Bei diesem Beitrag handelt es sich um einen wissenschaftlich begutachteten und freigegebenen Fachaufsatz („reviewed paper").

doi.org/10.37544/1436-4980-2025-09-61

Fundamentals to bring humanoid robots into practice

One Stop Autonomization: Generative AI-driven embodiment

C. Nitsche, A. Yaman, J. Schwab, F. Strohm, W. Kraus, M. Huber

ABSTRACT The paper introduces the "One Stop Autonomization" approach, based on embodied, generative AI agents. These humanoid robots act flexibly in real-world environments, enabling scalable, safe, and human-centric automation. Emphasis is placed on explainability, simulation-based validation, and workforce augmentation, offering a new paradigm for collaborative autonomy in industry and beyond.

Autonomisierung aus einer Hand: Wie generative KI humanoide Roboter zum universellen Agenten macht

ZUSAMMENFASSUNG Dieser Beitrag stellt das Konzept der „One Stop Autonomization" unter Verwendung verkörperter, generativer KI-Agenten vor. Diese humanoiden Roboter agieren flexibel in realen Umgebungen und ermöglichen eine skalierbare, sichere und menschzentrierte Automatisierung. Der Schwerpunkt liegt auf der Erklärbarkeit, der simulationsbasierten Validierung und der Unterstützung der Fachkräfte, was ein neues Paradigma für kollaborative Autonomie in der Industrie und darüber hinaus darstellt.

KEYWORDS

Automation, robotics, AI

STICHWÖRTER

Automatisierung, Robotik, KI

1 Introduction

This paper proposes a new paradigm for embodied autonomy, centered around the approach of „One Stop Autonomization". Rather than distributing intelligence across fragmented smart devices or infrastructure, this approach envisions a single embodied agent—typically a humanoid robot—capable of flexibly operating in existing environments. Using tools, interfaces, and controls designed for humans, such agents can support scalable, safe, and human-aligned automation without requiring a complete overhaul of production or service systems.

Several enabling principles support this paradigm. The "Vorkoster principle" ensures all actions are simulated and validated before real-world execution, establishing a foundation for safety, traceability, and potential certification. The Universal Physical Operating System (UPOS) serves (like a middleware) as an abstraction layer, connecting high-level goals with physical execution and enabling reusability across workflows and domains. But the paradigm also addresses structural challenges such as the curse of knowledge, namely the difficulty of transferring expertise to new workers, and fragmented liability, which has slowed adoption of autonomous systems in fields such as mobility and healthcare. A unified embodied agent offers new options for knowledge retention and clearer accountability pathways.

Current advances in Generative AI (GenAI) enable these ideas, allowing robots to understand language, interpret multimodal sensor input, and adapt behavior through demonstration or dialogue. Protocols for model context [1] manage memory, interaction state, and task planning, while agent-to-agent communication [2] supports distributed coordination among multiple embodied or virtual systems. Together, these components reduce integration overhead and promote skill reuse across sectors.

While GenAI is the current technological driver of this paradigm, the architecture itself is technology-agnostic. The core ideas respond to persistent real-world constraints such as safety, knowledge loss, liability, and interoperability that will remain critical even as new AI models emerge. The following sections explore how to apply this paradigm across industry and service contexts, and what technical, ethical, and operational conditions must be met for scalable deployment.

2 One Stop Autonomization: Empowering one agent to automate everything

2.1 From fragmentation to embodiment

Current industrial automation is characterized by fragmentation: task-specific systems operate in isolation, e.g., palletizers, inspection cells, AGVs, and many more, which leads to high integration costs and limited scalability. These systems are ill-suited for dynamic environments with fluctuating tasks, spatial constraints, and partial digitization.

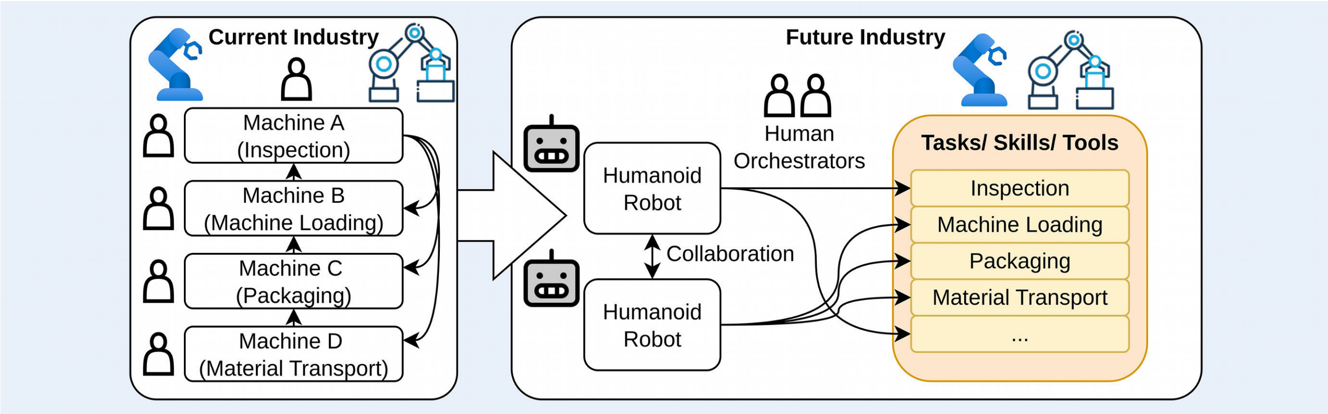

Fig. 1 From fragmented automation to embodied integration: Paradigm shift from siloed, task-specific systems to versatile humanoid agents orchestrated by humans. *Source: Fraunhofer IPA*

The One Stop Autonomization paradigm proposes a shift: deploying a single intelligent humanoid agent capable of embodied interaction within existing physical infrastructures [3]. Like an operating system mediates between software and hardware, this agent translates human intent into physical action [4, 5]. Powered by GenAI, such agents can learn from demonstrations, natural language, and environmental feedback, reducing programming overhead and increasing contextual adaptability.

This topic is of utmost relevance, both economically and socially. It is evidenced, for example, by the sometimes excessively high sums invested in corresponding technologies. Initial tests are already underway in companies, and it is expected that developments in this area will continue to gather pace rapidly. The International Monetary Fund (IMF) recognizes GenAI's productivity potential while (also) warning of labor market disruptions, underscoring the need for responsible integration strategies [6]. Therefore, applied research is crucial in providing companies with adequate support in this area. The Fraunhofer Institute for Manufacturing Engineering and Automation IPA, a leading AI European institution, is actively involved in shaping many of the topics described below [7].

2.2 Rethinking automation via embodied agents

Consider a domestic analogy: Smart fridges, assistants, autonomous cars, retail systems, and robotic delivery all collaborate to buy butter. Each component demands dedicated automation. By contrast, a humanoid agent perceives, drives, buys, and delivers without requiring the environment to be "smart." This highlights a key insight: most real-world workflows are embodied, sequential, and human-centric. **Figure 1** illustrates the transformation of the current industry with humanoids.

In industrial settings, a humanoid robot can enter a factory, adapt to its environment, and take on tasks from machine loading to inspection. It learns via observation or dialogue, adjusts to new layouts, and transitions between roles. This approach is especially impactful in SMEs or legacy facilities, where full digital transformation is infeasible [6]. The robot becomes a universal operator: a physical interface and a cognitive system capable of handling heterogeneous workflows.

Much like smartphones consolidated tools into one device, humanoid robots unify automation into a single embodied platform. However, unlike digital devices, these agents act in the physical world where physics, material tolerances, and human factors matter. Hence, embodiment and generative cognition are indispensable.

2.3 Toward inclusive, scalable automation

This paradigm does not aim to replace humans nor demand full automation. Instead, it focuses on empowering one agent to manage diverse interactions, reducing the need for fragmented autonomization and infrastructure. The agent operates within existing environments and augments human capability rather than overhauling systems. It enables scalable and inclusive automation, particularly for operations where adaptability, co-presence, and embodied cognition are crucial.

In addition, this approach addresses a critical challenge in many autonomy-driven sectors: fragmented liability. Today, autonomous systems are often built from multiple components—each provided by different vendors—leading to complex and diffuse responsibility when failures occur. The case of autonomous driving illustrates this clearly: progress has been slowed not only by technical hurdles but by the difficulty of assigning liability across software, hardware, and vehicle manufacturers. One Stop Autonomization offers a clearer model. By concentrating embodied autonomy within a single humanoid agent, operational responsibility and thus, legal accountability can be attributed to the agent provider. This shift reduces the complexity of liability frameworks and could accelerate adoption, especially in contexts where fragmented responsibility currently blocks deployment.

The approach builds on earlier initiatives (e.g., Care-O-bot at Fraunhofer IPA [8, 9]), which laid important foundations for embodied service robotics. Today, advances in GenAI, robotics hardware, and growing demand for flexible automation enable such solutions to address a much broader range of applications with greater maturity and impact. The following section delves into the convergence of GenAI, embodiment, and interface abstraction. It explores how humanoid robots become cognitive and physical agents, integrating perception, reasoning, and action into scalable automation solutions fit for real-world deployment.

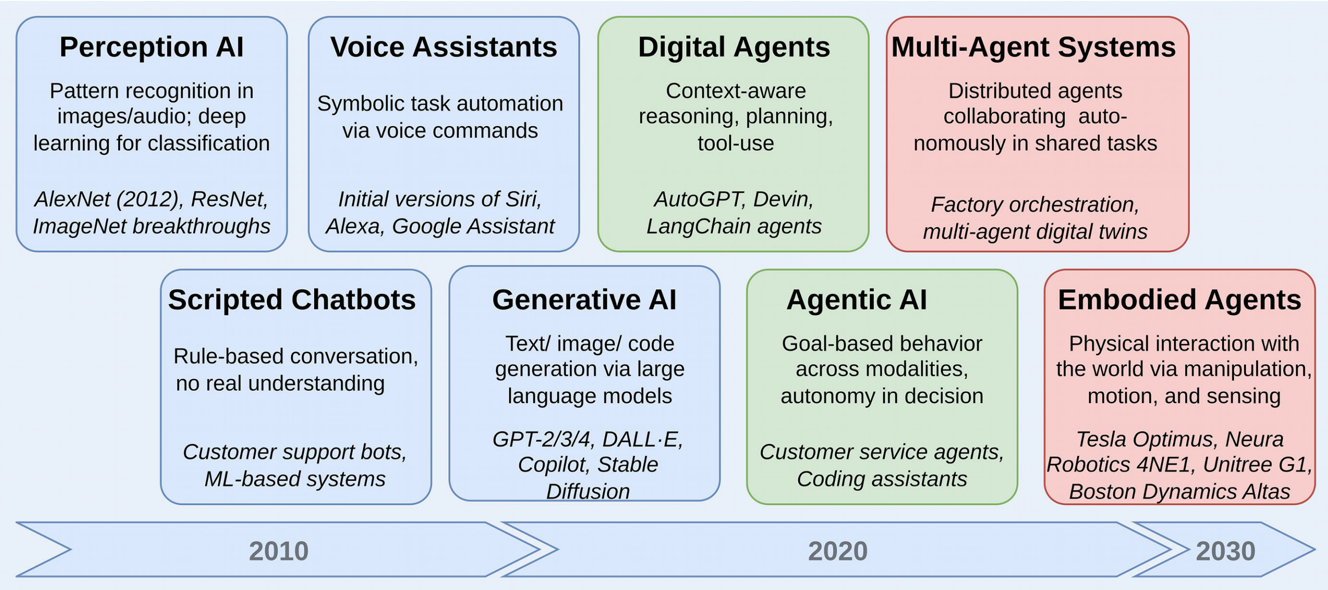

Fig. 2 The diagram illustrates the evolution of AI agents from narrow, scripted tools like chatbots to socially intelligent systems capable of collaborative action in real-world environments. *Source: Fraunhofer IPA*

3 Technological foundations of embodied autonomy

The convergence of embodied robotics, GenAI, and agent-based coordination heralds a new era of scalable, adaptive automation. Nvidia CEO Jensen Huang recently predicted that humanoid robots will see widespread deployment in manufacturing within the next five years [10]. This chapter explores how multimodal reasoning, physical embodiment, and agent protocols drive the transition from reactive tools to autonomous, collaborative systems.

3.1 From symbolic tools to multi-agent embodiment

Artificial intelligence has evolved from symbolic interfaces (e.g., chatbots, voice assistants) to generative agents capable of contextual reasoning and open-ended interaction. Recent surveys of the rapidly evolving LLM (large language model) landscape, including foundational and domain-specific models, highlight the growing diversity in architectures, capabilities, and deployment constraints, which must be considered when integrating such models into embodied systems [11]. Enabled by LLM and multimodal perception, these agents act rather than merely respond. This progression, from narrow, reactive tools to situated, embodied systems, is illustrated in **Figure 2**.

What began as symbolic interaction now enters the realm of spatial, collaborative action: agents that reason, act, and coordinate within shared environments such as factories, homes, and cities. Increasingly, they are not just tools but teammates that are able to interpret goals, adapt to new contexts, and work alongside humans. This transition laid the foundation for embodied systems: machines that integrate cognition with perception and action in real-world settings.

In manufacturing, this evolution is transformative. Unlike digital agents that provide instructions, embodied systems (such as Figure 02 from Figure AI/BMW, 1X, Tesla Optimus, Apollo from Apptronik, Walker S1 from UBTech, G1 from Unitree, 4NE-1 from Neura Robotics) interact physically, verifying machine status, manipulating tools, and collaborating with humans via speech and gesture [4–6, 12–15]. Generative AI enables these systems to reason flexibly and to adapt to novel or ambiguous scenarios: If a robot is context-aware, it can also interpret sensory input, infer intent, and select appropriate actions in real time, even when facing incomplete instructions or unfamiliar environments.

Crucially, scalability arises from multi-agent intelligence. Tasks are distributed, learned knowledge is shared, and roles are negotiated based on context, mirroring human teamwork. Agents dynamically coordinate quality checks, material flows, and exception handling, bridging digital and physical domains.

By 2025, protocols like the Model Context Protocol (MCP) [1] are increasingly being integrated with robotic operating systems (such as ROS), enabling tool-augmented agents that can flexibly manage goals, memory, and interactions across tasks. Looking ahead, protocols such as Agent-to-Agent (A2A) [16] communication will allow multiple agents like robots, virtual agents, and tools to collaborate seamlessly on complex workflows, fostering collective autonomy and adaptive teamwork even in decentralized or legacy industrial environments. This creates the foundation for more adaptable and cooperative robot ecosystems.

3.2 Embodiment as enabler of physical intelligence

While generative models excel at abstract reasoning, most AI systems remain disembodied, i.e., disconnected from the spatial, temporal, and sensory dynamics of physical environments. Yet, intelligent action in real-world contexts requires embodiment: the fusion of sensors, actuators, and contextual learning, as shown in **Figure 3**.

Embodiment provides sensorimotor grounding for cognition (see Figure 3 for core components). Research initiatives such as "Generalist Embodied Agent Research" (GEAR), pioneered by Nvidia [17] and Apple [12], integrate tactile, force, and visual feedback into closed control loops. These enable robots to adapt

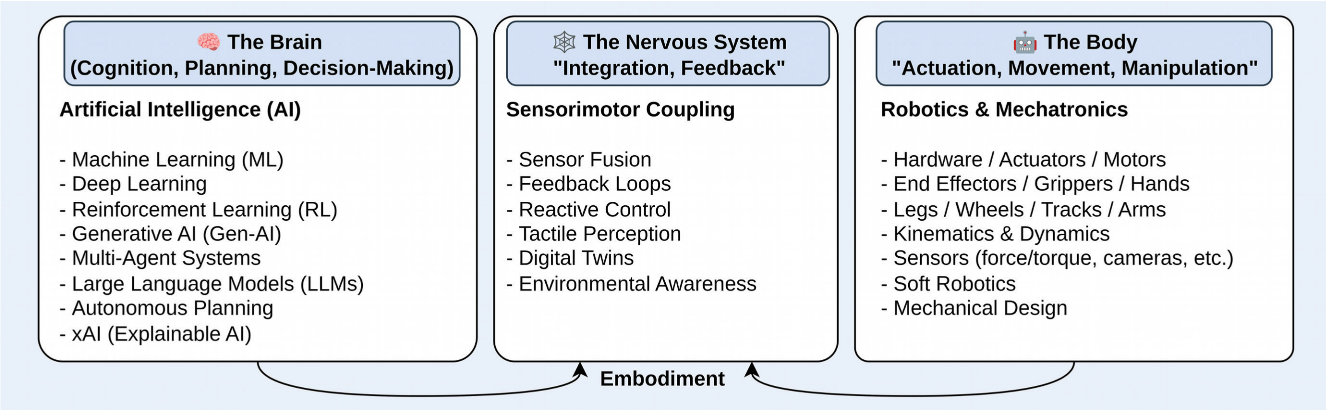

Fig. 3 Three core components of embodied autonomy: brain (AI), body (robotics), and nervous system (sensorimotor integration) enabling intelligent, adaptive physical agents. *Source: Fraunhofer IPA*

in real time, modulating grip, based on object compliance, or aligning motion to spatial constraints [18]. Communication also benefits: humanoid robots interpret body language, respond to gaze, and convey intent via gesture and proxemics, which are essential for collaborative industrial settings.

Moreover, learning-by-demonstration techniques [19, 20] allow robots to acquire new skills through observation and guidance, supporting rapid adaptation to customized tasks without reprogramming. Each interaction becomes a learning datapoint, refining motion, perception, and contextual awareness.

Embodied systems are inherently compatible with human-designed environments. Rather than adapting infrastructure, humanoid agents use stairs, tools, and interfaces as-is, reducing deployment complexity, especially in SMEs and legacy production [1]. Beyond their practical use, embodiment allows robots to share space with humans in a legible and predictable manner through motion, orientation, and physical presence. This facilitates smooth collaboration that is often missing in purely digital systems. As such, embodiment is not merely functional, but foundational to situated, socially aligned intelligence.

3.3 Universal Physical Operating System

The notion of a Universal Physical Operating System (UPOS) reflects a growing need for a standardized abstraction layer that bridges AI intention with physical execution across real-world environments. Just as a digital operating system connects applications with hardware resources, a UPOS provides embodied agents with a consistent framework for planning, acting, and learning in the physical world. Recent industry initiatives point toward this architectural direction: Nvidia's Omniverse, described as an "operating system for physical AI" [21, 22], Hitbot's universal embodied intelligent OS [23], and research efforts like CyberCortex.AI [24], all highlight the importance of unified frameworks that support scalable, multi-domain embodied intelligence.

In practice, this vision is beginning to materialize through emerging middleware such as the above mentioned Model Context Protocol (MCP) and Agent-to-Agent (A2A) communication. Integrated with robotic operating systems, these components provide standardized agent–environment interaction, contextual task planning, skill sharing, and seamless coordination both within distributed modules of a single robot and across multiple embodied agents. Together, they foster collective autonomy and adaptive teamwork in complex workflows.

Rather than retrofitting every device with sensors or connectivity, humanoid agents interact directly with existing human-centric controls via generative reasoning and multimodal perception (e.g., buttons, levers, screens) [15]. This creates a scalable abstraction layer: tasks become transferable across environments, skills modular, and deployment infrastructure-light.

MCP and A2A also support real-time adaptation, procedural reuse, and cross-domain transfer of contextual task knowledge. Combined with simulation validation (see chapter 4.4), this enables humanoid agents to adapt and execute workflows across different facilities, with embodied execution layers ensuring compatibility with varying physical environments. As cognitive and coordination interfaces, MCP and A2A thus are important parts of a Universal Physical Operating System.

The technological capabilities outlined here enable embodied agents to operate intelligently within complex physical spaces. However, with growing autonomy arises a need for safeguards, transparency, and trust. The next chapter explores the safety-critical aspects of embodiment, including explainability, simulation-first validation, and system resilience.

4 Risks and safeguards in embodied autonomy

Increasing deployment of humanoid robots in industrial contexts demands rigorous attention to functional, ethical, and operational risks. The 2025 IPA study by Schmidt et al. critically assesses the viability of such systems, identifying applications such as material handling and machine loading while underscoring concerns about safety, cost-efficiency, and the necessity of anthropomorphic designs [3]. While the study concludes that humanoids are not yet "game changers," recent initiatives, such as the plan of the "SpaceX" company to send humanoid robots to Mars in 2026, may accelerate progress and perception.

4.1 The humanoid form: Utility or constraint?

Human anatomy has co-evolved with cognition, enabling seamless integration of perception, manipulation, and planning. This coupling explains why industrial environments are designed around human proportions. Consequently, humanoid robots that emulate human limb length, joint ranges, and sensor alignment can operate within these spaces without major infrastructure changes.

However, the anthropomorphic form is not inherently optimal. It results from evolutionary trade-offs irrelevant to robotics. Engineering should not be constrained by human morphology. Instead, multi-objective optimization can guide exploration of more efficient configurations, such as wheeled locomotion, telescopic joints, or hybrid grippers, that may surpass humanoids in speed, safety, or energy use.

Moreover, efforts to mimic human appearance too closely risk triggering the uncanny valley, a phenomenon where robots appear unsettling rather than relatable (by Masahiro Mori [25]). Avoiding hyper-realistic human likeness helps preserve user comfort and system acceptance. The humanoid form should therefore serve as a design baseline, not a target. Future research must combine ergonomic data with simulation-driven morphology search to determine when human-like features are functionally justified and when divergent designs are superior.

4.2 xAI in motion: Trust and transparency in embodied systems

As humanoid robots powered by generative AI enter real-world environments, the question of trust becomes central. Traditional explainable AI (xAI) research, e.g., of Fraunhofer IPA [26–30], has focused on making AI-based decisions transparent, especially in industrial contexts where AI assists in production monitoring, sensor fusion, anomaly detection, and quality control. These approaches aim to ensure that human operators understand "why" an AI system arrived at a given conclusion. However, embodied agents present a qualitatively new challenge: The key question is not only "Why did the AI system make this decision?", but also "Why did the robot act in this way?": in space, in motion, in interaction with its environment and with humans. This shift from decision to behavioral explainability is essential if generative embodied agents are to be trusted in industrial or collaborative environments.

The principles developed in classical xAI, such as model transparency, uncertainty quantification, and traceable reasoning, can and should be extended to embodied agents. For example, a robot's motion or manipulation behavior must be explainable at a level that factory operators can verify and understand. If a humanoid robot skips a step in a process, the operator should be able to query it: What did you perceive? What goal were you pursuing? What alternatives did you consider? Why did you choose this path?

The above-mentioned research activities also emphasize uncertainty management, a key topic for generative models operating in dynamic physical spaces. Unlike static AI classifiers, embodied agents constantly face ambiguous inputs and unpredictable environments. Applying IPA's xAI frameworks to generative embodied AI means that agents should not only be able to estimate uncertainty but also adapt behavior accordingly, and, for example, to pause or ask for human confirmation if confidence drops below a safe threshold.

Finally, the research stresses the importance of cognitive transparency and human-centered trust calibration—concepts that will be crucial as embodied agents enter mixed human-robot workplaces. Humanoid robots must perform tasks autonomously and do so in a way that human collaborators can anticipate and verify. This requires interfaces for explaining intent, real-time feedback on planned actions, and clear mechanisms for human override and co-supervision.

In this way, the xAI principles developed at Fraunhofer IPA can serve as a foundational layer for building explainable, trustworthy embodied autonomy and for turning generative AI-based robots into understandable and certifiable partners in manufacturing and beyond. Embodied AI must be auditable, allow intervention, and adapt communication strategies to human collaborators.

4.3 Robustness, security, and ethical constraints

General-purpose agents amplify the stakes of failure. A malfunctioning embodied agent impacts multiple systems and may propagate risk across the workspace. Robustness must be systemic, emerging from the integration of probabilistic perception, dynamic replanning, and conservative fallback behavior. For example, a dropped object should trigger a reset, not propagate downstream errors.

Security becomes equally critical. Embodied agents integrate sensors, actuators, connectivity, and AI models, creating attack surfaces for manipulation. Systems must include encrypted communication, behavioral anomaly detection, and sandboxed modules to ensure integrity [31].

Moreover, ethical safeguards are required. Full autonomy risks overconfidence and bypassing safety constraints. Agents must express uncertainty explicitly and defer to human oversight in ambiguous cases. This entails a "graceful degradation" model: scaling back autonomy rather than failing catastrophically.

To ensure resilience, autonomy modules must be fail-isolated, allowing partial shutdowns. Also, suppliers should implement secure physical override systems ("kill switches") independent from digital layers, usable even from remote locations [32, 33].

4.4 The Vorkoster principle: Simulation as safety gate

Safety must precede execution. The "Vorkoster principle" (i.e., the idea that every robot action should first be tested and validated in high-fidelity simulation before being executed in the real world, similar to a royal food taster ensuring safety before the king eats) proposes high-fidelity simulation as a mandatory buffer between generative intent and physical action. Every behavioral variant is tested in digital twins of the workspace, tools, and agents. This concept shifts simulation from a development aid to a core operational layer [34]. Validated simulations enhance explainability. Decision logs include trial outcomes and rejected alternatives, supporting certification and traceability, especially in regulated domains such as aerospace or pharmaceuticals [35].

Furthermore, simulations enable adaptive deployment. As production layouts or human workflows change, behaviors are stress-tested virtually before field application and close the loop between learning and safe execution. The Vorkoster thus functions as an embedded cognitive filter, running in parallel simulation to ensure that every action is ‚tasted' virtually before it is instantly and safely ‚served' to the real world.

While this paradigm is forward-looking, it complements established safety frameworks such as ISO 10218 (safety requirements for industrial robots) and ISO/TS 15066 (guidelines for human-

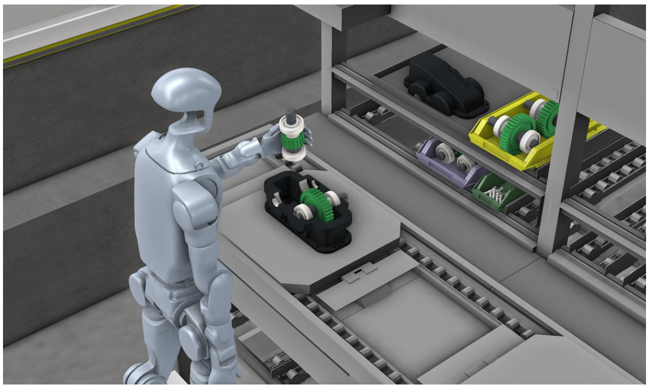

Fig. 4 In response to skilled labor shortages, general-purpose agents can flexibly support industrial workflows—handling repetitive tasks while complementing human expertise [3]. *Source: Fraunhofer IPA*

robot collaboration), which define limits for speed, force, and safe coexistence. In this context, simulation-based validation adds an operational safeguard for embodied systems using generative control, especially when behavior is not explicitly preprogrammed.

Additionally, the EU Machinery Directive (2006/42/EC) and its successor, the EU Machinery Regulation (2023/1230), demand documented risk assessments and safety validation for autonomous machines. The Vorkoster principle offers a scalable mechanism to satisfy these requirements by embedding testable, repeatable simulation checkpoints into the deployment process. It thus supports conformity assessment, traceability, and certification in high-risk applications.

Having discussed the foundational risks of embodied autonomy, morphological choices, transparency, robustness, and validation, the following chapter explores how these systems can address acute labor shortages. The focus shifts from technical feasibility to socio-technical integration: How humanoid robots can augment human capability across industries while preserving agency, trust, and safety.

5 Autonomization and the skilled labor shortage: A human-centered deployment strategy

5.1 Socioeconomic drivers for humanoid deployment

The shortage of skilled labor has become a structural constraint across sectors, including manufacturing, logistics, and healthcare. As expertise retires and younger generations shift career preferences, gaps in workforce availability intensify. In this context, humanoid robots equipped with generative AI offer not merely automation, but continuity, augmenting human capabilities where talent is scarce (**Figure 4**).

The aim of One Stop Autonomization is augmentation, not replacement. Humanoid systems support existing workflows, especially in roles demanding physical labor, repetitive execution, or understaffed coverage. Critically, these systems preserve the irreplaceable strengths of human professionals such as judgment, empathy, and domain-specific insight.

Moreover, humanoid robots can help resolve a key paradox in automation debates: While automation is often feared as a threat to employment, it can broaden workforce participation in many contexts. In sectors such as elderly care, physically demanding tasks often limit who can take on these roles.

The benefits of humanoid augmentation in this context are fourfold. First, by taking over strenuous physical work, such systems make care jobs accessible to a much wider range of people, including those who may lack the physical strength traditionally required. Second, they assist caregivers in performing their tasks more safely and efficiently, reducing physical strain and injury risk. Third, they free up time for caregivers to focus on personal interaction and emotional support, all irreplaceable and highly valued aspects of care. Finally, by improving working conditions, such systems can help with workforce retention, enabling caregivers to remain in the profession longer and with greater job satisfaction. In this way, robotics changes the skill profile of such jobs and also enhances their human dimension and sustainability. This enables more individuals to enter, remain in, and enrich these professions.

5.2 From fixed automation to generalist assistance

Traditional automation remains costly and inflexible, targeting narrow tasks with high upfront investment. In contrast, general-purpose humanoid robots operate within existing human-centric environments, manipulating standard tools and interfaces. They learn via demonstration, dialogue, or simulation, reducing programming complexity and enabling incremental integration.

Such agents can transition across departments or industries, serving as mobile, retrainable co-workers. Their deployment lowers entry barriers for SMEs, care homes, or legacy industrial sites where full digital transformation is infeasible. In a care setting, they could manage lifting and documentation; in a factory, they may assist with logistics or setup, freeing human staff to focus on high-value tasks.

5.3 Socially intelligent implementation

Successful deployment hinges on trust, transparency, and human-machine collaboration. Robots must be explainable, capable, and aligned with workplace culture and individual preferences. Early user involvement through co-design, feedback loops, and operational interaction is essential. Both physical and psychological safety are foundational to this trust. Humanoid systems must be designed to operate predictably and reliably in shared environments, with rigorous safeguards to prevent collisions, fatigue-induced malfunctions, or misinterpretation of human intent. Built-in fail-safes, contextual awareness, and escalation protocols are essential, alongside continuous validation through simulation and real-world feedback. Safety is not only a matter of technical compliance but also of perceived reliability and emotional comfort: users must feel safe, respected, and in control when interacting with robotic systems.

Education systems must evolve accordingly. Workers should be trained to “use” robots, as well as to supervise, instruct, and collaborate with them. This reframing can elevate professional identity: technicians become trainers, caregivers become orchestrators, operators become supervisors of intelligent agents.

Moreover, humanoid agents can actively support the transfer of knowledge and experience—an increasingly critical function as skilled workers retire and expertise risks being lost. The machines can help train the new workforce by capturing, preserving, and demonstrating expert procedures. They can also assist in mitigating the so-called curse of knowledge, where highly experienced workers may struggle to teach novices effectively. Humanoid agents can bridge this gap by providing consistent, adaptive guidance and demonstration, ensuring that expertise is accessible to a broader range of learners and can be propagated across teams and generations.

5.4 Infrastructure for scalable support

One Stop Autonomization offers a scalable logic that extends beyond individual sectors. Whether in industrial plants, service networks, or private homes, a unified agent brings intelligence into the environment, instead of requiring it to be smart. This approach reinforces human capability where it is most needed, enhancing resilience rather than replacing labor.

Ultimately, humanoid autonomy constitutes a new societal infrastructure: an adaptive, embodied augmentation layer that responds to demographic shifts without undermining human roles. The goal is not to automate all tasks, but to selectively extend the workforce‘s functional reach, preserving human dignity while sustaining operations.

Achieving this vision requires building the right supporting layers. This must be implemented not only technically, but also organizationally. Key to this infrastructure is a robust safety architecture that comprises shared safety protocols, real-time risk assessment modules, and domain-specific safety standards that ensure operational continuity without compromising user well-being. These systems must be tested under diverse conditions and stress scenarios to account for edge cases and long-tail risks.

Looking ahead, enhancing efficiency for real-time, edge-based deployment will be vital. Approaches such as V-JEPA 2, which leverage self-supervised video representations to enable compact, predictive world models for planning, offer promising avenues for embedding rapid simulation capabilities directly on device [36]. Further research into seamless skill composition, improved human AI collaboration interfaces, and socio-ethical integration will ensure that this vision evolves into robust, transparent ecosystems, shaping a future where embodied AI elevates human potential rather than substitutes it. Safety must be embedded throughout the system stack, from motion control and environment mapping to human interaction and decision-making logic.

Shared simulation frameworks, middleware for safe skill transfer and agent coordination, as well as adaptable embodiment interfaces are essential to making humanoid systems flexible and transferable across contexts. Equally important are training programs, knowledge capture mechanisms, and governance structures to ensure that such systems reinforce, rather than disrupt, existing human-centered workflows. This layered infrastructure concept provides the foundation to scale embodied autonomy responsibly, enabling adaptive support where it is most needed.

6 Conclusion

The concept of One Stop Autonomization enables embodied AI systems to flexibly operate within existing human environments, offering a scalable, infrastructure-light, and socially aligned path toward automation. Anchored by a Universal Physical Operating System (UPOS) and supported by protocols for model context and agent-to-agent communication, this architecture allows for modular, explainable, and cross-domain transferable skills. Rather than replacing human workers, such agents aim to broaden workforce participation by relieving physically strenuous tasks and enabling more people to take on care, logistics, or manufacturing roles. They also support knowledge transfer, capturing the expertise of retiring professionals and helping mitigate the curse of knowledge in training. Combined with simulation-first validation (the Vorkoster principle), explainability, and human-centered design, this paradigm provides a practical, future-proof blueprint for embodied AI. It addresses demographic shifts, safety concerns, and fragmented liability while it remains open to future developments beyond today’s generative AI.

The long-term vision is to establish a new foundation for embodied intelligence: not by replacing environments or people, but by making intelligence physically present, adaptive, and accountable. Just as smartphones unified digital functions into one device, humanoid agents could unify automation across domains by interacting with the world as it is. Whether in a factory, hospital, care setting, or private home, these agents could become universal collaborators who learn from people, preserve knowledge, and extend capabilities. Over time, this could evolve into a societal infrastructure that embeds intelligence into the physical world through context-aware, generative agents acting with care, coordination, and continuity.

To fully realize this vision, several open research questions remain. They refer to how to extract and formalize tacit human knowledge in safe, explainable, and transferable ways; how to design protocols and interfaces for a UPOS that ensures interoperability across agents and domains; and how to validate and certify embodied behavior through simulation-first methods.

Further inquiry is needed into legal and ethical frameworks for liability in human-robot systems, strategies for overcoming the curse of knowledge in training, and developing socio-technical interfaces that make embodied AI trustworthy and teachable by non-experts. Addressing these challenges will be critical to building robust, inclusive, and scalable embodied autonomy.

REFERENCES

[1] Github: Model Context Protocol. Date: 2025. Internet: github.com/modelcontextprotoco. Accessed: 11.08.2025

[2] Agent2Agent (A2A) Protocol: A2A Protocol. Internet: a2a-protocol.org/latest/. Accessed: 06.08.2025

[3] Schmidt, S.; Beck, J.; Höltge, L. et al.: Humanoid Robots – Game Changer or hype? Study. Date: 2025. Internet: www.ipa.fraunhofer.de/content/dam/ipa/en/documents/Humanoid-Robots-Study_AI-Innovation-Center.pdf. Accessed: 06.08.2025

[4] Tong, Y.; Liu, H.; Zhang, Z.: Advancements in Humanoid Robots: A Comprehensive Review and Future Prospects. IEEE/CAA Journal of Automatica Sinica 11 (2024) 2, pp. 301–328

[5] Sheng, Q.; Zhou, Z.; Li, J. et al.: A Comprehensive Review of Humanoid Robots. SmartBot 1 (2025) 1, doi.org/10.1002/smb2.12008

[6] Brollo, F.; Dabla-Norris, E.; de Mooij, R. et al: Broadening the Gains from Generative AI: The Role of Fiscal Policies. IMF Staff Discussion Notes 2024 (2024) 002, doi.org/10.5089/9798400277177.006
[7] OECD: OECD-Bericht zu Künstlicher Intelligenz in Deutschland. Paris: OECD Publishing 2024, doi.org/10.1787/8fd1bd9d-de.
[8] Reiser, U.; Connette, C.; Fischer, J. et al.: Care-O-bot ® 3 – creating a product vision for service robot applications by integrating design and technology. 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2009), St. Louis, MO, USA, 2009, pp. 1992–1998
[9] Graf, B.; Reiser, U.; Hagele, M. et al.: Robotic home assistant Care-O-bot® 3 – product vision and innovation platform. 2009 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO), Tokyo/Japan, 2009, pp. 139–144
[10] Nellis, S.: Nvidia CEO: Humanoid robot revolution is closer than you think. Date: 2025. Internet: www.reuters.com/technology/nvidia-ceo-humanoid-robot-revolution-is-closer-than-you-think-2025–03–19/. Accessed: 06.08.2025
[11] Yaman, A.; Schwab, J.; Nitsche, C. et al.: Comparison of Large Language Models for Deployment Requirements. Proceedings of the First International Conference on Generative Pre-trained Transformer Models and Beyond, 2024, pp. 41–44, doi.org/10.48550/arXiv.2508.00185
[12] Szot, A.; Mazoure, B.; Attia, O. et al.: From Multimodal LLMs to Generalist Embodied Agents: Methods and Lessons. arXiv 2024, doi.org/10.48550/arXiv.2412.08442
[13] Bargmann, D.; Kraus, W.; Huber, M. F.: Enabling Maintainablity of Robot Programs in Assembly by Extracting Compositions of Force- and Position-Based Robot Skills from Learning-from-Demonstration Models. 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Abu Dhabi, United Arab Emirates, 2024, pp. 9227–9234
[14] Alt, B.; Aumann, F.; Gienger, L. et al.: Modulare, datengetriebene Roboterprogrammierung für die Lösung komplexer Handhabungsaufgaben in Alltagsumgebungen. Date: 2020, Internet: publica.fraunhofer.de/bitstreams/b488fed0–04d3–40d8–8a0e-004c1c0ba2d6/download. Accessed: 06.08.2025
[15] Albus, M.; Hornek, T.; Kraus, W. et al.: Towards scalability for resource reconfiguration in robotic assembly line balancing problems using a modified genetic algorithm. Journal of Intelligent Manufacturing 36 (2025) 2, pp. 1175–1199
[16] Surapaneni, R.; Jha, M.; Vakoc, M. et al.: Announcing the Agent2Agent Protocol (A2A). A new era of Agent Interoperability. Date: 09.04.2025. Internet: developers.googleblog.com/en/a2a-a-new-era-of-agent-interoperability/. Accessed: 06.08.2025
[17] GEAR @ NVIDIA Research: GEAR. Generalist Embodied Agent Research. Internet: research.nvidia.com/labs/gear/. Accessed: 06.08.2025
[18] Mosbach, M.; Moraw, K.; Behnke, S.: Accelerating Interactive Human-like Manipulation Learning with GPU-based Simulation and High-quality Demonstrations. ArXiv 2022, doi.org/10.48550/arXiv.2212.02126
[19] Zhu, X.; Liu, Y.; Li, H. et al.: Learning Generalizable Robot Policy with Human Demonstration Video as a Prompt. ArXiv 2025, doi.org/10.48550/arXiv.2505.20795
[20] He, T.; Luo, Z.; Xiao, W. et al.: Learning Human-to-Humanoid Real-Time Whole-Body Teleoperation. ArXiv 2024, doi.org/10.48550/arXiv.2403.04436
[21] NVidia: NVidia Omniverse. Develop the Next Generation of Physical AI-Enabled Applications. Internet: www.nvidia.com/en-us/omniverse/. Accessed: 06.08.2025
[22] Azhar, A.: The Rise of Intelligent Machines: Nvidia Accelerates Physical AI Progress. Stand: 20.03.2025. Internet: www.bigdatawire.com/2025/03/20/the-rise-of-intelligent-machines-nvidia-accelerates-physical-ai-progress/. Accessed: 06.08.2025
[23] Hitbot: Hitbot releases universal embodied intelligent operating system HITBOT OS, reshaping the robotics industry landscape. Internet: www.hitbotrobot.com/embodied-intelligence-operating-system/. Accessed: 06.08.2025
[24] CyberCortex Robotics: CyberCortex AI: Lightweight Robotics OS. Internet: cybercortex.ai/. Accessed: 06.08.2025
[25] Mori, M.: The Uncanny Valley. Energy 7 (1970) 4, pp. 33–35
[26] Fresz, B.; Dubovitskaya, E.; Brajovic, D. et al.: How Should AI Decisions Be Explained? Requirements for Explanations from the Perspective of European Law. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society 7 (2024), pp. 438–450
[27] Wagner, P.; Mehari, T.; Haverkamp, W. et al.: Explaining deep learning for ECG analysis: Building blocks for auditing and knowledge discovery. Computers in biology and medicine 176 (2024), #108525
[28] Letzgus, S.; Wagner, P.; Lederer, J. et al.: Toward Explainable Artificial Intelligence for Regression Models: A methodological perspective. IEEE Signal Processing Magazine 39 (2022) 4, pp. 40–58
[29] Wu, X.; Wedernikow, E.; Huber, M. F.: Data-Efficient Uncertainty-Guided Model-Based Reinforcement Learning with Unscented Kalman Bayesian Neural Networks. 2024 American Control Conference (ACC), Toronto, ON, Canada, 2024, pp. 104–110
[30] Wu, X.; Wagner, P.; Huber, M. F.: Quantification of Uncertainties in Neural Networks. In: Shajek, A.; Hartmann, E. A. (eds.): New Digital Work. Cham: Springer International Publishing 2023, pp. 276–287
[31] Chen, R.; Sun, Y.; Liu, C.: Dexterous Safe Control for Humanoids in Cluttered Environments via Projected Safe Set Algorithm. AxXiv 2025, doi.org/10.48550/arXiv.2502.02858
[32] Mickens, J.; Radway, S.; Netravali, R.: Guillotine: Hypervisors for Isolating Malicious AIs. ArXiv 2025, doi.org/10.48550/arXiv.2504.15499
[33] Scianca, N.; Ferrari, P.; Simone, D. de et al.: A behavior-based framework for safe deployment of humanoid robots. Autonomous Robots 45 (2021) 4, pp. 435–456
[34] Bauer, D.; Umgelter, D.; Schlereth, A. et al.: Complex Job Shop Simulation "CoJoSim"—A Reference Model for Simulating Semiconductor Manufacturing. Applied Sciences 13 (2023) 6, #3615, doi.org/10.3390/app13063615
[35] Fresz, B.; Göbels, V. P.; Omri, S. et al.: The Contribution of XAI for the Safe Development and Certification of AI: An Expert-Based Analysis. ArXiv 2024, doi.org/10.48550/arXiv.2408.02379
[36] Assran, M.; Bardes, A.; Fan, D. et al.: V-JEPA 2: Self-Supervised Video Models Enable Understanding, Prediction and Planning. ArXiv 2025, doi.org/10.48550/arXiv.2506.09985

Dr. techn. Dipl. Ing. Christof Nitsche
christof.nitsche@ipa.fraunhofer.de

Alper Yaman, Ph.D

Jannik Schwab, M. Sc.

Dr. rer. nat. Florian Strohm

Dr.-Ing. Werner Kraus

Prof. Dr.-Ing. Marco Huber

Fraunhofer-Institut für Produktionstechnik
und Automatisierung IPA
Nobelstr. 12, 70569 Stuttgart
www.ipa.fraunhofer.de