

# Analysis of artificial intelligence models for the smart home industry

Yufeng Chen<sup>1</sup>, Yuheng Ren<sup>2,3,4,\*</sup>

<sup>1</sup>Uppsala University

<sup>2</sup>School of Digital Industry, Jimei University

<sup>3</sup>Xiamen Jianpan Kunlu Internet of Things Technology Research Institute Ltd

<sup>4</sup>hoture@126.com

\*corresponding author

**Abstract.** Since ChatGPT's release, large AI models have gained significant attention, particularly in the integration with Smart Home 3.0, marking a new research direction. This paper reviews cutting-edge research on the intersection of smart homes and large AI models, highlighting challenges and trends. We focus on data, models, and execution to explore the advancement of smart home platforms globally. We discuss data collection and feature research progress within smart homes, using intelligent voice assistants like Amazon Alexa and Google Bard as examples. We examine the applications, potentials, and challenges of AI models in smart homes and offer insights into future applications of large AI models in this field.

**Keywords:** Smart Home, Artificial Intelligence, Large Language Models.

## 1. Introduction

With the rise of artificial intelligence, IoT, big data, and 5G, coupled with the COVID-19 pandemic's push towards home-centric living, smart homes are becoming a pivotal aspect of intelligent living. Utilizing IoT, cloud computing, and mobile internet, they provide control over devices, health monitoring, and security, enhancing comfort and safety. The global smart home market is expected to reach \$182.3 billion by 2025, spanning entertainment to health solutions. As AI advances, smart homes evolve to offer innovative living experiences.

Smart homes' evolution can be traced through three stages: early 20th-century automation, mid-20th-century electronic advancements, and the 21st-century digital revolution. Despite earlier limitations, recent progress in IoT and cloud computing has enabled integrated connectivity and improved interoperability, transforming smart homes from standalone devices to comprehensive ecosystems. Advancements in AI, especially voice recognition and deep learning, have enabled smarter, more personalized smart home experiences. Interaction methods have evolved from button pressing to natural language and gesture-based interactions. Since ChatGPT release, global companies have launched large-scale AI models, enhancing automation and user experience in smart homes through NLP and multimodal capabilities. These models act as interpreters between humans and machines, driving innovations in smart home control, safety, and energy efficiency.

This paper consolidates cutting-edge research from domestic and international sources, identifying three pivotal elements for contemporary smart homes: Data, Models, and Execution. Their integrated synergy forms the basis for smarter, more convenient, and personalized home experiences. Data serves as the foundation, encompassing information from various sensors, devices, and user interactions. This data includes environmental metrics like temperature and humidity, alongside user behavior patterns and preferences. Collecting relevant data lays the groundwork for user profiling and machine learning model development; Models analyze the collected data using big data and AI technologies. These models enable intelligent systems to recognize user behaviors, adapt to living habits, and predict potential events, offering advanced decision-making capabilities. Continuous model optimization enhances adaptability to user needs and changing environments; Execution involves smart appliances and intelligent hardware, translating model decisions into actionable tasks such as adjusting temperatures and controlling lighting. This execution optimizes user experience, tailoring the home environment to users' needs and habits. The interplay between Data, Models, and Execution is essential; none can be overlooked. Following the logic of "Data-Models-Execution," we will review and analyze research on data collection in smart home environments and large AI models tailored for smart homes. We will also intersperse discussions on multimodal human-machine interaction and user experience research, combining them with models. Finally, we will offer insights into the future of the smart home industry.

## **2. Current Status of Advanced Smart Home Platforms**

Smart homes are pivotal applications within the Internet of Things (IoT) ecosystem, with platforms acting as essential frameworks. These platforms integrate system integration, artificial intelligence, cloud computing, edge computing, and federated learning technologies to offer comprehensive solutions at the application layer. They create automated, personalized ecosystems catering to diverse smart home needs, including scene linkage, personalized settings, privacy, health monitoring, decision-making, social interactions, and connectivity.

The advent of AI large models is driving the smart home industry towards transformative changes, transitioning from single-function devices to integrated ecosystems. Companies are adopting multimodal large model approaches to enhance security, reduce costs, and improve user interactions. These large models find applications across sectors like finance, healthcare, and e-commerce, providing adaptable AI solutions for various needs. Leading home appliance manufacturers are investing more in manufacturing and R&D to enhance supply chain stability, cost-effectiveness, and innovation, pushing the industry towards increased intelligence and efficiency.

## **3. Data research in the smart home**

Large AI models and data have a mutually beneficial relationship. Deep learning depends on data quality and diversity, directly impacting a model's effectiveness. A diverse dataset helps models understand problems, recognize features, and improve generalization. During training, models adjust parameters to learn the input-output relationships, aiming for robust generalization to predict unseen data accurately. This requires enough parameters to detect subtle features and diverse datasets to avoid overfitting.

Similarly, smart home development relies on data integration. Combining artificial intelligence with big data enables customized user experiences. In smart homes, data types vary from human posture to voice information. Utilizing this data helps smart homes understand user preferences, offering more personalized and intelligent services.

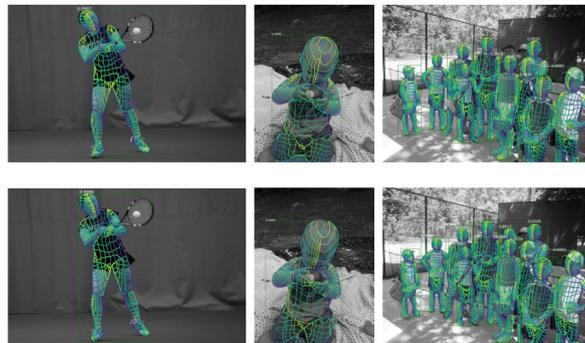
### *3.1. Human posture data*

Human pose encompasses various computer vision tasks, including human pose estimation, human action recognition, and human action generation. In the context of smart homes, human pose estimation and human action recognition are predominant, while human action generation is more common in generative AI applications. It can be categorized by application scenarios. The most advanced form is

single-view estimation, predicting key point positions based solely on a 2D RGB image. This estimation further divides into 2D skeleton estimation (Figure 1), and 3D shape estimation (Figure 2).



**Figure 1.** Estimated model output for a 2D skeleton [1]



**Figure 2.** Estimated model output map for 3D skeleton [2]

Human pose estimation technology has advanced significantly. The PoseNet model by J sun et al.[1] can predict the three-dimensional pose from 2D images with an average precision of 82.8% on the DensePose-COCO dataset. And Jie Yang et al.'s [2] end-to-end multi-person pose estimation model achieved 82.9% precision, and improve prediction speed by 60%. Human pose data holds significant potential in smart homes, aiding in activity analysis and behavior tracking. Tracking body key points provides insights into speed and acceleration, enabling behavior pattern analysis for health monitoring. In elder care, computer vision-based fall detection is common. Additionally, research is exploring emotion recognition based on pose to better understand users' emotional needs.

### 3.2. Voice data

Speech data is essential for speech recognition and dialogue models, primarily collected through microphones in devices like smart speakers, TVs, and tablets. Processing involves speech recognition, which understands spoken content, and speech emotion analysis, which identifies emotions from tone and intonation. In the speech field, speech recognition models are the most used in smart homes, interpreting voice commands to control devices. While large-scale speech generation and dialogue models are emerging, their application in smart homes is limited but growing in research interest.

## 4. Research on AI Modelling in Smart Homes

### 4.1. AI Models for Voice Assistant

Voice-based smart home models have emerged as a significant innovation in the smart home sector due to the rapid advancement of voice technology. These models use voice as the main interaction method, offering users a convenient and natural control experience. Many smart home devices now integrate advanced voice-based AI models, with Amazon Alexa being a prominent example. Amazon Alexa [3,4],

introduced in 2014 as Amazon's cloud-based voice service, initially featured in the Amazon Echo smart speaker. Today, it's accessible on over 100 million Amazon and third-party devices. Alexa enables natural voice interactions, providing an intuitive way to engage with daily technology. Since its inception, Alexa has become a leading voice assistant in the smart home industry, continuously evolving and improving.

In voice activation, Che-Wei Huang et al. [5] and Bowen Shi et al. [6] improved voice assistant accuracy by marking dialogue turns and using a Fourier transform-based signal representation. They found that confusion networks, representing word probabilities from continuous speech positions, were highly predictive; In speech recognition, Prakhar Swarup et al. [7] reduced error rates by 6.5% by incorporating embeddings of raw sound and estimated word sequences into the confidence scoring network. Harish et al. [8] achieved comparable results with shorter training times. Real-time adjustments based on the target speaker's context further enhanced recognition accuracy.

Furthermore, Lopatovska et al. [3] conducted experiments on Alexa's task completion and learning capabilities. Over a four-day experimental period, participants of various age groups tested Alexa using over a dozen different types of interactions: weather, facts, news, controlling other home devices, playing music, setting reminders/calendars/timers, telling jokes, playing games, among others, as shown in Figure 3. As indicated in Figure 3, overall, the task completion rate of the smart voice assistant Alexa reached 80%, with user satisfaction exceeding 80%. However, Alexa's daily tasks are generally basic, mostly limited to playing music (29 times), controlling other devices (18 times), etc.

Interaction type/Age group	Total frequency	Satisfaction/Dissatisfaction Fr (%)	Complete/Incomplete Fr (%)
Weather	39		
Ages 4-10	5	4/1 (80%/20%)	4/1 (80%/20%)
Ages 20-39	19	19/0 (100%/0)	19/0 (100%/0)
Ages 40-60	15	14/1 (93%/7%)	14/1 (93%/7%)
Music	29		
Ages 4-10	12	11/1 (92%/8%)	10/2 (83%/17%)
Ages 20-39	6	5/1 (83%/17%)	3/3 (50%/50%)
Ages 40-60	11	10/1 (91%/9%)	9/2 (82%/18%)
Device	18		
Ages 4-10	0		
Ages 20-36	8	7/1 (87%/13%)	6/2 (75%/25%)
Ages 40-60	10	9/1 (90%/10%)	7/3 (70%/30%)
Facts	10		
Ages 4-10	6	4/2 (67%/33%)	3/3 (50%/50%)
Ages 20-39	0	0	0
Ages 40-60	4	1/3 (25%/75%)	4/0 (100%/0)
News	9		
Ages 4-10	3	3/0 (100%/0)	3/0 (100%/0)
Ages 20-39	3	3/0 (100%/0)	3/0 (100%/0)
Ages 40-60	3	3/0 (100%/0)	3/0 (100%/0)
Timer	9		
Ages 4-10	1	1/0 (100%/0)	1/0 (100%/0)
Ages 20-39	7	6/1 (86%/14%)	6/1 (86%/14%)
Ages 40-60	1	1/0 (100%/0)	1/0 (100%/0)
Joke	6		
Ages 4-10	1	1/0 (100%/0)	1/0 (100%/0)
Ages 20-39	0	0	0
Ages 40-60	5	5/0 (100%/0)	4/1 (80%/20%)
Game	2		
Ages 4-10	0	0	0
Ages 20-39	0	0	0
Ages 40-60	2	1/1 (50%/50%)	2/0 (100%/0)
Reminder	2		
Ages 4-10	0	0	0
Ages 20-39	2	2/0 (100%/0)	2/0 (100%/0)
Ages 40-60	0	0	0
Telling time	4		
Ages 4-10	2	2/0 (100%/0)	2/0 (100%/0)
Ages 20-39	0	0	0
Ages 40-60	2	2/0 (100%/0)	2/0 (100%/0)
Others	8		
Total:	136	114/14 (84%/16%)	109/19 (80%/20%)

**Figure 3.** Tasks completed by Alexa during the four-day study by participants [3]

#### 4.2. AI Large Models for Smart Home

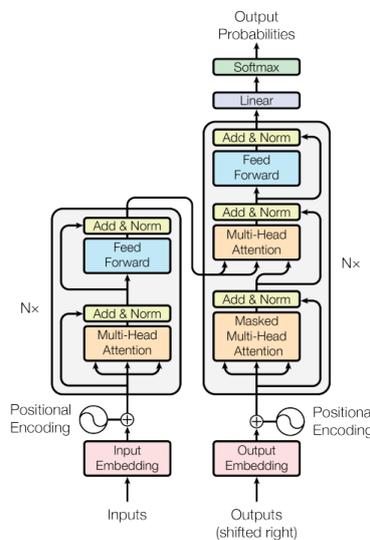
Today's AI large-scale models primarily fall into two categories: large language models and multimodal large models. Large language models, like ChatGPT [9], are widely used in dialogue and natural

language processing, gaining popularity. Multimodal large models integrate multiple perceptual modalities, surpassing single-input constraints.

In the smart home sector, large language models enhance natural language understanding, intelligent assistants, and personalized services, improving product interaction and user experience. Unlike traditional smart home systems limited to a single modality, modern AI models process voice, image, and other sensory data simultaneously. This enables systems to interpret voice commands and analyze images for precise control and feedback, driving innovation in smart homes and paving the way for future advancements.

Among the many smart home large-scale models, Google Bard was one of the earliest large models applied to smart homes. As a generative AI model, it integrates its functionalities with Google Assistant, Google's language assistant model, enabling the application of generative large models in the smart home domain. Designed as an interface to interact with a large language model, unlike ChatGPT which can only interact with users using text, Google Bard allows users to collaborate with generative artificial intelligence through text, voice, and images, helping users unleash their potential, enhance imagination, expand curiosity, and improve productivity.

Google Bard, launched in March 2023, has seen rapid iterations and updates since its inception. The main architecture of Bard is the Transformer structure, initially proposed by Vaswani et al.[10] in 2017. This architecture includes key components like self-attention mechanisms, encoder-decoder structures, multi-head attention, feedforward neural networks, residual connections, layer normalization, and positional encoding, as depicted in Figure 4. Today, Transformer has become a foundational element for various AI tools. Notably, the core of models like GPT-3 and GPT-4.1 also adopts the Transformer structure.



**Figure 4.** Schematic diagram of Transformer structure

Google Bard [11] operates in three stages: pre-training, response generation, and human feedback. During pre-training, the model learns to predict word sequences in context, aiming for accuracy and creativity. It generates engaging responses by blending unexpected yet plausible answers. When prompted, Bard drafts multiple responses, which undergo safety and quality checks. Approved responses are ranked based on quality. Human reviewers further refine Bard's responses, improving its learning dataset. Bard also uses "Reinforcement Learning from Human Feedback (RLHF)"[12] to align its decision-making with human goals, enhancing its performance.

Today, Google has integrated Bard with its smart home voice assistant, Google Assistant. The core idea of this innovative integration is to introduce Bard's powerful language comprehension, generation capabilities, and multimodal understanding into smart home scenarios, enhancing Google Assistant's

intelligence and flexibility in interactions. Compared to traditional assistants, Google Assistant, powered by the large language model Bard, offers more flexibility, intelligence, and convenience. The key lies in Bard's pre-training capability [13]. By pre-training on a broader and richer dataset of voice, text, and image data in the smart home domain, Bard can understand user natural language inputs more comprehensively. This enables Google Assistant to interpret user commands and queries more accurately and precisely, offering users more personalized and tailored services.

For instance, during interactions with Google Assistant, Bard's pre-training capability enables it to discern user intentions in intricate contexts, catering to diverse needs with smarter and more insightful responses. The deep integration of Bard [13,14,15,16] with Google Assistant has elevated voice interactions in smart homes. Users can command and query smart devices more naturally, with Google Assistant executing these tasks intelligently. This facilitates convenient scenarios like voice-controlled lighting and temperature adjustments, or accessing daily life information such as weather and news. Overall, integrating large-scale AI models enhances voice assistant intelligence and fosters natural, personalized user experiences, driving innovation in smart home technology.

#### *4.3. Existing Problems and Future Prospects*

While AI large-scale models in the smart home domain are advancing rapidly, several challenges persist, notably in bias. Model bias is a significant concern as AI models are trained on datasets that may contain gaps, biases, and stereotypes inherent in human-created data. This can lead to biased predictions, reflecting cultural, demographic, gender, religious, or racial biases. In the context of smart homes, where large language models predominantly learn from English-based data, there's a risk of Western cultural bias. Given the differences between Eastern and Western cultures, additional fine-tuning and human feedback are essential to mitigate such biases. Recent studies also highlight that large language models can exhibit emotions or viewpoints, further emphasizing the need for careful deployment and refinement [14,17]. However, such responses can sometimes be inappropriate. Therefore, establishing guidelines on how large language models might express themselves (i.e., personality) and continuing to fine-tune the models to provide objective and neutral responses are crucial steps before deployment.

### **5. The Future and Prospects of AI Big Models in Smart Homes**

Large AI models are reshaping the smart home sector, ushering in advanced intelligent living possibilities. Their integration promises to transform various aspects of smart homes. AI models can personalize devices based on user behaviors and preferences, automating adjustments for lighting, temperature, and audio systems to enhance comfort. Their advanced speech and natural language processing capabilities enrich user-device interactions, improving responsiveness to voice commands. Furthermore, AI's image recognition abilities strengthen smart home security with features like anomaly detection and facial recognition. Crucially, large AI models are becoming pivotal in the smart home ecosystem, enabling interoperability among devices for a more integrated experience.

In summary, large AI models hold tremendous potential for intelligent, convenient, and safe living environments. Leading companies are driving innovations in this space. For instance, Midea's deep learning-based model enhances user experiences with precise appliance control. AISpeech's DFM-2 model refines voice interactions for natural communication, while Haier's HomeGPT integrates natural language processing and image recognition to advance the smart home ecosystem. These developments indicate a future where large AI models play a central role in delivering a comprehensive and improved smart home experience.

### **6. Conclusion**

The article provides a comprehensive review of the development and future prospects of large AI models in the field of smart homes, focusing on the current status of research platforms both domestically and internationally, data research within smart homes, and AI model research in smart homes. The research discussed covers three key factors: data, models, and execution, addressing the main pain points, challenges, and mathematical and model issues in the study of large AI models in smart homes. By

summarizing cutting-edge research both domestically and internationally, the article outlines potential issues and future directions for research related to large AI models in smart homes: (1) Future smart home platforms should further integrate resources to form an IoT central platform with high integration, compatibility, and stability. (2) Data security and privacy in smart home environments should be better protected and risk-controlled. (3) Issues such as data bias and accuracy persist in the deployment of large AI models in smart homes. Techniques like reinforcement learning using human feedback should be employed for fine-tuning in application scenarios. (4) The integration of AI language models with smart voice assistants is a future development trend in smart homes. (5) Research should shift from a single-modal to a multi-modal human-computer interaction approach.

## References

- [1] Arsikere, Harish, Ashtosh Sapru, and Sri Garimella. "Multi-dialect acoustic modeling using phone mapping and online i-vectors." (2019).
- [2] Rink K, Knight I, et al. Talk to me: Exploring user interactions with the Amazon Alexa[J]. *Journal of Librarianship and Information Science*, 2019, 51(4): 984-997.
- [3] Lopatovska I, Rink K, Knight I, et al. Talk to me: Exploring user interactions with the Amazon Alexa[J]. *Journal of Librarianship and Information Science*, 2019, 51(4): 984-997.
- [4] Sinha, Atul, et al. "S407 Do Voice Assistants and Chatbots Provide Equally Appropriate Advice Regarding Colorectal Cancer Screening?." *Official journal of the American College of Gastroenterology| ACG 118.10S (2023): S297.*
- [5] Wu, Tianyu, et al. "A brief overview of ChatGPT: The history, status quo and potential future development." *IEEE/CAA Journal of Automatica Sinica* 10.5 (2023): 1122-1136.
- [6] Swarup, Prakhar, et al. "Improving ASR confidence scores for Alexa using acoustic and hypothesis embeddings." (2019).
- [7] Al-Sharif, Eman M., et al. "Evaluating the Accuracy of ChatGPT and Google BARD in Fielding Oculoplastic Patient Queries: A Comparative Study on Artificial versus Human Intelligence." *Ophthalmic Plastic & Reconstructive Surgery* (2024): 10-1097.
- [8] Lopatovska, Irene, et al. "Talk to me: Exploring user interactions with the Amazon Alexa." *Journal of Librarianship and Information Science* 51.4 (2019): 984-997.
- [9] Vakilzadeh, Ali, and Sara Pourahmad Ghalejoogh. "Evaluating the Potential of Large Language Model AI as Project Management Assistants: A Comparative Simulation to Evaluate GPT-3.5, GPT-4, and Google-Bard Ability to pass the PMI's PMP test."
- [10] Kepuska, Veton, and Gamal Bohouta. "Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home)." 2018 IEEE 8th annual computing and communication workshop and conference (CCWC). IEEE, 2018.
- [11] Chowdhery, Aakanksha, et al. "Palm: Scaling language modeling with pathways." *Journal of Machine Learning Research* 24.240 (2023): 1-113.
- [12] Zoujian Li, Zifei Li, Yuheng Ren. The Impact of Industrial Internet and the Digital Economy on the Management and Development of Manufacturing Information Systems Triggering Digitization as IoT and Artificial Intelligence[J]. *Journal of Information Systems Engineering and Management*, 2023
- [13] Zheng, Rui, et al. "Secrets of rlhf in large language models part i: Ppo." *arXiv preprint arXiv:2307.04964* (2023).
- [14] Chowdhery, Aakanksha, et al. "Palm: Scaling language modeling with pathways." *Journal of Machine Learning Research* 24.240 (2023): 1-113.
- [15] Thoppilan, Romal, et al. "Lamda: Language models for dialog applications." *arXiv preprint arXiv:2201.08239* (2022).
- [16] Suman, Shashi, Ali Etemad, and Francois Rivest. "Potential impacts of smart homes on human behavior: A reinforcement learning approach." *IEEE Transactions on Artificial Intelligence* 3.4 (2021): 567-580.
- [17] Liu, Xiao, et al. "GPT understands, too." *AI Open* (2023).