CogAAL: A Privacy-Preserving, Full-Stack Cognitive Architecture Integrating Local VLMs, Heterogeneous Robotics, and IoT for Proactive Ambient Assisted Living

CogAAL: A Privacy-Preserving, Full-Stack Cognitive Architecture Integrating Local VLMs, Heterogeneous Robotics, and IoT for Proactive Ambient Assisted Living

Karameldeen Omer
Department of Information Engineering
Università Politecnica delle Marche
Ancona, Italy
k.i.m.omer@univpm.it

Francesco Ferracuti
Department of Information Engineering
Università Politecnica delle Marche
Ancona, Italy
f.ferracuti@univpm.it

Andrea Monteriù
Department of Information Engineering
Università Politecnica delle Marche
Ancona, Italy
a.monteriu@univpm.it

***Abstract*—We introduce CogAAL, a unified and privacy-preserving cognitive architecture for proactive Ambient Assisted Living (AAL). Existing AAL frameworks are typically fragmented across perception, reasoning, and actuation layers, and frequently depend on cloud-based inference pipelines that conflict with the stringent data-sovereignty and confidentiality requirements of residential care environments. In contrast, CogAAL integrates three core modules via a shared semantic representation:**

(i) a four-layer Multimodal Cognitive Intelligence module that fuses data from wearable devices, Internet-of-Things (IoT) sensors, and RGB-D vision streams to enable robust health-state inference and longitudinal health monitoring;

(ii) a three-tier Semantic Robotic Navigation module that semantically grounds a heterogeneous fleet of cleaning, delivery, and mobility-assistance robots in a Building Information Modeling (BIM)–based digital twin, which is continuously updated using real-time occupancy patterns and health-related events; and

(iii) an Agentic IoT Orchestration module that translates natural-language intents into executable task graphs and orchestrates adaptive environmental comfort control (including lighting, HVAC, and acoustic conditions).

All inference processes are executed on-premises entirely using Gemma 4 MoE models with 2–3.8B active parameters, requiring 8–16 GB of VRAM at 4-bit precision and licensed under Apache 2.0, thereby eliminating the need for cloud-based data transmission. A federated feedback loop enables continuous, privacy-preserving personalization across deployments.

***Index Terms*—VLM in Ambient Assisted Living, Multimodal Sensor Fusion, Agentic AI, Heterogeneous Robotics, Internet of Things, Human-Robot Interaction, BIM Digital Twin, Adaptive Personalisation, Edge Intelligence, Privacy-Preserving AI, Local LLM Deployment, Semantic Mapping, Comfort Control**

I. Introduction

By 2050, over 21.6% of the global population will exceed 60 years of age [1]–[3], intensifying demand for Ambient Assisted Living (AAL) systems that sustain autonomy and safety for older adults and individuals with disabilities [4], [5]. Four structural limitations currently prevent effective deployment. Fragmented pipelines: most systems treat perception, reasoning, and actuation as isolated stages, blocking cognitive inferences from propagating to downstream actuators [6]. Privacy-incompatible inference: cloud-based VLMs transmit intimate physiological and visual data to third-party servers, violating GDPR in the majority of residential care contexts [7], [8]–[11] IoT complexity: proliferating heterogeneous devices impose unsustainable management burden without unified, intent-driven orchestration [12]. Weak contextual discrimination: existing activity and fall detection models lack the spatial context needed to distinguish routine events from critical incidents [13], [14].

A. Contributions

CogAAL addresses these limitations through five principal contributions: (1) a unified semantic architecture that integrates multimodal cognitive processing, robotic navigation, IoT orchestration within a shared latent representation space; (2) on-device deployment of Gemma 4, thereby eliminating the need for cloud-based transmission of sensitive sensor data streams; (3) coordination of a heterogeneous robot fleet—comprising cleaning, delivery, and mobility-assistance platforms—coupled with BIM–based digital twin that is continuously updated via real-time activity inference; (4) activity-contingent environmental comfort regulation that adaptively modulates lighting, heating, ventilation, and air conditioning (HVAC), as well as acoustic conditions, in accordance with occupancy dynamics and physiological indicators; and (5) a federated feedback mechanism that enables privacy-preserving, continual personalization of system behavior.

B. Related Work and Research Gaps

Robot-assisted and AAL frameworks [15] advance support for activities of daily living; however, challenges related to the coordinated orchestration and management of heterogeneous robotic fleets remain largely unresolved. Recent developments in semantic path planning [16] and multi-layer reactive control architectures [17] offer robust mechanisms for safe indoor navigation, yet these methods do not explicitly address the concurrent operation and coordination of multiple, semantically diverse robotic platforms. Concurrently, Edge-AI-enabled digital-twin-based object recognition [18] enables real-time map updates on resource-constrained edge devices.

Perception and fall detection. Multimodal datasets integrating LiDAR, radar, and thermal imaging [4], together with visual- and pose-based fall detection approaches [19], underscore the paucity of privacy-preserving, synchronised benchmark suites. Recent work on pose estimation for automated medication dispensing [20] further extends perception capabilities toward manipulation-oriented tasks.

Vision–language models and language models. Vision–language models (VLMs) for human-state recognition [21], [22] and medical imaging analysis [23] are rapidly maturing, yet they are still constrained by non-deterministic outputs and limited interpretability due to their black-box nature. NavMorph [24] improves generalisation in vision–language navigation, while systematic reviews of large language models (LLMs) in healthcare [25] reveal persistent gaps in security, robustness, and evaluation methodologies.

Agentic systems. Autonomous agentic workflows for cognitive assessment [26], multimodal clinical prediction [27], and smart-home automation [12], [28] demonstrate effective task decomposition and orchestration. Nevertheless, these systems do not incorporate robotic fleet dispatch mechanisms or privacy-preserving sensor fusion across heterogeneous modalities. Emerging Sensor–LLM fusion frameworks [6] and comprehensive surveys of AAL technologies [7] consistently identify the lack of a unified, end-to-end integration framework as a central limitation of the field. The proposed CogAAL framework is designed to address this gap.

II. Privacy-Preserving Local VLM: Gemma 4 Selection

AAL sensors collect sensitive data, including physiological signals, sleep patterns, and mobility routines. Sending this information to cloud inference services can lead to violations of GDPR, expose data to third-party retention policies, and create significant availability issues when monitoring is most crucial. Therefore, CogAAL requires that all VLM inference be conducted locally.

We select Gemma 4 [29] (Google DeepMind, April 2026, Apache 2.0) as the on-premises backbone: the 26B MoE variant for the central reasoning node and the E2&E4B edge variant for robot-local perception. Table I compares leading open-weight candidates.

Gemma 4 26B MoE (and edge variants E2B/E4B) are selected on six grounds. Efficiency: 128-expert routing activates only 3.8B parameters per pass, delivering ≈97% of dense-31B quality at ≈12% of its compute; at 4-bit quantisation (GGUF/AWQ) VRAM drops to 8–16 GB, fitting a single

TABLE I: Comparison of open-weight VLMs for local AAL deployment.

Model	Active Params	MMLU Pro	Multimodal Native	License
Gemma 4 31B Dense	31B	85.2%	✓	Apache 2.0
Gemma 4 26B MoE	3.8B	84.1%	✓	Apache 2.0
Gemma 4 E4B	4.5B	69.4%	✓	Apache 2.0
Gemma 4 E2B	2.3B	60.0%	✓	Apache 2.0
Qwen 3.5-32B	32B	86.1%	✓	Apache 2.0
Llama 4 Scout	109B	79.6%	✓	Restricted
Gemma 3 27B	27B	71.4%	Partial	Custom

RTX 3090/4090 or personal pc [29]. Native multimodality: across the model family, text, variable-resolution images (70–1,120 tokens/image), and 30 s audio via USM conformer (E-series) are processed natively without separate encoder pipelines.
Extended context: 256K-token windows (26B/31B) enable longitudinal reasoning over full-day sensor histories, medication logs, and activity records.
Function calling: built-in structured JSON output and multi-step planning supply directly executable task graphs to Module III, eliminating fragile post-processing parsers [12]. Tiered edge deployability E4B (≈4.5B effective parameters) runs on Jetson Orin Nano-class hardware, enabling robot-local perception without central round-trips. Furthermore, the ultra-lightweight E2B model (≈2B effective parameters) is deployed for extreme low-power scenarios, enabling basic semantic filtering directly on IoT sensor nodes and wearables with minimal latency and battery drain [29]. Open licensing: Apache 2.0 imposes no MAU caps or acceptable-use restrictions, permitting unrestricted institutional deployment across all model tiers.

III. CogAAL System Architecture

A. Overview

CogAAL is structured as a closed-loop, three-module architecture sharing a common semantic representation space, as illustrated in Fig. 1. The three modules are: (i) Multimodal Cognitive Intelligence, responsible for perception, representation, and reasoning; (ii) Semantic Robotic Navigation, responsible for grounding heterogeneous robotic agents in the environment; and (iii) Agentic IoT Orchestration, responsible for translating cognitive inferences into coordinated device and robot actions.

All modules are connected through a shared semantic latent space. The cognitive module generates semantic descriptors of the health state, which the navigation module interprets to update risk fields in the environment map. Additionally, the orchestration module uses these descriptors to create and dispatch task graphs. A federated feedback loop sends the execution outcomes back to the cognitive module for model adaptation.

The complete architecture is illustrated in Fig. 1 and the extended heterogeneous robot integration view is provided in Fig. 2.

B. Module I: Multimodal Cognitive Intelligence

1) Data Acquisition: Three sensor classes feed the pipeline: *wearables* (EEG, IMU, heart rate, skin conductance); *ambient IoT* (PIR, pressure mats, door contacts, temperature/CO_2 sensors); and *vision* (ceiling- and robot-mounted RGB-D cameras). All streams are timestamped and synchronised via MQTT or ROS 2.

2) Representation Learning and Alignment: Each modality is encoded by a TCN or a Transformer for physiological time-series, CNN/ViT for visual frames, and text encoder for structured logs, then aligned into a shared latent space via a cross-modal Transformer with contrastive loss. Temporal drift correction maintains coherence across asynchronous streams.

3) Cognitive Reasoning and Interpretation: The fused representation is serialised into a structured prompt and passed to locally deployed Gemma 4 26B MoE, which performs: (a) *activity classification* over a daily-living ontology; (b) *anomaly detection* against personal behavioural baselines; and (c) *health-state summarisation* with uncertainty estimates. Outputs are translated into prioritised caregiver alerts, semantic state descriptors written to the shared space (triggering Modules II and III), and longitudinal health log entries.

C. Module II: Semantic-Aware Robotic Navigation with Heterogeneous Fleets

CogAAL coordinates three concurrent robot classes: *cleaning robots*, dispatched to occupancy-clear zones; *delivery robots*, triggered by detected need states (medication schedule, meal time, object retrieval); and *mobility-aid platforms* (motorised wheelchairs, ambulatory supports), which carries a robot-local Gemma 4 E4B for on-board visual perception.

1) Offline Semantic Mapping via BIM Digital Twin: A BIM-derived digital twin supplies the foundational semantic map: spatial topology, room semantics, furniture locations, and safety constraints. Gemma 4 E4B performs *automatic object labelling* on robot-detected scenes, storing functional attributes ("medication cabinet", "fall-risk zone") in a graph knowledge base accessible to all fleet members via the ROS 2 map server.

2) Dynamic Geofencing and Real-Time Control: Health events detected by Module I generate repulsive *dynamic geofences* that re-route mobile robots and dispatch the appropriate platform. Occupancy inference maintains zone-level activity tags "food-preparation" blocks cleaning dispatch; "vacant" enables it and propagates updates fleet-wide. Online trajectory planning runs concurrently across all active platforms, with a priority-based task allocator and human-in-the-loop (HiL) voice or tablet override.

D. Module III: Agentic IoT Orchestration and Comfort Control

1) Intent Parsing and Task Decomposition: Natural-language intents, explicit "warm up the living room" or inferred from activity recognition, are processed by Gemma 4's native function-calling interface, which emits directly executable structured JSON task graphs (DAGs), eliminating post-processing parsers [12].

2) Hierarchical Task Memory and Comfort Actuation: Completed task graphs are cached in a DAG-based memory for reuse across recurring routines; a priority resolver arbitrates conflicts using device-agnostic resident preference profiles. Occupancy and activity inferences directly drive comfort actuation: lighting (colour temperature and luminance by activity type and occupant count), HVAC (zone setpoints from occupancy; physiological signals modulate cooling demand), and acoustics (attenuation during sleep, alert level modulation). Specialised sub-agents monitor execution across perception, planning, and actuation layers; failures trigger retry or fallback, with all decisions logged in an auditable trace.

E. Federated Feedback and System Integration

Execution outcomes (robot task completion, alert confirmation, comfort override) are collected as labeled feedback and used to fine-tune classification and anomaly detection models via on-device learning. FL aggregates gradient updates, not raw data, across institutionally managed edge nodes, with differential privacy bounding information leakage. The full pipeline is closed-loop: Module I fuses sensors and writes semantic descriptors; Module II reads them to update risk fields and dispatches robots; Module III generates and executes device commands; FL updates cycle back to Module I.

IV. Discussion

CogAAL achieves GDPR compliance in AAL and assistive environments [8]–[11], which requires that personal and health-related data collected through sensors, wearable devices, assistive robots, smart homes, and monitoring systems are processed lawfully, transparently, and securely. Our proposed design achieves its goals by executing all VLM inference locally with 4-bit quantization, while the FL loop enhances models without the need to centralize raw data. The modular structure allows for incremental deployment: Module I can function independently for monitoring, Module II can be added for robotics, and Module III can be included for orchestration. Additionally, implementing E4B on Jetson-class hardware helps minimize inter-robot bandwidth.

However, three limitations need to be addressed in future work: the semantic map labeling tends to degrade in low-light conditions or with heavy occlusion, which could be improved through depth and thermal fusion; the FL process assumes sufficient data heterogeneity across nodes, which might necessitate simulation augmentation in deployments with small populations; and the current design does not resolve conflicts arising from concurrent preferences among multiple residents. Our next iteration will prioritize addressing simultaneous health events.

V. Conclusion

We present CogAAL, a privacy-preserving, full-stack cognitive architecture designed for proactive AAL. CogAAL

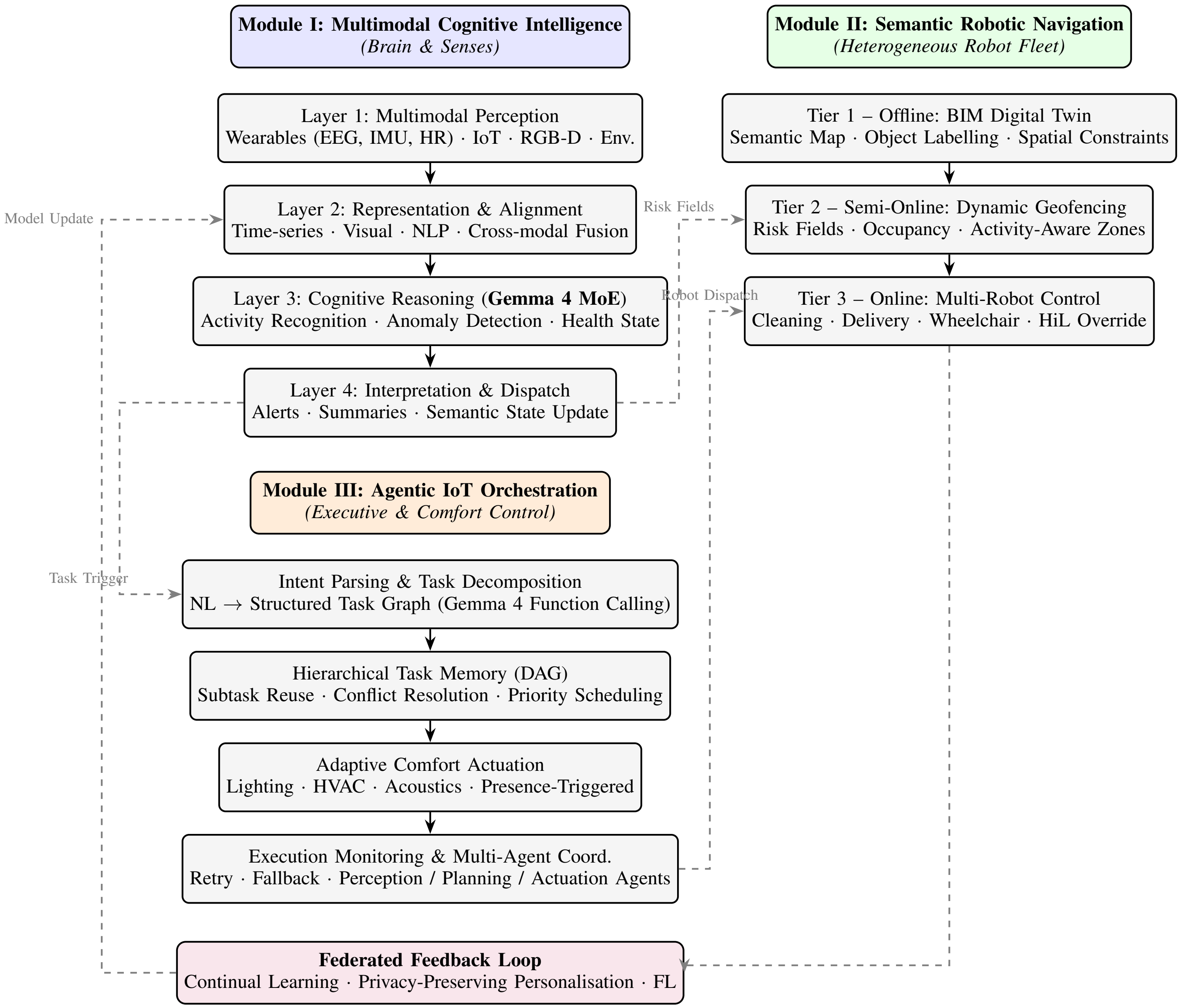

Fig. 1: CogAAL full-stack architecture integrating multimodal cognitive intelligence (Module I), semantic heterogeneous robotic navigation (Module II), and agentic IoT orchestration with comfort control (Module III), connected via a shared semantic space and a federated feedback loop.

employs Gemma 4 as its central multimodal reasoning engine to coordinate a heterogeneous ensemble of assistive robots via a BIM-based semantic map. Through agentic orchestration of IoT devices, CogAAL establishes an integrated linkage between health monitoring and environmental comfort control and can be extended to interoperate with the majority of existing indoor comfort and air quality management systems.

The proposed architecture demonstrates how advanced artificial intelligence capabilities can be deployed in residential care environments while satisfying stringent privacy constraints. CogAAL is modular, supports incremental deployment, and remains continuously adaptable through a federated learning feedback loop. As such, CogAAL constitutes a promising framework for the development of the next generation of trustworthy, personalized AAL systems.

References

[1] OECD, "Elderly population," Paris, France, 2022.

[2] R. A. Pecchenino and K. R. Utendorf, "Social security, social welfare and the aging population," *Journal of Population Economics*, vol. 12, no. 4, pp. 607–623, 1999.

[3] M. Higo and H. T. Khan, "Global population aging: Unequal distribution of risks in later life between developed and developing countries," *Global Social Policy*, vol. 15, no. 2, pp. 146–166, 2015.

[4] C. Taramasco, M. Pineiro, P. Ormeño-Arriagada, D. Robles, and D. Araya, "Multimodal dataset for sensor fusion in fall detection," *PeerJ*, vol. 13, p. e19004, 2025.

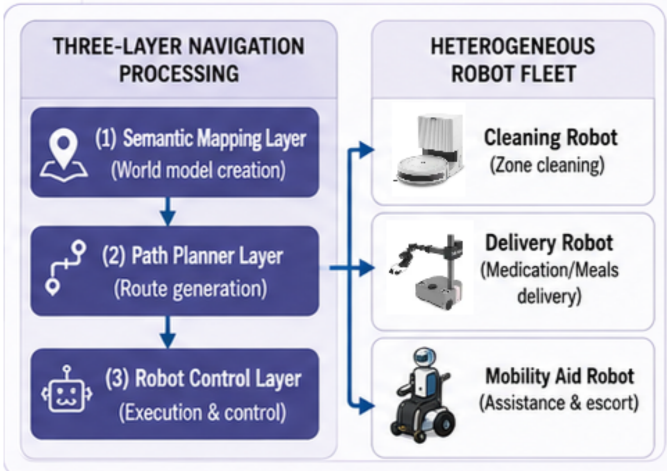

Fig. 2: Heterogeneous robot fleet integration within the CogAAL semantic navigation module.

[5] J. Roca-González, F. J. Ortiz-Zaragoza, J. A. Vera, and I. Méndez, "Convergence of science and technology for wellbeing: A case study on aal systems for older adults," *Sensors*, 2024.

[6] Z. Momand, P. Mongkolnam, J. H. Chan, and N. Charoenkitkarn, "Integrating sensor data with large language models for enhanced elderly care: A methodological framework," *Sensors and Materials*, vol. 37, no. 3, p. 1099–1138, 2025.

[7] P. Rashidi and A. Mihailidis, "A survey on ambient-assisted living tools for older adults," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 3, pp. 579–590, 2013.

[8] European Parliament and Council, "Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (general data protection regulation)," Official Journal of the European Union, L119, 1–88, 2016.

[9] European Data Protection Board, "Guidelines 03/2020 on the processing of data concerning health for the purpose of scientific research in the context of the covid-19 outbreak," 2020.

[10] European Union Agency for Cybersecurity (ENISA), "Privacy and data protection by design: From policy to engineering," ENISA, Tech. Rep., 2017.

[11] B. D. Mittelstadt, "From individual to group privacy in big data analytics," *Philosophy & Technology*, vol. 30, no. 4, pp. 475–494, 2017.

[12] C. Yu, C. Choi, S. Lee, H. Kim, and et al., "Leveraging llms for efficient and personalized smart home automation," *arXiv Preprint*, 2026. [Online]. Available: arXiv:2601.04680v1

[13] M. M. Shahid, P. Machado, J. J. Bird, and I. K. Ihianle, "Position-aware indoor human activity recognition and fall detection," *IEEE Sensors Journal*, 2026.

[14] E. Tirziu, A.-M. Vasilevschi, A. Alexandru, and E. Tudora, "Enhanced fall detection using yolov7-w6-pose for real-time elderly monitoring," *Sensors*, 2024.

[15] P. Toupas, G. Tsamis, A. Kargakos, and D. Giakoumis, "A framework towards ambient assisted living enhanced by service robots," in *Proceedings of PETRA '23*, 2023.

[16] K. Omer, E. Torta, and A. Monteriù, "Semantic-enhanced path planning for safety-centric indoor robots navigation," in *Proceedings of ICARA 2024*, 2024.

[17] K. Omer and A. Monteriù, "Multi-layer robotic controller for enhancing the safety of mobile robot navigation in human-centered indoor environments," *Frontiers in Robotics and AI*, vol. 12, p. 1629931, 2025.

[18] K. Omer and A. Monteriù, "Edge ai for object recognition in digital twins: Enhancing indoor navigation and bim systems for living environments," in *2025 IEEE International Workshop on Metrology for Living Environment (MetroLivEnv)*. IEEE, 2025, pp. 490–494.

[19] Y. Qiu, J. Meng, and B. Li, "Automated falls detection using visual anomaly detection and pose-based approaches: Experimental review and evaluation," *Journal of Biomedical Research & Environmental Sciences*, vol. 5, no. 1, pp. 055–063, 2024.

[20] H. Tian, Y. Dong, and J. Wang, "6d pose estimation of medicine bottles via rgb-d fusion and sc-yolov8 instance segmentation for service robots," *Engineering Research Express*, 2026.

[21] C. Florea, C.-B. Popescu, A. Racovițeanu, A. Nițu, and L. Florea, "From context to human: A review of vlm contextualization in the recognition of human states in visual data," *Mathematics*, vol. 14, no. 1, p. 175, 2026.

[22] A. Abid, T.-C. Ho, and F. Karray, "Vision language models for dynamic human activity recognition in healthcare settings," *arXiv Preprint*, 2025. [Online]. Available: arXiv:2510.21424v1

[23] B. C. Kalpélbé, A. G. Adaambiik, and W. Peng, "Vision language models in medicine," *arXiv Preprint*, 2025. [Online]. Available: arXiv:2503.01863v1

[24] X. Yao, J. Gao, and C. Xu, "Navmorph: A self-evolving world model for vision-and-language navigation in continuous environments," *Preprint*, 2025.

[25] S. Maity and M. J. Saikia, "Large language models in healthcare and medical applications: A review," *Bioengineering*, vol. 12, no. 6, p. 631, 2025.

[26] J. Tian, P. Fard, C. Cagan, N. Rezaii, and et al., "An autonomous agentic workflow for clinical detection of cognitive concerns using large language models," *NPJ Digital Medicine*, vol. 9, p. 51, 2026.

[27] L. P. Pulavarthy, R. Muthyala, A. V. Kuruvikkattil, Z. Yin, R. Kudamala, and S. Purkayastha, "Human-guided agentic ai for multimodal clinical prediction: Lessons from the agentds healthcare benchmark," *arXiv Preprint*, 2026. [Online]. Available: arXiv:2602.19502v1

[28] J. Aguirre and W. C. Cha, "Javis chat: A seamless open-source multi-llm/vlm deployment system to be utilized in single computers and hospital-wide systems with real-time user feedback," *Applied Sciences*, vol. 15, no. 4, p. 1796, 2025.

[29] Google Gemma Team, "Gemma 4 technical report & model cards," April 2026.