Enhancing Intelligent HVAC Management Using Large Language Models: A Case Study of ChatGPT

Fan, Qifeng[1], Deng, Suming[2], Lin Yanzheng[2], Gao feng[2]
[1] GD Midea Air-conditioning Equipment Co., Foshan, China
[2] GD Midea Air-conditioning Equipment Co., Shanghai, China

Abstract

In the context of technological advancement, intelligent HVAC (Heating, Ventilation, and Air Conditioning) managers assist in the operation of various HVAC equipment by analyzing users' historical behavior and conversation data. However, in traditional intelligent HVAC management systems, customer complaints are often raised due to misunderstandings of user intent or vague, or even incorrect, responses, and imprecise recommendations of HVAC actions. To solve this problem, this study applies large language model and prompt engineering techniques, integrating the large language model API and prompt into the intelligent HVAC system to achieve precise control over HVAC equipment. The contributions of this study are mainly reflected in three aspects: first, it improves the semantic understanding and latent semantic understanding of the intelligent HVAC manager; second, it enhances the accuracy of HVAC action recommendations; and third, it develops a human-computer dialogue system based on large models that can be demonstrated. Experimental results show that ChatGPT performs well in terms of accuracy in two experimental datasets. Although ChatGPT may face challenges in handling certain complex situations and specific domain knowledge, its powerful natural language understanding and generation capabilities bring new possibilities to the intelligent HVAC field, providing potential solutions for residents to achieve convenient, energy-saving, comfortable, and safe living experiences.

Highlights

- Improved the smart HVAC manager's semantic understanding by 9% utilizing the ChatGPT-API and prompt engineering techniques.
- Boosted the accuracy of the system's operation suggestions by 10% through the integration of the same methods into the HVAC system.
- Development of a Demonstrable Human-Computer Dialogue System
- Give residents a more comfortable, convenient, power-saving, and intelligent HVAC experience.

Introduction

With the rapid development of China's economic level, residents are paying increasing attention to smart home systems, including Smart HVAC Control Systems. According to data from the China Finance and Economics Report (He,2022), the size of China's smart home appliance market exceeded 600 billion CNY in 2022. Smart HVAC Control Systems can continuously provide automation and intelligent services, eliminating the need for manual intervention, and improving the convenience, comfort, interactivity, and safety of residents' lives. The emergence of Smart HVAC Control Systems is an inevitable step in the integration of modern digital technology into daily life, which will bring more convenience.

Intelligent HVAC control systems represent the cutting edge of modern air conditioning technology. The system utilizes advanced algorithms and sensors to optimize indoor air quality and environment to ensure comfort and efficiency. These systems do more than simply switch or adjust the temperature; They can analyze environmental data, user preferences and other relevant factors and automatically fine-tune to create the best indoor environment for users. Today, many intelligent HVAC control systems heavily rely on dialogue system modules, including NLU (Natural Language Understanding), DST (Dialogue State Tracking), DPL (Dialogue Policy), NLG (Natural Language Generation) (Saha, 2020). The dialog system allows users to interact with the system in a more intuitive and natural way. Further, the DPL part of the dialog system heavily relies on deep learning models. Its architecture usually follows the input-model-output process, so the design and optimization of the model becomes particularly important. In the field of human-computer interactive control systems, some deep learning models stand out in the application of natural language processing. For example, (Wen, 2016) LSTM (Long-Short-Term Memory Network), RNN (recurrent neural network), convolutional neural network (CNN) and other models play an important role in language understanding and sequence prediction in dialogue systems in DPL. However, research (Yang, 2013) shows that at this stage, 60% of users have complaints about their intelligent air-conditioning dialogue system: misinterpreting user intentions leading to wrong responses and inappropriate automatic controls. This paper focuses on these two major technical limitations of traditional HVAC dialogue system in DPL mentioned above: 1. Low prediction performance of intentions in equipment operation suggestions. 2. Misinterpretation of intentions in human-computer dialogue

1. Low Accuracy in Equipment Operation Predictions (Isinkaye, F ,2015) refers to a low accuracy in equipment operation predictions means that when a smart device makes operation predictions based on user commands or habits, there is a discrepancy between its predictions and the user's actual needs or expectations. Such discrepancies can result from imprecise predictive algorithms of the device, changes in user behavior patterns, or external environmental factors.

2. Misinterpretation of Intentions (Nasukawa, T, 2001) refers to instances when, during interactions between humans and computers or smart devices, the device fails to accurately identify and interpret the user's true intentions or needs. This misunderstanding may arise from ambiguous user commands, inaccurate contextual recognition, or limited algorithmic interpretation capabilities of the device. Such misinterpretations can lead to irrelevant or incorrect feedback, diminishing the user experience and satisfaction.

In the intelligent control of HVAC, there are two ways to improve the above two technical difficulties in the dialogue system. Firstly, an encoder method is adopted, which uses the encoder structure and self-attention mechanism to solve the long-distance dependent information problem. It can better capture the context relationship and increase the prediction effect of sequence data. Secondly, increasing the number of corpus and model parameters, which enable language models to learn more abundant corpus information. Therefore, this way increases the effect of model intention recognition.

1. To improve the understanding of the dialogue system of intelligent HVAC control systems (Jeon,2022), Hyunsik Jeon and colleagues developed an approach called "SmartSense", which uses an architectural model like ChatGPT. It captures remote dependencies of sequence data through a self-attention mechanism. It also supports parallel computing to improve computing speed. In detail, this method uses a two-encoder system to improve the accuracy of the appliance reaction. Its self-attention mechanism encodes the device control and its context, effectively reflecting the interrelationship. Therefore, it is good at processing and understanding historical intentions by analyzing the historical behavior of users. The steps are as follows: the device operation is first coded. The motion encoder then establishes an attention score related to the device's motion and time. The action sequence encoder then establishes an attention score between actions. This structure ensures a coherent context between device operations, thus predicting device operations with greater accuracy. It's worth mentioning that this self-attention mechanism isn't unique to SmartSense. In fact, ChatGPT (evolution of the GPT model) uses a similar architecture. This highlights the key role of self-attention mechanisms in improving model predictions. The experimental data further underscores that SmartSense outperforms its peers in smart home recommendations, highlighting the efficacy of self-attention mechanisms in identifying remote dependencies and contextual connections.

2.The adage "data is the new oil" holds particularly true for the realm of language models. Accumulation of vast and diverse corpora is pivotal to the prowess of predictive models, particularly in the intricate and multifaceted domain of natural language understanding and generation. Tom B. Brown and his team demonstrated the power of larger data and model parameters with GPT-3. Similarly, Google's BERT (Gao, 2019), trained on the vast English Wikipedia (2.5 billion words), set new standards in multiple natural language tasks. OpenAI's Codex (Finnie-Ansley,2022), enriched by extensive code from public repositories, excels in understanding programming languages. Facebook's M2M-100 model (Fan, 2022), trained on data from 100 languages, can translate even between unseen language pairs. Clearly, while architecture matters, the volume and diversity of data are pivotal. From these examples, it's clear that while model architecture, like attention mechanisms, plays a crucial role, the sheer volume, and diversity of the corpus remain an instrumental factor. For applications like Smart HVAC Control Systems, where precision is paramount, marrying a robust model structure with an extensive and varied dataset is the recipe for success. The consistent accumulation of varied and high-quality data is poised to drive further breakthroughs in the effectiveness and applicability of language models in myriad domains. For improving the natural language understanding ability of Smart HVAC Control Systems, the current model attention mechanism model structure and the scale of the corpus have become key.

This paper explores the scheme of combining the large language model ChatGPT (GPT-3.5-turbo) with Smart HVAC Control Systems, showing the improvement of ChatGPT in the HVAC field in terms of semantic understanding ability and action recommendation ability. It demonstrates its precise adaptation to the ability of semantic understanding dialogue in vertical scenes, and precise prediction of the next user's device operation, to a large extent solving the technical limitations of traditional HVAC systems.

This ChatGPT application also demonstrates the actual application effect in the field of HVAC systems: human-computer dialogue, and the ability of ChatGPT to understand semantics and recommend actions has been significantly improved. In the realm of smart home human-computer dialogue systems, four primary modules typically govern the interaction (Figure 1): NLU converts user's natural language input into structured data for machines, DST retains the current state of the conversation, DPL determines the next action based on user needs and conversation state, and NLG translates machine actions back into natural language for the user. Traditional systems face challenges: NLU might struggle with polysemy and ambiguous inputs, DST can falter with extended dialogues, DPL, often rule-based, requires defining all conversation pathways, and NLG might sound robotic. ChatGPT offers significant enhancements: a broader understanding of diverse user inputs, more effective long conversation state tracking, a more flexible dialogue strategy, and a more human-like language generation. These improvements simplify development, enhance response time, and boost accuracy, positioning ChatGPT as a transformative tool in human-computer dialogue for smart homes.

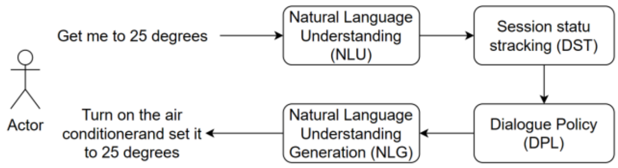

Figure 1: Traditional dialogue system

Methods

When constructing efficient and intelligent natural language processing system, it is very important to introduce multi-dimensional technology and method. A central hub technology architecture (see the Figure 2) is adopted to replace the traditional human-computer dialogue system, which is composed of four core components: ChatGPT API (see chapter 1.1), prompt engineering (see chapter 1.2), and control protocol (see chapter 1.3), vector database (see chapter 1.4). Firstly, the ChatGPT API provided by OpenAI provides efficient natural language understanding and generation. It provides users with a natural and smooth interactive experience. Secondly, the technical architecture uses embedding model to transform the local knowledge base into a vector database Pinecone for efficient data retrieval and management. To further optimize the natural language understanding ability of the system, COT and Few shot technologies are used in the prompt word to assist the accuracy and efficiency of the model in processing various natural language requests. Lanchain provides the prompt template. Finally, a set of natural language generation and control protocol is designed. It not only provides users with clear and easy to understand responses, but also enables real-time control of downstream HVAC systems by issuing control protocols to achieve seamless integration of intelligent control and user interaction. Through the integration of these four core technologies, the technology architecture provides a comprehensive and efficient solution for the development and application of natural language processing systems.

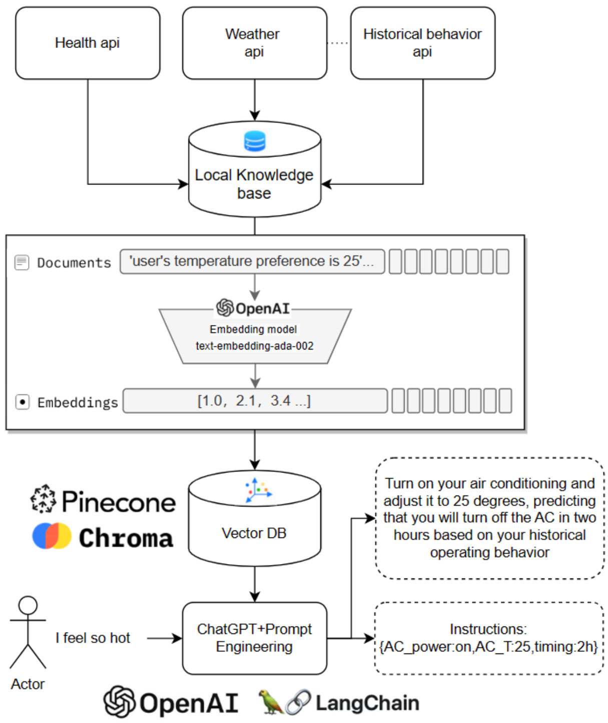

Figure 2: Dialogue system based on ChatGPT.

1.1 ChatGPT technical principles:

Open AI company provides ChatGPT API. ChatGPT is a large-scale language model based on GPT-3.5. Its architecture is a deep learning model based on Transformer (Vaswani, 2017) (Figure 3), which based on attention mechanism that shows remarkable performance for processing natural language tasks, especially text generation tasks. Transformer model (Vaswani,2017) is a key technology for processing sequence data and has been widely used in natural language processing. The core idea of this mechanism is to enable the model to pay targeted attention to relevant parts of the input sequence by calculating the weight of attention between different positions (see the formula (1)) (Vaswani,2017). The training stage of this model is divided into two parts, unsupervised pre-training, and fine-tuning [9]. In the pre-training phase, GPT is trained using a large Internet text dataset, including Wikipedia, books, news articles, web content, and so on. The goal of pre-training is let model learn and predict the content of the next word or paragraph of text through self-supervised learning. Specifically, GPT divides the input text sequence into fragments, and then, through a masking mechanism, masks some of these words, and the model needs to predict these obscured words. In this way, GPT can learn the semantic relationships between words and the grammatical structure of sentences. This approach is especially useful in the absence of large-scale annotated data and can help models achieve better performance on limited annotated data. After the pre-training is completed, the fine-tuning stage is entered. In the fine-tuning phase, the pre-trained model is further trained using task-specific labeled data. The process of fine-tuning generally uses supervised learning to update the parameters of the model by minimizing the difference between the generated and real responses. For ChatGPT, the data in the fine-tuning phase is usually a dataset about the chat conversation, which contains the pairing of questions and answers. By performing reinforcement learning training on these conversation data or adding additional rule constraints, the model can learn the ability to generate reasonable, accurate responses. This makes the model adaptable to specific task requirements at the same time.

$$\text{Attention}\ (Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (1)$$

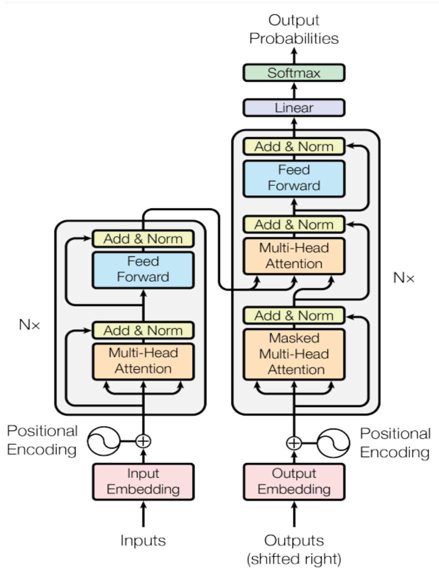

Figure 3: Transformer - Model structure (Vaswani,2017)

1.2 Prompt Engineering + Large Language Model -API development protocol

Prompt engineering combined with the large language model - API development protocol significantly improves the capacity of the language model, enabling its better application across diverse scenarios and research fields. A prompt could be a question, a description, or a complete instruction for a task, depending on the specific application scenario.

Prompt engineering encompasses various techniques, including Chain of Thought (COT) (Diao, 2023), Generated Knowledge Prompting, Few Shot (Diao, 2023), Zero Shot (Diao, 2023), Automatic Prompt Engineering (APE) (Zhou, 2022), and so on. Users can enhance the security, accuracy, and consistency of large language models through prompt engineering. This study primarily employs the COT and Few Shot methods.

In our prompt phrasing, we assign the role of a "home brain" to the ChatGPT, presetting the home appliance scenario and explaining the functional concepts of home appliances. The large language model then formulates outputs with high consistency in a fixed format, aligning with the outcomes of common sense.

In traditional methodology, an algorithm engineer might need to train a large language model in four steps: collecting specific field corpus, evaluating the labelled corpus, training the language model, and pre-release testing. The addition of a new intent would necessitate retraining (Figure 4).

This research utilizes the ChatGPT large language model API and prompt engineering technology to integrate with various smart home systems, thus achieving natural language understanding. This approach eliminates the need for model retraining (Figure 4).

In human-computer interaction, this method completes the task of developing protocol mapping - for example, translating natural language to a specific protocol, which will be sent to the device for execution. For instance, a user might say to the air conditioner: "My room is very hot, please turn on the air conditioner and adjust to my usual comfortable environment. I plan to go out tonight, my child will sleep in my room, please set the air conditioner to turn on at 10 PM and turn off at 7 AM the next day." The protocol generated is illustrated in the figure (Figure 5).

Through this process, we develop a unified control interface, and ChatGPT can realize intelligent control of the entire HVAC system based on the control protocol generated from natural language.

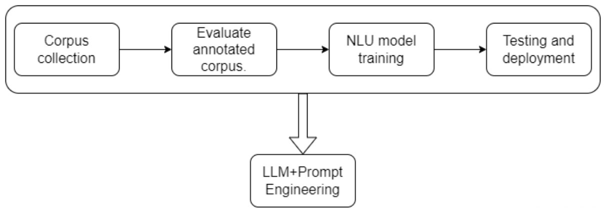

Figure 4: Large language model & Prompt Engineering.


```
{
   "classifier": "midea",
   "domain": "device_control",
   "utterance": "My room is very hot now，now help me to turn on the air conditioning to adjust my usual comfortable environment，I want to go out at night，my child sleeps in my room，set a time for the air conditioning，open at 10 PM，close at 7 am the next day。",
   "intent": {
      "devicename": "bedroom_AC",
      "power": "on",
      "temperature_set": "26",
      "humid_set": "60",
      "mode": "cool",
      "location": "bedroom",
      "action": "on",
      "timing_on": "on",
      "timing_off": "on",
      "timing_on_time": "22:00",
      "timing_off_time": "07:00"
   }
}
```


Figure 5: Designed protocol developed by Chat GPT

1.3: Explicit semantics understanding and action recommendations

In terms of semantic understanding, ChatGPT learns to understand and generate text by training data.

Specifically, ChatGPT enables deep-level semantic understanding. This means that it doesn’t only understand the explicit semantics of text, but also capture deeper implicit information. For example, the user says, "Turn down the temperature." Traditional voice control systems may only operate according to the two keywords "turn down" and "temperature". But with ChatGPT, the system can understand the user's instructions more deeply, such as determining which room AC the user wants to operate, and how much temperature the user wants to operate. ChatGPT is able to make inferences and judgments based on historical conversations, contextual information, and common-sense knowledge. It is able to understand the user's intentions and provide personalized responses.

In the field of whole-room intelligence, the running order of devices is contextual. [5] Device operations is translated into semantic representations, enabling ChatGPT to process and understand these operations and then predict the next device operation. For example, turning up/down temperature: "T_{up}", "T_{down}"; turning on/off air conditioning: "AC_{on}", "AC_{off}"; and so on. In this way, the device operations are converted into a text composed of these tags. ChatGPT understands the relationships between these tags and the position of these tags in the sequence. Based on this context, ChatGPT predict the next most likely tags, the next device operation. For example, give ChatGPT an operation sequence "AC_{on}, T_{up}, T_{down}", it might predict that the next operation will be "AC_{off} ". The previous device control will affect the use of the next device control. The time context also affects the user's behaviour. This kind of prediction help smart home systems better adapt to users' habits and provide more personalized services.

1.4 Vector DB

Vector Database (Han, 2023) is a specialized type of database designed for handling vector data. In the context of intelligent control of air conditioning, a local database integrates expert knowledge of air conditioning control, including a range of indicators of human comfort such as PM2.5, TVOC (volatile organic compounds), temperature, humidity, wind speed, etc. To make knowledge searchable efficiently, we chose to use the vectorization model provided by OpenAI to transform the local knowledge base into a vector database and deploy it on the Pinecone platform. In this architecture, the user's query will be matched with the content in the vector database by similarity calculation. The matched correlation vector data is parsed into natural language text and, together with user input. It is used as input data for ChatGPT to process. Through this process, ChatGPT is able to provide more professional and human-oriented answers.

Results

In this experiment, ChatGPT's capabilities are assessed based on its performance in semantic understanding using the Smart Home Commands Dataset (Ceunen, 2020) and in action recommendation through the Samsung SmartThings Internet of Things (Jeon, 2022). Table 1 provides an overview of the model's input fields, prediction outcomes, and prompt words for both datasets. The performance evaluation for action sequence prediction is found in section 2.1. For a detailed examination of the results related to semantic understanding, refer to section 2.2. Section 2.3 shows the dialog system, which includes a full-stack solution for HVAC intelligent control. This consolidated information offers a comprehensive understanding of ChatGPT's effectiveness and adaptability across different data structures and tasks.

Table 1, Prompt phrase of ChatGPT on two datasets

Dataset	Prediction	Prompt phrase
Samsung SmartThings Internet of Things (IoT) datasets	Day Time Device Device-control	You are an intelligent home brain. Identify patterns and predict the next piece of data based on the provided data in the fields of 'day', 'time', 'device', and 'device control'. 'Day' denotes the day of the week, 'time' refers to the number of hours (0-24), 'device' includes various household devices to be selected from {device_list}, and 'device_control' represents the specific action to control the device, selected from {device_control_list}. The format of the output result is exactly the same as the provided Example, no explanation for the result is needed. Example： Input: [['day:Wed','time:(21~24)','Other','Other:switch on'], ['day:Thu', 'time:(0~3)', 'Other', 'Other:switch on'], ['day:Thu', 'time:(3~6)', 'Other', 'Other:switch on'], ['day:Thu', 'time:(6~9)', 'Other', 'Other:switch on'], ['day:Thu', 'time:(9~12)', 'Other', 'Other:switch on'], ['day:Thu', 'time:(12~15)', 'Other', 'Other:switch on'], ['day:Thu', 'time:(15~18)', 'Other', 'Other:switch on'], ['day:Thu', 'time:(18~21)', 'Other', 'Other:switch on'], ['day:Thu', 'time:(21~24)', 'Other', 'Other:switch on']] Output：['day:Fri', 'time:(0~3)', 'Other', 'Other:switch on'] Input：{input} Output：
Smart Home Comman d-s Dataset	Category Action Subcategory	You are a home brain, and your task is to control devices based on user demands. Generate four fields - 'category', 'action', 'subcategory', and 'time' - based on the user input. 'Category' represents the device name and should be selected from {category_list}; 'action' stands for the operation of the device and should be selected from {action_list}; 'subcategory' denotes the location of the device and should be selected from {subcategory_list}. No explanation is needed for the output, just output the result. example: Input: Illuminate the kitchen today. Output: {'category': 'light','action': 'on','subcategory': 'kitchen'} Input: {sentence} Output:

2.1 Samsung SmartThings (IoT) datasets

dataset (recommended) from SmartThings platform, a global platform for the Internet of things, with 62 million users. The dataset provides equipment control history from four countries (South Korea, the United States, Spain and France) and equipment operation data from three regions (Asia Pacific, North America and Europe). The dataset is a matrix of N*10*4, where N represents N examples, 10 represents ten actions, and 4 represents the four fields: day, time, device, and device control. Using ChatGPT API and prompt engineering predict the dataset. A loop is used in the code to make predictions for every 9 pieces of data in the dataset. By defining an input template prompt, as shown in the figure, insert the values of the current nine data into the template and call ChatGPT API to generate the corresponding results. (table2). The table presents the performance metrics of various models on mean Average Precision at k (mAP@k) and Hit Rate at k (HR@k). These metrics are crucial for gauging the effectiveness of recommendation systems. Upon examination: The Gpt-4-prompt engineering consistently outperforms all other models across the metrics, followed closely by Gpt-3.5turbo-prompt engineering. Their superiority is evident across all levels of k, indicating a strong recommendation accuracy not just for the top recommendation but extending to the top 5. Observing the mAP@k, Gpt-4-prompt engineering excels, especially in single recommendation precision (mAP@1). Its lead is maintained even when the number of recommendations increases to 3 and 5. In terms of HR@k, Gpt-4-prompt engineering remains robust, showcasing its ability to provide not only accurate but also relevant recommendations to users. There's a noticeable enhancement from Gpt-3.5turbo-prompt engineering to Gpt-4-prompt engineering, suggesting meaningful model improvements in the GPT series with each iteration. Compared to the other models, the latter two, especially. Gpt-4-prompt engineering, demonstrate unparalleled dominance, emphasizing the potential superiority of the GPT series for such tasks. Gpt-4-prompt engineering emerges as the top contender in recommendation accuracy and relevancy, underscoring its potential for real-world applications.

Table 2, performance of ChatGPT on the Samsung SmartThings IOT dataset

Model	mAP@k			HR@k		
	@1	@3	@5	@1	@3	@5
POP	0.18	0.3	0.54	0.23	0.65	0.70
FMC	0.50	0.63	0.65	0.50	0.79	0.88
TransRec	0.38	0.56	0.58	0.38	0.75	0.83
Caser	0.56	0.70	0.72	0.56	0.87	0.91
SASRec	0.57	0.70	0.72	0.57	0.86	0.92
BERT4Rec	0.57	0.72	0.72	0.59	0.87	0.94
CA-RNN	0.59	0.73	0.74	0.57	0.84	0.90
SIAR	0.59	0.72	0.73	0.59	0.88	0.93
SmartSense	0.65	0.76	0.77	0.65	0.89	0.90
Gpt3.5turbo	**0.67**	**0.82**	**0.85**	**0.68**	**0.92**	**0.93**
Gpt-4	**0.73**	**0.85**	**0.88**	**0.69**	**0.94**	**0.95**

2.2 Smart Home Command-s Dataset

Semantic understanding dataset s a total of 664 data including data sets, what part of the data represents the user command in the field of smart home. The other fields category, action, subcategory, and time together form the intent part. Semantic understanding is carried out according to sentence. Prediction category, action, subcategory, and time are four fields. Correct prediction means correct understanding of intention. The study calls OpenAI API cycle each predict what field of the dataset, prompt as shown in table3 generate the corresponding results. OpenAI's API and prompt engineering methods' accuracy is 0.94 (table3).

Table 3, Comparison of ChatGPT and traditional machine learning models on the Smart Home Commands Dataset

Model	Accuracy
Random forests	0.88
Support vector machines	0.86
Xgboost	0.89
Multi-layer perceptrons	0.87
Lghtgbm	0.716
GPT-3.5-turbo	**0.94**
GPT-4	**0.98**

2.3 HVAC man-machine dialogue system

The research presents a full solution from modelling to deployment based on ChatGPT and prompt engineering method. Leveraging the power of large language models in home automation presents a novel approach to streamlining human-device interactions. The capacity of these models to interpret and act upon natural language commands enhances user experience by making smart HVAC truly "smart".

Table 3 further illustrates that on Sat 9am, when the statement "I'm so hot" is made, ChatGPT can comprehend the implicit semantic meaning. Even though there is no explicit request to raise the temperature, ChatGPT can understand the user's need and may suggest lowering the indoor temperature. On Mon at 13pm, based on the user's past behaviour, ChatGPT can predict the user's next device operation and prompt when to trigger it.

Table 3 ChatGPT-based human-machine HVAC dialog system parameters demo

Time	Input	User Status	GPT replies	Execution	History
Sat9am	I'm so hot	User intention: want to turn on the AC, Tasks to be completed: [], User preference: []	The bedroom HVAC in the living room has been turned on for you	[Sat 9am, AC: on]	[ ]
Sun10am	I'm so bored right now, do something I usually like, I want to listen to music	User intent: want to listen to music, Tasks to be completed: [], User preference: Listen to music	Turn on music for you	[Sun10am, Music: on]	[Sat 9am, AC: on],
Sun11am	I think the temperature is a bit high	User intention: want to reduce the temperature, Tasks to be completed: [], User preference: Listen to music	The temperature of the HVAC in the bedroom is lowered for you	[Sun11am, T: -1]	[Sat 9am, AC: on], [Sun10am, Music: on],
Sun12am	My ears are a little tired, turn off the music for me	User intention: want to turn off the TV, Tasks to be completed: [], User preference: Listen to music	The music has been turned off for you	[Sun12am, Music: off]	[Sat 9am, AC: on], [Sun10am, Music: on], [Sun11am, T: -1],
Mon13pm	I'm so bored	User intent: users want to be entertained, Tasks to be completed: [[Mon14pm, H: on]], User preference: Listen to music	Music has been turned on for you, according to the historical operation database: it is recommended that the next time you operate the device, turn on the HVAC humidification mode at 14pm on Monday	[Mon13pm, Music: on], [Mon14pm, H：on]	[Sat 9am, AC: on], [Sun10am, Music: on], [Sun11am, T: -1], [Sun12am, Music: off]
Mon14pm		User intent: [], Tasks to be completed: [], User preference: Listen to music			[Sat 9am, AC: on], [Sun10am, Music: on], [Sun11am, T: -1], [Sun12am, Music: off], [Mon13pm, Music: on], [Mon14pm, H: on]

Figure 6 shows the interactive dialogue system that using the PyQt5 third-party library in Python, integrated with the OpenAI API. This system interfaces with a layout representing various rooms inside a house, each equipped with multiple devices, including HVAC systems. Within the dialogue interface, users can input their requirements or needs. Once entered, these inputs are processed through the large language model, resulting in an output response that aligns with the user's demand. The graphical representation makes it intuitive for users, helping them visualize the effect of their commands in real-time, and witnessing the seamless interaction between natural language inputs and automated device operations in the household. In this case, the system is able to accurately identify the user's intent, execute the device in parallel, and give the user appropriate feedback.

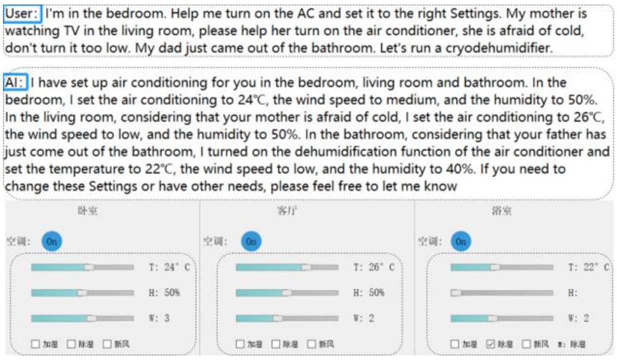

Figure 6 ChatGPT-based human-machine HVAC dialog system demo

Discussion

Advancements in machine learning and natural language processing are truly epitomized by the performance of the ChatGPT model across various experimental settings, particularly in the domains of IoT datasets and semantic understanding tasks.

Firstly, ChatGPT's performance on the Samsung SmartThings IoT dataset is exceptional, particularly with the Gpt-4-prompt iteration. It performs better than others in mAP@k and HR@k metrics, leading to better user experience and efficiency. Secondly, In the Smart Home Commands Dataset, ChatGPT excels at understanding both intent and context. This strength is crucial for HVAC applications and can lead to streamlined processes and cost savings. Although ChatGPT does not perform as well as gpt-4 in both datasets, it is more than sufficient in the HVAC intelligent control area.

In the field of HVAC intelligent control, conversation systems, especially based on advanced technologies, have become the core of a new generation of interactive interfaces. It provides users with a more natural and intuitive way to interact with HVAC systems. In the mobile phone app end, small program end, centre control screen end, line controller end, the user through natural language to the HVAC language control. The dialog system analyses policies (DPL) by parsing what the user says. It helps the HVAC intelligent control system to issue commands to control the HVAC and return the dialogue results. ChatGPT plays a key role in the dialog system, especially in the dialog policy segment (DPL). ChatGPT's powerful semantic understanding ability can more accurately identify the user's intention for the instructions of home appliances, and then generate dialogue and control strategies that meet the needs. The ChatGPT dialogue system enables the HVAC intelligent control system to precisely issue instructions to the HVAC for control and can also return natural language expression in the human-machine dialogue. ChatGPT can also predict the next device operation desired by the user based on the sequence of device operations regulated by the user. It is worth noting that the device action can be a whole-house smart device, or it can be an HVAC device.

While ChatGPT performs well in many areas, its reliance on prompt construction is a challenge. Effective prompts can be time-consuming and may need domain expertise, highlighting the need for domain-specific fine-tuning. Additionally, the robustness of ChatGPT, though an advantage, could also be a double-edged sword. While the model is efficient at generating human-like responses, ensuring that it doesn't generate misleading or incorrect information under ambiguous inputs remains a challenge. The illusion of model output is not resolved. A fail-safe or an overseeing mechanism might be necessary in critical applications to ensure optimal outcomes.

ChatGPT showcases remarkable promise in both recommendation and semantic understanding tasks, presenting a potential game-changer for sectors relying on these domains. While it brings forth considerable advantages in accuracy and cost-efficiency, ensuring its proper application with the right checks and balances will be the key to unlocking its full potential.

Conclusion

The intelligent capability of HVAC has always faced the challenges of insufficient semantic understanding ability and poor action recommendation accuracy. To address this issue, this paper adopted the ChatGPT and prompt engineering approach, and conducted semantic understanding ability tests on the home appliance command dataset, as well as action recommendation ability tests on the home appliance equipment action dataset. The research results indicate that, compared with traditional models, the ChatGPT and prompt engineering method performed better in terms of semantic understanding ability and device action recommendation prediction capability. This validates that integrating this technology with HVAC systems will undoubtedly enhance their intelligent capabilities.

In addition, this study developed a dialogue system based on ChatGPT, specifically highlighting its superior capabilities in semantic understanding and action recommendation. This research outcome provides significant prerequisites for the development of HVAC and future whole-house intelligence.

References

Adam Zielonka, Marcin Wozniak, Sahil Garg, Georges Kaddoum, Md. Jalil Piran, and Ghulam Muhammad. 2021. Smart Homes: How Much Will They Support Us? A Research on Recent Trends and Advances. IEEE Access (2021).

Alaa, M., Zaidan, A. A., Zaidan, B. B., Talal, M., & Kiah, M. L. M. (2017). A review of smart home applications based on Internet of Things. Journal of Network and Computer Applications, 97, 48-65.

Ceunen, B. (2020, May 22). Dataset for Smart Home. Retrieved from Kaggle : https://www.kaggle.com/datasets/shegguy87/dataset-for-smart-home

Ceunen, B. (2021, May 25). NLP Smart Home Commands Classification. Retrieved from Kaggle: https://www.kaggle.com/code/bouweceunen/nlp-smart-home-commands-classification

Chan, M., Campo, E., Estève, D., & Fourniols, J. Y. (2009). Smart homes - current features and future perspectives. Maturitas, 64(2), 90.

Chen Ma, Liheng Ma, Yingxue Zhang, Jianing Sun, Xue Liu, and Mark Coates. 2020. Memory Augmented Graph Neural Networks for Sequential Recommendation. In AAAI. AAAI Press.

Chu, S. and A. Majumdar (2012). Opportunities and challenges for a sustainable energy future. Nature 488, 294–303.

Clarke, J. (2015). A vision for building performance simulation: a position paper prepared on behalf of the IBPSA Board. Journal of Building Performance Simulation 8(2), 39–43.

Diao, S., Wang, P., Lin, Y., & Zhang, T. (2023). Active prompting with chain-of-thought for large language models. arXiv preprint arXiv:2302.12246.

Fan, A., Bhosale, S., Schwenk, H., Ma, Z., El-Kishky, A., Goyal, S., ... & Joulin, A. (2021). Beyond english-centric multilingual machine translation. The Journal of Machine Learning Research, 22(1), 4839-4886.

Finnie-Ansley, J., Denny, P., Becker, B. A., Luxton-Reilly, A., & Prather, J. (2022, February). The robots are coming: Exploring the implications of openai codex on introductory programming. In Proceedings of the 24th Australasian Computing Education Conference (pp. 10-19).

Gao, Z., Feng, A., Song, X., & Wu, X. (2019). Target-dependent sentiment classification with BERT. Ieee Access, 7, 154290-154299.

Han, Y., Liu, C., & Wang, P. (2023). A Comprehensive Survey on Vector Database: Storage and Retrieval Technique, Challenge. arXiv preprint arXiv:2310.11703.

He, Z., & Wei, W. (2022). China's financial system and economy: a review. Annual Review of Economics, 15.

Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. Egyptian informatics journal, 16(3), 261-273.

Jeon, H., Kim, J., Yoon, H., Lee, J., & Kang, U. (2022, October). Accurate Action Recommendation for Smart Home via Two-Level Encoders and Commonsense Knowledge. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management (pp. 832-841).

Mcloughlin, Ian & Sharifzadeh, Hamid. (2008). Speech Recognition for Smart Homes. 10.5772/6363.

Nasukawa, T., & Nagano, T. (2001). Text analysis and knowledge mining system. IBM systems journal, 40(4), 967-984.

Saha, T., Saha, S., & Bhattacharyya, P. (2020). Towards sentiment aided dialogue policy learning for multi-intent conversations using hierarchical reinforcement learning. PloS One, 15(7), e0235367–e0235367. https://doi.org/10.1371/journal.pone.0235367

Souza Dutra, Michael David, Miguel F. Anjos, and Sébastien Le Digabel. “A Realistic Energy Optimization Model for Smart‐home Appliances.” International journal of energy research 43.8: 3237–3262. Web.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

Wen, T. H., Vandyke, D., Mrksic, N., Gasic, M., Rojas-Barahona, L. M., Su, P. H., ... & Young, S. (2016). A network-based end-to-end trainable task-oriented dialogue system. arXiv preprint arXiv:1604.04562.

Yang, R., & Newman, M. W. (2013, September). Learning from a learning thermostat: lessons for intelligent systems for the home. In Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing (pp. 93-102).

Zhang, H., Zeng, Z., Lu, K., Wu, K., & Zhang, S. (2022, June). Efficient dialog policy learning by reasoning with contextual knowledge. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 36, No. 10, pp. 11667-11675).

Zharova, A., & Lee, H. E. (2022). Understanding User Perception and Intention to Use Smart Homes for Energy Efficiency: A Survey. arXiv preprint arXiv:2212.05019.

Zhou, Y., Muresanu, A. I., Han, Z., Paster, K., Pitis, S., Chan, H., & Ba, J. (2022). Large language models are human-level promp