Prospects and Challenges of Large Language Models in the Field of Intelligent Building

Wu Yang, Wang Junjie, Li Weihua

Vocational-Technical Training Schools of Northmount Inc., Beijing, China

Email address:
wienerblut@163.com (Wu Yang), 48355611@qq.com (Wang Junjie), weihua8105@163.com (Li Weihua)

Abstract: At the end of November 2022, the ChatGPT released by OpenAI Inc. performed excellently and quickly became popular worldwide. Despite some shortcomings, Large Language Models (LLM) represented by Generative Pre-trained Transformer (GPT) is here to stay, leading the way for the new generation of Natural Language Processing (NLP) technique. This commentary presents the potential benefits and challenges of the applications of large language models, from the viewpoint of intelligent building. We briefly discuss the history and current state of large language models and their shortcomings. We then highlight how these models can be used to improve the daily maintenance of intelligent building. With regard to challenges, we address some vital problems to be solved before deployment and argue that large language models in intelligent building require maintenance staff to develop sets of competencies and literacies necessary to both understand the technology as well as the maintenance and maneuver of intelligent building. In addition, a clear strategy within intelligent building troops with a strong focus on AI talents construction and training dataset annotation are required to integrate and take full advantage of large language models in the daily maintenance. We conclude with recommendations for how to address these challenges and prepare for further applications of LLM in the field of intelligent building in the future.

Keywords: Artificial Intelligence, Large Language Models, Intelligent Building

1. Introduction

Natural language understanding has always been one of the most concerned technical fields in the field of artificial intelligence. Alan Mathison Turing, the founder of artificial intelligence, used natural language processing capabilities as a symbol of whether a machine possesses human intelligence, known as the Turing test. Several Turing Prize winners, such as Geoff Hinton, the father of deep learning, and Yann LeCun, Facebook's vice president and chief scientist in artificial intelligence, have predicted that natural language understanding will have great development and application prospects [1]. Natural language processing is also known as the "pearl in the crown of artificial intelligence".

On November 30, 2022, ChatGPT, a large language model application developed by OpenAI, was launched, and the number of active users soared to 1 million within five days. The surging global access requests repeatedly caused server downtime. Despite multiple restrictions on registered user IP addresses by OpenAI, the number of active users in two

months still exceeded 100 million, and both the growth rate and total number of active users have created a history [2]. From the perspective of performance, ChatGPT has relatively mature and excellent language generation and knowledge reasoning capabilities, which can better understand user intentions and provide complete and coherent answers through multiple rounds of communication. Despite some flaws, ChatGPT has transformed the general public's understanding of chat robots from "artificially retarded" to "interesting", and has also attracted widespread attention in the global market. Major domestic and foreign technology companies in the field of artificial intelligence, such as Google, Meta (formerly Facebook), Apple, Baidu, Alibaba, Tencent, Huawei, and iFLYTEK, have either rapidly launched similar applications, or indicated that they have actively deployed GPT (Generated Pretrained Transformer) technology [3]. ChatGPT has also attracted close attention from governments, policy circles, and even the military. Science and technology policy scholars predict that in the near future, generative large-scale language models will

replace the work of many people, and will even disrupt many fields and industries.

Undoubtedly, the large-scale language model represented by ChatGPT has a broad military application prospect. However, the current huge hardware computational power requirements of large-scale language models limit their battlefield use, especially if they are difficult to deploy near the forefront, but they are suitable for deployment in intelligent building. This article focuses on the future application prospects of large-scale language models in the operation and maintenance of intelligent building, briefly introduces its development history and key technologies adopted by ChatGPT, focusing on analyzing its current universality issues. Then, from the perspective of intelligent building operation and maintenance, it proposes possible future application methods, analyzes possible problems, and finally puts forward corresponding countermeasures and suggestions.

2. History and Feature of the Large Language Models

Since 2010, significant improvements in chip computing power have led to significant progress in artificial intelligence technology based on neural network algorithms. A neural network based algorithm called deep learning has emerged, especially in image recognition and natural language processing. Neural network algorithms have enabled natural language processing technology to initially take on a "large-scale" scale. Currently, large-scale language models have become the core technology in the field of natural language processing [4], and will continue to develop in the direction of large models, big data, and large computing power in the foreseeable future.

2.1. Development History and Key Technologies

The Artificial Neural Network (ANN) [5] was first proposed by the American physicist John Hopfield in 1982, and is expected to mimic the way neural networks work in the biological brain to endow computers with human intelligence. The neural network algorithm simulates neural synapses in the brain into undetermined computational parameters, and connects these parameters into a multi-level network structure that is similar to synapses and nerve cells in the brain that are interconnected. Before using, first input a dataset with a correct correspondence from input to output for training, and determine each calculation parameter to complete the training. At this point, the answer can be obtained by solving the problem in the input band. Generally, the more parameters, the larger the training dataset, the more sufficient the training, and the more accurate the model. However, the total number of synapses in the human cerebral cortex exceeds 100 trillion. In the 30 years since the introduction of neural network algorithms, due to the limitations of hardware computing power, it is difficult to have supercomputers capable of running hundreds of millions of parameter algorithm models.

After 2010, deep learning technology based on neural networks has made significant progress compared to previous shallow machine learning algorithms, but deep learning algorithms rely too heavily on manually annotated training datasets [6]. Typically, such datasets are large in scale and the cost of manually tagging one by one is too high. In 2018, OpenAI launched the first generation of GPT (Generic Pretrained Transformer) model [7], using the "pre training" technology for the first time, that is, first training an initial model (called pre training) based on a common training dataset, and then conducting "fine tuning" for specific tasks. Pre training datasets can be used in public publications or even on web pages, reducing the need for manually annotated datasets. GPT-1 pioneered the use of the Transformer model with multiple attention heads [8], which can perform tasks similar to tracking multiple pronouns in a sentence referring to different sentence components, better realizing computer understanding of complex language phenomena in natural language.
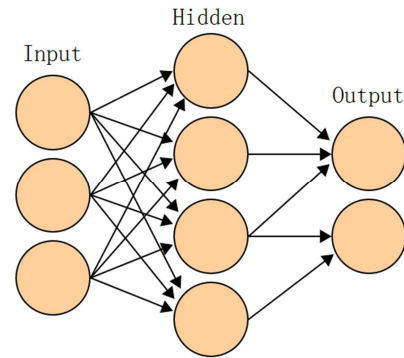

Figure 1. *A brief sketch of ANN.*

In order to improve the reasoning ability of language models, ChatGPT has also introduced the Chain of Thought (COT) [9] technology. Programmers decompose some basic simple problems into multiple intermediate steps that can be understood by computers, and convert them into code to be incorporated into the model, enabling ChatGPT to correctly solve some elementary school math application problems.

At the end of model training, ChatGPT also introduced the Reinforcement Learning with Human Feedback (RLHF) [10] technology, which provides feedback to the model on human preferences for different outputs through manual scoring of the output, and tries to correct negative elements such as aggressive language, racial discrimination, and abusive language in the final output.

ChatGPT adopts many pioneering technologies, making it capable of excellent contextual understanding, language generation, thinking chain reasoning, zero sample ability (answering questions that have no reference content in training data), and even preliminary understanding of the emotions of discourse, and writing computer code according to user requirements. Many fields are actively exploring the possible application and future impact of ChatGPT [11].

2.2. Existing Deficiencies and Problems Being Improved

ChatGPT has become a global sensation in a short period of time, triggering domestic and foreign technology giants to compete to develop similar types of products. However, large-scale language models still need to be continuously improved to truly enter the commercial operation stage, and the following issues need to be urgently addressed.

2.2.1. High Computational Power Demand and Energy Consumption

ChatGPT has 175 billion model parameters, and it is said that the model parameters of the next generation GPT-4 exceed 100 trillion. OpenAI did not disclose the energy consumption data of ChatGPT, while the literature [12] studied GPT-3 with a similar scale to ChatGPT. According to research estimates, if V100 GPU and the cheapest cloud computing package are used, training a GPT-3 model requires 355 GPUs to run for a full year, with a cost of $4.6 million. The cost of long-term operation and maintenance of ChatGPT is far greater than the "learning and training" stage. A study in December 2022 [13] estimated that the cost of running ChatGPT when the active users reach 1 million is about 100000 to hundreds of thousands of dollars per day. Too high costs have allowed large-scale language models to be monopolized by a few financially strong companies, which is not conducive to technological progress and popularization.

2.2.2. A Large Amount of Manual Participation Is Still Required in the Early Stage

Although ChatGPT's adoption of the Transformer model and pre training technology has significantly reduced the need for models to manually annotate datasets, complex instructions still need to be manually written during the instruction fine-tuning stage, the chain of thought (COT) requires manual listing of the process of thinking reasoning, and reinforcement learning based on artificial feedback (RLHF) still requires a significant amount of manually annotated data. Large language models are far from escaping their dependence on massive high-quality manual annotation data.

2.2.3. Closed-Source Nature of the Model

OpenAI released ChatGPT without opening the source code, only opening the application program interface (API) for remote access. The lack of open source code makes it difficult for users to obtain customized services that are suitable for certain specific fields and needs through secondary development, and also raises concerns about user security.

2.2.4. Weak Symbolic Reasoning Ability

ChatGPT's answers are clear and demonstrate a certain level of linguistic and logical abilities, but there are still shortcomings in symbolic reasoning. Currently, it is not possible to perform slightly more complex arithmetic operations and logical reasoning. Some studies [14] have attempted to compensate through external calls to computer APIs, but they are still unable to handle more complex tasks.

In addition, the use of large-scale language models can also lead to issues related to intellectual property rights, technological ethics, personal privacy protection, and even be used to commit various criminal acts [15]. These questions have nothing to do with the main thrust of this article and will not be listed one by one.

3. Application Prospect of Large Language Model in the Field of Intelligent Building

According to industry analysts, the launch of ChatGPT, an OpenAI company, is just the prelude to a new round of natural language understanding technology competition. Google, Meta, Baidu, Alibaba, Tencent, and other well-known enterprises in the field of artificial intelligence at home and abroad have already laid out in this direction for many years, and are developing large-scale language models in a strictly confidential state. On the other hand, hardware technology is also continuously improving, and the computing power of chips is constantly increasing Power consumption is continuously decreasing, and in the near future, there will be more large-scale language model technologies and products with better performance in various aspects. These products may be difficult to deploy in harsh battlefield conditions due to hardware computing power and energy consumption requirements, but it is completely feasible to deploy them within a relatively stable environment of intelligent building. The use of large-scale language model technology in the field of intelligent building maintenance and management has the inherent advantage of "building near water first served".

3.1. Typical Application Scenarios

In addition to the various general functions of ChatGPT listed above, combined with the actual situation of intelligent building, this article proposes several possible typical application scenarios for future large-scale language models in this field. These application scenarios obviously cannot cover all the functions and possible methods of large-scale language models, but are currently predictable and most valuable for the operation, maintenance, and use of intelligent building.

3.1.1. Troubleshooting Assistance

Information such as fault phenomena, fault types, and emergency repair steps for equipment and facilities within the protection project will be documented in advance, and provided to the GPT language model for learning during the training phase. Voice recognition software will be connected to the front end of the GPT application interface. When encountering a fault, the operation and maintenance personnel can even command the GPT to assist in finding the cause of the fault through voice dialogue, automatically generate emergency repair steps, and remind operation

precautions, If the corresponding intelligent terminal equipment is linked, it can also push relevant pictures and photos stored in advance or emergency repair video tutorials to speed up the troubleshooting process and improve the self-sustainability of the intelligent building battlefield.

3.1.2. Training Assistance

Similar functions such as fault repair assistance are applied to the training process of operation and maintenance personnel. During specific subject training, the operation steps, precautions, and evaluation criteria are pushed through the linkage of voice, text, pictures, animations, and videos, which can accelerate the ability generation process of operation and maintenance personnel.

3.1.3. Code Generation

In the battlefield operation, it is possible to assist in rapid code generation, rapid code comparison, and error correction for internal system functional requirements of intelligent building temporarily proposed by commanders based on battlefield conditions, or for replacement of alternative functional modules, to improve the battlefield adaptability of intelligent building.

3.1.4. Auxiliary Generation of Documents Such as Scheme Plans

Daily maintenance and wartime operations require the generation of a large number of documents such as plans, plans, and measures. Such structured documents usually have fixed content and format, and are particularly suitable for GPT, a generative artificial intelligence application. As long as a sufficient number and rich content of manually annotated solution planning document data are provided for the GPT language model in the pre training stage, it is only necessary for the operation and maintenance personnel to input the different areas and details that require special attention into the GPT application interface through induced conversations during the action, and mature documents can be generated, which can be issued after review and simple modification by the operation and maintenance personnel, It can liberate operation and maintenance personnel from time-consuming and laborious text editing and formatting, and focus their attention on wartime support of intelligent building.

3.2. Possible Problems

Unlike Section 2.3 of this article, the main issues mentioned here are issues that may have significant impact on users in the actual use of large-scale language models in the future. Currently, similar to ChatGPT, large-scale language models built using the "pretraining+instruction fine tuning" approach have problems. Some of the root causes are the technical system used by the model, while others are problems with the dataset itself, which may be the embodiment of social ideology in language, It may also be caused by the fuzziness of human language itself. The industry generally believes that these problems can be improved with technological progress, but may not be eradicated for some time.

3.2.1. Real-Time Issues

Currently, most AI technologies adopt the mode of "learning and training" first and deploying applications later. The deadline of the dataset used in the "learning and training" phase determines the timeliness of the application's "mastery" of knowledge. As shown in Figure 2, the training dataset used by ChatGPT is cut to September 2021 and cannot "master" information after September 2021 [16]. A large amount of new knowledge will emerge in the operation and maintenance of intelligent building, such as equipment upgrades, new operational styles, and new threat response plans. To "master" new knowledge, it is necessary to conduct "learning and training" again. Considering the large scale of large-scale language models, such training requires huge power consumption, long time, and high costs. It is usually necessary to wait until a certain amount of new knowledge is accumulated, and then carry out timely training, Large language models must have a trade-off between real-time performance and running costs.
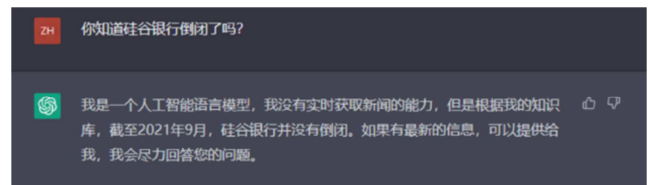

Figure 2. *ChatGPT didn't know information after September 2021.*

3.2.2. Uncertainty Issues

The same question model may not always provide accurate answers. Sometimes the semantics of the question do not change, but the wording is changed or word order is adjusted, and the model cannot correctly understand the question or what could not have been correctly understood will become normal after adjustment. Uncertainty may be caused by errors, noise, and ambiguity in human language contained in the training dataset itself, or it may be caused by "insufficient learning and training" and insufficient training data.

3.2.3. Credibility Problem

The credibility issue of the application output of large language models, that is, the response of language models to users may be "serious nonsense," which can be caused by various reasons. As shown in Figure 3, ChatGPT "fabricates" under user guidance, which is difficult to avoid in current GPT language models [17]. In order to make the output more suitable for natural language and user requirements, generating applications will make full use of the guidance information entered by users. In Figure 3, ChatGPT successfully identified the key information provided by users such as "唐三藏" (Tang Sanzang), "葬花吟" (Buried Flower Chant) and "情感" (Emotion). Combining the results of the pre training dataset "Learning and Training", it provides a structured, complete, and almost perfect answer, but only didn't realize that "Tang Sanzang" and "Burial Flower Chant" belong to different literary masterpieces, and should not be confused.

Figure 3. ChatGPT can be misguided by user's prompts.

3.2.4. Robustness Issues

Robustness (also known as robustness) of large language models refers to the ability of models to resist the impact of erroneous data, outdated information, and still output correct results. The model may be misused if its robustness is poor or certain specific laws are used by malicious elements to set up backdoors in the dataset. Similar to credibility issues, the reasons for robustness issues in large language models can be manifold, partly related to the dataset used in the "learning and training" phase. As shown in Figure 4 [16], although there are user induced factors, ChatGPT is able to confirm that "林黛玉倒拔垂杨柳" (Lin Daiyu hangs over the willows upside down) is one of the famous plots of "红楼梦" (A Dream of Red Mansions), probably because the Chinese dataset used in the "learning and training" stage of ChatGPT contains the hot content of parody literature creation on the Chinese network in previous years.
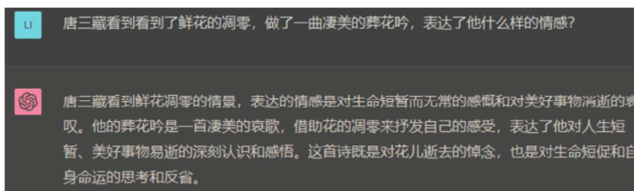

Figure 4. ChatGPT made an incorrect conclusion.

3.2.5. Bias from the Dataset

If the training dataset of large language models is inherently biased, it is also difficult for the output of language models to avoid factors in these "genes". For example, an article published in the sub journal "Machine Intelligence" of the authoritative journal "Nature" pointed out that LLM "consciously" associate Muslims with violence to a large extent, which is essentially a projection of mainstream social ideology on language texts in the English language environment [18]. Literature [19] also mentions that GPT-3 reached a completely wrong conclusion due to a misunderstanding of the keywords extracted from the dataset. Due to the large amount of pre training data, it is difficult to correct this bias in subsequent steps, and pre training dataset filtering for this type of bias is currently basically not feasible. Firstly, the cost of manual filtering is too high. If a certain rule is set to pass the program automatic screening, it is likely to cause more data to be mistakenly screened, making the diversity of the dataset not guaranteed, leading to defects in other aspects of the language model. It seems that bias issues such as large language models can only be resolved through output shielding.

4. Countermeasures and Suggestions

The pace of technological progress in the era of artificial intelligence is becoming faster and faster. According to the research [6] by Professor Che Wanxiang of the Natural Language Processing Research Institute of Harbin Institute of Technology, the duration of mainstream natural language processing technologies for each generation will be shortened by nearly 1/2 compared to the previous generation. Currently, the fifth generation technology represented by ChatGPT is expected to be popular for only 2.5 years, and will be replaced by new technologies by 2025. Continuing to focus on artificial intelligence technology and the possible future military application of large language models has become a very urgent task. A clear strategy within intelligent building troops with a strong focus on AI talents construction and training dataset annotation are required to integrate and take full advantage of large language models in the daily maintenance and wartime operations.

4.1. Establish a Unified Data Collection Framework

The final performance of NLP is largely determined by the volume and quality of datasets by which the LLM is trained. In terms of intelligent building maintenance and management, operation plans, training documents, status parameters, environmental parameters, fault information, and even monitoring videos are the best materials for future AI programs to carry out "learning" in this specific field. It is recommended to establish a unified maintenance and maneuver data collection framework in this field, and expand the sources and total amount of data. In pre-training, data sets in the field of military facilities are used, and the trained algorithms can also meet the needs of the field of public infrastructure diluting development costs.

4.2. Master Core Technologies and Establish Autonomous and Controllable Artificial Intelligence Algorithm Models

First of all, the release of ChatGPT by OpenAI only opens the application program interface (API), and domestic users log in to use ChatGPT, essentially helping OpenAI to conduct public intelligence collection in China. Secondly, language models developed in an English language environment naturally give priority to highlighting English language proficiency. Some studies [17] indicate that ChatGPT is not performing as well as similar domestic products in some aspects. Thirdly, due to the widespread interpretative difficulties of artificial intelligence technology, even if OpenAI company discloses the source code of ChatGPT, it is still impossible to clarify with current scientific knowledge what functions tens of billions or hundreds of billions of nodes actually achieve. Therefore, human intelligence technology used in the field of national

defense must achieve autonomous control. The issues listed in Section 3.2 of this article also prove that it is necessary to introduce personnel from user units to participate in key aspects of algorithmic "training", such as the selection of training datasets, manual annotation of datasets, fine tuning of instructions, and human feedback in reinforcement learning, and is crucial to further improve the accuracy, robustness, and controllability of the model.

4.3. Emphasize Talent Cultivation and Gradually Incline to Cultivating Intelligent Program Use, Operation and Maintenance Talents

The skills of the traditional intelligent building maintenance and management talent team mainly emphasize the operation and daily maintenance of mechanical equipment and facilities, as well as emergency repair during wartime. With the development of artificial intelligence technology and robotics technology, devices will generally achieve modularization and support "hot plugging". Robot patrol inspection will be popularized, equipment emergency repair will focus on module replacement, and talent team skills will gradually shift from focusing on the operation and maintenance of mechanical equipment to focusing on the operation and maintenance of artificial intelligence network information systems. On the one hand, it is necessary to strengthen the reserve of artificial intelligence knowledge and application skills of the talent team. Currently, it is necessary to accelerate the cultivation of a group of talents who understand both intelligent building and operational support, as well as data and algorithms, to participate in the development of artificial intelligence applications in specific fields in a timely manner. On the other hand, it is necessary to build a intelligent building maintenance and management talent team with big data and artificial intelligence, understanding intelligent algorithm logic, and mastering intelligent system regulation and operation and maintenance skills based on the long-term perspective.

References

[1]   Z. Feng, D. Zhang, and G. Rao, "From Turing test to ChatGPT: a milestone of man-machine interaction and its enlightenment," Chinese Journal of Language Policy and Planning, vol. 8, 2023, pp. 20-24.

[2]   H. Lu, "ChatGPT Brings Huge Challenges to Traditional Search Engines," China Economic Times, 2023-02-14, pp. 1.

[3]   Q. Yang, Y. Zhao, "Domestic and foreign manufacturers accelerate their efforts to compete with AIGC," The 21st Century Business Herald, 2023-02-08, pp. 11.

[4]   T. Brown, B. Mann, and N. Ryder, "Language models are few-shot learners," Advances in neural information processing systems, vol. 33, 2020, pp. 1877-1901.

[5]   J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," Proceedings of the National Academy of Sciences of the United States of America, vol. 8, 1982, pp. 79.

[6]   W. Che, T. Liu, "New Paradigm of Natural Language Processing: A Method Based on Pre-Trained Models," ZTE Technology Journal, vol. 28, 2022, pp. 3–9.

[7]   A. Radford, K. Narasimhan, and T. Salimans, "Improving language understanding by generative pre-training," OpenAI blog, 2018.

[8]   A. Vaswani, N. Shazeer, and N. Parmar. "Attention is all you need," Advances in neural information processing systems, vol. 30, 2017, pp. 48.

[9]   J. Wei, X. Wang, and D. Schuurmans, "Chain of thought prompting elicits reasoning in large language models," arXiv preprint, 2022, arXiv: 2201.11903.

[10] L. Ouyang, J. Wu, and X. Jiang "Training language models to follow instructions with human feedback," arXiv preprint, 2022, arXiv: 2203.02155.

[11] P. Jie, "Large language model for molecular chemistry," Nature Computational Science, vol. 1, 2023, pp. 3.

[12] X. Qiu, T. Sun, and Y. Xu, "Pre-trained models for natural language processing: A survey," Science China (Technological Sciences), vol. 10, 2020, pp. 63.

[13] S. Kalyan, A. Rajasekharan, S. Sangeetha, "Ammus: A survey of transformer-based pretrained models in natural language processing," arXiv preprint, 2021, arXiv: 2108.05542.

[14] T. Schick, J. Dwivedi-Yu, and R. Dessì, "Toolformer: Language models can teach themselves to use tools," arXiv preprint, 2023, arXiv: 2302.04761.

[15] A. Tamkin, M. Brundage, and J. Clark, "Understanding the capabilities, limitations, and societal impact of large language models," arXiv preprint, 2021, arXiv: 2102.02503.

[16] S. Zheng, "Measured ChatGPT-4: Improved mathematical skills, poor temporal reasoning, difficult to understand the 'Daiyu inverted pull willow' network stem", Sohu blog, 2023, https://www.sohu.com/a/654613394_115565

[17] H. Zhang, L. Li, Chun. Li, "ChatGPT Performance Evaluation on Chinese Language and Risk Measures," Data Analysis and Knowledge Discovery, vol. 3, 2023, pp. 1-14.

[18] A. Abubakar, F. Maheen, Z. James, "Large language models associate Muslims with violence," Nature Machine Intelligence, vol. 6, 2021, pp. 3.

[19] S. Wu, X. Wang, R. Xiao, "Problem and Perspective for Battery Researcher Based on Large Language Model," Energy Storage Science and Technology, vol. 3, 2023, pp. 1-7.