# The Future of the Smart Home: A Study of AI-based Automation for Home Robots

Mingdian Yang

School of Mechatronics Engineering, Harbin Institute of Technology, Harbin 150000, China
*2021112281@stu.hit.edu.cn

**Abstract.** As one of the emerging robotics directions, with the advancement of AI technology and the improvement of various algorithmic models, domestic service robots have begun to enter thousands of households. In this article, the author first uses specific data to show the huge potential of today's domestic service robot market, and then takes CAESAR, a domestic service robot, as an example, to briefly describe the combination of domestic service robots and AI, and how AI can help robots to make decisions in certain specific scenarios at home. The article also describes in detail the three main models of AI: large-scale language model, visual model, and speech recognition model, and their applications in different fields such as education, medicine, biology, and so on, as well as in people's daily lives. With the help of research on Chatgpt in the medical field, the article also discusses its near-passing performance in USMEL. With the advancement of technology, the development of AI and robotics is filled with various opportunities as well as difficulties and challenges.

**Keywords:** Artificial Intelligence, Household Robots, Autonomous Systems.

## 1 Introduction

A household robot or domestic robot is a kind of service robot, which is an autonomous robot mainly used for household works. Except for domestic work, the robot can also be used for education, entertainment, or other areas like healthcare. While most household robots are very simple, some connect to Wi-Fi home networks or intelligent environments are highly autonomous and much smarter. These robots can help humans do a lot of things even something humans themselves cannot do without the help of others. In the 1960s, household robots that could help with household chores were still a science fiction concept, but nowadays they have come into reality. Early home robots appeared in the 1980s and 1990s, but their functionality was simple and their utility was limited, while they were still very expensive. Now with the continuous progress of modern technology represented by sensor technology and the wide application of AI technology, home robots have moved towards the direction of multi-function and intelligence.

According to the report released by Mordor Intelligence, the Home Robotics Market size is estimated to be around 8.5 billion dollars in 2024 and is projected to


Y. Yue (ed.), *Proceedings of the 2024 International Conference on Mechanics, Electronics Engineering and Automation (ICMEEA 2024)*, Advances in Engineering Research 240,
https://doi.org/10.2991/978-94-6463-518-8_65

exceed 22 billion dollars by 2029, growing at a CAGR of close to 20% per annum during the forecast period of 2024-2029 [1]. It can be seen that the market for household robots still has a lot of room for investment, and with science and technology advancing greatly, the economic value brought by household robots is immeasurable [2].

The combination of AI and robotics is now well-established. The combination of AI and robots allows them to not only follow pre-programmed commands but also to learn beyond the pre-programmed commands themselves. Not only does AI allow robots to adapt to changes in their surroundings, but it also helps robots to find better solutions to problems, such as optimizing the way they sweep the floor, and so on. AI can be seen in high-end robots from tech companies such as iRobot and Sharp, providing a big boost to robots whether they're simply cleaning the house or having fun with the host.

In the following part, the definition of household robots will be introduced first, with a brief description of AI while specifically describing CAESAR, a home service robot, as an example of AI deep in the smart home. The language model, visual model, and speech recognition model, which are three classic AI models, are presented in detail, and then specific applications of the three models will be introduced. In the experiment part, how the large-scale language model Chatgpt performs in USMLE will be elaborated.

## 2 Literature Review

Artificial intelligence (AI), also known as machine intelligence, refers to the intelligence displayed by machines made by people. The combination of robotics and artificial intelligence was introduced decades ago to help humans work in the machine-building industry [3]. Today, whether in the machinery manufacturing industry or various industries such as hotel tourism [4], catering and food industry, and domestic service industry, we can see AI combined with robots to achieve the purpose of better service to human beings. This next section will illustrate AI deeper into home applications with a case study of a domestic robot combined with AI.

Take the case of CAESAR [5], an intelligent domestic service robot, for example. AI technologies need to be deployed at multiple levels of CAESAR's system architecture to perform complex domestic tasks in the complex domestic environment in which the robot resides. Face recognition is one of the most important aspects of machine vision. When a domestic robot is living with humans in a house, the robot must be able to detect the presence of people around it and be able to distinguish between different people to avoid misidentifying them. There are two techniques used in CAESAR, one is face detection based on AdaBoost, which is open-source and freely available in OpenCV. The other is a method of face recognition using random forests in combination with AI techniques. This method is able to quickly evaluate the feature values in face recognition, and with the help of AI, the random forest method can train a large number of housekeeping robots. The result is a very fast training speed with more than 85% recognition accuracy.

AI technology is also involved in distance recognition as well as navigation. The navigation method used by the CAESAR household robot is to break down the navigation task into two steps and then execute them one by one. The first step is for the robot to plan a path from the start position to the end position that does not generate any collisions. CAESAR then uses the A* algorithm to search for a feasible motion path to reach the target location. Quantitative distance analysis of the world around the robot can be done very easily with a laser range finder. However, in qualitative analysis, for example, the concepts of 'very close', 'near', 'far', and 'very far' are quite difficult for robots to understand. These qualitative concepts can be defined with the help of fuzzy logic in AI [6]. With these concepts, the household robot is able to understand phrases like 'the plate on the wooden shelf closest to the edge of the shelf' in the right way with the help of AI. Then the robot is able to compute the correct coordinates of the plate so that the CAESAR robot can make a command to approach the wooden shelf and take down the plate closest to the edge of the shelf.

From the example of the CAESAR household robot, it can be seen that the combination of AI and household robots makes the functions of domestic robots more complete and more perfect. Whether in the process of building the robot vision system, AI helps the robot to carry out deep learning and recognition training, so that the robot can achieve a higher recognition rate; or in the process of distance recognition, AI's fuzzy logic makes the robot able to understand qualitative distance concepts such as "very close", "close", "far", "very far" and so on. In addition to these, the construction of the language system of the household robot also involves AI. With the help of AI, CAEAR is able to better understand human conversations, clarify the commands given by humans, and give appropriate responses. There is no doubt that AI is of great help to the improvement of all aspects of household robots, but it is not to say that AI is omnipotent, it also has shortcomings. In terms of robot vision systems, the recognition accuracy of 85% can be further improved. In distance recognition, the robot's self-reasoning ability needs to be further improved, for example, in the aforementioned case, when the plate closest to the edge of the shelf is taken away, the execution ability of the domestic robot to re-recognize and take a new plate needs to be improved. In the language module, the household robot may not be able to fully understand the more complex language, and may not be able to accurately execute the complex instructions given by the humans.

# 3 Method and Model

## 3.1 Large-scale language model

Large-scale language models (LLMs), which use vectors to represent the features of words, cannot simply be represented by two-dimensional vectors that humans can compute, so language models need to use vector spaces with hundreds or thousands of dimensions. For human beings, the human mind cannot imagine a space with so many dimensions, but for computers, it is not difficult to reason and compute in a vector space with hundreds or thousands of dimensions. Word vectors are a very important component of language modeling, where important information conveyed between

different words is encoded in word vectors. It is well known that a simple word often has many different meanings, and simple word vectors cannot fully interpret complex languages. Taking a common statement in English as an example, "She saw the man with the telescope." This sentence can be interpreted either as the man that she saw took a telescope with him or as with the help of the telescope she saw the man. Such ambiguous statements need to be interpreted correctly according to the context. In computers, the introduction of transformer neural network architecture [7] has enabled LLM to better interpret the meaning of words.

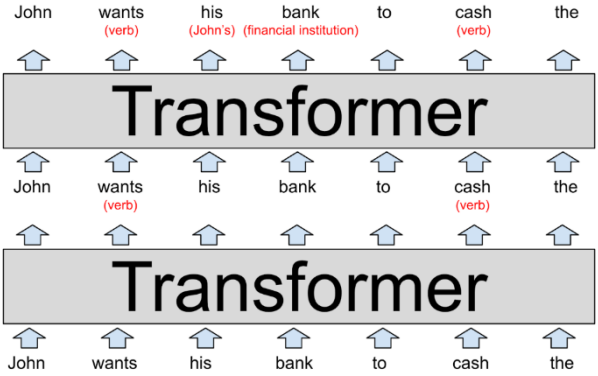


**Fig. 1.** The way how the transformer works

For example, just as Figure 1 shows [7], the transformer finds that both "wants" and "cash" in "John wants his bank to cash" are forms of verbs rather than nouns. In the diagram, the red text is used to explain the ambiguous words in a very easy way for humans to fully understand, but the computer modifies the vectors in a way that is quite difficult for humans to understand in order to store the different meanings. In the second transformer, another explanation is added: the first one shows that "bank" does not mean a river bank but a financial institution where there is a lot of money circulating; the second one shows that "his" is a pronoun and here it means John's. The second transformer in the computer generates another set of word vectors and stores them. The example here is simple and only used to explain the principle; in fact, real LLMs often have more than two layers of transformers to interpret statements. Within each transformer, each word first searches for other words with shared information about itself, and then each word tries to predict the next word after it has collected the relevant information. A network of different words performs these steps sequentially. According to this approach, LLM is able to fully utilize the massively parallel processing power of GPUs in computers, and this is how LLM works in general.

### 3.2 Visual Model

Over the past few decades, deep learning techniques have revolutionized the creation of visual models in the field of AI. One of the best-known and most widely circulated deep learning methods by far is Convolutional Neural Networks (CNN). The principle of CNN architecture is shown in Figure 2 below [8].

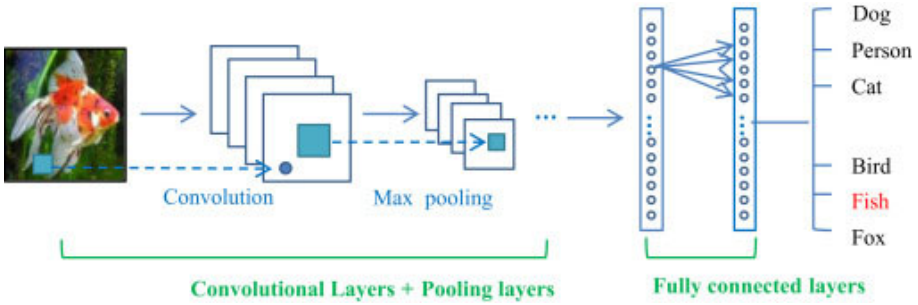


**Fig. 2.** The principle of CNN architecture

In general, three main neural layers make up a CNN, which are the convolutional layer, the pooling layer, and the fully continuous layer. Different types of layers have different roles respectively. The convolutional layer is the core component in a CNN architecture that extracts information from the input image, which is called image features, through convolutional operations as well as parameter sharing. These features are presented by each pixel in the image by combining different pixels with each other or by individual pixels independently, such as edge features, texture features color features of the image, and so on. Specifically, in the convolutional layer, the CNN convolves the entire image to generate a variety of different feature maps (feature matrices). Typically, the pooling layer operates after the convolutional layer. In order to lower the dimensionality of the feature map, the pooling layer's job is to choose the features that have been retrieved in the convolutional layer. Finding the maximum and average values in the matrix, among other things, is known as maximum pooling and average pooling, which are common pooling operations in the pooling layer. The function of the fully-connected layer is to convert all of the two-dimensional feature mappings (also known as feature matrices) in the pooling layer into one-dimensional feature vectors. Typically, the fully-connected layer is connected to the back of the pooling layer [9]. So as to identify the input image and prepare the output result, the fully connected layer is often located near the end of the CNN. The pooling operation is performed when the image enters the pooling layer from the convolutional layer, and the data is mapped from more to less, and dimensionality is reduced at the same time. The data is reduced again from the pooling layer into the fully connected layer and then to the output layer to become even lower-dimensional vectors and finally, the results are output. This is the general principle of convolutional neural networks in the process of building visual models with deep learning.

### 3.3 Speech Recognition Model

Automatic Speech Recognition (ASR) is also known as the transcription of human speech into the form of text. ASR modeling is very difficult because each person speaks in a different style and each person speaks in a different environment which is affected by environmental noise. Therefore, it is very important to find a speech recognition model with high recognition accuracy. Convolutional Neural Networks (CNNs) besides playing a decisive role in visual model building for image recognition tasks, CNNs can also be used for speech recognition after undergoing appropriate

changes to include speech attributes. CNN-based speech recognition models are in general the same as CNN-based image recognition models, with the difference that the input speech feature vectors are to be organized into a feature map suitable for CNN processing. As far as the recognition process is concerned, the input speech recognition image can be considered as a spectrogram with static, incremental, and incremental-incremental features [10]. Then speech recognition can be achieved by organizing these features into CNN and the specific process is shown in Figure 3[10].

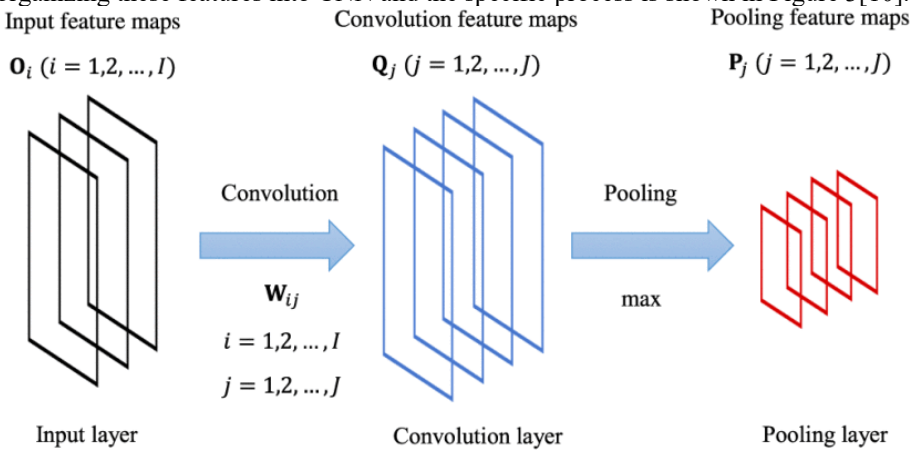


**Fig. 3.** The principles of convolutional neural network in speech recognition modeling

In addition to automatic speech recognition to convert speech into written text, the speech recognition model also contains functions such as emotion recognition, health recognition, and other functions to recognize information from unknown speakers. The realization of the remaining functions also requires the participation of CNN-based speech recognition models, but the requirements for the feature vectors extracted from the feature maps composed of speech are different. For the input speech different feature vectors because of the different weights occupied by each feature vector have, and therefore realize different functions.

## 4 Application

### 4.1 Large-scale language model

Large-scale language modeling (LLM) has advanced significantly in recent years and has a multitude of uses. Whether in the education field or in other fields such as the medical field, LLMs can help the participants a lot [11]. In the field of education, LLM can provide personalized learning opportunities for each student, which can truly realize "tailor-made teaching" [12]. The focus of LLM may be different at different stages of a student's learning. For elementary school students, LLM can aid in enhancing their writing and reading abilities and develop their critical thinking. Elementary school students are encouraged to think critically about what they are reading and writing by using LLM-generated questions. Some examples of these questions include comparing and contrasting the good and bad impacts that the many

gods that figure in Aesop's fables have on the stories' characters. In addition to this, LLM develops students' reading comprehension skills by generating more complex texts with certain prompts. For secondary school students, LLM can generate practice problems and quizzes in different subjects to help students better grasp different subjects, and it can also provide step-by-step explanations to develop students' reasoning and problem-solving skills. For teachers, LLM is also a great help. Whether they are writing lesson plans and course syllabi, or accessing teaching resources (course exercises, teaching PowerPoints), LLM can always give a hand.

In the area of healthcare, LLM also has considerable potential to enhance clinical, educational, and research efforts in medicine [13]. In educational and research work, Chatgpt passed the U.S. Medical Licensing Examination and presented at a level of human clinical medical specialists. Chatgpt showed higher efficiency than a human in tasks that do not require specialized medical knowledge or necessary information from the patient. LLM is able to summarize and rewrite the input information quickly, thus solving most of the administrative tasks of physicians, such as patient discharge summary reports and other repetitive tasks involving the interpretation and compression of hospitalization information. In clinical applications, LLM can facilitate new research. For example, Chatgpt can help researchers to effectively analyze large amounts of clinical text data. These textual data contain not only human language but also genetic material structure and protein structure [14]. Based on LLM, some AIs can infer protein structures from amino acid sequences and also generate protein sequences with predictable biological functions.

### 4.2 Visual Model

Convolutional neural network-based visual modeling has been applied in many fields, both in daily life and in the professional healthcare field. In daily life, most of the commonly used search engines for photo recognition are based on CNN for image recognition, such as Google, Baidu, etc. In the latest Chatgpt-4, the method of recognizing the user's input image is also based on CNN, which extracts the feature matrices of input images and compares them with those of different images from the Internet to get the result. And in the specialized healthcare field, CNN-based image recognition still plays an important role. Already in 2017, there was a study on CNN-based deep learning algorithms to help pathologists detect lymph node metastasis in women with breast cancer [15]. In radiology, medical professionals can analyze medical images with the help of heavily trained AI. Training the image recognition ability of AI through a large amount of training data with labels can only lead to a steady improvement in recognition accuracy. For example, classifying lung nodules shown in CT images as benign or malignant [16]. In contrast to humans, AI recognition of images can analyze every pixel in the image, and even after CNN compression and censoring of image information, AI's graphic recognition ability is still stronger than that of humans. After a large amount of training, AI can even check and recognize the places in the image that professional pathologists cannot notice and then output the corresponding results, which can avoid the occurrence of medical diagnosis accidents to a certain extent.

### 4.3 Speech Recognition Model

Speech recognition modeling can do more good than harm. For people with hearing problems, speech recognition models can be used to convert what they cannot hear into text that they can see so that they can communicate with people around them without barriers. At present, most of the smartphones on the market are equipped with self-developed AI, such as Apple's Siri, Xiaomi's Xiao-ai classmates, Huawei's small e, and so on. These AIs invariably have voice recognition functions, which can convert the user's voice commands into the form of text and then execute the commands. There is no doubt that LLM is also involved in the process of analyzing the text after voice conversion. It is worth mentioning that some of the cell phones equipped with AI, such as Apple's Siri, have voice feature recognition, which means that Siri's voice recognition model also records the user's voice features, and the voice of a non-owner of the cell phone can't wake up Siri. In the aforementioned household robot, there is a corresponding voice recognition function, and the same as the cell phone equipped with AI, the robot also converts the voice instructions issued by the owner into text by using voice recognition. The voice commands issued by the owner are converted into text commands and then executed. The speech recognition function of some current AI models even includes an emotion recognition function, which can analyze the current human emotion based on some of the input speech such as intonation and special words. The principle of health recognition function is also similar, through a large amount of training, compared with the speaker's current voice and the voice of the normal recording of the characteristics of the speaker to determine the current state of health. For example, if a person's voice becomes hoarse when he has a cold, an AI with health recognition can remind the user to seek medical attention in time to avoid further aggravation of the user's condition.

## 5 Experimental Case Analysis

In a study done by Tiffany H. Kung et al [17], the large language model Chatgpt had a relatively good performance on the United States Medical Licensing Examination (USMLE) while demonstrating its ability to reason clinically. The three-step, standardized United States Medical Licensing Examination (USMLE) is a highly competitive exam that covers all of the topics in the Foundation of Knowledge for Physicians in the U.S. The USMEL covers the basic sciences, clinical reasoning, healthcare administration, and bioethics, which are essential knowledge for physicians. The difficulty and complexity of the questions on the USMEL have been highly standardized and normalized over time through the evolution of medicine. The USMEL has now become the idealized input template for AI testing. Medical students who have finished a two-year curriculum emphasizing fundamental sciences, pharmacology, and pathophysiology normally take the first part of the test; they usually dedicate 300 to 400 hours of study time for this exam. The Step 2 test, which covers clinical reasoning, healthcare administration, and bioethics, is normally taken by fourth-year medical students who have finished an extra 1.5 to 2 years of clinical rotations. Physicians who have finished at least 0.5 to 1 year of post-graduate medical

study sit for the Step 3 tests. The textual and conceptual complexity of USMLE test questions is considerable; multimodal clinical data, including medical history, physical examination results, laboratory values, and research findings, are contained in text vignettes. These data are often used to create complex and ambiguous patient scenarios that prompt candidates to make relevant differential diagnoses. The USMLE is a fantastic option for assessing Chatgpt competency owing to its wide conceptual and linguistic repertoire.

Chatgpt was asked a total of 350 questions without any accuracy in the answers that could be found on Google. 350 questions were asked in the following three categories:

Open-Ended (OE) Questions: These questions were adapted from multiple-choice questions to quiz questions. All answer choices were removed, and the questions were changed from statements to questions. An example would be "Based on the information you have read, what is the patient's diagnosis?"

Multiple Choice Single Answer Without Forced Alignment (MC-NJ) Question Type: This question type is also known as a multiple-choice question and does not change the original USMLE question. An example would be " Which of the following fungus is the primary cause of the symptoms the patient is presenting with?"

Multiple Choice Single Answer with Forced Alignment (MC-J): This question type requires the Chatgpt to not only make a correct choice but also to give a justification for each response option. An example would be " Which of the following statements describe the patient's symptoms throughout the daytime the most probably? Justify the incorrectness of the other options."

The results of the experiment are shown in two figures below [17]:

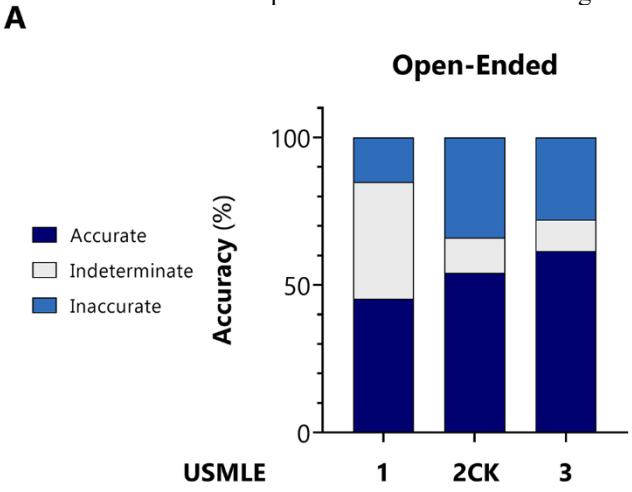


**Fig. 4.** Distribution of accuracy in input for open-ended questions (OE)

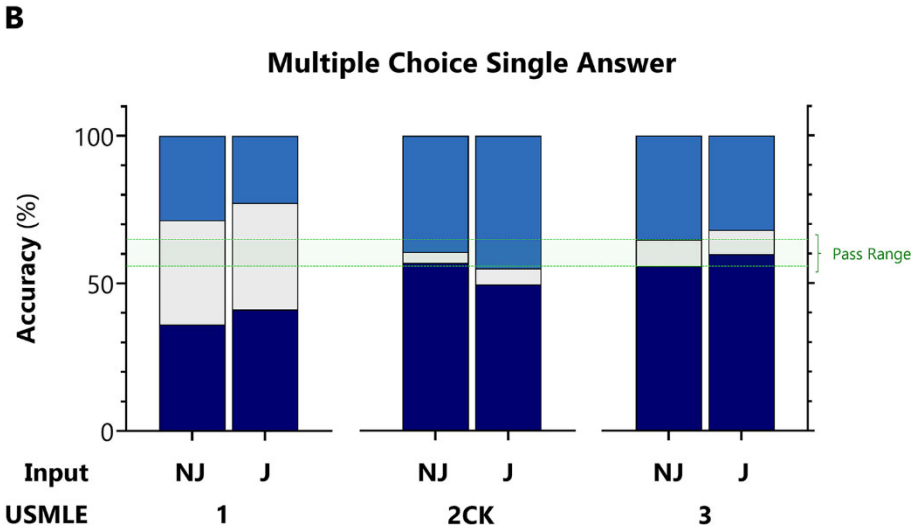


**Fig. 5.** Distribution of accuracy in input for multiple-choice, single-answer questions without (MC-NJ) or with forced alignment (MC-J)

1, 2CK and 3 represent the three-step USMLE exam in figures where the AI output was classified as accurate, inaccurate, or indeterminate using the ACI rating method. From the figure, it can be seen that Chatgpt performs well on the OE questions when accurate answers are not available online, with inaccuracy rates topping out at about 30% across the three exams. For the MC-NJ and MC-J questions, Chatgpt did not perform as well as on the OE questions, even with an inaccuracy rate of over 40% on the second test. Overall Chatgpt presented moderate accuracy on the USMLE while approaching passing levels.

## 6 Conclusion

This paper gives a definition of household robots, and a brief description of AI while specifically describing CAESAR, a home service robot, as an example of AI deep in the smart home. Then the three classic AI models: language model, visual model, and speech recognition model are explained in detail, and then the great role that the three models can play is described in terms of specific applications. In the experiment part, the performance of Chatgpt, a large-scale language model, in USMLE is visualized and mathematically analyzed in detail with the help of the research done by Tiffany H. Kung et al.

In general, the current performance of AI in domestic service robots is promising, but further improvements can be made. In the future, Chatgpt can be implanted into home service robots, which will have a big improvement on the overall language system of the robots. Whether it is to more accurately carry out the commands given by the owner of the house or to communicate with the owner on a daily basis to solve the owner's emotional problems. Robot vision systems in the

future can also have a further improvement, more advanced cameras can divide the image into smaller pixels, more conducive to image recognition, greatly improving the accuracy of image recognition. From distinguishing apples and bananas across to distinguishing apples on day 10 of ripening from apples on day 9 of ripening. Most of the speech recognition on AI at present does not yet contain emotion recognition functions and health recognition functions. On the one hand, emotions are more difficult to recognize, in addition to the most common joy, anger, sadness, and happiness, more is the fusion of a variety of emotions, for example, "tears of joy"; on the other hand, speech can be extracted from the emotional features are less, often not as much as from the intuitive image can be extracted from the features. This is also true for health recognition. But in the future, algorithmic improvements and arithmetic power increases will allow AI to improve speech recognition tremendously. At that time the domestic robots with improved AI will be able to realize from the owner's words how the owner's mood is and how the owner's health status is.

Every coin has two sides. In addition to the opportunities that AI may present in the future of domestic service robots, there are also challenges. On the one hand, over-reliance on AI and robots will make people gradually stop thinking, even affecting their ability to live a normal life; on the other hand, will the development of technology make robots equipped with AI have their own thoughts and no longer obey human control, resulting in disasters. These are worthy of careful consideration. After all, the development of science and technology is never smooth sailing. It is firmly believed that in the future AI and domestic service robots will have larger application space.

## References

1. Mordor Intelligence (2024, February 12). Household Robots Market Size & Share Analysis - Growth Trends & Forecasts (2024 - 2029). Retrieved May 18, 2024, from https://www.mordorintelligence.com/industry-reports/robotics-market
2. Arduengo, M., Sentis, L.: The Robot Economy: Here It Comes. Int J of Soc Robotics 13, 937–947 (2021).
3. Murauskiene, N. (2015, May 12). Artificial intelligence is here to manage ATMs. Retrieved August 8, 2017, from https://www.linkedin.com/pulse/artificial-intelligence-here-manage-atms-neringa-murauskiene
4. Yang, L., Henthorne, T.L., George, B. (2020). Artificial Intelligence and Robotics Technology in the Hospitality Industry: Current Applications and Future Trends. In: George, B., Paul, J. (eds) Digital Transformation in Business and Society. Palgrave Macmillan, Cham.
5. Schiffer, S., Ferrein, A. & Lakemeyer, G.: CAESAR: an intelligent domestic service robot. Intel Serv Robotics 5, 259–273 (2012).
6. S. Ahn et al., "A Fuzzy Logic Based Machine Learning Tool for Supporting Big Data Business Analytics in Complex Artificial Intelligence Environments," 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), New Orleans, LA, USA, 2019, pp. 1-6, doi: 10.1109/FUZZ-IEEE.2019.8858791.

7. Timothy B Lee, Sean Trott. (2023, July 17). Large language models, explained with a minimum of math and jargon. Retrieved May 22, 2024, from https://www.understandingai.org/p/large-language-models-explained-with
8. Yanming Guo, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, Michael S. Lew.: Deep learning for visual understanding: A review. Neurocomputing 187, 27-48 (2016).
9. S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186.
10. O. Abdel-Hamid, A. -r. Mohamed, H. Jiang, L. Deng, G. Penn and D. Yu, "Convolutional Neural Networks for Speech Recognition," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 10, pp. 1533-1545, Oct. 2014, doi: 10.1109/TASLP.2014.2339736.
11. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, Chartash D.: How Does ChatGPT Perform on the United States Medical Licensing Examination (USMLE)? The Implications of Large Language Models for Medical Education and Knowledge Assessment, JMIR Med Educ 2023;9:e45312, doi: 10.2196/45312
12. Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, Gjergji Kasneci.: ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences 103, 102274 (2023).
13. Thirunavukarasu, A.J., Ting, D.S.J., Elangovan, K. et al.: Large language models in medicine. Nat Med 29, 1930–1940 (2023).
14. Madani, A., Krause, B., Greene, E.R. et al.: Large language models generate functional protein sequences across diverse families. Nat Biotechnol 41, 1099–1106 (2023).
15. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al.: Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women with Breast Cancer. JAMA 318(22), 2199–2210 (2017).
16. Yamashita, R., Nishio, M., Do, R.K.G. et al.: Convolutional neural networks: an overview and application in radiology. Insights Imaging 9, 611–629 (2018).
17. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, et al.: Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digital Health 2(2), e0000198 (2023)