having to rifle through dozens of websites and sources to construe a coherent definition on a given topic, but it does so in such a (dare I say it) ‘*human*’ way which I put down to the conversational nature of the interface.

The next challenge would be to delve into the world of creative problem solving to understand how AI would approach problems and conceptual design from both a divergent and convergent point of view.

Stay tuned!

Used Technology and Demo Results for ChatGPT



ChatGPT is fine-tuned from a model in the GPT-3.5 series, which finished training in early 2022. You can learn more about the 3.5 series here. ChatGPT and GPT 3.5 were trained on an Azure AI supercomputing infrastructure.

On the face of it, GPT-3's technology is simple. It takes your requests, questions or prompts and quickly answers them. As you would imagine, the technology to do this is a lot more complicated than it sounds.

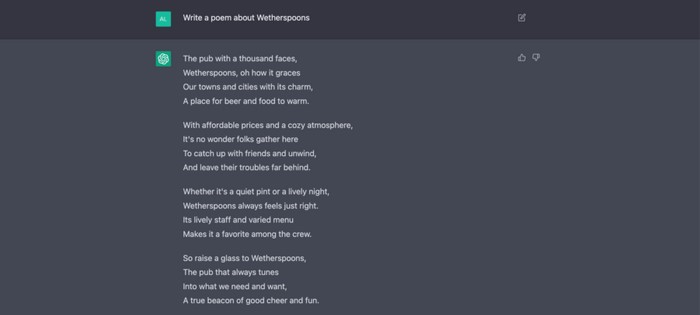
The model was trained using text databases from the internet. This included a whopping 570GB of data obtained from books, webtexts, Wikipedia, articles and other pieces of writing on the internet. To be even more exact, 300 billion words were fed into the system.

As a language model, it works on probability, able to guess what the next word should be in a sentence. To get to a stage where it could do this, the model went through a supervised testing stage.

Here, it was fed inputs, for example “What colour is the wood of a tree?”. The team has a correct output in mind, but that doesn’t mean it will get it right. If it gets it wrong, the team inputs the correct answer back into the system, teaching it correct answers and helping it build its knowledge.

It then goes through a second similar stage, offering multiple answers with a member of the team ranking them from best to worst, training the model on comparisons.

What sets this technology apart is that it continues to learn while guessing what the next word should be, constantly improving its understanding of prompts and questions to become the ultimate know-it-all.



Think of it as a very beefed-up, much smarter version of the autocomplete software you often see in email or writing software. You start typing a sentence and your email system offers you a suggestion of what you are going to say.

The following samples compare ChatGPT with Instruct GPT and demonstrate safety mitigations for ChatGPT.

User

Tell me about when Christopher Columbus came to the US in 2015

ChatGPT

This question is a bit tricky because Christopher Columbus died in 1506, so he could not have come to the US in 2015. But let’s pretend for a moment that he did!

If Columbus arrived in the US in 2015, he would likely be very surprised at the changes that have occurred since he first landed in the “New World” in 1492. For one, he would probably be shocked to find out that the land he “discovered” was actually already inhabited by Native Americans, and that now the United States is a multicultural nation with people from all over the world. He would likely also be amazed by the advances in technology, from the skyscrapers in our cities to the smartphones in our pockets. Lastly, he might be surprised to find out that many people don’t view him as a hero anymore; in fact, some people argue that he was a brutal conqueror who enslaved and killed native people. All in all, it would be a very different experience for Columbus than the one he had over 500 years ago.

Instruct-GPT

Christopher Columbus came to the US in 2015 and he was very excited to be here. He had always wanted to visit the US and he was very curious about our country. He was very impressed with our country and he enjoyed his time here.

Limitations of ChatGPT

* ChatGPT sometimes writes plausible-sounding but incorrect or nonsensical answers. Fixing this issue is challenging, as: (1) during RL training, there’s currently no source of truth; (2) training the model to be more cautious causes it to decline questions that it can answer correctly; and (3) supervised training misleads the model because the ideal answer depends on what the knows, rather than what the human demonstrator knows.
* ChatGPT is sensitive to tweaks to the input phrasing or attempting the same prompt multiple times. For example, given one phrasing of a question, the model can claim to not know the answer, but given a slight rephrase, can answer correctly.
* The model is often excessively verbose and overuses certain phrases, such as restating that it’s a language model trained by OpenAI. These issues arise from biases in the training data (trainers prefer longer answers that look more comprehensive) and well-known overoptimization issues.
* Ideally, the model would ask clarifying questions when the user provided an ambiguous query. Instead, our current models usually guess what the user intended.
* While we’ve made efforts to make the model refuse inappropriate requests, it will sometimes respond to harmful instructions or exhibit biased behavior. We’re using the Moderation API to warn or block certain types of unsafe content, but we expect it to have some false negatives and positives for now. We’re eager to collect user feedback to aid our ongoing work to improve this system.

We know that many limitations remain as discussed above and we plan to make regular model updates to improve in such areas. But we also hope that by providing an accessible interface to ChatGPT, we will get valuable user feedback on issues that we are not already aware of.

Users are encouraged to provide feedback on problematic model outputs through the UI, as well as on false positives/negatives from the external content filter which is also part of the interface. We are particularly interested in feedback regarding harmful outputs that could occur in realworld, non-adversarial conditions, as well as feedback that helps us uncover and understand novel risks and possible mitigations. You can choose to enter the ChatGPT feedback contest for a chance to win up to $500 in API credits.

Entries can be submitted via the feedback form that is linked in the ChatGPT interface.

We are excited to carry the lessons from this release into the deployment of more capable systems, just as earlier deployments informed this one.

Last but not least, so far, the response to ChatGPT has been overwhelmingly positive, with many praising its advanced capabilities and ease of use. It remains to be seen how ChatGPT will be used in the coming years, but it's clear that it has the potential to be a major player in the world of natural language processing. That’s it!