Bridging Mind and Matter: A Taxonomy of Embodied Generative AI

Research Paper

Jan Laufer[1], Leonardo Banh[1], Gero Strobel[1]

[1] University of Duisburg-Essen, Rhine Ruhr Institute of Information Systems, Essen, Germany
{jan.laufer,leonardo.banh,gero.strobel}@uni-due.de

Abstract. The rapid advancements in generative artificial intelligence (GenAI) are reshaping industries and redefining AI's role in society. Moving beyond purely generating content, GenAI-enabled agents now demonstrate autonomous, goal-driven behavior. The next frontier in GenAI lies in embodiment, enabling real-time perception, reasoning, and action in physical systems. While prior research focuses on the embodiment potential and real-world applications of generative models, systematic classification and evaluation remain scarce. By developing a taxonomy based on a systematic literature review and 40 real-world instances this paper addresses this gap. Comprising three meta-characteristics, 16 dimensions, and 50 characteristics, our taxonomy serves as a foundation for future research and offers practitioners a framework to analyze and optimize emerging applications. By shedding light on key characteristics, this study paves the way for informed decision-making in fields ranging from service robotics to industrial automation.

Keywords: Generative Artificial Intelligence, Embodied AI, Autonomous Agents, Human-GenAI Collaboration.

1 Introduction

The recent developments in the field of generative artificial intelligence (GenAI) have disrupted businesses, academia, as well as our daily lives (Chui et al., 2023; Susarla et al., 2023). Large language model (LLM) based GenAI applications are capable of generating marketing content, democratize education, or enable users to quickly summarize and translate complex information, empowering its users with simple natural language prompts (Banh & Strobel, 2023; Ooi et al., 2025). Studies project that the global GenAI market will reach $356.05 billion by 2030, significantly impacting the productivity of workers and organizations (Chui et al., 2023; Noy & Zhang, 2023; Statista Market Insights, 2024). However, GenAI has been traditionally operated within digital confines accessible through web applications and mobile devices (Strobel et al., 2024). With the advancements of more sophisticated GenAI-enabled agents, the focus shifts from purely generating content to agentic behavior where systems can reason to auton-

20th International Conference on Wirtschaftsinformatik,
September 2025, Münster, Germany

omously accomplish goals in an adaptive manner by interacting with the (physical) environment (Acharya et al., 2025; Nwankwo & Rueckert, 2024). For instance, systems like OpenAI's Operator are able to navigate software interfaces, do internet research, book services, and coordinate with applications and the internet (OpenAI, 2025).

To bridge the physical-digital gap, research and development have pursued the field of embodied AI, an idea that stems from embodied intelligence, where intelligence arises from the integration of perception, reasoning, learning, (physical) action, and real-time interaction (Duan et al., 2022; Paolo et al., 2024; Pfeifer & Iida, 2004). Here, AI has already been successfully integrated into physical bodies (i.e., robots) for stabilization, object recognition, and navigation (Bai et al., 2020; Radosavovic et al., 2024; Zhu & Zhang, 2021). Nonetheless, these systems have so far relied on human control or predefined rule-based scenarios, missing the essence of embodied intelligence.

Emerging from the interplay between agentic behavior of GenAI and embodied AI in robotics research, academia and practice both have catalyzed a paradigm shift which results in GenAI extending its capabilities to interact with the physical world (e.g., Da Silva et al., 2024; Fan et al., 2025). The integration of GenAI into physical embodiments allow for a naturalistic interaction schema with human users through the LLM's inherent natural language processing capabilities (Nwankwo & Rueckert, 2024). Furthermore, the large corpus of real-world data generative models have been trained on enable an advanced understanding of the physical environment (J. Gao et al., 2024; W. Lai et al., 2024). Companies such as Boston Dynamics have already developed embodied GenAI robots that leverage GenAI to create value to their users through a novel step of intelligence. However, the heterogenous field of artifacts leads to a gap in understanding the defining characteristics of embodied GenAI, particularly regarding the relationship between GenAI-driven intelligence, robotic embodiment, and overall system functionality. Previous studies have focused on either technically developing and evaluating different kinds of generative models for robotic use cases, such as LLMs and vision-language models (VLM) or on conceptual overviews like surveys (Brohan et al., 2023; Kim et al., 2024). However, an examination of real-world use cases of embodied GenAI as well as a conceptualization of characteristics and capabilities is yet to be conducted to structure the fast-changing landscape and guide researchers and practitioners in making informed decisions. This leads to our research question: *How can embodied GenAI be classified in terms of its core dimensions and attributes?*

To address this question, we develop a taxonomy following the approach of Nickerson et al. (2013). Our methodology combines insights from a systematic literature review and 40 empirical cases of embodied GenAI systems. We define embodied GenAI, outline our two-step taxonomy development process, present and validate the taxonomy using the identified cases, and conclude with a discussion of findings, implications, and limitations.

2 Embodied Generative Artificial Intelligence

GenAI refers to the capability of AI models, to generate novel data rather than merely classifying existing data (Banh & Strobel, 2023; Feuerriegel et al., 2024). It serves as both a general-purpose technology and a domain-specific solution, continuously expanding into new fields while transforming established domains such as healthcare and manufacturing (Ooi et al., 2025). For instance, chatbots, which were initially rule-based interaction partners, have evolved into comprehensive copilots based on GenAI capable of processing and generating multimodal inputs and outputs, including text, images, and audio (Banh & Strobel, 2023). Recent advancements in GenAI can be classified into two pathways. First, research aims to deploy GenAI systems as adaptive autonomous agents in digital environments. In this context, adaptivity refers to the system's ability to adjust itself to varying environments and situations. Second, efforts are being made to enhance GenAI's understanding of physics and real-world objects, enabling effective interaction with its physical environment (Ahn et al., 2022; W. Lai et al., 2024). As a result, new approaches have emerged, such as VLMs and world foundation models extending beyond text-based LLMs (Agarwal et al., 2025; J. Gao et al., 2024).

Simultaneously, robotics research strives to automate physical-world tasks and support humans in various capacities through robots enhancing productivity, working conditions (Chugh et al., 2024; Fan et al., 2025; Nwankwo & Rueckert, 2024). AI-powered capabilities such as body stabilization, navigation, and collision prevention play a crucial role (Radosavovic et al., 2024; Wiedebach et al., 2016; Zhu & Zhang, 2021). Despite these advancements, robots still largely remain passive entities, executing predefined actions or requiring manual human control. A promising solution is to integrate advancements from robotics, GenAI, and agentic AI research into embodied GenAI.

Embodied GenAI ca be defined as the integration of GenAI into physical systems (e.g., robots, drones, vehicles) to enable intelligent, autonomous, and interactive agents as a holistic information system (Abeyruwan et al., 2025; Azzolini et al., 2025). *Autonomy* refers to the system's ability to operate independently and in a goal-directed manner without user intervention enabling task automation and decision-making (Acharya et al., 2025). A physical system is one that perceives and engages (i.e., *interactivity*) with the environment (i.e., physical world and its entities) through sensors and actuators, and can physically manipulate or respond to elements within its surroundings (Duan et al., 2022; Pfeifer & Iida, 2004). GenAI provides intelligence to the physical system. In this case, *intelligence* refers to general understanding of the physical world and the generation of appropriate embodied decisions, such as positioning within an environment, executing manipulation tasks, handling open-vocabulary instructions, and facilitating dynamic interactions and continuous learning (Abeyruwan et al., 2025; Kato et al., 2024; Paolo et al., 2024).

3 Research Design

Based on our research question, we build a taxonomy following Nickerson et al. (2013) with to goal of identifying and understanding the key attributes of embodied GenAI. To establish a reliable foundation for the taxonomy, we follow a combined approach that incorporates both academic literature (conceptual-to-empirical) and empirical data from existing systems (empirical-to-conceptual). This approach has already proven itself in the literature (e.g., Börner, 2024; Brogt & Strobel, 2020; Gimpel et al., 2018).

Our objective is to develop a scientifically grounded and practical framework that clearly defines the key dimensions and characteristics of embodied GenAI. The scope of the taxonomy is to classify a specific instantiation of an embodied GenAI system, which is characterized by a particular embodiment and a specific GenAI model, functioning together as a unified entity. To achieve this, we first decompose the concept of embodied GenAI into two core elements: *embodiment*, rooted in robotics research, and *intelligence*, derived from GenAI. Additionally, considering its integration into a holistic system completes this perspective, leading to the three overarching meta-characteristics: *embodiment, intelligence, and system*. The corresponding ending conditions of our taxonomy development are based on Nickerson et al.'s (2013) subjective ending conditions, along with the following objective ending conditions: "*Every dimension is unique and not repeated; Every characteristic is unique within its dimension*" (Nickerson et al., 2013, p. 344).

The first iteration **(conceptual-to-empirical)** establishes the foundation of the taxonomy by identifying an initial set of dimensions and characteristics based on existing literature. To achieve this, we perform a systematic literature review following the methodology of Webster and Watson (2002) and vom Brocke et al. (2009), creating a comprehensive knowledge base as the starting point for taxonomy development as shown in Figure 1.

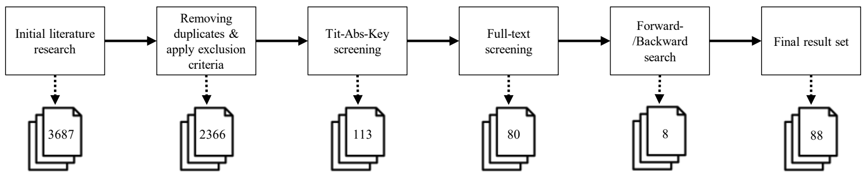

Figure 1. Systematic Literature Review

Our search query is guided by our meta-characteristics and integrates *GenAI* with *embodiment*, incorporating relevant synonyms: *"("generative artificial intelligence" OR "GenAI" OR "LLM" OR "language model" OR "chatbot" OR "conversational agent") AND ("embodi*" OR "robot*" OR "drone")"*. The search was conducted based on title, abstract, and keyword in six established academic databases (e.g., *AISeL*, *IEEE*, *ACM*, etc.). To ensure a high-quality selection, we applied exclusion criteria, restricting the dataset to peer-reviewed articles in English in accordance with our understanding of embodied GenAI. We excluded purely technical paper, e.g., benchmarks of GenAI models. The initial search yielded 3,687 results, which were reduced to 2,366 after removing duplicates and non-peer-reviewed papers. A title-abstract-keyword screening narrowed the selection to 113 papers, which were full screen, resulting in a set of 80 relevant papers. Forward and backward searches for the selected full texts resulted in a final dataset of 88 papers (see online appendix: https://bit.ly/embodied_genai_taxonomy). The final dataset includes papers on generative AI models developed for embodiment, applications of LLMs tested in physical contexts, and robotic systems, such as drones and autonomous vehicles, that integrate generative AI. During the first iteration, we assessed the identified dimensions and characteristics against the meta-characteristics and ending conditions, which were not yet fully met, particularly regarding conciseness and comprehensiveness. Thus, we conduct a second iteration **(empirical-to-conceptual)** by analyzing embodied GenAI systems. We triangulated publicly available data sources (e.g., websites, blog posts, product presentations) of organizations which develop and/or sell embodied GenAI systems, resulting in 40 real world cases. These sources provide insights into the characteristics, functionalities, and applications of embodied GenAI (see online appendix). The case analysis was discussed among the authors to ensure reliability. Finally, the classification of the real-world cases confirms that both objective ending conditions and the subjective ending conditions are met.

4 A Taxonomy of Embodied Generative Artificial Intelligence

Our taxonomy of embodied GenAI comprises three meta-characteristics, 16 dimensions, and 50 associated characteristics (Figure 2). These characteristics are classified as either mutually exclusive (ME) or non-exclusive (N).

	Dimension	Characteristics						Mutual
Embodiment	**Interface**	Gesture (19)	Verbal (35)	Touch (6)	Remote Controller (11)	Imitation (8)	Code (23)	N
	Perception	Functional (9)	Human-Like (28)	Superhuman (3)				ME
	Actions	Manipulation (27)	Communication (31)	Repositioning (28)				N
	Movement	Static (1)	Functional (35)	Expressive (16)				N
	Appearance	Functional (12)	Zoomorphic (4)	Anthropomorphic (24)				ME
Intelligence	**Autonomy**	Heteronomous (2)	Semi-Autonomous (13)	Autonomous (25)				ME
	Agency	Reactive (31)	Proactive (27)					N
	Capability	Analyzing (34)	Planning (28)	Execution (35)				N
	Knowledge	Pre-Trained (13)	Memory (12)	Learning (15)				ME
	Functionality	Core (5)	Fixed (11)	Universal (24)				ME
	Location	On-Premise (32)	Edge (3)	Cloud (30)				N
	Embodiment Dependency	Tied (31)	Agnostic (9)					ME
	Openness	Closed (15)	Semi-Open (16)	Open (9)				ME
System	**Interaction**	Isolated (6)	Dyadic (12)	Group (22)				ME
	Collaboration	Solo (36)	With Human(s) (32)	With other System(s) (25)				N
	Perceived Value	Operational (34)	Psychological (7)	Societal (22)	Aesthethic (9)			N

Figure 2. Embodied Generative Artificial Intelligence Taxonomy

4.1 Embodiment

The *embodiment* serves as the physical foundation of embodied GenAI, anchoring the intelligence provided by GenAI in the real world (Ahn et al., 2022; Paolo et al., 2024). It encompasses design as well as input and output capabilities.

Interface is the non-mutually exclusive input layer between humans and embodied GenAI. It encompasses various interaction modalities, including *gesture-based control* through camera-detected movements, *verbal interaction* via speech or text, and *touch-*

based engagement with physical components or embedded touchscreens (Yuhan Hu et al., 2025; Y. Z. Lai et al., 2024; Nwankwo & Rueckert, 2024). Additionally, *remote control* methods include gamepads and smartphone apps for manual navigation. Some systems employ *imitation*, where AI mimics human actions using vision-based tracking or teleoperation, while others allow *code-based* input for fine-grained control.

Perception in embodied GenAI encompasses its multi-modal ability to interpret sensory data (Firoozi et al., 2024; Kim et al., 2024; Paolo et al., 2024). It can be categorized into three mutually exclusive levels: *Functional* perception encompasses fundamental environmental sensing, including visual, auditory, and haptic input (Al Moubayed et al., 2012; Chamiti et al., 2024). *Human-like* perception extends functional perception by integrating facial and mental state recognition (Abe et al., 2012; Sievers & Russwinkel, 2024), emotion tracking (Cherakara et al., 2023; Etesam et al., 2024), object and voice recognition, language understanding (Chen et al., 2023; Cui et al., 2024), and reasoning about physics and world concepts (B. Shi et al., 2024; Hwang et al., 2024; W. Lai et al., 2024). *Superhuman* perception surpasses human capabilities by integrating digital data from external sources such as the Internet, IoT devices, and advanced sensory inputs like 360-degree 3D point clouds, ultrasonic frequencies, and gas detection, enabling a comprehensive understanding of the environment (Akl et al., 2025).

Actions define the non-mutually exclusive output capabilities of embodied GenAI (Paolo et al., 2024). This includes *communication* actions by embodied GenAI systems through, e.g., gestures, facial expressions, activation of signal lights, and verbal interaction via speech or text (Da Silva et al., 2024; B. Ding et al., 2024; Holgado et al., 2024). Beyond communication, actions involve *manipulation* of the environment, such as picking, placing, or pushing objects (Y. Ding et al., 2023; Jeong et al., 2024), as well as self-*repositioning*, where the system moves through space to enhance its perception or complete tasks (W. Z. Cai et al., 2024; Dorbala et al., 2024).

Movement is characterized by the motivation behind motions of embodied GenAI. If the system does not have the ability to move, we categorize it as a *static* systems (Cherakara et al., 2023). If the system is capable of movement, a distinction is made between functional and expressive movements, which can occur simultaneously, making this dimension non-mutually exclusive. *Functional* movement includes self-repositioning and object manipulation for executing tasks or improving perception. *Expressive* movement enhances communication by conveying intention, attention, attitude, and emotion. Intention helps users anticipate actions, while attention is signaled through gaze or gestures like waving (Gasteiger et al., 2024; Serfaty et al., 2023). Here, attitude is expressed through gestures such as nodding, while emotion is simulated via movement patterns, e.g., bouncy motions for happiness, slow motions for relaxation, lowered head for sadness, and abrupt movements for fear (Yuhang Hu et al., 2024; Sievers & Russwinkel, 2024; Wang et al., 2024). Effective expressiveness aligns with the interaction goal, be convincing, and maintain character consistency (Nichols et al., 2023; Paplu et al., 2021), contributing to a credible personality (Makkonen et al., 2022).

Appearance of embodied GenAI varies mutually exclusively from functional to zoomorphic and anthropomorphic designs (Yuhan Hu et al., 2025). *Functional* embodiments like industrial machines or autonomous vehicles are optimized for specific tasks. For example, a robotic vacuum cleaner is designed to be slim and flat to clean under furniture, while an industrial system prioritizes degrees of movement and strength of its manipulators (Choi et al., 2021; Jia et al., 2024; Wu et al., 2023). In contrast, *zoomorphic* and *anthropomorphic* designs are modeled after animals (Sathyamoorthy et al., 2024; Schnitzer et al., 2024; Zhang et al., 2024) or humans (Antikatzidis et al., 2024; Seppelfelt et al., 2022), respectively, to enhance trust and acceptance through familiar appearances (Janeczko & Foster, 2022). These designs are practical for systems operating in human-optimized environments, as natural counterparts demonstrate both free mobility (zoomorphic designs) and tool usage (anthropomorphic designs). While sub forms may exist (e.g., anthropomorphic upper bodies mounted on wheels) they are still inspired by human or animal forms and do not represent purely functional designs.

4.2 Intelligence

The *intelligence*, powered by the latest GenAI models, operates within the embodiment, shaping its capabilities and interactions.

Autonomy determines how independently and goal-directed embodied GenAI can operate without user intervention, divided into three successive level (Makkonen et al., 2022). *Heteronomous* systems require active human control and direct commands to operate. *Semi-autonomous* systems leverage GenAI to interpret open-vocabulary commands, autonomously reason intermediate steps to execute tasks (Y. Ding et al., 2023; Li et al., 2023; Tsushima et al., 2025). *Autonomous* systems function without additional user input, independently pursuing goals.

Agency in embodied GenAI defines the system's intelligence in terms of its ability to take initiative and respond to external stimuli, with both characteristics potentially occurring simultaneously (Chugh et al., 2024; Yuhan Hu et al., 2025; Kim et al., 2024). *Proactive* intelligence involves anticipating future situations as well as independently initiating interactions and actions driven by high-level objectives (Paolo et al., 2024). This can be enhanced through knowledge integration, allowing the system to autonomously guide conversations by recalling past interactions. *Reactive* intelligence refer to a embodied GenAI system's ability to respond to environmental changes in real time (Makkonen et al., 2022). Similar to human reflexes, these immediate responses ensure stability and contribute to self-preservation (Paolo et al., 2024).

Capability encompasses the embodied system's cognitive abilities enabled by GenAI models. In some systems, only a single capability is present, while others combine multiple capabilities within their intelligence. In this context, GenAI may also function as a shared intelligence, coordinating multiple embodied GenAI systems within a network. *Analyzing* involves real-time processing and interpretation of sensory input (Fang et al., 2024; Firoozi et al., 2024; Jeong et al., 2024). It involves context understanding and task grasping to infer user intent from incomplete instructions (Ahn et al., 2022). Key aspects include recognizing world concepts (B. Shi et al., 2024), such as zero-shot object detection (Firoozi et al., 2024), and material properties like fragility, mass, etc. (J. Gao et al., 2024). *Planning* involves organizing tasks and actions, including decision-making (Dalal et al., 2024; Holgado et al., 2024; Qi & Jing, 2024) and movement planning (Ahn et al., 2022; B. Shi et al., 2024; W. H. Cai et al., 2024). Zero-shot task planning plays a key role, enabling embodied GenAI to plan solutions without prior examples (Kapelyukh et al., 2024; Vemprala et al., 2024). Key aspects include assessing task feasibility through goal imagination (Aregbede et al., 2024), recognizing required subtasks (B. Shi et al., 2024), and simulating action outcomes (Cao et al., 2024; Chu et al., 2024). Safety considerations are critical, particularly in human-robot collaboration (Varley et al., 2024). *Execution* includes low-level control of actuators triggered by using API calls (Chu et al., 2024; Sun et al., 2024; Yoshida et al., 2023)

Knowledge in embodied GenAI describes the system's ability to acquire and utilize knowledge beyond its pre-trained understanding. The individual levels of knowledge build upon each other being mutually exclusive. *Pre-trained* knowledge represents the foundation based on embedded information from GenAI model training, including domain expertise in certain cases (Banh & Strobel, 2023; Brohan et al., 2023; Hasko et al., 2023). Specialized multimodal vision and task-planning models that enable real-world understanding are particularly relevant in the context of embodied GenAI (Driess et al., 2023; Firoozi et al., 2024; J. Gao et al., 2024). *Memory* extends pre-trained knowledge by acquired knowledge through perception and processing (Da Silva et al., 2024; Hasko et al., 2023; Kang et al., 2024). It also includes the capacity to retain past experiences (Paolo et al., 2024), which is particularly relevant for multitasking, object state tracking, and remembering conversational context (Ali et al., 2024; Saravanan et al., 2022). *Learning*, builds on pre-trained knowledge and memory to develop new skills (Paolo et al., 2024). While learning relies on memory, it is not synonymous with it; the use of memory does not necessarily lead to learning. One application of learning is adapting to user preferences to personalize systems (Wu et al., 2023).

Functionality in embodied GenAI ranges mutually exclusive from *core* over *fixed* to *universal*, reflecting both the system's purpose and adaptability. Systems with *core* functionality can adapt slightly to new contexts, being capable of executing essential baseline tasks such as perception, environmental interaction, and navigation (Anwar et al., 2024; Dorbala et al., 2024). *Fixed* functionality describes domain-specific systems with moderate adaptability, allowing adjustment within a predefined purpose. These systems are optimized for specific applications, for example service (e.g., receptionists, waiters, chefs) (Antikatzidis et al., 2024; Gasteiger et al., 2024; Kanazawa et al., 2024), industrial use (Choi et al., 2021; Fan et al., 2025), etc. At the highest level, *universal* systems are multi-functional capable of dynamically adapting to diverse environments and tasks, independent of a fixed application domain (Makkonen et al., 2022). They generalize across use cases, making them general-purpose GenAI-driven embodiments.

Location of the intelligence in embodied GenAI systems varies depending on computational requirements and latency constraints but can also be distributed and combined as needed. *On-premise* deployment runs GenAI directly on the embodiment's hardware, but limited processing power restricts the use of large models (Strobel et al., 2024). *Edge* computing places GenAI near the embodiment, utilizing more powerful hardware while maintaining low latency (Dong et al., 2024; Grumeza et al., 2024; Sundaravadivel et al., 2024). *Cloud*-based deployment offers scalability and high computational power introducing latency due to remote processing (Dong et al., 2024).

Embodiment Dependency describes two distinct characteristics of the relationship between intelligence and physical form. While GenAI models are inherently independent, their deployment within an embodiment requires adaptation to its specific characteristics, to effectively analyze, plan, and execute actions (*tied* intelligence) (Qi & Jing, 2024; Sun et al., 2024). *Agnostic* intelligence can operate through standardized frameworks like ROS2, enabling seamless integration across various hardware platforms by adapting to different embodiments (Yuhan Hu et al., 2025; Koubaa et al., 2025).

Openness encompasses both the customizability and the transparency of GenAI within an embodied system, classified into three mutually exclusive levels (Serfaty et al., 2023; Strobel et al., 2024). *Open* intelligence enables users to freely select GenAI models and reprogram hardware through frameworks like ROS2, aligning with a high level of transparency by providing open-source code and full documentation (Koubaa et al., 2025). These systems are often offered as educational or research versions of embodied GenAI. *Semi-open* intelligence offers limited transparency, disclosing partial details such as the AI model provider, with restricted customization options. In contrast, *closed* intelligence is fully vendor-controlled, consisting of proprietary, pre-installed solutions that can only be updated by the provider and lack technical disclosure.

4.3 System

The meta-characteristic *system* integrates embodiment and intelligence, classifying dimensions that emerge from their interaction as a coherent system.

Interaction defines the extent to which embodied GenAI supports interactions, categorized into mutually exclusive, sequentially building levels. *Isolated* interaction applies to systems that operate independently, engaging with the environment and non-living objects while avoiding interaction with humans, animals, or other systems. *Dyadic* interaction enables communication with a single individual or other AI system (Qi & Jing, 2024; Seppelfelt et al., 2022). *Group* interaction allows engagement with multiple individuals simultaneously, such as participating in group conversations (Addlesee et al., 2024; Al Moubayed et al., 2012). This characteristic is often combined with expressive movements to enhance communication.

Collaboration in embodied GenAI distinguishes whether a system operates independently or cooperates with humans or other AI systems to achieve an overall goal (Makkonen et al., 2022). As it can engage in both collaboration and independent operation at different times, this dimension is not mutually exclusive. *Solo* operation means the system functions without collaboration. Collaboration *with other systems* involves multi-agent cooperation, such as sharing sensory data for navigation (Akl et al., 2025) or coordinating in fleets (e.g., drone swarms) (Pueyo et al., 2024). Embodied GenAI may also respond cooperatively to environmental challenges like route obstructions or work collectively toward a shared goal (Kato et al., 2024). Collaboration *with humans* includes task execution based on human instructions (Choi et al., 2021; Farooq et al., 2024; Gkournelos et al., 2023), such as identifying and delivering objects (F. Gao et al., 2024). In human-in-the-loop approaches, the human acts as an additional control or decision-making unit (Hunt et al., 2024; Vemprala et al., 2024), or embodied GenAI requests human assistance to achieve a shared goal (Mandi et al., 2024).

Perceived Value of embodied GenAI reflects the system's impact as recognized by society, encompassing four non-mutually exclusive characteristics. *Operational* value describes the system's impact on efficiency and effectiveness in workflows. Embodied GenAI enhances productivity, performance, and cost efficiency by automating routine tasks and facilitating human-AI collaboration (Fan et al., 2025; Holgado et al., 2024). *Psychological* value relates to the system's contribution to mental and emotional well-being, offering emotional support, empathetic interactions, and engaging experiences, particularly beneficial for elderly, disabled, or socially isolated individuals (Schnitzer et al., 2024; Sievers & Russwinkel, 2024). *Societal* value reflects broader social benefits, such as democratizing education, providing public services, and offering entertainment through embodied GenAI (Abe et al., 2012; Antikatzidis et al., 2024; Borg et al., 2024). *Aesthetic* value concerns the visual design and appeal of the system including style of voice, which influences user acceptance, trust, and integration into daily life (Liberman-Pincu et al., 2023). While less critical in manufacturing, it plays a significant role in household and public service systems, ensuring seamless adaptation to home, workplace, or public environments (Buchem et al., 2024; Janeczko & Foster, 2022).

5 Evaluation and Application

We demonstrate the taxonomy's practical relevance and robustness by applying 40 cases (see Section 3) to the taxonomy. The cases were divided among the team of authors for classification. The final taxonomy, including the descriptions of dimensions and characteristics, served as the basis for classification. The classifications were reviewed and discussed internally to ensure reliability. Numbers in brackets (see Figure 2) indicate the embodied GenAI cases assigned to a specific characteristic. Since some dimensions are not mutually exclusive, the total count within a dimension may exceed 40. In the following, we present two representative cases in detail (Figure 3). Our online appendix includes a detailed classification of all 40 embodied GenAI systems.

Figure 02, developed by Figure AI Inc., is an anthropomorphic embodied GenAI system. Equipped with speech interface, human-like perception, and hand-like manipulators, it enables universal actions such as manipulation, repositioning, and communication. Integrating OpenAI's LLM and NVIDIA's Cosmos World Foundation Model (Agarwal et al., 2025; FigureAI, 2024), Figure 02 autonomously analyzes its environment, plans tasks, and executes actions. Its intelligence is distributed between on-premise computation and the cloud, extending pre-trained knowledge with memory. Figure 02 supports solo and group operation, with a focus on task automation and assistance.

Spot, a zoomorphic quadruped robot by Boston Dynamics Inc., operates flexibly across industries like manufacturing and construction by integrating GenAI. It supports various control interfaces, as well as full autonomy. Spot enables both functional and expressive movement. Its open-source SDK allows integration of various GenAI models, running locally or via the cloud, making its intelligence open and adaptable. Spot exhibits both reactive and proactive behavior, while learning from experience.

In summary, the two respective cases illustrate different implementations of embodied GenAI, representing the current state-of-the-art in systems that integrate GenAI with robotics. Figure 02 follows a GenAI-first approach aimed at physically embodying

GenAI, whereas Spot originates from robotics research and is subsequently enhanced through the latest GenAI capabilities.

	Dimensions	**Figure 02**	**Spot**
Embodiment	*Interface*	Gesture, Verbal, Code	Verbal, Controller, Code
	Perception	Human-Like	Superhuman
	Actions	Manipulation, Communication, Repositioning	
	Movement	Functional	Functional, Expressive
	Appearance	Anthropomorphic	Zoomorphic
Intelligence	*Autonomy*	Autonomous	
	Agency	Proactive	Reactive, Proactive
	Capability	Analyzing, Planning, Execution	
	Knowledge	Memory	Learning
	Functionality	Universal	
	Location	On Premise, Cloud	
	Embodiment Dependency	Tied	Agnostic
	Openness	Semi-Open	Open
System	*Interaction*	Group	
	Collaboration	Solo, With Human(s), With other System(s)	
	Perceived Value	Operational, Societal	

Figure 3. Taxonomy Classification of Two Respective Cases

Our analysis shows that 60% of embodied GenAI systems resemble the human body. Despite the uncanny valley effect (Rosenthal-von der Pütten & Krämer, 2014), anthropomorphic designs are advantageous for assistive systems, especially in human-designed environments where they can utilize human-made tools. A clear trend is also observable: the majority of systems support communication (78%) and rely on verbal input as their primary interface (88%). Nevertheless, manipulation actions (68%) and repositioning (70%) are also frequently observed. System perception is typically human-like (70%). In contrast, superhuman perception remains uncommon (8%) and is primarily used for specialized applications, such as detecting hazardous situations (e.g., gas leaks in factories). 97% of the systems analyzed are capable of movement, either functional (88%) and/or expressive (40%), primarily focusing on facial expressions.

Regarding intelligence, 63% of embodied GenAI systems are fully autonomous. Furthermore, 60% are designed for general-purpose applications, 28% are restricted to fixed use cases, and 12% provide only core functionality. A high number of systems supports reactive (78%) and proactive (68%) agency, while 45% support both. This is enabled by the fact that GenAI allows task planning in over 70% of cases, while at least 85% of systems leveraging GenAI to both analyze their environment and execute actions. Approximately each third of the analyzed systems build their knowledge differently: 32% rely on pre-trained models, 30% of systems extend them through a memory function, and 38% support learning. The deployment of intelligence is distributed between on-premise hardware (80%) and cloud infrastructure (75%), with 55% of systems utilizing a hybrid approach. Edge deployment remains relatively rare, occurring in only 8% of cases. For most systems, not all information about the intelligence architecture is accessible (40% are classified as semi-open, 38% as closed), which overall aligns with tied embodiment dependency present in 78% of all systems.

Only 12% of the analyzed systems do not support collaboration with humans or other systems. The remaining systems explicitly incorporate collaboration into their intended purpose, with an additional focus on interaction, either dyadic (30%) or in groups (55%). Most systems demonstrate operational value (85%) and societal value (55%), whereas 23% of system developers prioritize the aesthetic aspects of their designs.

6 Conclusion, Limitations, Outlook

Recently, generative AI has evolved beyond its role as a copilot tool and now serving as the foundation for Agentic AI, enabling autonomous AI agents to solve digital tasks (Acharya et al., 2025). The next step in this evolution extends from the digital to the physical world, driving the emergence of a new type: Embodied GenAI. By integrating GenAI into physical embodiments such as robots (i.e., generative physical AI), GenAI is enabled to perform actions in the physical world (Agarwal et al., 2025; Liu et al., 2025) and introduces a new dimension of human-AI collaboration. To shed light on this emerging field, we identify the characteristics of embodied GenAI.

From a **scientific perspective**, our taxonomy offers a first conceptual outline of key dimensions and characteristics, enabling the classification of embodied GenAI while serving as a foundation for further research. Our taxonomy is not constrained to specific domains, enabling future studies to explore embodied GenAI in various application areas, such as service robots and industrial automation. For **practitioners**, the taxonomy functions as a tool for systematically analyzing existing embodied GenAI and as a reference framework for the development and optimization of new applications.

Despite its contributions, our work has certain **limitations.** *First*, as embodied GenAI is still in its early stages, both conceptual research contributions and empirical systems will continue to grow and evolve as they progress toward market maturity. This ongoing development necessitates continuous reassessment and refinement of our taxonomy to ensure its future alignment. To establish a robust foundation for the current state of research, we incorporated both conceptual insights, identified through a systematic literature review, and empirical data from existing embodied GenAI cases into our analysis. *Second*, the classification of dimensions and characteristics is inherently interpretative, as our analysis of empirical cases relies on the triangulation of publicly available data about embodied GenAI. Other researchers may identify alternative aspects, potentially leading to modifications or extensions of our framework. This results in implications for **future research**. To gain insights beyond publicly available data, experts such as embodied GenAI developers could be involved. Their perspectives can deepen the understanding of systems still under development and contribute to a more refined taxonomy. Another possible direction would be to examine specific domains, such as the care sector, and develop a corresponding specialized taxonomy. This could highlight the potential of embodied GenAI in addressing societal challenges such as workforce shortages, loneliness, and the growing demand for care due to an aging population.

References

Abe, K., Iwasaki, A., Nakamura, T., Nagai, T., Yokoyama, A., Shimotomai, T., Okada, H., & Omori, T. (2012). Playmate Robots That Can Act According to a Child's Mental State. In *International Conference on Intelligent Robots and Systems,* Vilamoura-Algarve, Portugal.

Abeyruwan, S., Ainslie, J., Alayrac, J.-B., Arenas, M. G [Montserrat Gonzalez], Armstrong, T., Balakrishna, A., Baruch, R., Bauza, M., Blokzijl, M., Bohez, S., Bousmalis, K., Brohan, A [Anthony], Buschmann, T., Byravan, A., Cabi, S., Caluwaerts, K., Casarini, F., Chang, O., Chen, J. E., . . . Zhou, Y. (2025). *Gemini Robotics: Bringing AI into the Physical World*. http://arxiv.org/pdf/2503.20020.

Acharya, D. B., Kuppan, K., & Divya, B. (2025). Agentic AI: Autonomous Intelligence for Complex Goals—A Comprehensive Survey. *IEEE Access*, **13**, 18912–18936. https://doi.org/10.1109/ACCESS.2025.3532853

Addlesee, A., Cherakara, N., Nelson, N., García, D. H., Gunson, N., Sieinska, W., Romeo, M., Dondrup, C., & Lemon, O. (2024). A Multi-party Conversational Social Robot Using LLMs. In *HRI '24, Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 1273–1275). Association for Computing Machinery. https://doi.org/10.1145/3610978.3641112

Agarwal, N., Ali, A., Bala, M., Balaji, Y., Barker, E., Cai, T., Chattopadhyay, P., Chen, Y [Yongxin], Cui, Y., Ding, Y [Yifan], Dworakowski, D., Fan, J., Fenzi, M., Ferroni, F., Fidler, S., Fox, D., Ge, S., Ge, Y., Gu, J., . . . Zolkowski, A. (2025). *Cosmos World Foundation Model Platform for Physical AI*. https://arxiv.org/abs/2501.03575.

Ahn, M., Brohan, A [A.], Brown, N., Chebotar, Y., Cortes, O., David, B., Finn, C., Fu, C. Y., Gopalakrishnan, K [K.], Hausman, K., Herzog, A [A.], Ho, D., Hsu, J., Ibarz, J., Ichter, B., Irpan, A., Jang, E., Ruano, R. J., Jeffrey, K., . . . Zeng, A. (2022). Do As I Can, Not As I Say: Grounding Language in Robotic Affordances. In *6th Conference on Robot Learning, CoRL 2022*.

Akl, J., Gning, A., Omrani, H., Caspary, O., Blansche, A., & Abdallah, F. (2025). Opportunistic collaboration between heterogeneous agents using an unstructured ontology via GenAI. *International Journal of Intelligent Robotics and Applications.* https://doi.org/10.1007/s41315-024-00411-5

Al Moubayed, S., Beskow, J., Skantze, G., & Granström, B. (2012). Furhat: A back-projected human-like robot head for multiparty human-machine interaction. In *Cognitive behavioural systems: Lecture Notes in Computer Science* (7403 LNCS, pp. 114–130). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-34584-5_9

Ali, H., Allgeuer, P., Mazzola, C., Belgiovine, G., Kaplan, B. C., Gajdošech, L., & Wermter, S. (2024). Robots Can Multitask Too: Integrating a Memory Architecture and LLMs for Enhanced Cross-Task Robot Action Generation. In *2024 IEEE-RAS 23rd International Conference on Humanoid Robots (Humanoids)* (pp. 811–818). https://doi.org/10.1109/Humanoids58906.2024.10769803

Antikatzidis, A., Feidakis, M., Marathaki, K., Toumanidis, L., Nikolaou, G., & Patrikakis, C. Z. (2024). Enrich Humanoids With Large Language Models (LLM). In *2024 IEEE Global Engineering Education Conference (EDUCON)*.

Anwar, A., Welsh, J., Biswas, J., Pouya, S., & Chang, Y. (2024). *ReMEmbR: Building and Reasoning Over Long-Horizon Spatio-Temporal Memory for Robot Navigation*. https://arxiv.org/abs/2409.13682.

Aregbede, V., Abraham, S. S., Persson, A., Langkvist, M., & Loutfi, A. (2024). Affordance-Based Goal Imagination for Embodied AI Agents. In *International Conference on Development and Learning (ICDL)*.

Azzolini, A., Bai, J., Brandon, H., Cao, J., Chattopadhyay, P., Chen, H., Chu, J., Cui, Y., Diamond, J., Ding, Y [Yifan], Feng, L., Ferroni, F., Govindaraju, R., Gu, J., Gururani, S., Hanafi, I. E., Hao, Z., Huffman, J., Jin, J., . . . Zhang, Z. (2025). *Cosmos-Reason1: From Physical Common Sense To Embodied Reasoning*. http://arxiv.org/pdf/2503.15558.

B. Shi, Y. Ou, H. Zhao, K. Cao, & D. Wang (2024). Research on Robot's Understanding of Different Concepts and Autonomous Planning. In *2024 IEEE International Conference on Real-time Computing and Robotics (RCAR)*.

Bai, Q., Li, S., Yang, J., Song, Q., Li, Z [Zhiang], & Zhang, X [Xingxing] (2020). Object Detection Recognition and Robot Grasping Based on Machine Learning: A Survey. *IEEE Access*, **8**, 181855–181879. https://doi.org/10.1109/ACCESS.2020.3028740

Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, **33**(1), 1–17. https://doi.org/10.1007/s12525-023-00680-1

Borg, A., Parodis, I., & Skantze, G. (2024). Creating Virtual Patients using Robots and Large Language Models: A Preliminary Study with Medical Students. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*.

Börner, E. (2024). Unlocking Value: Towards a Comprehensive Taxonomy for Commercializing Open Data in the Business Context. *Wirtschaftsinformatik 2024 Proceedings*.

Brogt, T., & Strobel, G. (2020). Service Systems in the Era of the Internet of Things: A Smart Service System Taxonomy. *ECIS 2020 Research Papers*.

Brohan, A [A.], Brown, N., Carbajal, J., Chebotar, Y., Chen, X [X.], Choromanski, K [K.], Ding, T [T.], Driess, D [D.], Dubey, A., Finn, C., Florence, P., Fu, C., Arenas, M. G [M. G.], Gopalakrishnan, K [K.], Han, K., Hausman, K., Herzog, A [A.], Hsu, J., Ichter, B., . . . Zitkovich, B. (2023). RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control. In *7th Conference on Robot Learning, CoRL 2023*. https://arxiv.org/abs/2307.15818.

Buchem, I., Bonga, G. A., & Tutu, R. (2024). From Screens to AI-Driven Human-Robot Interaction: Comparing Students' Perceptions of the Robots Pepper and Furhat for ChatGPT-Based Exam Preparation. In *International Conference on Robotics in Education (RiE)*.

Cai, W. H., Mori, Y., & Shimada, N. (2024). Generating Robot Action Sequences: An Efficient Vision-Language Models with Visual Prompts. In *2024 International Workshop on Intelligent Systems (IWIS)*.

Cai, W. Z., Huang, S. Y., Cheng, G. R., Long, Y. X., Gao, P., Sun, C. Y., & Dong, H. (2024). Bridging Zero-shot Object Navigation and Foundation Models through Pixel-Guided Navigation Skill. In *IEEE International Conference on Robotics and Automation (ICRA)*.

Cao, Y. J., Zhao, H., Cheng, Y. H., Shu, T., Chen, Y [Y.], Liu, G. L., Liang, G. Q., Zhao, J. H., Yan, J. Y., & Li, Y. (2024). Survey on Large Language Model-Enhanced Reinforcement Learning: Concept, Taxonomy, and Methods. *IEEE Transactions on Neural Networks and Learning Systems.* https://doi.org/10.1109/TNNLS.2024.3497992

Chamiti, T., Passalis, N., & Tefas, A. (2024). *Large Models in Dialogue for Active Perception and Anomaly Detection.* https://doi.org/10.1007/978-3-031-78447-7_25

Chen, X [X.], Luo, K., Gee, T., & Nejati, M. (2023). Does ChatGPT and Whisper Make Humanoid Robots More Relatable? *Proceedings of the 2023 Australasian Conference on Robotics and Automation*. https://arxiv.org/abs/2402.07095.

Cherakara, N., Varghese, F., Shabana, S., Nelson, N., Karukayil, A., Kulothungan, R., Farhan, M. A., Nesset, B., Moujahid, M., Dinkar, T., Rieser, V., & Lemon, O. (2023). *FurChat: An Embodied Conversational Agent using LLMs, Combining Open and Closed-Domain Dialogue with Facial Expressions*. https://arxiv.org/abs/2308.15214.

Choi, D., Shi, W., Liang, Y. S., Yeo, K. H., & Kim, J. J. (2021). Controlling Industrial Robots with High-Level Verbal Commands. In *Social Robotics: 13th International Conference, ICSR 2021,* Singapore, Singapore.

Chu, K., Zhao, X., Weber, C., Li, M., Lu, W., & Wermter, S. (2024). Large Language Models for Orchestrating Bimanual Robots. In *IEEE-RAS 23rd International Conference on Humanoid Robots (Humanoids)*.

Chugh, T., Tyagi, K., Srinivasan, P., & Challagundla, J. (2024). State-Based Dynamic Graph with Breadth First Progression For Autonomous Robots. In *14th Annual Computing and Communication Workshop and Conference (CCWC)*.

Chui, M., Hazan, E., Roberts, R., Singla, A., Smaje, K., Sukharevsky, A., Yee, L., & Zemmel, R. (2023). *The economic potential of generative AI: The next productivity frontiert*. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#introduction.

Cui, C., Ma, Y., Cao, X., Ye, W., & Wang, Z [Ziran] (2024). Drive as You Speak: Enabling Human-Like Interaction with Large Language Models in Autonomous Vehicles. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*.

Da Silva, C. B., Ramirez, J., Mastelari, N., Lotufo, R., Pereira, J., & Rohmer, E. (2024). Robotic Action Planning Using Large Language Models. In *2024 Latin American Robotics Symposium (LARS).* https://doi.org/10.1109/LARS64411.2024.10786412

Dalal, M., Chiruvolu, T., Chaplot, D. S., & Salakhutdinov, R. (2024). *Plan-Seq-Learn: Language Model Guided RL for Solving Long Horizon Robotics Tasks*. https://arxiv.org/abs/2405.01534.

Ding, B., Kirtay, M., & Spigler, G. (2024). Imitation of Human Motion Achieves Natural Head Movements for Humanoid Robots in an Active-Speaker Detection Task. In *2024 IEEE-RAS 23rd International Conference on Humanoid Robots (Humanoids)* (pp. 645–652). https://doi.org/10.1109/Humanoids58906.2024.10769814

Ding, Y [Yan], Zhang, X. H., Paxton, C., & Zhang, S. Q. (2023). Task and Motion Planning with Large Language Models for Object Rearrangement. In *International Conference on Intelligent Robots and Systems (IROS)*.

Dong, Q., Chen, X [X.], & Satyanarayanan, M. (2024). Creating Edge AI from Cloud-based LLMs. In *HOTMOBILE '24: Proceedings of the 25th International Workshop on Mobile Computing Systems and Applications*.

Dorbala, V. S., Mullen, J. F., & Manocha, D. (2024). Can an Embodied Agent Find Your "Cat-shaped Mug"? LLM-Based Zero-Shot Object Navigation. *IEEE Robotics and Automation Letters*, **9**(5), 4083–4090. https://doi.org/10.1109/LRA.2023.3346800

Driess, D [D.], Xia, F [F.], Sajjadi, M., Lynch, C., Chowdhery, A., Ichter, B., Wahid, A [A.], Tompson, J [J.], Vuong, Q., Yu, T., Huang, W., Chebotar, Y., Sermanet, P [P.], Duckworth, D., Levine, S., Vanhoucke, V [V.], Hausman, K., Toussaint, M., Greff, K., . . . Florence, P. (2023). PaLM-E: An Embodied Multimodal Language Model. In *Proceedings of the 40 th International Conference on Machine Learning,* Honolulu, Hawaii, USA.

Duan, J., Yu, S., Tan, H. L., Zhu, H., & Tan, C [Cheston] (2022). A Survey of Embodied AI: From Simulators to Research Tasks. *IEEE Transactions on Emerging Topics in Computational Intelligence*, **6**(2), 230–244. https://doi.org/10.1109/TETCI.2022.3141105

Etesam, Y., Yalçin, O. N., Zhang, C., & Lim, A. (2024). Contextual Emotion Recognition using Large Vision Language Models. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 4769–4776). https://doi.org/10.1109/IROS58592.2024.10802538

Fan, H. L., Liu, X [Xuan], Fuh, J. Y., Lu, W. F., & Li, B. B. (2025). Embodied intelligence in manufacturing: leveraging large language models for autonomous industrial robotics. *Journal of Intelligent Manufacturing*, **36**(2), 1141–1157. https://doi.org/10.1007/s10845-023-02294-y

Fang, J., Tan, X., Lin, S., Vasiljevic, I., Guizilini, V., Mei, H., Ambrus, R., Shakhnarovich, G., & Walter, M. R. (2024). Transcrib3D: 3D Referring Expression Resolution through Large Language Models. In *International Conference on Intelligent Robots and Systems (IROS)*.

Farooq, M. U., Kang, G., Seo, J., Bae, J., Kang, S., & Jang, Y. J. (2024). DAIM-HRI: A new Human-Robot Integration Technology for Industries. In *IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO)*.

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, **66**(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7

FigureAI. (2024). *We're developing the next generation of embodied artificial intelligence.* https://www.figure.ai/ai. Accessed: 13.03.2025.

Firoozi, R., Tucker, J., Tian, S., Majumdar, A [A.], Sun, J., Liu, W., Zhu, Y., Song, S., Kapoor, A., Hausman, K., Ichter, B., Driess, D [D.], Wu, J [Jiajun], Lu, C., & Schwager, M. (2024). Foundation models in robotics: Applications, challenges, and the future. *The International Journal of Robotics Research.* https://doi.org/10.1177/02783649241281508

Gao, F., Xia, L., Zhang, J [Jianjing], Liu, S., Wang, L., & Gao, R. X. (2024). Integrating Large Language Model for Natural Language-Based Instruction toward Robust Human-Robot Collaboration. In *Procedia CIRP*.

Gao, J., Sarkar, B., Xia, F [F.], Xiao, T [Ted], Wu, J [Jiajun], Ichter, B., Majumdar, A [A.], & Sadigh, D [D.] (2024). Physically Grounded Vision-Language Models for Robotic Manipulation. In *IEEE International Conference on Robotics and Automation (ICRA)*.

Gasteiger, N., Zhu, T., Broadbent, E., Lim, J., MacDonald, B. A., & Ahn, H. S. (2024). Social Robots in Hospital Settings: An Initial Exploration of the Services Provided, Interaction Style and in the Field Evaluation. In *2024 ACM/IEEE International Conference on Human-Robot Interaction,* Boulder, CO, USA.

Gimpel, H., Rau, D., & Röglinger, M. (2018). Understanding FinTech start-ups – a taxonomy of consumer-oriented service offerings. *Electronic Markets*, **28**(3), 245–264. https://doi.org/10.1007/s12525-017-0275-0

Gkournelos, C., Konstantinou, C., Angelakis, P., Michalos, G., & Makris, S. (2023). Enabling Seamless Human-Robot Collaboration in Manufacturing Using LLMs. In *Advances in Artificial Intelligence in Manufacturing. ESAIM*.

Grumeza, T. R., Lazar, T. A., & Fortis, A. E. (2024). Social Robots and Edge Computing: Integrating Cloud Robotics in Social Interaction. In *Advanced Information Networking and Applications. AINA*.

Hasko, R., Hasko, O., & Kutucu, H. (2023). Teaching Assistant Robots in Various Fields: Natural Sciences, Medicine and Specific Non-Deterministic Conditions. In *IDDM* (Vol. 3609, pp. 303–309).

Holgado, A. C., Ah Kun, L., & Di Paola, D. (2024). On the Future of Robotic Industrial Inspection: Challenges and Opportunities in the Era of Generative AI. In *2024 Eighth IEEE International Conference on Robotic Computing (IRC)* (pp. 272–274). https://doi.org/10.1109/IRC63610.2024.00054

Hu, Y [Yuhan], Huang, P., Sivapurapu, M., & Zhang, J [Jian]. (2025). *ELEGNT: Expressive and Functional Movement Design for Non-anthropomorphic Robot*. https://arxiv.org/abs/2501.12493.

Hu, Y [Yuhang], Wang, Y., Liu, R., Shen, Z., & Lipson, H. (2024). Reconfigurable Robot Identification from Motion Data. *Proceedings of the 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 14133–14140. https://doi.org/10.1109/IROS58592.2024.10801809

Hunt, W., Godfrey, T., & Soorati, M. D. (2024). *Conversational Language Models for Human-in-the-Loop Multi-Robot Coordination*. https://arxiv.org/abs/2402.19166.

Hwang, H., Kwon, S., Kim, Y., & Kim, D. (2024). Is It Safe to Cross? Interpretable Risk Assessment with GPT-4V for Safety-Aware Street Crossing. In *21st International Conference on Ubiquitous Robots (UR)*.

Janeczko, Z., & Foster, M. E. (2022). A Study on Human Interactions With Robots Based on Their Appearance and Behaviour. In *CUI '22, Proceedings of the 4th Conference on Conversational User Interfaces.* Association for Computing Machinery. https://doi.org/10.1145/3543829.3544523

Jeong, H., Lee, H., Kim, C., & Shin, S. (2024). A Survey of Robot Intelligence with Large Language Models. *Applied Sciences-Basel*, **14**(19). https://doi.org/10.3390/app14198868

Jia, S., Li, R., & Shi, Y. (2024). Exploring the Application of Large Language Models: Opportunities and Challenges. In *Proceedings of 2024 Chinese Intelligent Systems Conference. CISC*.

Kanazawa, N., Kawaharazuka, K., Obinata, Y., Okada, K., & Inaba, M. (2024). Recognition of Heat-Induced Food State Changes by Time-Series Use of Vision-Language Model for Cooking Robot. In *International Conference on Intelligent Autonomous Systems 18*.

Kang, H., Ben Moussa, M., & Thalmann, N. M. (2024). Nadine: A large language model-driven intelligent social robot with affective capabilities and human-like memory. *Computer Animation and Virtual Worlds*, **35**(4). https://doi.org/10.1002/cav.2290

Kapelyukh, I., Ren, Y., Alzugaray, I., & Johns, E. (2024). Dream2Real: Zero-Shot 3D Object Rearrangement with Vision-Language Models. In *IEEE International Conference on Robotics and Automation (ICRA)*.

Kato, Y., Yoshida, T [Takahiro], Sueoka, Y., Osuka, K., Yajima, R., Nagatani, K., & Asama, H. (2024). Design of a Multi-robot Coordination System based on Functional Expressions using Large Language Models. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 3447–3454). https://doi.org/10.1109/IROS58592.2024.10802571

Kim, Y., Kim, D., Choi, J., Park, J., Oh, N., & Park, D. (2024). A survey on integration of large language models with intelligent robots. *Intelligent Service Robotics*, **17**(5), 1091–1107. https://doi.org/10.1007/s11370-024-00550-5

Koubaa, A., Ammar, A., & Boulila, W. (2025). Next-generation human-robot interaction with ChatGPT and robot operating system. *Software - Practice and Experience*, **55**(2), 355–382. https://doi.org/10.1002/spe.3377

Lai, W., Zhang, T [Tianwei], Lam, T. L., & Gao, Y. (2024). Vision-Language Model-based Physical Reasoning for Robot Liquid Perception. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 9652–9659). https://doi.org/10.1109/IROS58592.2024.10801833

Lai, Y. Z., Radke, M., Nassar, Y., Gopal, A., Weber, T., Liu, Z. H., Zhang, Y. H., & Rätsch, M. (2024). Intuitive Multi-modal Human-Robot Interaction via Posture and Voice. In *International Conference on Robotics, Computer Vision and Intelligent Systems* (Vol. 2077, pp. 441–456). https://doi.org/10.1007/978-3-031-59057-3_28

Li, G., Han, X., Zhao, P., Hu, P., Nie, L., & Zhao, X. (2023). RoboChat: A Unified LLM-Based Interactive Framework for Robotic Systems. In *5th International Conference on Robotics, Intelligent Control and Artificial Intelligence (RICAI)*.

Liberman-Pincu, E., Parmet, Y., & Oron-Gilad, T. (2023). Judging a Socially Assistive Robot by Its Cover: The Effect of Body Structure, Outline, and Color on Users' Perception. *ACM Transactions on Human-Robot Interaction*, **12**(2), 1–26. https://doi.org/10.1145/3571717

Liu, D., Zhang, J [Junyu], Dinh, A.-D., Park, E., Zhang, S., Mian, A., Shah, M., & Xu, C. (2025). *Generative Physical AI in Vision: A Survey*. http://arxiv.org/pdf/2501.10928.

Makkonen, M., Salo, M., & Pirkkalainen, H. (2022). What Makes a (Ro)bot Smart? Examining the Antecedents of Perceived Intelligence in the Context of Using Physical Robots, Software Robots, and Chatbots at Work. *MCIS 2022 Proceedings. 3.*

Mandi, Z., Jain, S., & Song, S. (2024). RoCo: Dialectic Multi-Robot Collaboration with Large Language Models. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*.

Nichols, E., Szapiro, D., Vasylkiv, Y., & Gomez, R. (2023). How to Make a Robot Grumpy Teaching Social Robots to Stay in Character with Mood Steering. In *International Conference on Intelligent Robots and Systems (IROS)*.

Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems*, **22**(3), 336–359. https://doi.org/10.1057/ejis.2012.26

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science (New York, N.Y.)*, **381**(6654), 187–192. https://doi.org/10.1126/science.adh2586

Nwankwo, L., & Rueckert, E. (2024). The Conversation is the Command: Interacting with Real-World Autonomous Robot Through Natural Language. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*.

Ooi, K.-B., Tan, G. W.-H., Al-Emran, M., Al-Sharafi, M. A., Capatina, A., Chakraborty, A., Dwivedi, Y. K., Huang, T.-L., Kar, A. K., Lee, V.-H., Loh, X.-M., Micu, A., Mikalef, P., Mogaji, E., Pandey, N., Raman, R., Rana, N. P., Sarker, P., Sharma, A., . . . Wong, L.-W. (2025). The Potential of Generative Artificial Intelligence Across Disciplines: Perspectives and Future Directions. *Journal of Computer Information Systems*, **65**(1), 76–107. https://doi.org/10.1080/08874417.2023.2261010

OpenAI. (2025). *Introducing Operator*. https://openai.com/index/introducing-operator/. Accessed: 13.03.2025.

Paolo, G., Gonzalez-Billandon, J., & Kégl, B. (2024). Position: A Call for Embodied AI. In *Forty-first International Conference on Machine Learning.* (Vol. 235, pp. 39493–39508).

Paplu, S. H., Arif, M. N., & Berns, K. (2021). Adapting Behaviour of Socially Interactive Robot Based on Text Sentiment. In Zeghloul, S., Laribi, M.A., Sandoval, J. (Ed.), *Advances in Service and Industrial Robotics. RAAD 2021: Mechanisms and Machine Science* (Vol. 102, pp. 221–228). https://doi.org/10.1007/978-3-030-75259-0_24

Pfeifer, R., & Iida, F. (2004). Embodied Artificial Intelligence: Trends and Challenges. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, F. Iida, R. Pfeifer, L. Steels, & Y. Kuniyoshi (Eds.), *Lecture Notes in Computer Science. Embodied Artificial Intelligence* (Vol. 3139, pp. 1–26). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-27833-7_1

Pueyo, P., Montijano, E., Murillo, A. C., & Schwager, M. (2024). CLIPSwarm: Generating Drone Shows from Text Prompts with Vision-Language Models. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* Abu Dhabi, United Arab Emirates.

Qi, Z., & Jing, X. (2024). Advances in Large Language Models for Robotics. In *7th International Conference on Mechatronics, Robotics and Automation (ICMRA),* Wuhan, China,

Radosavovic, I., Xiao, T [Tete], Zhang, B., Darrell, T., Malik, J., & Sreenath, K. (2024). Real-world humanoid locomotion with reinforcement learning. *Science Robotics*, **9**(89), eadi9579. https://doi.org/10.1126/scirobotics.adi9579

Rosenthal-von der Pütten, A. M., & Krämer, N. C. (2014). How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior*, **36**, 422–439. https://doi.org/10.1016/j.chb.2014.03.066

Saravanan, A., Tsfasman, M., Neerincx, M. A., & Oertel, C. (2022). Giving Social Robots a Conversational Memory for Motivational Experience Sharing. In *31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN),* Napoli, Italy.

Sathyamoorthy, A. J., Weerakoon, K., Elnoor, M., Zore, A., Ichter, B., Xia, F [F.], Tan, J [J.], Yu, W [W.], & Manocha, D. (2024). CoNVOI: Context-aware Navigation using Vision Language Models in Outdoor and Indoor Environments. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Schnitzer, B [Benjamin], Vural, U. C., Schnitzer, B [Bastian], Sardar, M. U., Fuerst, O., & Korn, O. (2024). Prototyping a Zoomorphic Interactive Robot Companion with Emotion Recognition and Affective Voice Interaction for Elderly People. *Proc. ACM Hum.-Comput. Interact.*, **8**(EICS). https://doi.org/10.1145/3660244

Seppelfelt, G. D., Asaka, T., Nagai, T., & Yukizaki, S. (2022). HumanoidBot: Full-Body Humanoid Chitchat System. In *21st International Conference on Humanoid Robots (Humanoids),* Ginowan, Japan.

Serfaty, G. J., Barnard, V. O., & Salisbury, J. P. (2023). Generative Facial Expressions and Eye Gaze Behavior from Prompts for Multi-Human-Robot Interaction. In *UIST '23 Adjunct, Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology.* Association for Computing Machinery. https://doi.org/10.1145/3586182.3616623

Sievers, T., & Russwinkel, N. (2024). Interacting with a Sentimental Robot - Making Human Emotions tangible for a Social Robot via ChatGPT. In *2024 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO)*.

Statista Market Insights. (2024). *Generative AI - Worldwide*. www.statista.com/outlook/tmo/artificial-intelligence/generative-ai/worldwid. Accessed: 13.03.2025.

Strobel, G., Banh, L., Möller, F., & Schoormann, T. (2024). Exploring Generative Artificial Intelligence: A Taxonomy and Types. In T. Bui (Ed.), *Proceedings of the Annual Hawaii International Conference on System Sciences, Proceedings of the 57th Hawaii International Conference on System Sciences.* Hawaii International Conference on System Sciences. https://doi.org/10.24251/hicss.2024.546

Sun, S., Li, C., Zhao, Z., Huang, H., & Xu, W. (2024). Leveraging large language models for comprehensive locomotion control in humanoid robots design. *Biomimetic Intelligence and Robotics*, **4**(4), 100187. https://doi.org/10.1016/j.birob.2024.100187

Sundaravadivel, P., Roselyn, P. J., Narayanaswamy, V., Jeyaraj, V. I., Ramesh, A., & Khanal, A. (2024). Integrating Image-based LLMs on edge-devices for Underwater Robotics. In *SPIE Defense + Commercial Sensing, 2024, National Harbor, Maryland, United States* (Vol. 13034). https://doi.org/10.1117/12.3014446

Susarla, A., Gopal, R., Thatcher, J. B., & Sarker, S. (2023). The Janus Effect of Generative AI: Charting the Path for Responsible Conduct of Scholarly Activities in Information Systems. *Information Systems Research*, **34**(2), 399–408. https://doi.org/10.1287/isre.2023.ed.v34.n2

Tsushima, Y., Yamamoto, S., Ravankar, A. A., Luces, J. V., & Hirata, Y. (2025). Task Planning for a Factory Robot Using Large Language Model. *IEEE Robotics and Automation Letters*, **10**(3), 2383–2390. https://doi.org/10.1109/LRA.2025.3531153

Varley, J [J.], Singh, S [S.], Jain, D [D.], Choromanski, K [K.], Zeng, A., Chowdhury, S., Dubey, A., & Sindhwani, V [V.] (2024). Embodied AI with Two Arms: Zero-shot Learning, Safety and Modularity. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* Abu Dhabi, United Arab Emirates.

Vemprala, S. H., Bonatti, R., Bucker, A., & Kapoor, A. (2024). ChatGPT for Robotics: Design Principles and Model Abilities. *IEEE Access*, **12**, 55682–55696. https://doi.org/10.1109/ACCESS.2024.3387941 vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing The Giant: On The Importance Of Rigour In Documenting The Literature Search Process. *ECIS 2009 Proceedings*.

Wang, Z [Zining], Reisert, P., Nichols, E., & Gomez, R. (2024). Ain't Misbehavin' - Using LLMs to Generate Expressive Robot Behavior in Conversations with the Tabletop Robot Haru. In *HRI '24, Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 1105–1109). Association for Computing Machinery. https://doi.org/10.1145/3610978.3640562

Webster, J., & Watson, R. (2002). Analyzing The Past To Prepare For The Future: Writing A Literature Review. *Management Information Systems Quarterly*, **26**(2).

Wiedebach, G., Bertrand, S., Wu, T., Fiorio, L., McCrory, S., Griffin, R., Nori, F., & Pratt, J. (2016). Walking on partial footholds including line contacts with the humanoid robot atlas. In *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids),* Cancun, Mexico.

Wu, J [Jimmy], Antonova, R., Kan, A., Lepert, M., Zeng, A., Song, S., Bohg, J., Rusinkiewicz, S., & Funkhouser, T. (2023). TidyBot: Personalized Robot Assistance with Large Language Models. In *International Conference on Intelligent Robots and Systems (IROS),* Detroit, MI, USA.

Yoshida, T [Takahide], Masumori, A., & Ikegami, T. (2023). *From Text to Motion: Grounding GPT-4 in a Humanoid Robot "Alter3"*. https://arxiv.org/abs/2312.06571.

Zhang, X., He, R., Tong, K., Man, S., Tong, J., Li, H., & Zhuang, H. (2024). Complex Motion Planning for Quadruped Robots Using Large Language Models. In *2024 IEEE International Symposium on Circuits and Systems (ISCAS), Singapore.* https://doi.org/10.1109/ISCAS58744.2024.10558349

Zhu, K., & Zhang, T [Tao] (2021). Deep reinforcement learning based mobile robot navigation: A review. *Tsinghua Science and Technology*, **26**(5), 674–691. https://doi.org/10.26599/TST.2021.9010012