A Chemical Engineer's Introduction to ChatGPT

Sean J. Dee, P.E. ▪ **Sahand Faraji** ▪ Exponent, Inc.

As OpenAI's ChatGPT gains popularity across different professions, chemical engineers have questions about if (and how) the tool works in practice.

Chatbots have captured the imagination of chemical engineers as they debate how artificial intelligence (AI) technologies can be leveraged to enhance industry practices. Many companies are exploring the deployment of AI-based tools such as OpenAI's ChatGPT or Microsoft Copilot, evaluating whether these technologies can improve the quality and efficiency of their employees' workstreams. As chemical engineers begin to experiment with these tools, they uncover interesting insights into the strengths, weaknesses, and potential applications of the technology.

However, users in the engineering space have reported a broad spectrum of first impressions, ranging from impressed to underwhelmed to highly skeptical. These tools are often "black box" technologies, and users without deep knowledge of the underlying research may struggle to recognize the technology's limitations, as well as the strategies for optimizing its output. This article aims to provide an introduction to these aspects of chatbots. The authors have used a variety of chatbot technologies, but this article focuses on OpenAI's ChatGPT. It reviews the recent updates and changes to ChatGPT announced by OpenAI in May 2024, compares paid and free ChatGPT plans, and discusses key questions facing the deployment of chatbot technologies such as ChatGPT. Finally, two illustrative case studies showcase the potential applications of the technology and its performance as a process safety tool.

How does ChatGPT work?

ChatGPT is based on a technology referred to as a large language model (LLM) *(1)*. As shown in Figure 1, when a user types an input prompt into ChatGPT, the technol-

ogy breaks the string of text into smaller pieces of data called tokens, which can be as short as one character or as long as one word. The model then analyzes the relationships between these tokens based on an internal analysis framework referred to as its architecture, which is typically a transformer model. This architecture uses a mechanism called "attention" to weigh the relevance of different tokens in the context of the input.

The architecture in ChatGPT is OpenAI's generative pre-trained transformer (GPT) *(2)*. This architecture is a type of artificial neural network that is inspired by biological neural networks where layers of interconnected nodes (*e.g.,* neurons connected by synapses) work together to relay and analyze information. OpenAI's GPT uses weighting factors and biases, referred to as parameters, to characterize the relationships between tokenized data from the input prompt. These parameters are developed by training GPT on large datasets of text to help the technology recognize underlying patterns and relationships between words within the dataset.

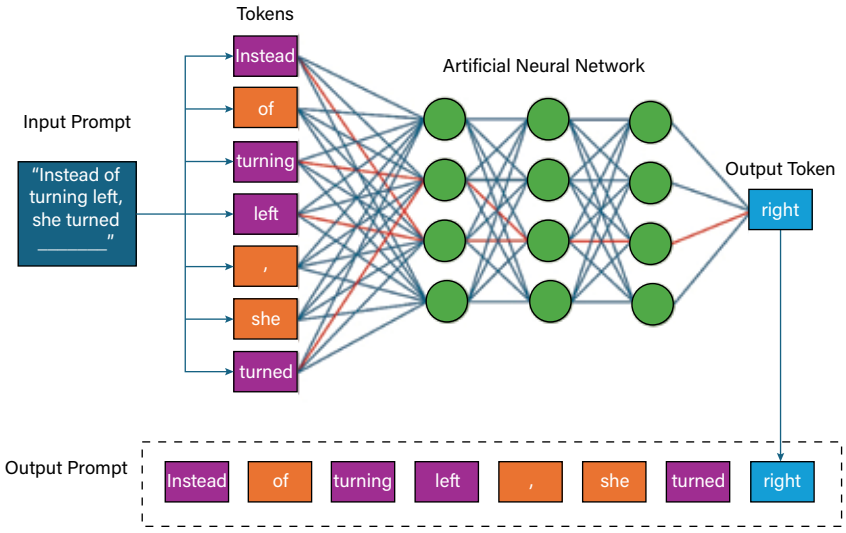
GPT is trained using both supervised and unsupervised machine learning *(3)*, with the difference being the use of labeled input and output datasets in supervised training (*e.g.,* providing a series of spam emails will help the tool recognize spam-specific patterns) and unlabeled datasets for unsupervised training. After training, the architecture can also be further fine-tuned through reinforcement learning from human feedback (RLHF), where human trainers rank or evaluate the technology's outputs to further refine the internal parameter set.

In the example prompt in Figure 1, the key relationships would likely be between the words "instead," "turned" and "turning," and "left," which establish the next token would likely be a directional word other than "left." The output from the analysis is the most likely next token in the sequence of text. This process is then iterated to continually append words onto the output. In practice, ChatGPT builds its output one word at a time based on its internal architecture and the context provided by the text in the input prompt.

Increasing and refining the parameters within the architecture is a key feature of improving the performance and functionality of the technology. The number of parameters in OpenAI's GPT architecture has expanded over time. In 2018, the first iteration of the GPT architecture included approximately 117 million parameters, while current versions boast 375 billion parameters in GPT-3.5 and 1.7 trillion in GPT-4 *(4–6)*.

Before training *(1)*, the tool may provide random outputs for the next word. After training on a text dataset and developing parameters to model and predict patterns, the technology is better equipped to predict more accurate outputs. For example, "around," "right," or "back" could be alternate output tokens for the example prompt shown in Figure 1. There are often multiple acceptable outputs when predicting the next word in a text sequence. As a result, the output from the technology may vary even when prompted with the exact same input text.

According to OpenAI *(1)*, there are three primary sources of data used for model training: publicly available information on the internet, licensed information from third parties, and information provided by users or human trainers. OpenAI states that the training dataset is filtered for inappropriate content (*e.g.,* hate speech, adult content, spam, personal information) and does not include information behind paywalls. Through training, the architecture develops and modifies its parameter set, and once training is completed, the architecture no longer has access to its training dataset. In other words, when a user interacts with



◀ **Figure 1.** The architecture of ChatGPT analyzes the relationships between the words in an input prompt based on an artificial neural network that is developed through training on large datasets. These relationships are used to determine the output.

ChatGPT, the tool generates text based on its architecture's understanding of the relationships between tokens in the input prompt and the context it has learned during training. It does not search, find, or retrieve specific pieces of information in its training data to generate its outputs, which is a common misconception of those unfamiliar with the underlying technology. Instead, it uses patterns and knowledge embedded in the model during training to produce coherent and contextually relevant responses.

Fundamentally, ChatGPT is a conversational text-generating tool designed to mimic human language. However, as the technology has developed over time, it is apparent that the applications and potential benefits extend beyond simply predicting the next word in a sentence.

What is GPT-4o?

In May 2024 *(7)*, OpenAI announced several key changes to ChatGPT, including the release of the newest version of its architecture, GPT-4o, which stands for "omni" or "omnimodel." GPT-4o touts an enhanced ability to process across various inputs, including text, speech (audio), and vision (images). This processing reportedly occurs within a single architecture, a significant improvement over prior versions that required multiple models for processing inputs that were not text-based. In previous models, audio had to be transcribed to text; the text was then processed in GPT-4 and then converted to an audio output using text-to-speech processing.

OpenAI states that GPT-4o also significantly improves analysis involving non-English-based languages. Demonstrations of the new GPT-4o model highlighted applications based on conversational speech prompts and visual prompts (*e.g.,* interacting with the camera or the visual display of a computer screen). OpenAI also notes that since GPT-4o is the first model with end-to-end modalities across text, vision, and audio, the company is still uncovering the potential capabilities and limitations of GPT-4o.

Concerns regarding data security, copyright, and accuracy

Data security, copyright, and accuracy are three very important concerns regarding the use of ChatGPT. LLMs like ChatGPT can be trained on the data generated by users interacting with the platform to improve the model's accuracy and capabilities. However, many users and businesses may be concerned about how their data will be used or whether training on sensitive data (*i.e.,* personal or confidential trade secrets) could result in that information being provided in a subsequent ChatGPT output to other users. OpenAI provides information on these topics on its website, and a summary of key features is provided here. However, appropriate caution should still be exercised, and many companies, governments, and regulators are addressing this issue as they develop and update policies, regulations, and guidance.

Beginning in April 2023 *(8)*, ChatGPT users were given the option to opt out of allowing their conversations to be used for model training. According to OpenAI, when a user opts out of this setting, new conversations generated in the platform are not used for model training. The conversation will still appear in the user's chat history unless the user deletes it. ChatGPT also has a "temporary chat" feature located in the interface's model selection drop-down menu. Temporary chats are also not used for training but do not appear in the user's chat history (although a copy of the chat may be maintained by OpenAI for 30 days for safety reasons). OpenAI also offers Team and Enterprise plans that exclude data from training as a default setting and offer additional data security measures (*i.e.,* administrative console and controls and custom data retention windows).

Several lawsuits have been filed alleging copyright law violations associated with OpenAI's use of copyrighted materials for training and ChatGPT's outputs containing copyrighted content *(9)*. In response, OpenAI has pointed to fair use policy, a legal doctrine focused on allowing the use of copyrighted materials without permission from the copyright holder for transformative purposes such as research or teaching *(10, 11)*. OpenAI's terms of use state that the output is owned by the user, and the user is responsible for ensuring both inputs and outputs do not violate applicable laws or infringe upon the rights of others *(12)*. Users should be cautious and mindful of these considerations when using ChatGPT and OpenAI's tools.

"ChatGPT can make mistakes" is a warning statement that is clearly placed at the bottom of the user interface, regardless of whether the user has selected the GPT-3.5, GPT-4, or GPT-4o model. While the potential capabilities of the tools can be impressive, it is important to remember that outputs should be verified when accuracy is important. This is particularly true in engineering applications, which may require a deep understanding of scientific and mathematic principles that may not be accurately represented in the large text database used to train ChatGPT.

Free and paid plan options

The corporate structure surrounding OpenAI is unique *(13)*. OpenAI started as a nonprofit with the goal of building safe artificial general intelligence (AGI) — artificial intelligence that is equivalent to, or exceeds, human intelligence — for the benefit of humanity. In 2019, a "capped-profit" entity (OpenAI LP) was created to help accelerate investment. Today, OpenAI is a combination of the nonprofit and capped-profit entities. As a result, offerings from OpenAI (like ChatGPT) may include a free version to provide AI

to the masses and paid options with additional features and functionality to generate capital to support further technological development.

OpenAI currently offers four plans for ChatGPT *(14)*. The features and offerings of the plans change over time, but this section presents a brief overview of the options available at the time of this writing.

Historically, the paid subscription plans provided users access to advanced tools and features, like GPT-4, attachment upload, and the GPT Store, that were not available to users on the Free plan. However, OpenAI announced in May 2024 that it was expanding ChatGPT capabilities for Free users, giving them access to the newly released GPT-4o architecture, file upload and vision capabilities, data analysis, and the GPT store *(7)*.

Paid subscriptions offer expanded limits for the advanced tools. For example, a Plus user on the $20 per month plan can submit 40 messages every three hours to GPT-4 and 80 messages every three hours to GPT-4o *(14)*. These limits may be adjusted dynamically by OpenAI based on available capacity in the platform. Both Free and Plus have unlimited access to the GPT-3.5 architecture when respective limits on other architectures are reached. Additionally, Plus users are given early access to new features, the ability to build, create, and use customized GPTs, and access to specialized GPTs, such as DALL·E, for image creation.

In addition to Free and Plus plans, OpenAI also offers Team and Enterprise plans. The Team plan includes higher message limits for GPT-4, GPT-4o, and tools like DALL·E, and excludes data from training as a default setting. The Team plan also provides a workspace where GPTs can be created and shared with other Team members and users with administrative console control. The Team plan is billed per user at either $25 per month (annually) or $30 per month (monthly).

The Enterprise plan opens unlimited access to restricted tools, additional security controls, custom data retention policies, and priority support. OpenAI does not publish rates for the Enterprise plan, and their website refers interested users to their sales team.

Case study examples with GPT-4o

The following two case studies are provided as examples of hypothetical applications for a chatbot technology. The case study input prompts were given to ChatGPT with the GPT-4o model in May 2024.

Case study 1: Reviewing regulatory standards. A safety professional at a chemical company is asked by maintenance staff if their Occupational Safety and Health Administration (OSHA) 40-hour hazardous waste operations and emergency response (HAZWOPER) training fulfills the requirements to perform confined space entry of an atmospheric storage tank at the facility (Figure 2). The following input prompt was provided to ChatGPT:

"You are an environmental, health, and safety professional. If a worker goes through a 40-hour HAZWOPER training course, is the worker considered appropriately trained according to OSHA regulations to perform confined space entry on an atmospheric storage tank? If you need more information before responding, please ask a question for clarification. Include references to the relevant citations in the OSHA regulations for both 'HAZWOPER' and 'confined space' in your response."

In response, GPT-4o noted that HAZWOPER training (29 CFR 1910.120) provides a broad overview of safety topics for workers involved in the cleanup of hazardous substances. Confined space training (29 CFR 1910.146) covers hazards, equipment, and rescue procedures for confined spaces. GPT-4o then provided a summary of key points and concluded that a worker with 40-hour HAZWOPER training

◀ **Figure 2.** In the first hypothetical case study, engineers asked ChatGPT whether a 40-hr hazardous waste operations and emergency response (HAZWOPER) training course met the Occupational Safety and Health Administration's (OSHA's) requirements for performing confined space entry. The large language model (LLM) identified that personnel would need additional training on confined-space safety, and the HAZWOPER module would not be sufficient. Image generated with OpenAI's DALL·E.

would not be considered appropriately trained for confined space entry. Workers would require both trainings if they were performing tasks that combined hazardous materials and confined spaces.

Case study 2: Mitigating corrosion of relief valves. An engineer at an acid production facility is asked for ideas on how to extend the service life of a relief valve on an acid reactor, which is a rated pressure vessel that operates at high pressure and temperature (Figure 3). The initial proposal from maintenance staff is to install a check valve between the reactor and the relief valve to prevent the reactor's corrosive vapors from entering the relief header. The engineer is concerned this modification would violate the American Society of Mechanical Engineers (ASME) Boiler and Pressure Vessel Code (BPVC). The following input prompt was provided to ChatGPT:

"You are the engineer at a facility that produces a corrosive acid product in a pressure vessel at high temperature and pressure. The pressure relief valve on the pressure vessel is corroding in service. Can you give me three options for modifications to the system that could be implemented to reduce the corrosion of the pressure relief valve and extend the service life of the relief valve? Could we install a check valve between the vessel and the relief valve? Or would that violate the ASME Boiler and Pressure Vessel Code?"

GPT-4o's first suggestion was to upgrade the relief valve to corrosion-resistant materials such as Hastelloy, Inconel, or other high-nickel alloys. Second, it proposed applying a protective coating or liner to the relief valve, such as Teflon, ceramic, or epoxy, to provide a barrier between the corrosive vapors and the metal surface. Third, it discussed a recirculating barrier fluid system, where water or a non-reactive oil could create a barrier between the valve and the corrosive acid. The response also stated that installing a check valve would not be recommended and typically violates the ASME BPVC, because the code states there should not be obstructions between the valve and the vessel it protects.

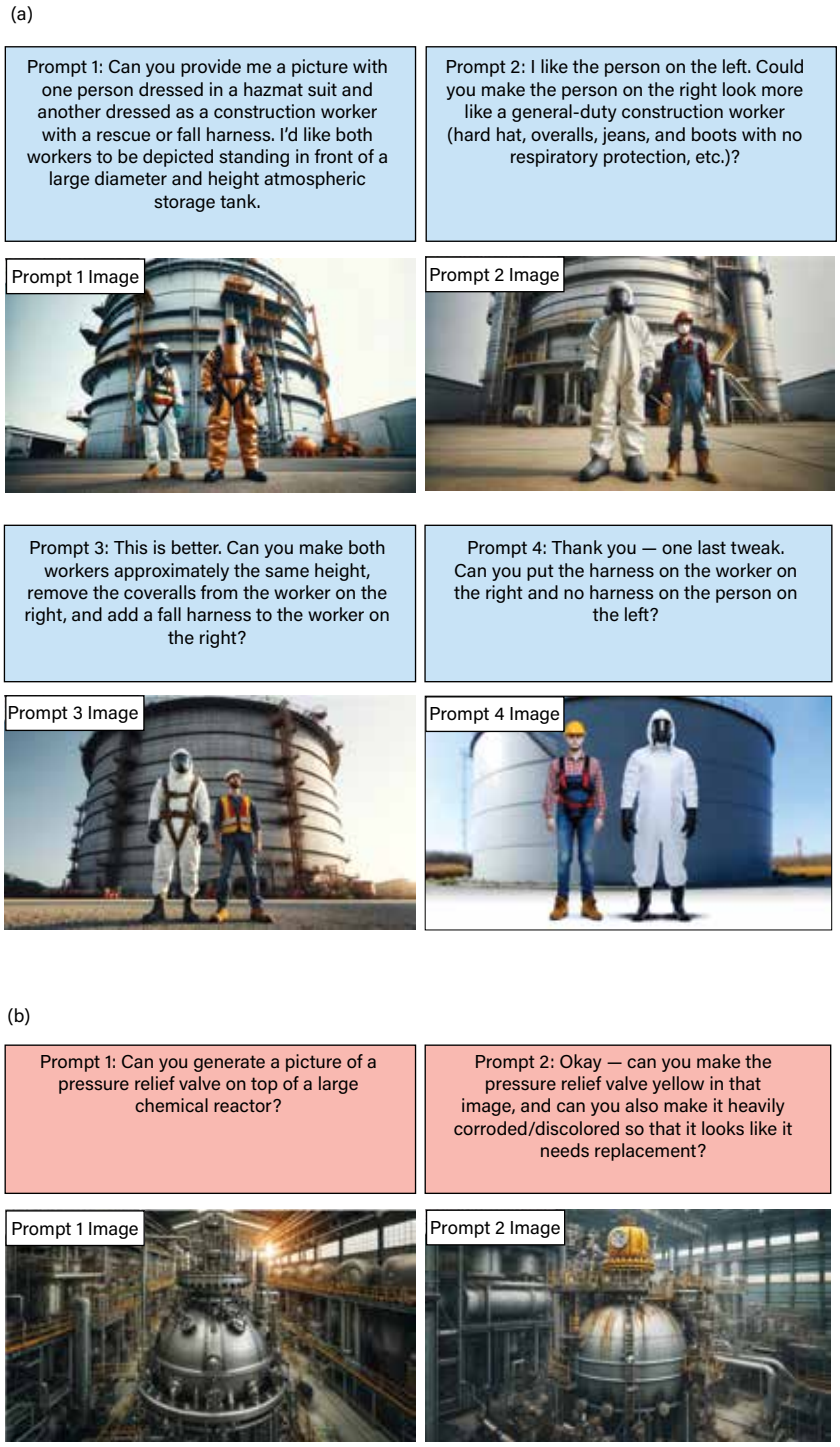
While this discussion includes several key themes related to corrosion, it failed to identify a rupture disc as a potential design solution based on the initial input prompt. Rupture discs are often used to protect relief devices from corrosion *(15)*.

This example highlights an important limitation of

SEAN J. DEE, P.E., PhD, CFEI is a Senior Managing Engineer in the Thermal Sciences practice at Exponent, Inc. (4580 Weaver Parkway, Suite 100, Warrenville, IL 60555; Email: sdee@exponent.com). He applies the fundamentals of chemical engineering to help clients address the risks associated with accidents, losses, and injuries. His core areas of expertise include chemical process safety, fire and explosion investigation, alternative energy, and high-purity oxygen systems. He services clients in many diverse industries, including insurance, legal, and chemical processing. He earned his PhD in chemical engineering from the Univ. of California, Berkeley, where his thesis focused on the catalytic production of biofuels from biomass. He also earned a BS in chemical engineering and a BA in sociology from Case Western Reserve Univ. He is a certified fire and explosion investigator and a professional engineer licensed in multiple states. He is also an elected Director of AIChE's Process Safety Division (formerly the Safety & Health Division).

SAHAND FARAJI, PhD, is a Senior Associate in the Thermal Sciences practice at Exponent, Inc. (15615 Alton Parkway, Suite 350, Irvine, CA 92618, Email: sfaraji@exponent.com). He has expertise in chemical engineering and industrial processing, hydrogen and clean energy production, nanomaterial synthesis, petrochemical and process safety, and experimental combustion processes. He earned his PhD in chemical engineering from the Univ. of California, Irvine, where he studied the synthesis and evaluation of nanomaterials for combustion technologies. Prior to attending graduate school, he received a MSc in chemical engineering from Istanbul Technical Univ. in Turkey and a BSc in chemical engineering from Univ. of Tabriz in Iran.

► **Figure 3.** In the second hypothetical case study, engineers asked ChatGPT how to address a corroding pressure relief valve under acidic conditions. The LLM analyzed the American Society of Mechanical Engineers (ASME) boiler and pressure vessel code to come up with a broad but not exhaustive list of possible solutions. Image generated with OpenAI's DALL·E.



▲ **Figure 4.** The images for the hypothetical case studies (Figures 2 and 3) were generated using DALL·E. (a) The first case study required four successive prompts; early outputs were fed back into DALL·E to further refine the image. (b) The image for the second hypothetical case study was also generated using DALL·E, and two of the outputs that DALL·E generated are shown here. Prior prompts in the second case study (not shown here) attempted to improve the accuracy of DALL·E's depiction of a pressure relief valve, to varying levels of success. Images generated with OpenAI's DALL·E.

ChatGPT. Recall that ChatGPT is a conversational tool that builds its output based on the relationships between the text in the input prompt. It is likely that the training database had very strong relationships for text regarding corrosion and the use of corrosion-resistant materials, protective coatings, and/or barriers. In practice, the problem discussed in the input prompt is a combination of two issues — corrosion and pressure relief — which both need to be addressed in parallel in the proposed solution. Research on prior models of ChatGPT (GPT-3 and GPT-4) suggests that the architecture struggles with problems that require advanced planning or problems where a seemingly discontinuous leap forward is needed to progress toward the solution of a task *(16)*. It is possible this complex relationship between the competing objectives of protecting the relief valve against corrosion and protecting the pressure vessel from overpressure required a so-called "Eureka" idea to arrive at a rupture disc as a solution.

Using DALL·E images. One of the advanced tools OpenAI offers is DALL·E, which can create images from text prompts within the ChatGPT user interface. DALL·E is a 12-billion parameter version of GPT-3 that was first released in 2021 and has subsequently been updated with DALL·E2 and DALL·E3 to improve the resolution, accuracy, and realism of the generated images *(17)*. The outputted images are owned by the user, including the right to reprint, sell, and merchandise the image, subject to OpenAI's Content Policy and Terms *(18)*. OpenAI's guidelines for DALL·E also state that users are encouraged to proactively disclose the use of AI in their work and cannot mislead others about AI's involvement *(19)*.

Figures 2 and 3 were generated with DALL·E. In general, several iterations are required to refine the image output to match a prompt's intent. Figure 4 demonstrates this process. In the pressure relief valve case study, DALL·E required five

prompts before the prompts shown in Figure 4b to refine the tool's depiction of a relief valve. These prompts included uploading example pictures of pressure relief valves into the user prompt, and DALL·E searching for photographs of relief valves across multiple search engines and stock photo websites. Ultimately, these iterations had varying degrees of success in improving the accuracy of the image.

Closing thoughts

This article should not be interpreted as an endorsement of ChatGPT or other AI tools for any application; however, significant progress is being made in LLM technology that is widening the technology's applicability in process safety. In the authors' experience, the GPT-4 model, released in 2023 to Plus users, displayed stark improvements in performance over the Free user's GPT-3.5 model. With the recent expansion of the functionality of ChatGPT's free platform, many casual users will likely also experience these performance improvements when experimenting with OpenAI's more advanced tools. However, users must remember that LLMs are text-generation tools designed to construct outputs based on the relationships between tokenized data. They can still provide inaccurate responses based on the limitations of their architecture, training, datasets, and context provided in the input prompt.

These caveats are not meant to diminish the significance of the more recent iterations of ChatGPT. Researchers are openly questioning whether GPT-4 represents an early but incomplete form of AGI, which has led to interesting debates regarding definitions, measures, and assumptions regarding the concept of "intelligence" *(20)*. This does not mean that ChatGPT has independent thoughts, emotions, or motivations, which some may consider central to a definition of intelligence. However, research has shown that GPT-4 demonstrates emergent behaviors and capabilities in other areas of intelligence, including reasoning, problem solving, abstract thinking, complex idea comprehension, and learning (within a user interaction session). While ChatGPT was originally developed to predict the next word in a sentence, the resulting technology has demonstrated much broader capabilities. As researchers and developers continue to improve the technology, its limitations may become less apparent. For these reasons, it is important that users, and engineers in particular, have a fundamental understanding of how these tools work and recognize the importance of verifying the accuracy of their output. CEP

Literature Cited

1. **OpenAI,** "How ChatGPT and Our Language Models are Developed," https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed (accessed May 31, 2024).
2. **Delua, J.,** "Supervised Versus Unsupervised Learning: What's the Difference?" IBM, https://www.ibm.com/think/topics/supervised-vs-unsupervised-learning (accessed May 31, 2024).
3. **Open AI,** "Introduction," https://platform.openai.com/docs/introduction (accessed May 31, 2024).
4. **Schriener, M.,** "GPT-4 Architecture, Datasets, Costs, and More Leaked," *The Decoder,* https://the-decoder.com/gpt-4-architecture-datasets-costs-and-more-leaked (July 11, 2023).
5. **Marr, B.,** "A Short History of ChatGPT: How We Got to Where We Are Today," *Forbes,* https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/?sh=47e0d6dd674f (May 19, 2023).
6. **Emmanuel, C.,** "GPT-3.5 and GPT-4 Comparison: Exploring the Developments in AI-Language Models," https://medium.com/@chudeemmanuel3/gpt-3-5-and-gpt-4-comparison-47d837de2226 (accessed May 31, 2024).
7. **OpenAI,** "Introducing GPT-4o and More Tools to ChatGPT Free Users," https://openai.com/index/gpt-4o-and-more-tools-to-chatgpt-free (May 13, 2024).
8. **OpenAI,** "New Ways to Manage Your Data in ChatGPT," https://openai.com/index/new-ways-to-manage-your-data-in-chatgpt (Apr. 25, 2023).
9. **Grynbaum, M. M., and R. Mac,** "The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work," *New York Times,* https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html (Dec. 27, 2023).
10. **O'Brien, M.,** "ChatGPT-Maker Braces for Fight with New York Times and Authors on 'Fair Use' of Copyrighted Works," *Seattle Times,* https://www.seattletimes.com/business/chatgpt-maker-braces-for-fight-with-new-york-times-and-authors-on-fair-use-of-copyrighted-works (Jan. 9, 2024).
11. **OpenAI,** "OpenAI and Journalism," https://openai.com/index/openai-and-journalism (Jan. 8, 2024).
12. **OpenAI,** "Terms of Use," https://openai.com/policies/terms-of-use (Nov. 14, 2023).
13. **OpenAI,** "Our Structure," https://openai.com/our-structure (accessed May 31, 2024).
14. **OpenAI,** "Pricing," https://openai.com/chatgpt/pricing (accessed May 31, 2024).
15. **Towler, G., and R. K. Sinnott,** "Chemical Engineering Design: Principles, Practice and Economics of Plant and Process Design," 2nd ed., Chapter 10, Elsevier, Amsterdam, Netherlands (2013).
16. **Bubeck, S., *et al.,*** "Spark of Artificial General Intelligence: Early Experiments with GPT-4," Microsoft Research, https://arxiv.org/abs/2303.12712 (accessed Feb. 27, 2024).
17. **OpenAI,** "DALL·E: Creating Images from Text," https://openai.com/index/DALL·E (accessed July 3, 2024).
18. **OpenAI,** "Can I Sell Images I Create with DALL·E?" https://help.openai.com/en/articles/6425277-can-i-sell-images-i-create-with-DALL·E (accessed July 3, 2024).
19. **OpenAI,** "How Should I Credit DALL·E in My Work?" https://help.openai.com/en/articles/6640875-how-should-i-credit-DALL·E-in-my-work (accessed July 3, 2024).
20. **Bubeck, S.,** "First Contact," presented at Massachusetts Institute of Technology, Cambridge, MA, https://www.youtube.com/watch?v=qbIk7-JPB2c (Mar. 22, 2023).