ChatGPT: An Adversary or Someone You Trust?

Arpan Dahal, 21422011

The University of Northampton, [arpan.2021136@nami.edu.np](mailto:arpan.2021136@nami.edu.np)

**Abstract**

This report outlines an overview of chatbots and chatGPT, numerous applications, including customer service, personal assistants, and language learning, have made extensive use of chatbots whereas ChatGPT has the ability to provide replies that are more nuanced and human-like and it may be applied to many different things, such as chatbots, virtual assistants, and translating languages. The usage of ChatGPT and chatbots, as well as cross-referencing information from other sources and expressing my own thoughts. I go over the history, working mechanism, the potential advantages, and drawbacks of these conversational ChatGPT as well as any potential moral and legal problems that may result from its deployment.

**Keywords:** AI, ML, NLU, Chatbots, GPT,  ChatGPT

**1. Introduction**

As intelligent software and hardware, also known as intelligent agents, are developed and analyzed, artificial intelligence (AI) becomes more and more integrated into our daily lives. One of the most simplest and well-known forms of intelligent Human-Computer Interaction (HCI), a chatbot is a typical example of an AI system[1]. A chatbot is defined as a dialogue-capable piece of software that can converse with people and interpret their natural language input. When communicating with a chatbot via text or speech, the computer software acts intelligently and uses Natural Language Processing (NLP) to understand one or more human languages[2].

Chatbots are often referred to as intelligent bots, conversational agents, dynamic assistants, or digital personal assistants. They are beneficial in applications like online commerce, the business department, education, and data searching[1].

**1.1 Brief History**

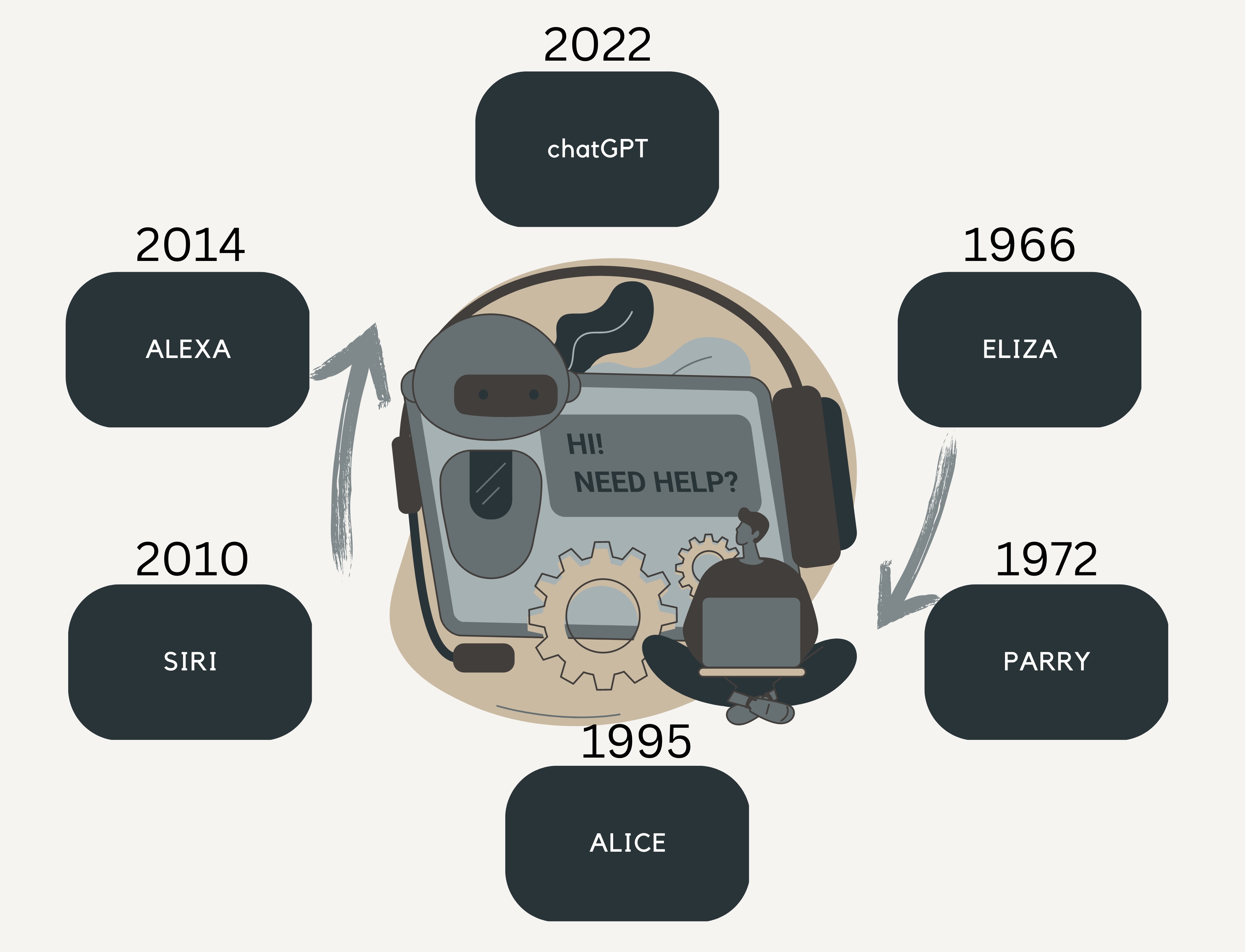
Chatbots have a long history that dates all the way back to Alan Turing in the 1950s. Professor Joseph Weizenbaum of MIT created the first chatbot ever in the 1960s, named ELIZA following the introduction of the Turing Test[2]. It employed a straightforward template-based response system and matching of patterns. Their conversational skills were poor, but they were effective to confuse humans during a period when they were not accustomed to talking with machines, which inspired them to begin creating more chatbots[4]. A chatbot designated as PARRY that was created in 1972 was an upgrade over ELIZA. It is a program that uses natural language and simulates human thought processes. After that, in 1995, Richard Wallace was a forerunner in the development of ALICE which was the first program to run on a computer called Alice, it was once known as Alicebot. In 2000, 2001, and 2004, Alice earned the Loebner Prize, which is awarded for passing the Turing Test. The term "almost human computer" was initially used for this machine[6].

Figure 1: Evolution of Chatbots

Likewise, in 2001, the invention of the SmartChild served as a forerunner to Siri with the ability to have engaging discussions while having easy access to other services' data. Later, the account suits a certain discourse. In 2010, Apple created Siri for iOS. A few instances of a dialogue between Siri and a user in Messages during which the user asks inquiries are shown in the patent. Amazon created Alexa, a highly intelligent virtual assistant in 2014[2]. The Alexa Skills Kit (ASK) from Amazon enables developers to create and publish skills for Alexa that can be added to any Alexa-enabled device to increase its functionality. Between 1960s-now, many more chatbots were introduced and some got highly appreciated when other bots stopped. Currently, the large-scale language model ChatGPT was developed by OpenAI. It is intended to help users create writing that appears human-like from the input[1].

**1.2 Relevance of Chatbots**

In the current digital world, chatbots are growing more and more significant for a number of objectives because they can connect with people quickly, effectively, and individually[4]. Here are some significances of them:

* Customers may start a chatbot discussion at any moment and get an immediate answer if they have any queries or concerns about a product or service. This has improved accessibility and efficiency for the customer service experience.
* A huge amount of simultaneous inquiries may be handled by chatbots, eliminating the need for human customer care agents. Businesses may be able to devote resources to other areas and save money as a result[3].
* In the healthcare sector, chatbots can offer medical advice and assistance, bridging the gap between the patient and the healthcare professional, particularly in rural and distant places where access to healthcare may be constrained[5].

**1.3 Variations of Chatbots**

Chatbots may be categorized using a variety of factors. At the moment, different sectors are using various forms of chatbots. Let’s talk about widely used chatbot types[1].

Voice chatbots: A vocal bot is an AI- and NLU-powered communication channel that can convert voice to text and text to speech. The entire speech understanding and response process is carried out by these bots in a manner that closely resembles that of a person.

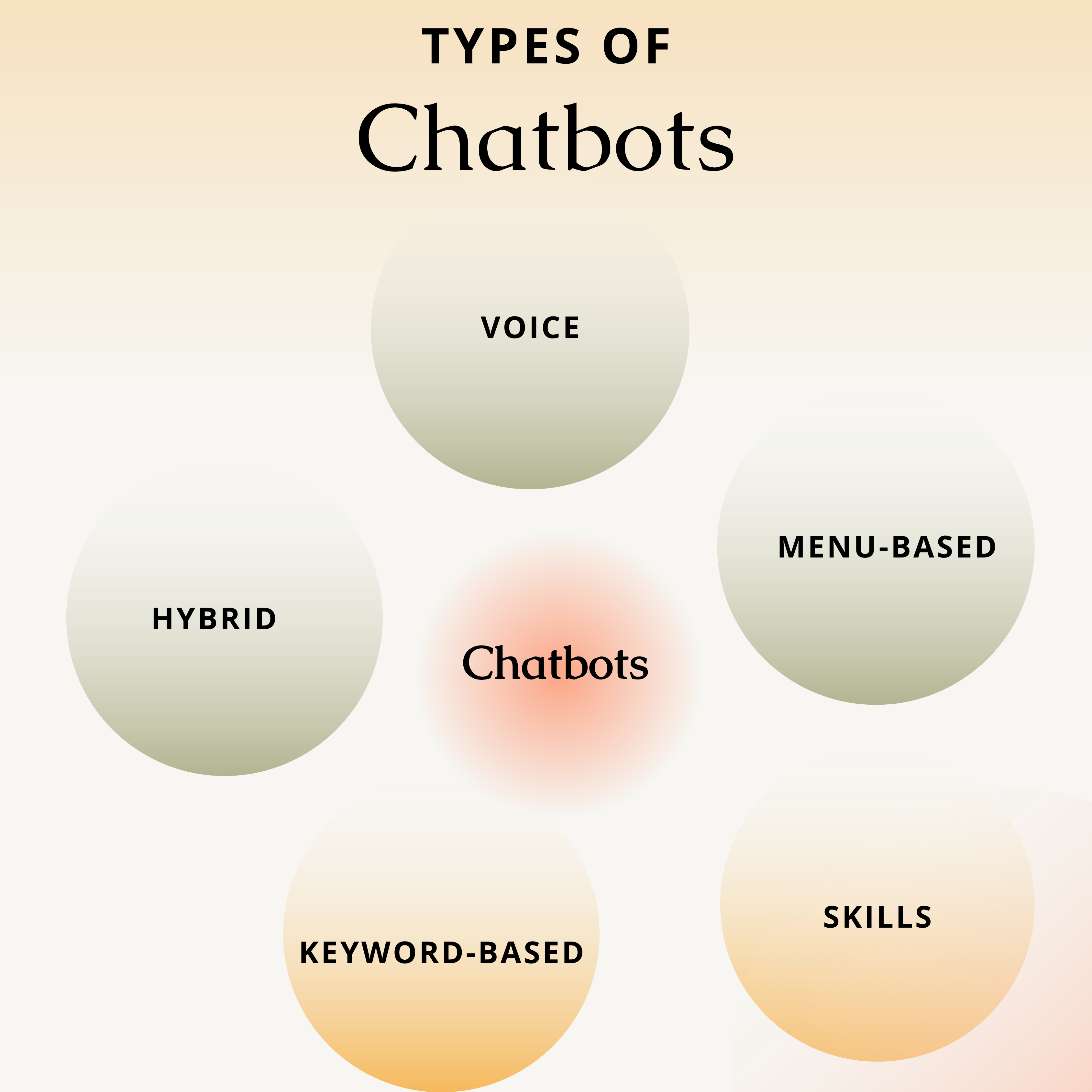
Hybrid chatbots: The greatest features of both chatbots and live chat are harmoniously combined in a hybrid chatbot. Any client questions that may be too complicated or nuanced for automation alone can be addressed by a customer support agent during live chat[3].

Figure 2: Types of Chatbots

Key-based chatbots: Keyword-based chatbots can hear what users put in and reply correctly, in contrast to those that rely on menus. These chatbots utilize natural language processing (NLP) and programmable keywords to recognize activity triggering in the conversation and decide how to respond to the audience[6].

Skills chatbots: A skills chatbot can carry out a certain set of activities after its capabilities have been expanded using pre-defined skills software. For instance, when linked to a smart home appliance, the chatbot may be able to switch down your room's lights, deliver weather information, shop for groceries online, and perform other functions[3].

**1.4 ChatGPT**

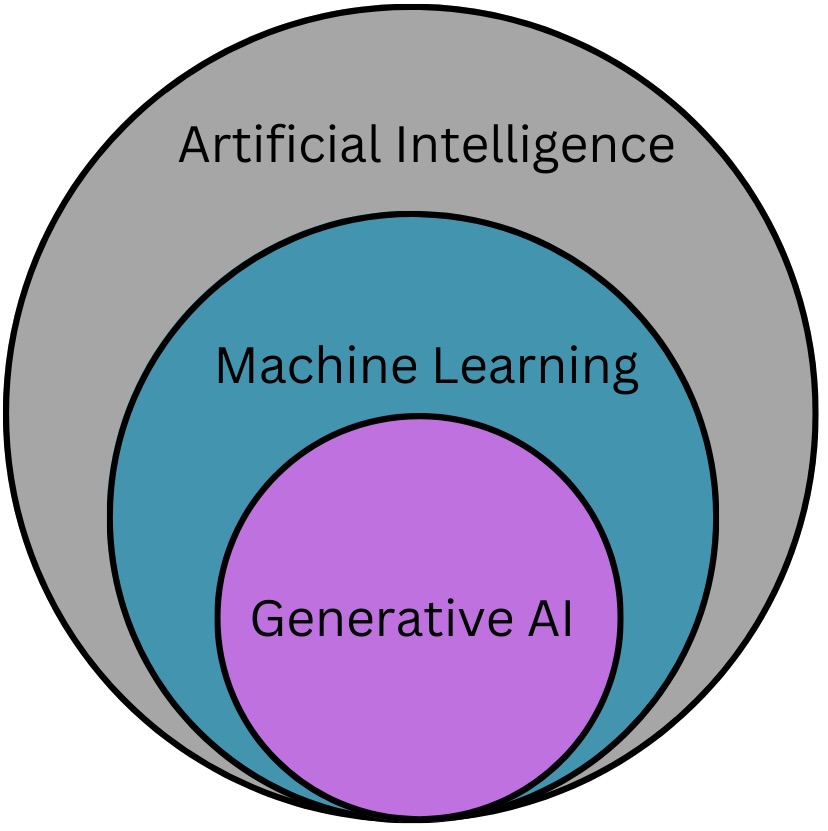
As a natural language processing model, ChatGPT from OpenAI can respond to queries and carry out activities that need text-based input using deep learning. ChatGPT is far more generalizable since it interprets the query or task using its linguistic knowledge to get the best possible answer[6].

Figure 3: Generative AI is a subset of AI and ML

With a history dating back to the introduction of GPT-1 in 2018, ChatGPT has grown to become one of the most prominent technological debuts in recent years. GPT-2, released in 2019, has 1.5 billion parameters and is considerably more potent. Conversely, ChatGPT has a bigger model with 175 billion parameters, enabling it to produce replies to user inquiries that are more precise and human-like. Although both ChatGPT and chatbots are interactive AI tools, they serve different functions. While ChatGPT employs a large language model trained on a ton of data, chatbots often use a rule-based or decision tree-based technique to create replies to user input. Compared to conventional chatbots, ChatGPT is more versatile and produces replies that sound more natural[1].

**1.5 Working Mechanism of ChatGPT**

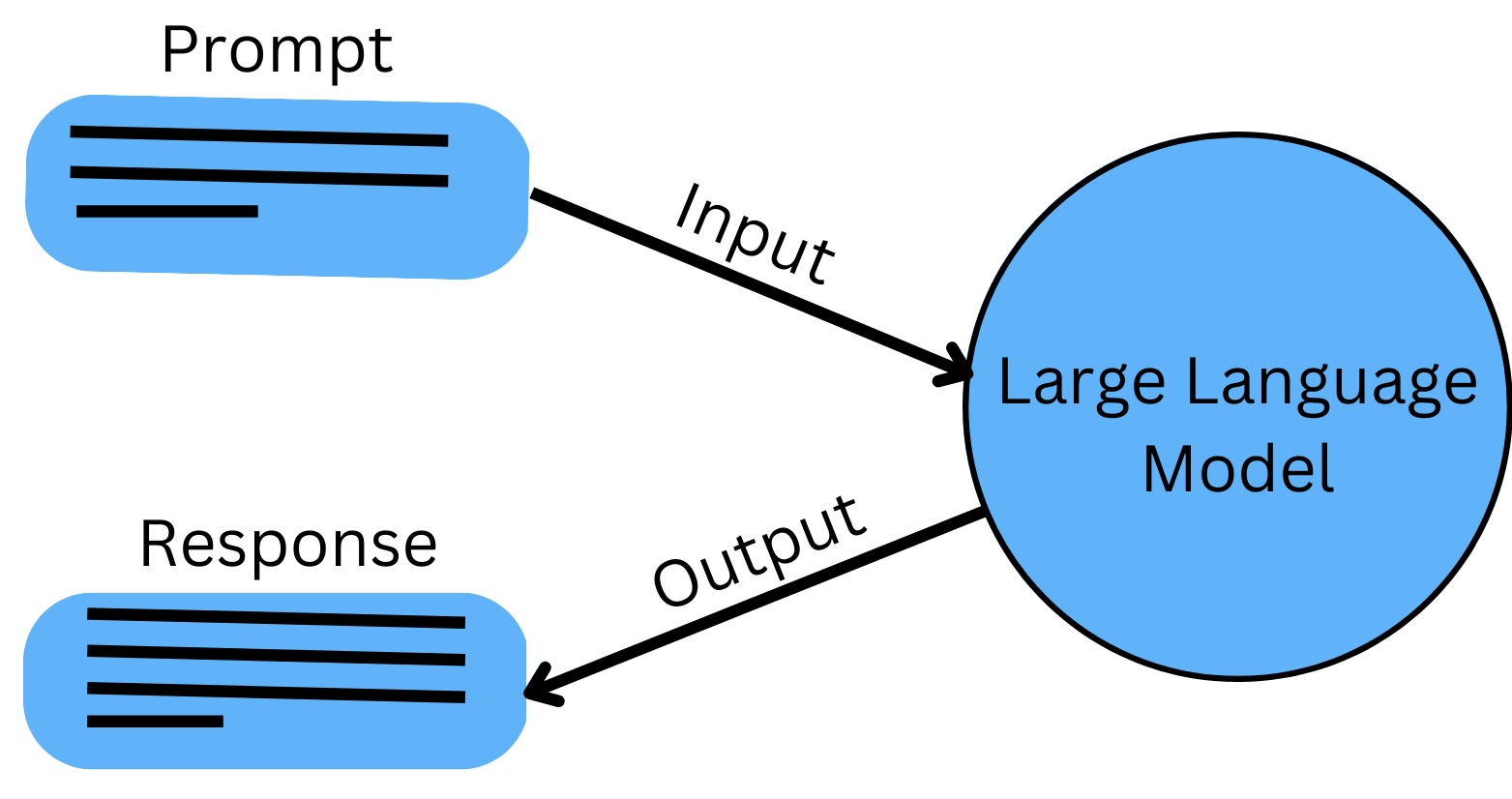
A good example of a generative AI model is ChatGPT. A model may produce new material using generative AI, a subset of artificial intelligence and machine learning, by drawing on patterns in data that it has already seen[2].

Figure 4: Simple Working Mechanism of ChatGPT

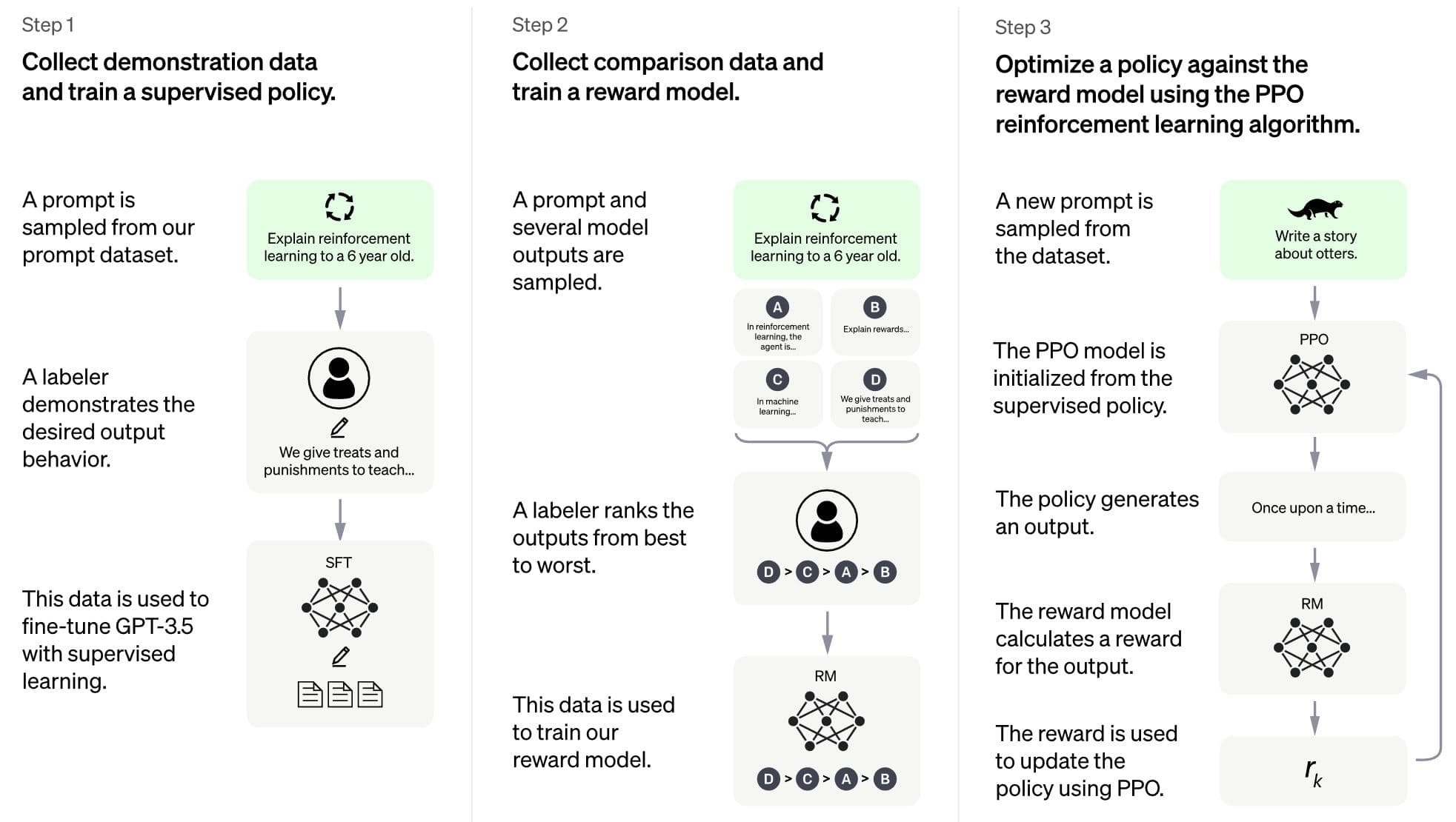
A generative language model called ChatGPT is built on the "transformer" architecture. GPT requires a lot of text to be "trained" on in order for it to function. Conversations between the user and the AI assistant were given by human AI trainers. It generates a response based on its linguistic expertise and comprehension of the situation.  It accomplishes this by dissecting the text input into its component words before processing and producing a response using a multi-layer neural network[4].

Figure 5: Work Method of ChatGPT[5]

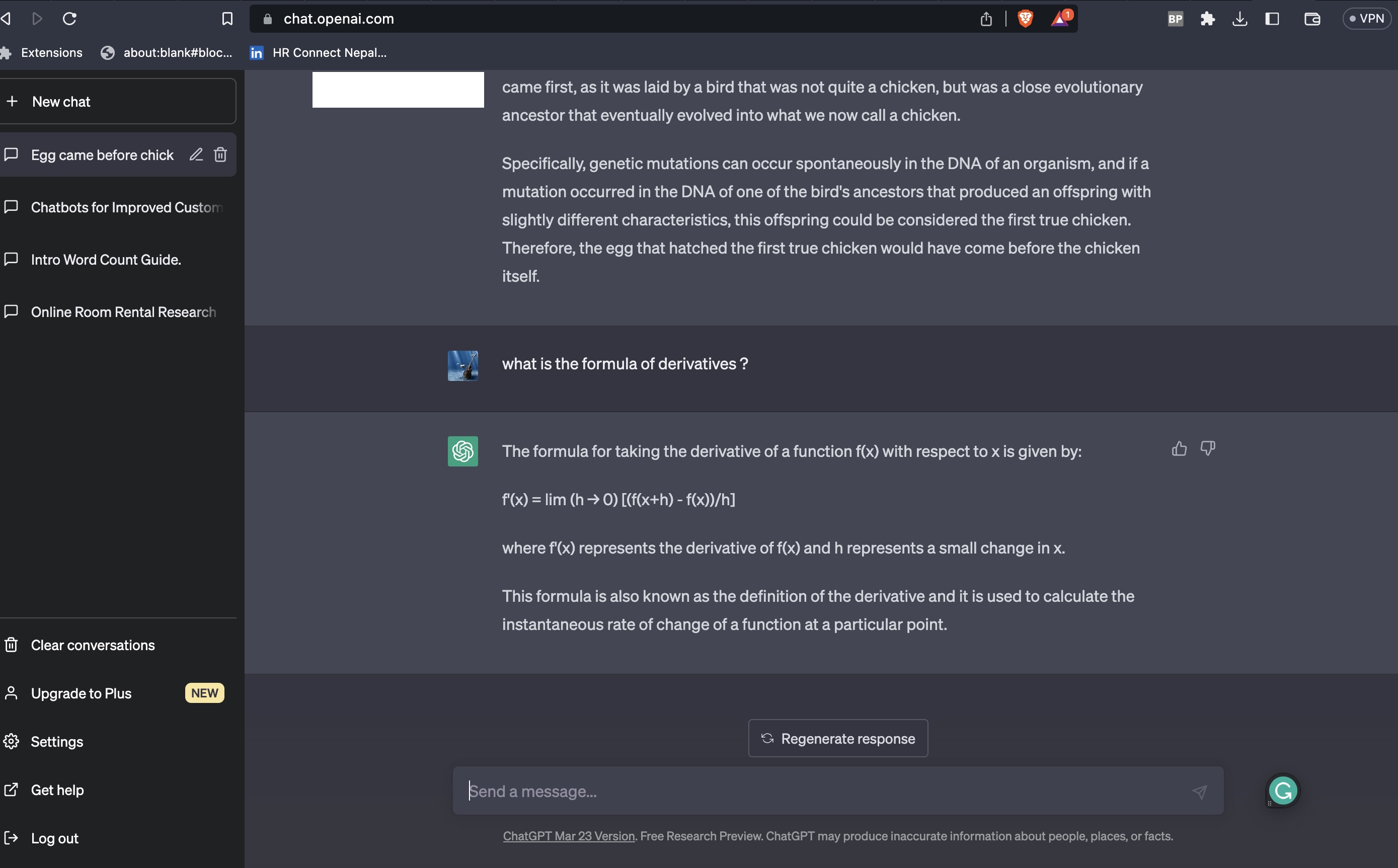
Because of the GPT design, ChatGPT may gradually produce more insightful replies by taking into account the conversation's context. It can comprehend and produce natural language, which enhances its ability to interact with people. The stimulus is then interpreted using these patterns, and fresh, pertinent text is produced in response. Finally, the user receives the created text or answer. Let's look at some instances of ChatGPT in use[4].

Figure 6: Response of ChatGPT

**2. Literature Review**

According to recent research, ChatGPT is a formidable chatbot powered by AI that has the capacity to revolutionize a number of areas, including education, health, medicine, business, and research[7].

This paper is presented from a research journal(Puranjay Savar Mattas 2023) that attempted to better understand the advantages and difficulties of AI language processing and its potential influence on many businesses via a thorough assessment of the literature and case studies[7]. Similarly, Lund and Ting(2023) observed that the advantages of ChatGPT, such as enhanced search and discovery, reference and information services, production of catalogs and metadata, and material creation, as well as the moral issues that must be taken into consideration, such as prejudice and security[8]. Both journals[7, 8] were conducted in the rush to produce new scholarly knowledge and train the next generation of professionals, it is crucial to think about how to use this technology responsibly and ethically and to discover how we, as professionals, can work alongside it to improve our work rather than abusing it or allowing it to abuse us.

**2.1 Industry Applications**

When it comes to education, in order to produce instructional content and improve student interactive learning, ChatGPT has attracted attention. According to the research by Lund and Ting(2023), high-quality instructional material that is tailored to the requirements of specific pupils may be produced using ChatGPT. In the medical sector, advanced analytics techniques like machine learning, and predictive modeling may be combined with digital health data to find patterns, trends, and insights that might enhance patient outcomes and lower costs. A study by Mijwil et al.(2023)  found that patients with anxiety and depression can benefit from effective assistance through ChatGPT. Within the scientific field, researchers have utilized ChatGPT to generate academic papers and support their studies. By researchers, Tiwary et al.(2023) conducted that researchers can benefit from using ChatGPT to develop hypotheses and plan studies. In the field of organizations, it may be used in business to enhance relations with clients, automate processes, and spark innovation. A study by Lund and Ting(2023) found that we may utilize ChatGPT to come up with fresh concepts for products or marketing plans.

**2.2 NLP Applications**

Numerous NLP applications, including content creation, text summarization, machine translation, and question-answering, have made use of ChatGPT[1]. When it comes to creating content, high-quality material has been produced with ChatGPT for a variety of uses, including social media, news stories, and blog entries. ChatGPT can provide engaging social media postings that effectively engage people which was conducted by Paul et al.(2023). Text summarization is a discipline in which long papers have been summarized accurately and succinctly using ChatGPT. Summaries produced by ChatGPT can be as good as those produced by humans, as studied by Lund and Ting(2023). Additionally, it has been demonstrated that ChatGPT's machine translation feature performs better than conventional machine translation techniques. They observed(Tiwary et al. 2023) observed that modern machine translation performance is possible using ChatGPT. in the area of query resolution, chatGPT has been used to generate natural language answers to inquiries. Both (Tiwary et al.2023; Paul et al.2023) had done research on chatGPT that can answer questions with a high degree of accuracy. But the researchers Mijwil et al.(2023) didn’t elaborate on it.

**2.3 Benefits, Drawbacks, and Limitations of ChatGPT**

The advantages of ChatGPT include its capacity to produce language that is human-like, flexibility in many areas, and potential to increase worker productivity and effectiveness, 24/7 availability,  personalization, and cost-effectiveness[13]. Researchers Ding and Lin(2023) observed that it provides a number of advantages, including higher efficiency, more accuracy, and cost savings. Nevertheless, there are also issues with ChatGPT's limitations, including its need for a lot of training data, inability to comprehend context or reason, and potential for prejudice. It's critical to establish strategies to address these possible ChatGPT ethical and societal problems. According to the research by Paul et al.(2023) said that a lack of emotional intelligence, biased answers, a lack of subject expertise, a susceptibility to adversarial attacks, and a lack of explainability, among others.

**2.4 Technical Implementation of ChatGPT**

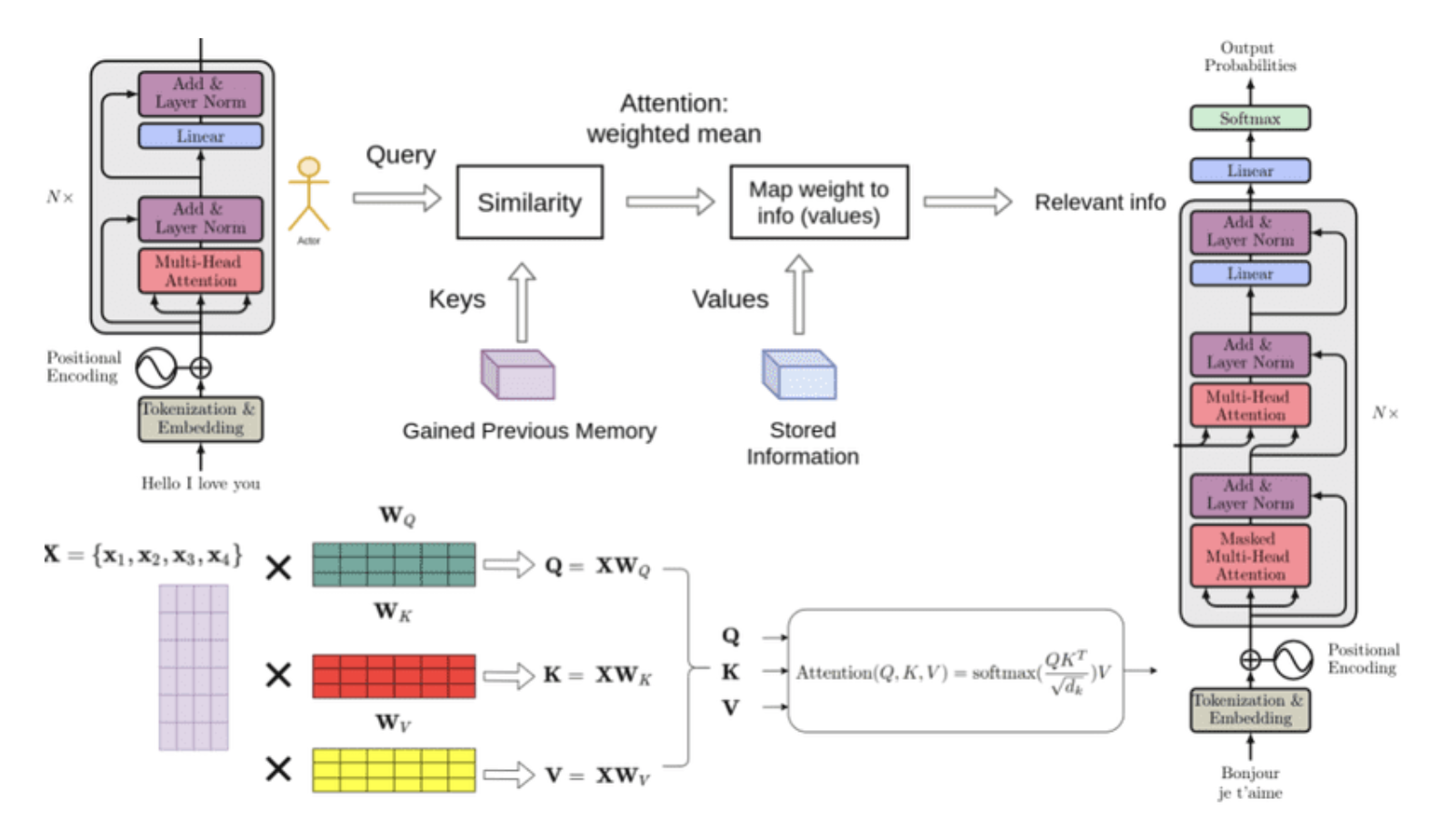
In order to build ChatGPT technically, a transformer-based architecture with self-attention techniques was used. The model's use of the transformer architecture's several layers enables efficient processing of lengthy sequences[4]. For language processing tasks like interpreting and producing natural language, the transformer model is a sort of neural network design. Deng and Lind(2023) introduced it is now a widely used technique for language modeling.

Figure 7: Working Mechanism of ChatGPT[12]

In the transformer architecture, an embedding layer places the input token sequence (words or subwords) into a high-dimensional space. The embedded sequence is processed by encoder and decoder layers. Each word in the input sequence is ranked by the transformer model encoder layer using a multi-head self-attention method. The output of the self-attention mechanism is converted into encoder output by a feedforward neural network. The decoder layer of the transformer model may attend to the output of the encoder thanks to multi-head self-attention and an attention mechanism. The output of the decoder is routed via a feedforward neural network to create the final output sequence. The model is trained unsupervised on a sizable corpus of text using the GPT (Generative Pre-trained Transformer) architecture[13]. During training, the model develops the ability to predict the next words based on previous words. The model gains a broad understanding of language from this pre-training, which may then be honed for particular language tasks like question-answering or summarizing[15].

**2.5 Comparison Between ChatGPT and Other AI-based Chatbots**

We will evaluate ChatGPT's performance, accuracy, and efficiency against those of other AI-based chatbots[16]. Aljanabi and M. (2023) compared several natural language processing (NLP) tasks including text categorization, question answering, and language production, ChatGPT competes with other cutting-edge language models, such as BERT and XLNet. According to the study, BERT and XLNet beat other models on classification and question-answering tasks whereas ChatGPT did better on language production tasks. According to research by Lee and H.(2023) differentiated on a healthcare application, ChatGPT was tested with a number of other chatbots to determine their efficacy and accuracy in giving patients medical advice. According to the study, ChatGPT is a potential tool for healthcare applications since it performs more accurately and with a faster reaction time than other chatbots[14].

**2.6 Performance Issues**

ChatGPT has received accolades for its capacity to make coherent replies in a variety of circumstances and generate natural-sounding English. It also has a number of performance problems, though, that must be resolved.

The computational demands of ChatGPT are one of the key performance challenges. ChatGPT demands a huge amount of computing resources to train and execute since it is a massive model with millions of parameters[9]. ChatGPT may be difficult for firms without access to high-performance computing resources, making it less available to people and smaller companies. A study by Deng and Lin(2023) showed a huge carbon footprint might result from training big models like ChatGPT, which can be problematic for the environment. ChatGPT's propensity to create biased or abusive language is another performance flaw. Large volumes of text data may have biases that ChatGPT picks up as it learns, which might result in the development of prejudiced or objectionable language. A study by Paul et al.(2023) found that ChatGPT's propensity to create biased or abusive language is another performance flaw. Large volumes of text data may have biases that ChatGPT picks up as it learns, which might result in the development of prejudiced or objectionable language[11].

**3. Analysis**

I believe that the model is built on the cutting-edge GPT-3.5 architecture, a language model that was trained on a sizable quantity of text data while GPT-4 only the premium ChatGPT Plus membership offers it and is more precise and competent[9].

From my point of view, one of ChatGPT's attractions is its capacity to provide replies to queries and prompts that are human-like. The flexibility of ChatGPT to accommodate various communication settings and styles is another asset. The model is capable of producing replies in conversational, instructive, and persuasive styles. Due to its adaptability, it may be utilized for a variety of purposes, including customer service, research, medical, and teaching[1]. The model may produce erroneous or biased results in some situations since its answers are dependent on statistical patterns in the training data. Furthermore, the model could not always comprehend the subtleties of spoken languages, such as paradox, comedy, and sarcasm. I find[10, 11] in order to keep stakeholders' trust and make sure that moral norms are respected, transparency and accountability should come first. I concur with the authors of both publications that it is crucial to take ethical considerations into account while doing research and development, especially in fields like social media and AI where using personal data is frequently involved[16]. I'll also provide my personal views on the raised concerns.

**3.1 Ethical Issues**

I admit that the ethical issues related to social media platforms, such as those with informed permission, privacy, and the use of individual data. We may give ChatGPT access to a lot of information about ourselves, including private discussions, and this information may be used to monitor us, send us advertisements that are relevant to us, or even hurt us[8]. From both papers[13, 14] I assume that the enormous text and code dataset used to train ChatGPT may be skewed. As a result, ChatGPT could produce text that is prejudiced or discriminating. Decisions made by ChatGPT have a genuine influence on the globe, and if they are made incorrectly, the results might be terrible. For instance, deep fakes or fake news might be produced using ChatGPT. In my opinion, it is crucial to make sure ChatGPT is trained on a variety of representative data sets and that the model is consistently validated for bias and accuracy in order to meet these ethical concerns. Users should also receive training on how to communicate with the model in an ethical and inclusive manner. Last but not least, adequate steps should be made to stop the dissemination of dangerous or false information via ChatGPT[13]. The authors[13, 14] contend that researchers must get participants' informed permission and be open and honest about their procedures.

**3.2 Trust Issues**

I think despite the fact that the model was trained on a sizable dataset and is intended to deliver logical and pertinent answers, there may be some circumstances in which ChatGPT's responses are incorrect, deficient, or deceptive. Users may become distrustful of the model and start to doubt its dependability as a result[4]. Papers from[8, 9] argue that machine learning models must be able to justify their choices if they are to be taken seriously. Users have the capacity to abuse ChatGPT. Users may lose faith in the model and may suffer harm if they knowingly use it to disseminate false information, advertise hazardous content, or participate in other illegal actions. The trust problems with ChatGPT may also get worse if the model is used to produce deep fakes or other kinds of altered material. It is crucial to make sure that ChatGPT is consistently checked for correctness and dependability and that safeguards are in place to stop user abuse[11]. This can entail putting in place user rules or placing limitations on specific kinds of information, as well as routinely updating and upgrading the model to increase its accuracy and guard against any weaknesses. Additionally, being open and honest about ChatGPT's limits and the kinds of replies it might provide will build user trust and reduce the likelihood of misunderstandings or disinformation[10].

**3.3 Accountability Issues**

In my opinion, who is accountable for the model's conduct is the subject of ChatGPT. As an AI language model, ChatGPT responds to human input rather than acting independently. It could be challenging to pinpoint who is ultimately in charge of these results in situations where the model's reactions cause injury or have undesirable effects[16]. The use of ChatGPT may have unforeseen implications. Although the model is meant to offer pertinent and beneficial replies, there may be times when such responses have unexpected or undesirable results. For instance, it may be difficult to foresee larger societal repercussions if the model is used to produce false information or fake news. It could be required to define clearly defined lines of accountability and responsibility for the use of ChatGPT, especially when its usage may have effects that go beyond the individual users[4]. These papers[15, 16] discuss that the designers and users of ChatGPT have a duty to make sure that they are acting in a way that is compatible with ethical and moral standards since ChatGPT has the power to influence social norms and values.

**Conclusion**

In conclusion, it can be said that some of the most ground-breaking and revolutionary innovations in the fields of artificial intelligence and natural language processing include chatbots and ChatGPT. ChatGPT takes things a step further by applying machine learning to construct more sophisticated and nuanced dialogues. Chatbots have been extensively used by corporations and organizations to offer customer support and automate regular operations. The possible uses for ChatGPT are numerous and diverse, ranging from chatbots and customer support tools to content creation and language translation. It has the potential to completely change how we communicate with computers and automate numerous jobs thanks to its real-time comprehension and linguistic analysis capabilities. Its continued research and improvement will surely result in brand-new, interesting applications that have the potential to greatly enhance our daily lives.

**References**

1.Huang, W., Fryer, L. K., & Hew, K. F. (2021). Chatbots for Language Learning—Are They Really Useful? *A Systematic Review of Chatbot‐Supported Language Learning, 237–257.* [https://doi.org/doi/epdf/10.1111/jcal.12610](https://onlinelibrary.wiley.com/doi/epdf/10.1111/jcal.12610)

2. (2020). *An Overview of Chatbot Technology*, *584*. [https://doi.org/chapter/10.1007/978-3-030-49186-4\_31](https://link.springer.com/chapter/10.1007/978-3-030-49186-4_31)

3. Liu, D., Huang, W., & Lv, Y. (2023). How Do Consumers React to Chatbots' Humorous Emojis in Service Failures, 73. [https://doi.org/chapter/10.1007/978-3-030-49186-4\_31](https://www.sciencedirect.com/science/article/abs/pii/S0160791X23000490?via=ihub)

4. Luo, B., Li, C., Lau, R. Y. K., & Si, Y.-W. (2021). A Critical Review of State-of-the-Art Chatbot Designs and Applications . [https://doi.org/chapter/10.1007/978-3-030-49186-4\_31](https://www.citationmachine.net/bibliographies/9e7e4ff5-c4a0-4a37-b262-063d7057ee16)

5. Fedewa, J. (2023). *What Is ChatGPT, and Why Is It Important?* <https://howtogeek.com/871071/what-is-chatgpt/>

*6.* Lee, J. Y. (2023). Can an Artificial Intelligence Chatbot Be the Author of a Scholarly Article? [https://doi.org/chapter/10.1007/978-3-030-49186-4\_31](https://synapse.koreamed.org/articles/1516081874)

7. Mattas, P. S. (2023). CHATGPT: A study of ai language processing and its implications. *International Journal of Research Publication and Reviews*, *04*(02), 435–440. [https://doi.org/10.55248/gengpi.2023.4218](https://ijrpr.com/uploads/V4ISSUE2/IJRPR9921.pdf)

8. Lund, B., & Ting, W. (2023). Chatting about chatgpt: How may AI and GPT Impact Academia and libraries? *SSRN Electronic Journal*. [https://doi.org/10.2139/ssrn.4333415](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4333415)

9. Mijwil, M., Aljanabi, M., & Ali, A. H. (2023). Chatgpt: Exploring the role of cybersecurity in the Protection of Medical Information. *Mesopotamian Journal of Cyber Security*, 18–21. <https://doi.org/10.58496/mjcs/2023/004>

10. A, S., A, V., & Tiwary, N. (2023, February 1). Netizens, academicians, and information professionals' opinions about AI with special reference to chatgpt. arXiv.org. Retrieved April 25, 2023, from <https://arxiv.org/abs/2302.07136>

11. Paul, J., Ueno, A., & Dennis, C. (2023). chatgpt and consumers: Benefits, Pitfalls and Future Research Agenda. *International Journal of Consumer Studies*. <https://doi.org/10.1111/ijcs.12928>

12. Kumar, S. (2023, April 4). *All about chat gpt(version, datasizes, working) explained*. Curious Steve. Retrieved April 25, 2023, from <https://curioussteve.com/tech/explained/chat-gpt-explained/>

13. Deng, J., & Lin, Y. (2023). The benefits and challenges of chatgpt: An overview. *Frontiers in Computing and Intelligent Systems*, *2*(2), 81–83. <https://doi.org/10.54097/fcis.v2i2.4465>

14. Lee, H. (2023). The rise of chatgpt : Exploring its potential in medical education. *Anatomical Sciences Education*. <https://doi.org/10.1002/ase.2270>

15. Omar, R., Mangukiya, O., Kalnis, P., & Mansour, E. (2023, February 8). Chatgpt versus traditional question answering for knowledge graphs: Current status and future directions towards knowledge graph Chatbots. arXiv.org. Retrieved April 25, 2023, from <https://arxiv.org/abs/2302.06466>

16. Aljanabi, M. (2023). Chatgpt: Future directions and open possibilities. *Mesopotamian Journal of Cyber Security*, 16–17. <https://doi.org/10.58496/mjcs/2023/003>