A Survey of Foundation Model-Based Robots in Patient and Elderly Care

Zhiwen Qiu
Information Science Department, Cornell University
Ithaca, NY, United States
zq76@cornell.edu

Wei Liu
Radiation Oncology Department, Mayo Clinic
Phoenix, AZ, United States
Liu.Wei@mayo.edu

Huangyuan Su
Computer Science Department, Harvard University
Boston, MA, United States
csu@g.harvard.edu

Yuexing Hao
EECS Department, MIT
Cambridge, MA, United States
yuexing@mit.edu

Abstract

As populations age worldwide, demands for older-adult and patient care is growing faster than the capacity of caregiver and clinical workforce. Foundation models are increasingly integrated into robots and interactive agents for patient and elderly care with the promise of more flexible communication and task guidance than conventional robotic systems. However, care settings requires reliability, safety, and workflow compatibility, and it remains unclear how current embodied systems can translate technical advances into real-world impacts. We present a targeted mini-survey of 22 studies from 2023-25 on foundation model–enabled care robots. We synthesize (i) integration patterns and design features, and (ii) empirical evidence on user experience and health-related outcomes. Across studies, most systems integrate foundation models as an interaction and reasoning layer through voice-first pipelines and socially assistive embodiments, with limited multimodal grounding for closed-loop autonomy. Evaluations are primarily feasibility-focused which emphasizes acceptability, usability, engagement, and qualitative feedback. We also highlight gaps that limit cross-study comparisons and real-world deployment challenges.

CCS Concepts

• **Human-centered computing** → *Human computer interaction (HCI)*; • **Computer systems organization → Robotics**; • **General and reference** → *Surveys and overviews.*

Keywords

Foundation Model, Embodied Agent, Robotics, Patient and Elderly Care

ACM Reference Format:
Zhiwen Qiu, Wei Liu, Huangyuan Su, and Yuexing Hao. 2026. A Survey of Foundation Model-Based Robots in Patient and Elderly Care. In *Proceedings of The ACM/IEEE International Conference on Human-Robot Interaction Workshops (Robots for Care @HRI '26)*. ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

The rise of foundation models, especially large language models (LLMs) and vision–language models (VLMs), has sparked interests in their applications to robotics for older-adult and patient care [2, 11, 37]. Compared to traditional rule-based dialogue and hard-coded interaction scripts, foundation models offer a practical way to support more flexible communication, personalization, and task guidance [32, 34] in messy real-world care contexts [10, 33]. In principle, these foundation models trained on large-scale multimodal data such as text, images, and sensors to ensure robots that can understand natural language instructions, adapt to unfamiliar situations, and interact more intuitively with people who need assistance [1, 13, 36].

The global aging population has increasing demand for care while caregivers and clinical staff face growing workload constraints [3, 8, 9]. A robot that can fetch medications or provide companionship without needing explicit programming for everyday scenario could make a real difference. However, deploying such systems in care settings raises practical questions about reliability, safety, and whether the complexity they introduce actually solves the problems caregivers and patients face [29]. Although foundation models bring new capabilities that can potentially generalize across interactive tasks, learn from demonstrations, and handle human environments without explicit feature engineering, they also introduce new challenges such as hallucinations and disruptions in conversations [14, 16]. Despite the rapid progress, much of the research focuses on technical capabilities and less is driven by evidences that foundation model-enabled robots can deliver reliable, workflow-compatible value in real-world care settings.

To address these gaps, a focused review is needed to synthesize how foundation models are actually integrated into embodied care systems, and to understand the impacts on user experiential factors, safety, and health outcomes [24]. In this mini-survey, we analyze 22 recent studies on foundation model–based robots and agents for older-adult and patient care. We organize findings around (i) how foundation models are integrated into embodied and semi-embodied care systems, (ii) the experiential evidence reported in user evaluations, and (iii) the types and strength of health outcomes measured.

Robots for Care @HRI '26, Edinburgh, UK

ACM ISBN 978-1-4503-XXXX-X/2026/03
https://doi.org/XXXXXXX.XXXXXXX

2 Methods

We conducted a survey to understand how foundation model-based robots for older-adult and patient care, focusing on recent work from 2023–2025. We searched Google Scholar, ACM Digital Library, IEEE Xplore, and arXiv using combinations of keywords such as (“foundation model” OR "LLM" OR “large language model” OR “vision-language model”) AND ("robot" OR “social robot” OR “assistive robot” OR "agent") AND ("older adult" OR "elderly" OR "patient" OR "caregiver" OR "dementia" OR "healthcare").

We included papers that (i) explicitly used a foundation model (e.g., LLM/VLM) as a central component of the system's interaction, reasoning, or tool-use; (ii) targeted older adults, patients, caregivers, or clinical staff in a healthcare or assistive-care context; and (iii) provided system details and empirical evaluations. We excluded papers that were (i) purely technical robotics demonstrations without a care setting, population, or care task framing, (ii) works without a foundation model component, and (iii) non-empirical studies.

For each included study, we extracted key attributes such as population, condition, model usage, types of robot, task and purpose, study setting, and summarized reported outcomes, with particular attention to experiential factors and any reported care/health impacts. 22 studies were included based on the above criteria and search process.

3 Design Features for Foundation Model-Based Robots

Across 22 studies in the review (Table 1), these robots largely represent systems where LLM is the primary interaction and reasoning layer such as free-form dialogue [27], structured coaching and assessment [7, 30], summarization and triage [35], or tool-use to trigger robot functions [18], often paired with speech pipelines and lightweight safety moderation. Many systems implement LLMs as open-ended conversational partners or well-being support agents for older adults (e.g., daily check-ins and reflective conversation) [17, 19]. Others emphasize LLM-mediated information exchange and care coordination in low-embodiment settings to reduce friction for older adults and providers [35]. The target population skew toward older adults (e.g., 65+), with several studies focusing on mild–moderate cognitive impairment [6, 22] and a smaller set targeting physical assistance needs (e.g., feeding support) [20, 27]. Some works also target caregiver and staff-facing interaction loops which suggests that patient-facing companionship is only one slice of the emerging design space [35].

Among all studies, the design space is largely defined by embodiment and control rather than model choices. Most robots attach an LLM to an existing platform as a conversational brain, with socially assistive robots (humanoid and tabletop) used to improve presence and engagement [16, 17], and a smaller subset of mobile companions extending interaction across home-like spaces [23]. Physically assistive manipulation robots remain less common, and are typically constrained to a narrow action library with explicit safeguards [5]. For instance, Miyake et al. [25] designed a range-of-motion training exercises using humanoid robots to emphasize communication that demonstrates motion using LLMs, and found that the system can increase the preference and experience of willingness to participate in the exercises. In terms of architecture, systems most often implement a voice-first loop for real-time speech recognition and translation, sometimes paired with a screen for confirmations [6, 22], while multimodal grounding such as vision, environment context, and user state signals is used less often and mainly for prompt conditioning rather than closed-loop autonomy [19].

Overall, these studies reveal that foundation models are currently most impactful as an interaction and text generation layer that expands what care and assistive robots can say and how flexibly they can respond. There currently lacks the study on reliability in user interaction, including turn-taking breakdowns, common sense knowledge, and hallucinations, which can directly decrease trust and engagement in older adult use cases. There is also limited integration of multimodal and contextual grounding into decision-making loops, which constrains many systems to conversation-first autonomy and makes action execution difficult to scale beyond predefined skill libraries. The reviewed literature also suggests an underexplored area in co-use settings for patient–caregiver–clinician loop, where LLM-enabled robots may provide the most immediate value as mediators of communication and care routines.

4 Health Outcomes

Empirical evidence is mostly based on the feasibility results. Most studies reported experiential outcomes such as acceptability, usability, engagement, and qualitative feedback, and fewer papers reporting validated health and behavioral outcomes [7, 28]. Studies are conducted across diverse contexts, including home-like contexts and senior living [26], clinical and hospital-oriented deployments [6], as well as lab studies and simulations [25, 30]. Across settings, user experience outcomes are strongly shaped by interaction reliability (e.g., speech recognition quality, response latency) and expectation management (what the robot can do and when it should pass to caregivers or clinicians).

A recurring theme is that LLM integration can improve perceived naturalness and reduce reliance on scripted dialogues, but it also introduces new risks like hallucinations and off-topic responses which causes diminished trust and increase the need for guardrails in aging population. For example, Blavette et al. [6] reported clear quantitative experiential gains after iterative refinement of a waiting area socially assistive humanoid robot, where acceptability scores rose from 15.4 to 22.5 (out of 30) and usability scale [21] improved from 47.9 to 69.3 (out of 100). The authors explicitly attributed the improved interaction quality to technical upgrades, including LLM integration that made responses more coherent and context-aware. On the other hand, Irfan et al. [16] noted the challenges of the integration of GPT-3.5 [1] to personalized companion robots for older adults with multi-turn conversations such as disruptions in conversations, disengaging responses, and outdated information that can lead to frustration, confusion, and concerns. This demonstrates that conversational flexibility alone is insufficient without reliability and safe grounding strategies [12].

In contrast, health and care outcome measures are reported less frequently and are typically proximal rather than long-term clinical results. Reported measures are often short-horizon improvements in functional behavior and care processes, including task performance [4], participation and adherence to activities [15], communication quality in care interactions [31]. A smaller subset explores

Table 1: Taxonomy of design features in foundation model-based robots.

Categories	Examples
Foundation Model's Role	Dialogue and companionship ([16, 19]); Coaching and structured guidance ([7, 25, 30]); Assessment and screening ([4]); Workflow mediation (triage, summarization, care coordination) ([31, 35]); Tool-use and skill triggering ([18, 20])
Embodiment Type	Socially assistive robots (humanoid/tabletop) ([6, 16]); Mobile companion/monitoring ([23]); Physically assistive/manipulation ([5, 27]); Voice-only or low-embodiment agents ([31, 35])
Interaction Modality	Voice-first Interface ([6, 22]); Voice and screen confirmations ([6, 22]); Text-first and chat ([35]); Multimodal cues for prompting (vision and user state) ([19])
Autonomy Level	Conversational autonomy (information and support only) ([16, 19]); Workflow autonomy (summarize, triage, coordinate) ([35]); Constrained action autonomy (bounded skill library) ([5, 18]); Physical assistance (task-specific actuation) ([27])
Grounding and Safety	Prompt constraints and moderation ([6, 22]); Retrieval and curated knowledge ([35]); Confirmation patterns ([5, 31]); Multimodal grounding ([19])
Deployment Setting	Homes and senior living centers ([23, 26]); Clinics and hospitals ([6, 31]); Labs and simulation ([25, 30])

cognition-related outcomes via repeated conversational tasks or screening-style estimates (e.g., improved performance trends over sessions, or better dementia-scale estimation accuracy) [4], but these are generally framed as monitoring and engagement support rather than validated diagnostic or intervention effects. For example, Blanco et al. [4] evaluated an LLM-enabled conversational approach for older adults that goes beyond user preference and reports outcome evidence. The results showed that repeated interaction can support measurable task-level improvement and LLM-based estimation can outperform a simpler baseline for dementia-related assessment. Notably, the authors still positioned these results as early evidence and emphasized the need for validation before clinical use.

These findings point to several existing research gaps. First, outcome reporting across studies is highly heterogeneous, with many papers using customized instruments or short study durations. The field would benefit from a more consistent set of core outcome measures aligned with common care goals such as engagement, adherence, caregiver burden, communication quality, and domain-specific functional outcomes. Second, many studies lack strong comparative baselines (e.g., LLM-based robots vs conversational agents), which makes it hard to attribute improvements specifically to foundation models or general system implementations. Lastly, more longitudinal and in-situ evaluations are needed to test whether short-term improvements in experience and participation translate into long-term behavioral or health benefits.

5 Conclusion

This mini-survey reviewed 22 studies on foundation model-based robots for patient care. Current systems mostly use LLMs as an interaction and conversational layer in voice-first pipelines and socially assistive embodiments, with limited multimodal grounding and autonomy. Health outcome evidence is largely focusing on feasibility related metrics and few report care outcomes. Progress toward reliable real-world deployment is limited by heterogeneous outcome reporting, lack of comparative baselines, and the need of longitudinal in-situ validation. Addressing these gaps will be essential for translating foundation model advances into dependable care robotics.

References

[1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).

[2] Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Chuyuan Fu, Keerthana Gopalakrishnan, Karol Hausman, et al. 2022. Do as i can, not as i say: Grounding language in robotic affordances. *arXiv preprint arXiv:2204.01691* (2022).

[3] Jotheeswaran Amuthavalli Thiyagarajan, Christopher Mikton, Rowan H Harwood, Muthoni Gichu, Victor Gaigbe-Togbe, Tapiwa Jhamba, Daniela Pokorna, Valentina Stoevska, Rio Hada, Grace Sanico Steffan, et al. 2022. The UN Decade of healthy ageing: strengthening measurement for monitoring health and wellbeing of older people. *Age and ageing* 51, 7 (2022), afac147.

[4] Antonio Blanco, Alicia Condón, Zoraida Clavijo, Trinidad Rodríguez, and Pedro Núñez. 2025. EBO Robot in Elderly Care: Interaction Styles and Multimodal Engagement Through Serious Games in Care Centers. In *International Conference on Social Robotics*. Springer, 79–91.

[5] Lauriane Blavette, Sébastien Dacunha, Xavier Alameda-Pineda, Jeanne Cattoni, Anne-Sophie Rigaud, and Maribel Pino. 2025. Integrating a Large Language Model Into a Socially Assistive Robot in a Hospital Geriatric Unit: Two-Wave Comparative Study on Performance, Engagement, and User Perceptions. *JMIR Human Factors* 12, 1 (2025), e81936.

[6] Lauriane Blavette, Sébastien Dacunha, Xavier Alameda-Pineda, Daniel Hernández García, Sharon Gannot, Florian Gras, Nancie Gunson, Séverin Lemaignan, Michal Polic, Pinchas Tandeitnik, et al. 2025. Acceptability and usability of a socially assistive robot integrated with a large language model for enhanced human-robot interaction in a geriatric care institution: mixed methods evaluation. *JMIR Human Factors* 12, 1 (2025), e76496.

[7] Ryan Browne, Mirza Mohtashim Alam, Qasid Saleem, Abrar Hyder, Tatsuya Kudo, Francesca D'Agresti, Martino Maggio, Keiko Homma, Eerik-Juhanna Siitonen, Naoko Kounosu, et al. 2024. Reflective Dialogues with a Humanoid Robot Integrated with an LLM and a Curated NLU System for Positive Behavioral Change in Older Adults. *Electronics* 13, 22 (2024), 4364.

[8] Leonid Grinin, Anton Grinin, and Andrey Korotayev. 2023. Global aging and our futures. *World Futures* 79, 5 (2023), 536–556.

[9] Alan S Gutterman. 2023. Caregiving and Families. *Available at SSRN 4610245* (2023).

[10] Yuexing Hao, Jason Holmes, Mark R Waddle, Brian J Davis, Nathan Y Yu, Kristin S Vickers, Heather Preston, Drew Margolin, Corinna E Löckenhoff, Aditya Vashistha, et al. 2025. Personalizing prostate cancer education for patients using an EHR-Integrated LLM agent. *NPJ Digital Medicine* 8, 1 (2025), 770.

[11] Yuexing Hao, Zeyu Liu, Robert N Riter, and Saleh Kalantari. 2024. Advancing patient-centered shared decision-making with AI systems for older adult cancer patients. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–20.

[12] Yuexing Hao, Zhiwen Qiu, Jason Holmes, Corinna E Löckenhoff, Wei Liu, Marzyeh Ghassemi, and Saleh Kalantari. 2025. Large language model integrations in cancer decision-making: a systematic review and meta-analysis. *NPJ Digital Medicine* 8, 1 (2025), 450.

[13] Jason Holmes, Yuexing Hao, Mariana Borras-Osorio, Federico Mastroleo, Santiago Romero Brufau, Valentina Carducci, Katie M Van Abel, David M Routman, Andrew YK Foong, Liv M Muller, et al. 2025. RadOnc-GPT: An Autonomous LLM Agent for Real-Time Patient Outcomes Labeling at Scale. *arXiv preprint arXiv:2509.25540* (2025).

[14] Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, et al. 2025. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems* 43, 2 (2025),

1–55.

[15] Sonabayim Huseynzade, Rainer Wieching, Toshimi Ogawa, Yoshio Matsumoto, Volker Wulf, and Yasuyuki Taki. 2025. When Robots Care: Elderly Reactions to Emotionally Intelligent Android. In *International Conference on Social Robotics*. Springer, 16–29.

[16] Bahar Irfan, Sanna Kuoppamäki, Aida Hosseini, and Gabriel Skantze. 2025. Between reality and delusion: challenges of applying large language models to companion robots for open-domain dialogues with older adults. *Autonomous Robots* 49, 1 (2025), 9.

[17] Bahar Irfan, Sanna Kuoppamäki, and Gabriel Skantze. 2024. Recommendations for designing conversational companion robots with older adults through foundation models. *Frontiers in Robotics and AI* 11 (2024), 1363713.

[18] Hangyeol Kang, Maher Ben Moussa, and Nadia Magnenat-Thalmann. 2024. Nadine: an LLM-driven intelligent social robot with affective capabilities and human-like memory. *arXiv preprint arXiv:2405.20189* (2024).

[19] Weslie Khoo, Long-Jing Hsu, Kyrie Jig Amon, Pranav Vijay Chakilam, Wei-Chu Chen, Zachary Kaufman, Agness Lungu, Hiroki Sato, Erin Seliger, Manasi Swaminathan, et al. 2023. Spill the tea: When robot conversation agents support well-being for older adults. In *Companion of the 2023 ACM/IEEE international conference on human-robot interaction*. 178–182.

[20] Kyungki Kim, John Windle, Melissa Christian, Tom Windle, Erica Ryherd, Pei-Chi Huang, Anthony Robinson, and Reid Chapman. 2024. Framework for integrating large language models with a robotic health attendant for adaptive task execution in patient care. *Applied Sciences* 14, 21 (2024), 9922.

[21] James R Lewis. 2018. The system usability scale: past, present, and future. *International Journal of Human–Computer Interaction* 34, 7 (2018), 577–590.

[22] Maria R Lima, Amy O'Connell, Feiyang Zhou, Alethea Nagahara, Avni Hulyalkar, Anura Deshpande, Jesse Thomason, Ravi Vaidyanathan, and Maja Matarić. 2025. Promoting Cognitive Health in Elder Care with Large Language Model-Powered Socially Assistive Robots. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–22.

[23] A Logeshwar, RM Manikandan, R Parvesh, A Ram Solaiappan, and L Anju. 2025. Smart Home Robotic Companion with AI-Driven Personalized Care for Elderly Assistance. In *The 2025 International Conference on Advanced Research in Electronics and Communication Systems (ICARECS-2025)*. Atlantis Press, 322–332.

[24] Nikita Mehandru, Brenda Y Miao, Eduardo Rodriguez Almaraz, Madhumita Sushil, Atul J Butte, and Ahmed Alaa. 2024. Evaluating large language models as agents in the clinic. *NPJ digital medicine* 7, 1 (2024), 84.

[25] Tamon Miyake, Yushi Wang, Pin-chu Yang, and Shigeki Sugano. 2023. Feasibility study on parameter adjustment for a humanoid using LLM tailoring physical care. In *International Conference on Social Robotics*. Springer, 230–243.

[26] Masayuki Numao and Masahiro Kawamura. 2025. An Interactive Monitoring Robot for Dementia Mitigation via Daily Conversations with Multiple LLMs. In *Proceedings of the AAAI Symposium Series*, Vol. 5. 250–255.

[27] Akhil Padmanabha, Jessie Yuan, Janavi Gupta, Zulekha Karachiwalla, Carmel Majidi, Henny Admoni, and Zackory Erickson. 2024. Voicepilot: Harnessing llms as speech interfaces for physically assistive robots. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–18.

[28] Maria Pinto-Bernal, Matthijs Biondina, and Tony Belpaeme. 2025. Designing Social Robots with LLMs for Engaging Human Interaction. *Applied Sciences* 15, 11 (2025), 6377.

[29] Dimitri Roustan, François Bastardot, et al. 2025. The clinicians' guide to large language models: A general perspective with a focus on hallucinations. *Interactive journal of medical research* 14, 1 (2025), e59823.

[30] Micol Spitale, Minja Axelsson, and Hatice Gunes. 2025. VITA: A Multi-Modal LLM-Based System for Longitudinal, Autonomous and Adaptive Robotic Mental Well-Being Coaching. *ACM Transactions on Human-Robot Interaction* 14, 2 (2025), 1–28.

[31] Hassam Tahir, Samina Ansari, Naeem Bushra, Aila Khan, Michael Lwin, Fady Alnajjar, and Omar Mubin. 2025. LLM Enabled Social Robots–Transforming Aged Care Through AI. In *2025 International Conference on Activity and Behavior Computing (ABC)*. IEEE, 1–8.

[32] Adam Villaflor, Brian Yang, Huangyuan Su, Katerina Fragkiadaki, John Dolan, and Jeff Schneider. 2024. Tractable Joint Prediction and Planning over Discrete Behavior Modes for Urban Driving. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 14785–14791.

[33] Xuan Xiao, Jiahang Liu, Zhipeng Wang, Yanmin Zhou, Yong Qi, Shuo Jiang, Bin He, and Qian Cheng. 2025. Robot learning in the era of foundation models: A survey. *Neurocomputing* (2025), 129963.

[34] Brian Yang, Huangyuan Su, Nikolaos Gkanatsios, Tsung-Wei Ke, Ayush Jain, Jeff Schneider, and Katerina Fragkiadaki. 2024. Diffusion-es: Gradient-free planning with diffusion for autonomous driving and zero-shot instruction following. *arXiv preprint arXiv:2402.06559* (2024).

[35] Ziqi Yang, Xuhai Xu, Bingsheng Yao, Ethan Rogers, Shao Zhang, Stephen Intille, Nawar Shara, Guodong Gordon Gao, and Dakuo Wang. 2024. Talk2care: An llm-based voice assistant for communication between healthcare providers and older adults. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (2024), 1–35.

[36] Jingyi Zhang, Jiaxing Huang, Sheng Jin, and Shijian Lu. 2024. Vision-language models for vision tasks: A survey. *IEEE transactions on pattern analysis and machine intelligence* 46, 8 (2024), 5625–5644.

[37] Brianna Zitkovich, Tianhe Yu, Sichun Xu, Peng Xu, Ted Xiao, Fei Xia, Jialin Wu, Paul Wohlhart, Stefan Welker, Ayzaan Wahid, et al. 2023. Rt-2: Vision-language-action models transfer web knowledge to robotic control. In *Conference on Robot Learning*. PMLR, 2165–2183.