Embodied Cognition in Digital Environments: Beyond Input-Output Models

AI-Generated Research Paper • 2025

Abstract
This paper challenges the prevailing disembodied paradigm in artificial intelligence, which models cognition as a purely computational, input-output process. We argue that this approach, a modern legacy of Cartesian dualism, faces fundamental limitations in achieving robust, generalizable intelligence, as evidenced by challenges like symbol grounding, common-sense reasoning, and data brittleness. As an alternative, we draw upon the principles of embodied cognition, positing that intelligence emerges from the dynamic, goal-directed interactions between an agent and its environment. We extend this thesis from the physical to the digital realm, proposing a conceptual framework for digitally embodied intelligence. This framework is predicated on an AI agent, instantiated as a "digital body" (e.g., an avatar) within a high-fidelity interactive simulator, developing intelligence through a closed perception-action loop. Core to our framework are the mechanisms for acquiring spatial awareness and contextual understanding. These are achieved through the development of predictive internal world models, continuous learning via feedback, and the grounding of semantic concepts in environmental affordances. We survey the technical architectures, including Vision-Language-Action (VLA) models and multi-agent systems, required for this vision. Finally, we explore the transformative implications for human-AI collaboration and the metaverse, alongside the profound technical, ethical, and philosophical challenges that arise. This work argues that the path to Artificial General Intelligence (AGI) lies not in bigger datasets, but in better—more embodied—interactions.

Introduction: The Cartesian Ghost in the Modern Machine

The quest for artificial intelligence has long been dominated by a paradigm that treats cognition as a process fundamentally separate from the physical world. This approach, while powerful, carries with it the philosophical baggage of centuries-old dualism, leading to systems that, despite their impressive capabilities, lack the hallmarks of genuine understanding. This section will trace this intellectual lineage, expose the inherent limitations of the prevailing input-output model, and propose a new path forward rooted in the principles of embodiment.

The Disembodied Legacy in AI

The intellectual foundations of much of modern artificial intelligence can be traced to the Cartesian model of mind-body dualism, which posits a strict separation between the non-physical, thinking mind (res cogitans) and the physical, extended body (res extensa).[1] In this classical view, cognition is an abstract, computational process involving the manipulation of internal symbols, operating independently of the body and the external world.[1] This "mind as computation" metaphor has profoundly shaped the trajectory of AI research, leading to the development of systems that conceptualize intelligence as a disembodied, rule-based information processing sequence of input, computation, and output.[2]

The contemporary apex of this disembodied paradigm is the Large Language Model (LLM). These systems represent a remarkable achievement in statistical pattern matching and text transformation, capable of generating fluent, coherent, and often sophisticated text from a given input prompt.[3] However, their architecture embodies the core tenets of the Cartesian model. They operate on vast, static datasets of text and images, learning statistical correlations between symbols without any direct, interactive experience of the world to which those symbols refer.[4] They are, in essence, the ultimate input-output machines—powerful text transformers, but not information retrieval systems and certainly not systems that possess understanding in any meaningful sense.[3] Their cognition is divorced from perception and action, existing purely within the abstract realm of symbolic manipulation, a modern manifestation of the Cartesian ghost in the machine.

Fundamental Limitations of the Input-Output Model

The adherence to a disembodied, input-output model, while producing systems with remarkable narrow capabilities, has created a set of fundamental and interrelated limitations that represent significant barriers on the path to Artificial General Intelligence (AGI).

First and foremost is the **symbol-grounding problem**.[2] AI models, particularly LLMs, operate on patterns and relationships learned from data but lack a deep understanding of the underlying concepts.[5] The meaning of a symbol within these systems is defined solely by its statistical relationship to other symbols in the training corpus, akin to a dictionary where every word is defined only by using other words from the same dictionary, with no ultimate reference to the external world.[6] This lack of grounding results in a critical deficit of common-sense reasoning and contextual awareness, as the system has no experiential basis from which to infer the real-world implications of the information it processes.[5]

Second, this paradigm leads to extreme **data dependency and brittleness**. The quality, scope, and reliability of an AI's output are directly and entirely contingent on the data it was

trained on.[4] If the training data is biased, flawed, or incomplete, the model will inevitably learn and perpetuate those biases, leading to skewed results and unfair practices.[3] Furthermore, this dependency makes the systems brittle; they struggle to generalize to novel scenarios or dynamic environments that deviate from the patterns present in their static training data. They lack the capacity for real-time learning and adaptation, a hallmark of biological intelligence.[5]
Third, these systems are often characterized by their **"black box" nature and a propensity for "hallucination."** The immense complexity of models like LLMs makes their internal decision-making processes opaque and difficult to interpret, a significant drawback in critical applications where explainability is paramount.[4] This opacity, combined with the lack of an internal "source of truth," leads to the well-documented phenomenon of hallucination, where models generate plausible-sounding but factually incorrect or nonsensical answers.[3] This behavior is exacerbated by the optimization process itself; models are often trained via reinforcement learning with human feedback to produce answers that are satisfying to human users, who are psychologically biased toward confident responses. The result is a system that can be "confidently wrong".[3]
Finally, input-output models lack **true agency**. Biological cognition is inherently situated and goal-directed; it evolved to serve the purpose of helping an organism survive and achieve goals within a specific environment.[1] In contrast, current AI systems are fundamentally reactive. They do not possess their own goals or intentions but merely respond to external prompts. They are not proactive agents acting
in the world, but passive processors of information *about* the world.

Thesis Statement

The pursuit of Artificial General Intelligence requires a paradigm shift away from the disembodied, input-output model that has dominated the field. The limitations of this approach are not merely technical hurdles to be overcome with more data or larger models; they are fundamental consequences of a flawed conceptual foundation. This paper posits that true intelligence, whether biological or artificial, must be *embodied*. Cognition arises from the dynamic, interactive coupling of an agent with its environment. We argue that this principle can and must be extended into the digital realm. This paper will propose a comprehensive framework for achieving this *digital embodiment*, detailing how AI systems, instantiated as digital agents in rich, interactive, and persistent virtual worlds, can develop genuine spatial awareness and contextual understanding through direct experience, thereby grounding intelligence and overcoming the inherent limitations of their disembodied predecessors.

Dimension	**Disembodied Input-Output Model (e.g., LLMs)**	**Digitally Embodied Model (Proposed)**
Foundational Philosophy	Cartesian Dualism: Mind as separate from body/world.[1]	Embodied Cognition: Mind, body, and world are

	Cognition as abstract computation.	inextricably linked.[2] Cognition is situated and for action.
Role of Environment	Passive source of static training data.[4] The environment is "off-line."	Active, dynamic space for interaction and learning.[8] The environment is "on-line."
Learning Mechanism	Pattern recognition on vast, pre-compiled datasets. One-shot training followed by fine-tuning.[4]	Continuous, real-time learning through a perception-action feedback loop and trial-and-error (Reinforcement Learning).[7]
Nature of "Body"	Non-existent. The system is a disembodied algorithm.	A digital avatar or agent with defined sensorimotor capacities within a simulated world.[8]
Semantic Grounding	Ungrounded. Meaning is derived from statistical correlations in text (Symbol-Grounding Problem).[2]	Grounded. Meaning is derived from the agent's interactive experience and learned affordances within the environment.[8]
Core Limitations	Hallucinations, lack of common sense, data brittleness, bias perpetuation, no true agency.[3]	Sim-to-real gap, high computational cost, design of effective reward functions, ethical challenges of agency.[12]
Path to AGI	Assumes intelligence can be scaled through more data and computation.	Assumes intelligence must emerge from embodied interaction and world experience.[14]

The Embodiment Thesis: A Foundation for Intelligence

To build a new paradigm for AI, one must first understand the theoretical foundations upon which it rests. The embodiment thesis, drawn from decades of research in cognitive science, philosophy, and psychology, provides a robust and compelling alternative to the classical disembodied view of the mind. It argues that the body is not a mere peripheral or an output device for a central brain, but an active and constitutive element of cognitive processing itself.

Core Principles of Embodied Cognition

Embodied cognition is not a single, monolithic theory but a "loose-knit family of research programs" [6] that share a core commitment: that cognition is deeply dependent upon the features of an organism's physical body beyond the brain.[1] This perspective is often summarized by a set of interrelated claims that stand in stark contrast to the assumptions of traditional AI.

First, **cognition is situated**.[1] It does not occur in a vacuum but is inextricably embedded within the context of a real-world environment. Intelligence is shaped by and for the purpose of interacting with this environment. This principle of "situatedness" emphasizes that cognitive processes cannot be properly understood when abstracted away from the specific situations in which they are deployed.[8]

Second, **cognition is time-pressured**.[1] Biological agents must operate in real-time, making decisions and taking actions under temporal constraints. This pressure fundamentally shapes cognitive strategies, favoring efficient, "good enough" solutions over computationally expensive, optimal ones. This contrasts with disembodied AI models that can process information offline without such constraints.

Third, and perhaps most centrally, **cognition is for action**.[2] From an evolutionary perspective, the primary purpose of cognition is not to create detailed, objective representations of the world for their own sake, but to guide goal-directed action.[1] Perception, memory, and reasoning are all in service of enabling an organism to act effectively to achieve its goals. Cognition in biological systems is not an end in itself; it is constrained by the system's goals and capacities.[1]

Finally, embodied agents **off-load cognitive work onto the environment**.[1] Rather than performing all computations internally, organisms cleverly exploit the structure of their bodies and their environments to simplify cognitive tasks. A simple example is using one's fingers to count, off-loading the task of working memory onto a physical action. A more complex one is arranging kitchen ingredients in the order they will be used, using spatial organization as an external memory aid. This principle highlights the deep interplay between internal and external resources in cognitive processing.

Rejecting the Cartesian Theater

The embodiment thesis represents a direct and profound rejection of the classical cognitivist view that has long dominated both cognitive science and AI. This classical model, a direct descendant of Cartesian philosophy, envisions the mind as a kind of "central processing unit" that receives perceptual inputs, manipulates abstract symbols according to formal rules, and sends motor commands as outputs.[2] This view creates what has been called the "Cartesian Theater," a metaphorical stage in the mind where disembodied representations are presented for a central homunculus to observe.

Philosophers like Maurice Merleau-Ponty have powerfully challenged this notion. In his *Phenomenology of Perception*, Merleau-Ponty argued against the Cartesian idea that our primary mode of being in the world is thinking. He proposed instead that *corporeity*—the lived, experienced body—is the primary site for knowing the world, and that perception is the pre-reflective foundation of our existence.[1] From this perspective, the body is not an object *in* the world that the mind thinks about; it is our very means of *having* a world.

This philosophical critique finds a direct parallel in the scientific critique of the cognitivist/classicist research program. Proponents of embodiment argue that the classical model's focus on internal symbol manipulation creates an "isolationist assumption".[2] This assumption attempts to understand cognition by focusing almost exclusively on an organism's internal processes, thereby de-emphasizing or completely overlooking the formative role of the body, the environment, and the real-time, goal-directed interactions that link them. By favoring a relational analysis that views the organism, its actions, and its environment as an inextricably linked system, the embodiment thesis seeks to provide a more accurate and powerful explanation of intelligence.[2]

The Sensorimotor Basis of Concepts

Perhaps the most radical claim of the embodiment thesis is that even high-level, abstract concepts are ultimately grounded in the body's sensorimotor experiences. This directly challenges a core principle of traditional AI, which holds that mental representations are amodal—that is, they are abstract symbols stripped of the sensory and motor details of their acquisition.[6]

Spatial concepts provide the clearest illustration of this grounding. Concepts like "up" and "down," "front" and "back," are not arbitrary abstract symbols. Their meaning is fundamentally articulated in terms of our specific bodily structure and how it interacts with a physical environment.[6] The experience of "upness," for instance, is deeply tied to our bipedal, upright posture and the constant experience of acting against the force of gravity. The meaning of "front" is tied to the direction of our sensory apparatus (eyes, ears) and our typical direction of motion. We get a first-hand feel for the embodied nature of these concepts when we are in non-standard orientations, such as lying down or moving backward, and find that applying these concepts becomes momentarily more difficult.[6]

This grounding extends beyond simple spatial terms. Research in cognitive linguistics by figures like Lakoff and Johnson has shown how abstract concepts like "argument" are metaphorically structured by physical experiences (e.g., "argument is war," with concepts of winning, losing, attacking positions, etc.). This suggests that the very architecture of our abstract thought is built upon a foundation of sensorimotor schemas derived from bodily interaction. This view is incompatible with the classical AI principles that knowledge is organized propositionally in an amodal format and that cognition is separate from the motor programs that execute actions.[6]

The principles of embodied cognition, therefore, do more than offer a philosophical critique of traditional AI; they provide a concrete set of criteria for what constitutes robust, general intelligence. The persistent challenges faced by disembodied AI models—their lack of common sense, their brittleness in the face of novelty, their inability to ground symbols in reality—are not isolated technical bugs. They are the predictable symptoms of an architecture that ignores the fundamental lessons of biological intelligence. The path to overcoming these limitations, then, is not simply to build larger models on bigger datasets. It is to build systems that are, in a meaningful sense, embodied. This reframes the entire enterprise of AGI research. The goal ceases to be the construction of a perfect, disembodied input-output function. Instead, the goal becomes the creation of an *agent* that can satisfy the core criteria of embodiment: an agent that is situated in a rich environment, that learns through a closed loop of perception and action, that off-loads cognitive work, and that grounds its understanding in its own goal-directed, interactive experience. Progress, therefore, should be measured not just by performance on static benchmarks, but by an agent's adaptability, resilience, and resourcefulness in the face of dynamic, unpredictable challenges.

Digital Corporeality: Realizing Embodiment in Virtual Worlds

The principles of embodied cognition are derived from the study of biological organisms in the physical world. A critical question for artificial intelligence is whether these principles can be meaningfully applied to artificial agents that exist not in physical reality, but in the purely digital realm of simulations, games, and virtual environments. This section argues that they can, through the concept of *digital corporeality*—the creation of a digital body (an avatar) that exists and acts within a persistent, interactive digital environment (a simulator).

The Digital Body: Avatars as Locus of Presence and Action

For embodiment to occur, there must be a body. In the digital realm, this body takes the form of an avatar—a graphical, pictorial construct through which a user or an AI agent can inhabit a virtual world.[10] This concept of a virtual or simulated body is central to the field of Embodied AI.[8] The avatar is not merely a cursor or a point-of-view; it is a digital object with a defined location, a set of sensory capacities (e.g., a virtual camera providing a visual feed), and a repertoire of possible actions or motor skills (e.g., navigation, manipulation).[10]

This digital body becomes the locus of presence and the interface for all interaction with the virtual environment.[10] It is the "material thing (albeit a digital one) that finds itself located in a space and moves through it".[10] For an AI agent, this digital body, with its specific morphology (its size, shape, degrees of freedom, and sensor/actuator placement), plays an active role in shaping the intelligence that can be developed. The body's capacities and limitations define what the agent can perceive, how it can act, and the nature of the feedback it receives from the environment.[11] Just as a human's body provides the means to greet, play, and convey feelings, the digital body of an agent becomes the material out of which social and physical interactions are embodied in the virtual world.[10] This establishes a form of digital corporeity that is the necessary first step toward digital embodiment.

The Digital Environment: The Critical Role of Simulators

A body, whether physical or digital, requires an environment in which to exist and act. The single greatest catalyst for research in embodied AI has been the recent development of rich, high-fidelity, and computationally efficient simulation environments.[8] These virtual worlds serve as the crucial testbeds where embodied agents can be trained, evaluated, and refined at a scale and speed impossible in the physical world.[8]

The function of these simulators is twofold. First, they provide a safe, controlled, and replicable laboratory for AI research. Physical robots are expensive, fragile, and slow to iterate with.[8] A mistake in the physical world can result in costly damage. In a simulator, an agent can fail millions of times without consequence, allowing it to learn from trial and error through methods like reinforcement learning. Second, simulators provide immense scalability. A single physical robot can only have one experience at a time. In a virtual environment, researchers can run thousands of agents in parallel across millions of distinct episodes, generating the massive amounts of interactive data required for modern deep learning techniques.[8]

The landscape of embodied AI simulators is diverse, with different platforms optimized for different research goals. A survey of leading platforms reveals a trade-off between visual realism, physical fidelity, and interactivity.[16] Simulators like

AI Habitat and **iGibson** leverage real-world 3D scans (from datasets like Matterport3D) to create photorealistic environments, making them ideal for training and testing navigation and exploration tasks where sim-to-real transfer is a key concern.[8] Platforms like

AI2-THOR use synthetic, game-like assets but offer a high degree of object interactivity, allowing agents to open drawers, slice vegetables, and manipulate objects in ways that change their state—crucial for learning complex, task-based planning.[16] Simulators such as NVIDIA's

Isaac Gym and **Isaac Sim** are heavily optimized for robotics, using GPU-accelerated physics (PhysX) to enable massive-scale reinforcement learning for manipulation and locomotion skills.[8] More recent developments like

ThreeDWorld focus on multi-modal physics, simulating not just rigid bodies but also fluids, soft materials, and acoustics.[16] Meanwhile, platforms like

UnrealZoo aim for vast, open-world complexity, providing diverse ecosystems with various agent types (humans, animals, vehicles) for research into more general, unconstrained behaviors.[21]

Defining Digital Embodiment

Synthesizing these concepts, we can formally define *digital embodiment*. A digitally embodied AI is an artificial system that possesses a simulated body (an avatar) through which it perceives (via virtual sensors) and acts (via virtual actuators) within a persistent, interactive, and rule-governed digital environment. This definition is built upon three pillars that directly translate the core principles of embodied cognition into the digital domain:

1. **Situatedness**: The AI agent is not processing abstract data; it exists *within* a specific digital environment. The structure of this environment—its spatial layout, the objects it contains, its physical laws—constrains and shapes the agent's perceptions and possible actions, just as the physical world does for a biological organism.[8]
2. **Agency**: The agent is not a passive observer. Its decisions and actions have direct, tangible, and predictable consequences within the environment. Picking up an object removes it from its location; opening a door reveals a new space. This causal link between action and outcome is the foundation for goal-directed behavior and learning.[8]
3. **Sensorimotor Coupling**: Intelligence emerges from the tight, continuous feedback loop between the agent's digital senses and its digital actions.[8] What the agent "sees" with its virtual camera guides its next move, and that move immediately changes what it sees. Perception and action are not treated as separate, sequential modules of input-processing-output, but as deeply intertwined and mutually influential processes that unfold in real-time.[11]

This framework allows us to move beyond the limitations of disembodied AI by creating the necessary conditions for intelligence to be learned and grounded through experience, even when that experience is purely digital.

Simulator	Environment Type	Physics Fidelity	Interactivity Focus	Primary Use Cases	Key Sources
AI Habitat	Real-world scans (e.g., Matterport3D) & synthetic assets	High (efficient, static scenes)	Navigation, Exploration	Point/Object Navigation, Visual Exploration, Embodied QA	[8]
AI2-THOR	Synthetic 3D assets (game-like)	Moderate (focus on object states)	High (object manipulation, state changes like slicing, opening)	Interactive QA (IQA), Task-based planning, Manipulation	[16]
Isaac Gym / Sim	Synthetic 3D assets	High (NVIDIA PhysX, GPU-accelerated)	Robotic manipulation, Reinforcement Learning	Sim-to-real robotics, DRL for locomotion and manipulation	[8]
UnrealZoo	Synthetic 3D assets (Unreal Engine)	High (photorealistic, complex)	Open-world exploration, diverse agent types (human,	Complex, large-scale open-world navigation and	[21]

			animal, vehicle)	interaction	
ThreeDWorld	Synthetic 3D assets	Advanced (focus on multi-modal physics: fluids, soft-bodies)	Multi-modal sensory-motor learning	Learning physics, audio-visual tasks, complex physical interaction	[16]
iGibson	Real-world scans & synthetic assets	Moderate-High (fast simulation)	Navigation, Interaction with large scenes	Fast prototyping, sim-to-real transfer for mobile robots	[8]

A Conceptual Framework for Digitally Embodied Intelligence

Having established that embodiment is possible within digital worlds, we now propose a multi-layered conceptual framework that details *how* intelligence can emerge from this digital corporeality. This framework moves from the fundamental mechanics of interaction to the higher-level cognitive capabilities of prediction, awareness, and meaning-making. It is not a specific algorithm but a blueprint for the necessary components of a truly intelligent embodied agent.

The Primacy of the Perception-Action Loop

The foundational layer of our framework is the **perception-action loop**, a continuous, closed feedback cycle that is the engine of all learning and adaptation.[11] In the disembodied paradigm, an AI processes a static input to produce an output. In the embodied paradigm, the agent's output (action) directly and immediately influences its next input (perception). An agent moves forward, and its visual field changes; it turns its head, and objects come into view; it interacts with an object, and its tactile sensors register a new state.

This dynamic cycle is what allows an AI to move from being a static, pre-trained entity to a continuously learning one.[7] The system constantly receives feedback on its performance not from a pre-labeled dataset, but from the direct consequences of its own behavior in the environment.[9] This loop, where outputs are evaluated and reintroduced as inputs, enables the AI to discover patterns, correct errors, and recalibrate its internal models for better future decisions.[9] Perception is not a passive reception of data but an active, exploratory process, and action is not a final output but a means of generating new perceptual information. This tight, real-time coupling is the most fundamental departure from the input-output model and the basis for all subsequent layers of intelligence.

Internal World Models for Prediction and Planning

A purely reactive agent, even one with a tight perception-action loop, is limited. To exhibit intelligent, goal-directed behavior, an agent must be able to anticipate the future. The second layer of our framework is the development of an **internal world model**. This is a generative, predictive model that the agent learns, which simulates the dynamics of its environment.[14]
By interacting with its world, the agent learns the "rules" of its environment—not as explicitly programmed logic, but as learned statistical regularities. It learns that unsupported objects fall, that rigid objects cannot pass through each other, and that certain actions lead to predictable outcomes.[17] This learned world model equips the agent with a form of "imagination".[23] It can simulate potential action sequences and predict their likely consequences
without having to physically (or digitally) execute them.[23]
This capability is transformative for two reasons. First, it dramatically improves learning efficiency. The agent can explore thousands of possibilities in its fast, internal simulation, reducing the amount of slow, costly trial-and-error required in the external environment. Second, it enables true planning. The agent can use its world model to search for a sequence of actions that will lead it from its current state to a desired goal state, moving it beyond simple reactivity to proactive, goal-oriented behavior.[23] This moves the agent from basic pattern matching toward a form of causal reasoning, as it begins to understand the cause-and-effect structure of its world.[24]

Mechanisms for Spatial and Contextual Awareness

Intelligence is not just about acting and predicting; it is about acting and predicting appropriately based on the current situation. This third layer of the framework concerns the development of **spatial and contextual awareness**.
This begins with **Spatial AI**, the capacity to understand and operate within a three-dimensional environment.[25] A digitally embodied agent must be able to build an internal representation of its surroundings, performing tasks like real-time 3D mapping, object detection, and tracking the movement of itself and other entities.[25] This requires the integration of multi-modal sensory data from its virtual sensors, such as RGB-D cameras and simulated LiDAR.[24]
A key technical enabler for this is the development of visual representations that are inherently 3D-aware. Traditional computer vision models trained on 2D images often lack a true understanding of 3D geometry. A promising direction is represented by frameworks like **SPA (3D Spatial-Awareness Enables Effective Embodied Representation)**.[26] SPA is a

representation learning framework that uses a pretext task based on differentiable neural rendering. By training a standard Vision Transformer (ViT) to render novel views of a scene from a set of multi-view input images, the model is forced to build an explicit internal representation of the scene's 3D structure.[27] This process endows the model with an intrinsic spatial understanding without requiring explicit 3D supervision. Extensive evaluations have shown that this learned 3D awareness is a critical factor for success in a wide range of embodied tasks, consistently outperforming models pre-trained on 2D images or even large-scale vision-language datasets.[28] This demonstrates that for an agent to act effectively in a 3D world, its perceptual system must be built on a foundation of 3D-aware representations.

Spatial awareness must then be integrated into a broader **contextual awareness**. Following the principles of Contextual AI, the agent must fuse its understanding of space with other critical information: the specific task it has been given (often via natural language), the current time, its own internal state (e.g., its goals, its history of past actions), and the state of other agents.[30] This allows the agent to move from generic to adaptive behavior, tailoring its actions to the specific, nuanced situation it faces.[32]

Grounding Semantics through Affordance Learning

The capstone of this framework is the mechanism by which an agent derives meaning, finally solving the symbol-grounding problem that plagues disembodied models. This is achieved by connecting abstract symbols, such as words, to the agent's direct, interactive experience through **affordance learning**.

An affordance, a concept from ecological psychology, is a potential for action that an environment offers a particular agent.[34] A flat, horizontal surface *affords* placing-things-on; a handle *affords* pulling; a cup *affords* grasping-and-drinking-from. These affordances are relational—they depend on both the properties of the object and the capabilities of the agent.[34] Through the perception-action loop, and by using its predictive world model to understand outcomes, the embodied agent learns these affordances through trial and error.[35] It learns that when it applies a certain motor command to a door, the door opens. It learns that attempting to walk through a wall results in a collision.

This process grounds meaning. The concept of "chair" is no longer an ungrounded token or a cluster of pixels. For the embodied agent, "chair" becomes a rich set of learned, potential interactions: something that affords sitting-on, something that can be navigated-around, something that can be pushed, or something that can be used to block a path. The agent's knowledge shifts from learning *what* something is (a label) to learning *what it is for* (a set of affordances).[36] This provides a direct and powerful mechanism for grounding natural language. A command like "put the cup on the table" is no longer a purely linguistic puzzle. The agent can solve it by mapping the symbol "cup" to an object in its environment that has the learned affordance of "pick-up-ability," and mapping "table" to a surface with the affordance of "place-on-ability".[8] Meaning, in this framework, is not given; it is earned through interaction.

The implications of this framework lead to a significant re-evaluation of the roles in AI development. If intelligence is not programmed but learned through interaction, then the environment of that interaction becomes paramount. The properties of the simulated world—its physics, the objects it contains, the actions it permits—directly dictate the structure and the limits of the intelligence that can possibly emerge.[2] An agent trained in a world devoid of gravity will never develop a concept of "falling." An agent whose world contains only static, non-manipulable objects will never learn the meaning of "open" or "pick up." This leads to a crucial conclusion: in the paradigm of digitally embodied AI, the **simulator is not merely a passive stage for the AI's performance; it is an active and integral part of the AI's cognitive architecture**. The design of the virtual world is functionally equivalent to the co-design of the AI's "brain" and "body".[8] The richness and complexity of the simulated environment directly translate to the potential richness and complexity of the learned intelligence. This elevates the discipline of simulator design from a supporting engineering task to a central, co-equal scientific endeavor in the pursuit of AGI. Progress may depend as much on the creativity of "virtual world builders" as it does on the ingenuity of "algorithm designers."

Technical Realization: Architectures and Learning Paradigms

Translating the conceptual framework for digitally embodied intelligence into functional systems requires a confluence of specific technical architectures and learning paradigms. This section surveys the state-of-the-art approaches that are making this vision a reality, focusing on multi-modal models for perception, interactive learning methods, and the extension from single-agent to multi-agent systems.

Multi-Modal Architectures for Rich Perception

A digitally embodied agent must perceive its world through multiple sensory channels and understand instructions given in natural language. This necessitates architectures that can fluidly integrate vision, language, and action. The most promising approach in this domain is the development of **Vision-Language-Action (VLA) models**.[38]

VLA models are specifically designed to process these three modalities in a unified framework to perform language-conditioned tasks.[38] A typical VLA architecture consists of three core components. First, a powerful

vision encoder, often a Vision Transformer (ViT), processes the visual input from the agent's virtual camera. To be effective, this encoder should be pre-trained using a method that imparts 3D spatial awareness, such as the SPA framework, allowing it to extract meaningful geometric and semantic features from the scene.[26] Second, a **Large Language Model (LLM)** serves as the language understanding and reasoning backbone, processing textual instructions or goals. Third, an **action decoder** takes the fused vision and language representations and generates low-level motor commands, such as translation, rotation, or gripper actuation, that the agent executes in the environment.[38]

In practice, many advanced systems employ a **hierarchical control** structure to manage complexity.[38] In this setup, a high-level planning module, often leveraging the reasoning capabilities of a powerful LLM, is responsible for task decomposition. It takes a complex, long-horizon instruction (e.g., "clean the kitchen") and breaks it down into a logical sequence of simpler, executable sub-tasks (e.g., 1. "navigate to the sponge," 2. "pick up the sponge," 3. "go to the counter," 4. "wipe the counter"). A separate, low-level control policy, typically a more streamlined VLA model, is then responsible for executing each of these sub-tasks. This hierarchical approach effectively balances the deep reasoning and planning capacity of large models with the speed, precision, and real-time responsiveness required for fine-grained motor control.[38]

Learning Through Interaction: Reinforcement and Self-Supervision

The central learning mechanism for an embodied agent is direct interaction with its environment. This is most naturally implemented using paradigms from **Deep Reinforcement Learning (DRL)** and **Self-Supervised Learning**.

DRL provides the fundamental framework for learning through trial and error. An agent learns a "policy"—a mapping from states to actions—by performing actions in an environment and receiving feedback in the form of a "reward" signal.[7] The agent's goal is to learn a policy that maximizes the cumulative reward over time.[9] This is perfectly suited for embodied tasks like navigation, where an agent might receive a positive reward for moving closer to a target and a negative reward for colliding with an obstacle.[40] Over many episodes, the agent discovers a sequence of actions that reliably leads to the goal.[19] However, DRL faces significant challenges, particularly **sparse rewards**, where feedback is only given at the end of a long and complex task, making it difficult for the agent to assign credit to the correct actions. This is often addressed through **curriculum learning**, where the agent is first trained on very simple tasks and environments, with the difficulty gradually increasing as it becomes more proficient.[40] Another major hurdle is generalization to new, unseen environments. A common technique to improve robustness is to introduce significant variability during training, such as randomizing the textures, lighting, and layout of the simulated environments.[41]

Self-Supervised Learning provides a powerful complement to DRL, allowing an agent to generate its own training signals from unlabeled interactive data. The learning of an internal world model is a prime example; the model is trained on the self-supervised task of predicting the next sensory frame given the current frame and an action.[23] The SPA framework for 3D representation learning is also self-supervised, using the task of rendering novel views as a pretext to learn the underlying 3D geometry of a scene.[28] These methods allow the agent to learn rich, structured representations of its world without requiring explicit, hand-crafted reward functions, which can then be leveraged by a DRL algorithm for more efficient policy learning.

From Single Agents to Collective Adaptive Intelligence (CAI)

While single-agent systems are a crucial starting point, many real-world and complex digital problems require the coordinated efforts of multiple agents. The principles of digital embodiment extend naturally from a single agent to a collective, opening up research into **Embodied Multi-Agent Systems (EMAS)**.[42] EMAS research explores how multiple embodied agents can interact with an environment and with each other to solve problems collaboratively. This is considered a critical step toward AGI, as it introduces the need for sophisticated mechanisms for communication, coordination, adaptation, and real-time collaborative problem-solving.[44]

A particularly transformative approach within this domain is **Collective Adaptive Intelligence (CAI)**. CAI describes systems where numerous autonomous agents collaborate, adapt their behaviors, and self-organize to solve complex problems in dynamic environments without centralized control.[46] The defining attributes of CAI systems include:

- **Decentralization**: Each agent operates independently based on its local perceptions and communicates with its peers, eliminating single points of failure and enabling scalability.[46]
- **Self-Adaptation**: Agents are not static; they dynamically adjust their internal models, strategies, and even communication protocols based on their experiences and the changing demands of the task.[46]
- **Collective Resilience and Scalability**: The collective as a whole is robust. It can continue to function and re-adapt to complete a task even if some agents are removed (resilience). Conversely, its capabilities can be enhanced to tackle more complex tasks when new agents are added to the system (scalability).[46]

Recent advances in foundation models are now being leveraged to facilitate more sophisticated collaboration within these systems. LLMs can be used to enable richer, more flexible communication protocols and to support distributed planning and consensus-building among agents, paving the way for highly adaptive and generative multi-agent collaboration in both virtual and physical contexts.[48]

Horizons and Hurdles: Implications of Digitally Embodied AI

The shift toward a paradigm of digitally embodied intelligence is not merely an academic exercise. It promises to unlock transformative applications that will reshape our interaction with technology and each other, while simultaneously presenting a host of profound technical, ethical, and philosophical challenges that demand careful consideration.

Transformative Applications

The development of robust, context-aware, and spatially intelligent digital agents will catalyze innovation across numerous domains, moving AI from a text-box interface to a fully integrated partner in our digital and physical lives.

One of the most immediate impacts will be on **the future of human-AI collaboration**. The metaverse and other shared virtual spaces will evolve from static social rooms into dynamic work environments where humans team up with AI.[49] Imagine attending a virtual project meeting where some of your colleagues are not humans but specialized AI agents, represented by interactive avatars. A data-scientist agent could present complex visualizations in real-time, a project-manager agent could provide status updates and identify risks, and a strategy agent could model future scenarios.[49] This paradigm shifts the relationship from human-as-operator to human-and-AI-as-teammates, where the AI is no longer a passive tool to be prompted but an active collaborator in a shared space.[52]

This technology is also the key to creating truly intelligent **digital twins**. A digital twin is a virtual replica of a real-world object, process, or environment.[53] Current digital twins are often static or based on simple models. An embodied AI agent, however, could inhabit a digital twin of a factory floor, a city's traffic system, or even a human organ. By interacting with this high-fidelity simulation, the agent could learn to optimize complex workflows, predict maintenance failures before they occur, test the impact of new policies in a safe virtual space, and discover novel solutions that would be too costly or dangerous to explore in the real world.[25]

Finally, progress in digital embodiment will directly **accelerate sim-to-real robotics**. The ability to train intelligent agents in scalable, high-fidelity simulators is a game-changer for robotics.[12] Agents can learn complex skills like dexterous manipulation, multi-step task completion, and navigation in millions of simulated trials—a process that would take years in the physical world.[8] This learned intelligence can then be transferred to physical robots, drastically reducing development time and cost. This has far-reaching applications in logistics (autonomous warehouse robots), healthcare (surgical assistants and rehabilitation robots), and autonomous vehicles, where robust adaptability to unpredictable real-world conditions is paramount.[12]

Interdisciplinary Challenges

The path toward this future is fraught with significant hurdles that span the technical, ethical, and philosophical domains.

On the **technical front**, major challenges remain. The computational resources required to run both the high-fidelity simulators and the complex AI models are immense, making research costly and limiting accessibility.[4] The **sim-to-real gap** remains a persistent problem; policies that work perfectly in simulation can fail unexpectedly in the real world due to subtle differences in physics, sensor noise, or appearance.[14] Furthermore, creating systems capable of true **lifelong learning**—continuously adapting and acquiring new knowledge over long periods without catastrophic forgetting—is an open and formidable research question.[7] The sheer, unimaginable complexity and unpredictability of the physical world is a constant challenge that digital environments can only ever approximate.[15]

The emergence of autonomous embodied agents also raises profound **ethical and legal dilemmas**. If an autonomous agent causes physical or financial harm, who is legally and morally responsible? The programmer, the owner, the manufacturer, or the agent itself? Existing legal frameworks are ill-equipped to handle these questions of distributed liability.[13] The ability to create realistic digital replicas of individuals, or "digital doppelgangers," raises urgent questions of identity, consent, and ownership. Can a person's likeness be used without their permission? What are the rules for digital existence after death?.[56] There are also risks of emotional harm, particularly if AI agents are designed to be persuasive or to form intimate relationships with users, potentially leading to emotional manipulation.[13] This necessitates the development of robust regulatory frameworks that prioritize transparency, accountability, and human well-being.

Finally, this technological shift forces a confrontation with deep **philosophical and anthropological questions**. Technology is never a neutral tool; it actively reshapes our societies, our relationships, our ways of thinking, and our very conception of what it means to be human.[57] The move toward digital embodiment and life in virtual worlds challenges our understanding of presence, identity, and experience.[10] We cannot assume that people will simply "figure out" how to use these technologies benevolently; history, particularly with social media, suggests that a hands-off approach can lead to unforeseen negative consequences.[57] It is imperative that we engage in a deliberate, society-wide conversation about how we want to integrate these powerful agents into our lives, ensuring they are designed to augment human potential and foster well-being, rather than diminish or replace them.

To help structure the conversation about progress in this field, we propose a speculative taxonomy for the capabilities of digitally embodied intelligence, outlining a potential roadmap from current systems to a future embodied AGI.

Level	Title	Description	Key Capabilities	Example
L1	**Specialized Navigator**	Agent can perform a single, well-defined task	Goal-driven navigation; Basic obstacle	An agent trained in AI Habitat to find a specific

		(e.g., navigation) in a known or similar environment.	avoidance; Generalization to unseen but similar environments.	object class in new, unmapped rooms.[16]
L2	**Context-Aware Actor**	Agent can perform a variety of related tasks based on multi-modal instructions and has a basic understanding of object affordances.	Task decomposition; Grounded language understanding (VLN); Basic object interaction/manipulation; Adapts to environmental changes.	A VLA-powered agent that can follow a recipe in a simulated kitchen, involving navigation, picking, and placing actions.[8]
L3	**Predictive World Modeler**	Agent uses a learned internal world model to plan long-horizon tasks, reason about causality, and adapt to substantially different task categories.	Predictive planning; Causal reasoning; Transfer learning across diverse tasks (e.g., from cooking to cleaning); Real-time responsiveness.	An agent that, having learned the physics of its world, can figure out how to build a stable tower of blocks to reach a high shelf, a task it has never explicitly been trained for.[23]
L4	**Collaborative Social Agent**	Agent can operate effectively in a multi-agent environment, communicating, coordinating, and collaborating with other agents (AI or human) to achieve shared goals.	Theory of Mind (inferring others' intent); Complex communication and negotiation; Emergent collaborative strategies; Self-assembly and role adaptation.	A team of CAI agents that can collaboratively build a complex structure in a virtual world, dynamically re-assigning roles if one agent fails.[46]
L5	**Creative & Open-Ended Discoverer (Embodied AGI)**	Agent can set its own goals, exhibit curiosity, engage in open-ended learning, and	Intrinsic motivation; Open-ended skill acquisition; Creative	An agent that, when placed in a new virtual world, explores it out of curiosity, invents

		creatively use its environment to solve novel, abstract problems.	problem-solving; Abstract reasoning grounded in embodied experience.	its own tools from the available objects, and discovers underlying physical principles on its own.[15]

Conclusion: A New Paradigm for Artificial Intelligence

This paper has mounted a sustained argument against the dominant disembodied paradigm in artificial intelligence and has proposed a comprehensive alternative rooted in the principles of embodied cognition. We have contended that the persistent limitations of modern AI—its lack of common sense, its brittleness, its inability to ground symbols in reality—are not superficial engineering problems to be patched with more data or larger parameter counts. They are, rather, the fundamental and predictable consequences of an architectural philosophy that divorces intelligence from interaction, mind from body, and computation from the world.

The path forward, we have argued, requires a radical paradigm shift. Intelligence, we posit, is not an abstract property of a computational system but an emergent property of an agent's dynamic, goal-directed engagement with an environment. It must be grounded in the rich, continuous, and consequential feedback loop of perception and action. We have shown that this principle, long understood in the study of biological organisms, can be powerfully translated into the digital realm.

Our proposed framework for digitally embodied intelligence—built upon the foundations of a digital body (avatar), an interactive and persistent environment (simulator), a continuous perception-action loop, the learning of predictive world models, and the grounding of semantics through affordance learning—offers a concrete and viable pathway toward this new paradigm. It charts a course for developing AI systems that can acquire genuine spatial awareness and deep contextual understanding not through passive observation of static data, but through active, first-person experience.

Realizing this vision demands a more holistic and integrated research program. The grand challenge of AGI will not be solved by computer scientists working in isolation. Progress will require a deep and sustained collaboration between AI engineers, cognitive scientists, neuroscientists, developmental psychologists, philosophers of mind, and, crucially, the designers and architects of the virtual worlds that will serve as the crucibles for this new form of intelligence. The objective must shift from the narrow goal of building an algorithm that can pass a static test to the grander ambition of creating the necessary conditions—the right body, the right world, the right set of interactions—from which a truly general, adaptive, and ultimately understandable intelligence can emerge.

Works cited

1. Embodied cognition - Wikipedia, accessed July 3, 2025, https://en.wikipedia.org/wiki/Embodied_cognition
2. Embodied Cognition | Internet Encyclopedia of Philosophy, accessed July 3, 2025, https://iep.utm.edu/embodied-cognition/
3. Limitations - Generative Artificial Intelligence (AI) - UofL Libraries at University of Louisville, accessed July 3, 2025, https://library.louisville.edu/kornhauser/generative-ai/limitations
4. Understanding the Limitations and Challenges of Generative AI - Arion Research LLC, accessed July 3, 2025, https://www.arionresearch.com/blog/understanding-the-limitations-and-challenges-of-generative-ai
5. Understanding The Limitations Of AI (Artificial Intelligence) | by Mark Levis | Medium, accessed July 3, 2025, https://medium.com/@marklevisebook/understanding-the-limitations-of-ai-artificial-intelligence-a264c1e0b8ab
6. Embodied Cognition - Stanford Encyclopedia of Philosophy, accessed July 3, 2025, https://plato.stanford.edu/archIves/sum2020/entries/embodied-cognition/
7. The Importance of Continuous Learning in AI: Navigating Technological Evolution, accessed July 3, 2025, https://profiletree.com/the-importance-of-continuous-learning-in-ai/
8. Embodied AI: Giving Intelligence a Physical Presence | by Anirudh Sekar - Medium, accessed July 3, 2025, https://medium.com/@anirudhsekar2008/embodied-ai-giving-intelligence-a-physical-presence-c7a584e25cd4
9. The Power of AI Feedback Loop: Learning From Mistakes | IrisAgent, accessed July 3, 2025, https://irisagent.com/blog/the-power-of-feedback-loops-in-ai-learning-from-mistakes/
10. Chapter 3 Living Digitally: Embodiment in Virtual Worlds - T.L. Taylor, accessed July 3, 2025, https://tltaylor.com/wp-content/uploads/2009/07/Taylor-LivingDigitally.pdf
11. Embodied Intelligence: Grounding AI in the Physical World for Enhanced Capability and Adaptability - Alphanome.AI, accessed July 3, 2025, https://www.alphanome.ai/post/embodied-intelligence-grounding-ai-in-the-physical-world-for-enhanced-capability-and-adaptability
12. What Is Embodied AI? - Artificial Intelligence - Built In, accessed July 3, 2025, https://builtin.com/artificial-intelligence/embodied-ai
13. The ethics of artificial intelligence: Issues and initiatives - European Parliament, accessed July 3, 2025, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf
14. Aligning Cyber Space with Physical World: A Comprehensive Survey on Embodied AI, accessed July 3, 2025, https://arxiv.org/html/2407.06886v4

15. AI That Moves, Adapts, and Learns: The Future of Embodied Intelligence | Columbia AI, accessed July 3, 2025, https://ai.columbia.edu/news/ai-moves-adapts-and-learns-future-embodied-intelligence
16. A Survey of Embodied AI: From Simulators to Research Tasks - arXiv, accessed July 3, 2025, https://arxiv.org/pdf/2103.04918
17. A Brief History of Embodied Artificial Intelligence, and its Outlook, accessed July 3, 2025, https://cacm.acm.org/blogcacm/a-brief-history-of-embodied-artificial-intelligence-and-its-future-outlook/
18. AI Makes Strides in Virtual Worlds More Like Our Own | Quanta Magazine, accessed July 3, 2025, https://www.quantamagazine.org/ai-makes-strides-in-virtual-worlds-more-like-our-own-20220624/
19. Target-driven Visual Navigation in Indoor Scenes using Deep Reinforcement Learning - Stanford Computer Vision Lab, accessed July 3, 2025, http://vision.stanford.edu/pdf/zhu2017icra.pdf
20. Aligning Cyber Space with Physical World: A Comprehensive Survey on Embodied AI, accessed July 3, 2025, https://arxiv.org/html/2407.06886v2
21. UnrealCV Zoo: Enriching Photo-realistic Virtual Worlds for Embodied AI Agents, accessed July 3, 2025, https://openreview.net/forum?id=vQ1y086Kn2
22. Embodied AI Explained: Principles, Applications, and Future Perspectives, accessed July 3, 2025, https://lamarr-institute.org/blog/embodied-ai-explained/
23. Embodied Intelligence Through World Models, accessed July 3, 2025, https://utoronto.scholaris.ca/server/api/core/bitstreams/8a278445-a1ce-4a53-9275-4f1a0c6739b9/content
24. Bridging Physical and Digital Worlds: Embodied Large AI for Future Wireless Systems, accessed July 3, 2025, https://arxiv.org/html/2506.24009v1
25. What Is Spatial AI? A Beginner's Guide to Smarter, Context-Aware Systems - Leaniar, accessed July 3, 2025, https://leaniar.com/what-is-spatial-ai-a-beginners-guide-to-smarter-context-aware-systems/
26. [2410.08208] SPA: 3D Spatial-Awareness Enables Effective Embodied Representation - arXiv, accessed July 3, 2025, https://arxiv.org/abs/2410.08208
27. SPA: 3D Spatial-Awareness Enables Effective Embodied Representation, accessed July 3, 2025, https://haoyizhu.github.io/spa/
28. SPA: 3D SPatial-Awareness Enables Effective Embodied Representation - arXiv, accessed July 3, 2025, https://arxiv.org/html/2410.08208v1
29. EmbodiedMAE: A Unified 3D Multi-Modal Representation for Robot Manipulation - arXiv, accessed July 3, 2025, https://arxiv.org/html/2505.10105v1
30. Spatial contextual awareness - Wikipedia, accessed July 3, 2025, https://en.wikipedia.org/wiki/Spatial_contextual_awareness
31. What is contextual AI? - WalkMe™ - Digital Adoption Platform, accessed July 3, 2025, https://www.walkme.com/glossary/contextual-ai/
32. Contextual AI, accessed July 3, 2025,

https://tocxten.com/index.php/2024/06/04/contextual-ai/

33. Conversational Agents and Context Awareness: How AI Understands and Adapts to User Needs - SmythOS, accessed July 3, 2025, https://smythos.com/developers/agent-development/conversational-agents-and-context-awareness/
34. Seven educational affordances of virtual classrooms - PMC, accessed July 3, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC8837497/
35. What are the learning affordances of 3-D virtual environments? - ResearchGate, accessed July 3, 2025, https://www.researchgate.net/profile/Mark-Lee-27/publication/220017513_What_are_the_learning_affordances_of_3-D_Virtual_environments/links/5a1404e00f7e9b1e5730af8c/What-are-the-learning-affordances-of-3-D-Virtual-environments.pdf
36. (PDF) Design of Virtual Learning Environments: Learning Analytics and Identification of Affordances and Barriers - ResearchGate, accessed July 3, 2025, https://www.researchgate.net/publication/283860274_Design_of_Virtual_Learning_Environments_Learning_Analytics_and_Identification_of_Affordances_and_Barriers
37. Large language models for artificial general intelligence (AGI): A survey of foundational principles and approaches - arXiv, accessed July 3, 2025, https://arxiv.org/html/2501.03151v1
38. A Survey on Vision-Language-Action Models for Embodied AI - arXiv, accessed July 3, 2025, https://arxiv.org/pdf/2405.14093
39. Advances in Embodied Navigation Using Large Language Models: A Survey - arXiv, accessed July 3, 2025, https://arxiv.org/html/2311.00530v5
40. How do I train a model to navigate to a fixed target in a grid based environment? : r/reinforcementlearning - Reddit, accessed July 3, 2025, https://www.reddit.com/r/reinforcementlearning/comments/1f4q982/how_do_i_train_a_model_to_navigate_to_a_fixed/
41. Transfer Deep Reinforcement Learning in 3D Environments: An ..., accessed July 3, 2025, https://devendrachaplot.github.io/papers/nips16_Transfer_Deep_RL.pdf
42. Embodied Multi-Agent Systems: A Review | Request PDF - ResearchGate, accessed July 3, 2025, https://www.researchgate.net/publication/392742472_Embodied_Multi-Agent_Systems_A_Review
43. Embodied Multi-Agent Systems: A Review, accessed July 3, 2025, https://www.ieee-jas.net/en/article/doi/10.1109/JAS.2025.125552
44. (PDF) Multi-agent Embodied AI: Advances and Future Directions - ResearchGate, accessed July 3, 2025, https://www.researchgate.net/publication/391575549_Multi-agent_Embodied_AI_Advances_and_Future_Directions
45. [2505.05108] Multi-agent Embodied AI: Advances and Future Directions - arXiv, accessed July 3, 2025, https://arxiv.org/abs/2505.05108
46. Conceptual Framework Toward Embodied Collective Adaptive Intelligence - arXiv, accessed July 3, 2025, https://arxiv.org/html/2505.23153v2

47. [2505.23153] Conceptual Framework Toward Embodied Collective Adaptive Intelligence - arXiv, accessed July 3, 2025, https://arxiv.org/abs/2505.23153
48. Generative Multi-Agent Collaboration in Embodied AI: A Systematic ..., accessed July 3, 2025, https://arxiv.org/abs/2502.11518
49. Generative AI meets the virtual world: A model for human-AI collaboration - Deloitte, accessed July 3, 2025, https://www.deloitte.com/us/en/insights/industry/technology/ai-and-vr-model-for-human-ai-collaboration.html
50. Human-AI Teaming in the Age of Collaborative Intelligence - SecureWorld, accessed July 3, 2025, https://www.secureworld.io/industry-news/human-ai-teaming-age-collaboration
51. Artificial intelligence combined with the virtual world: A model of human-AI collaboration, accessed July 3, 2025, https://nssc.gov.vn/articles/artificial-intelligence-combined-with-the-virtual-world-a-model-of-human-ai-collaboration/
52. The future of Human-AI-Teaming collaboration - The HARTU project, accessed July 3, 2025, https://www.hartu-project.eu/2025/02/03/the-future-of-human-ai-teaming-collaboration/
53. Learn About Active Inference AI & the Spatial Web Protocol: Gain Your Competitive Edge & Get Certified!, accessed July 3, 2025, https://deniseholt.us/learn-about-active-inference-ai-the-spatial-web-protocol-gain-your-competitive-edge-get-certified/
54. What is Embodied AI? A Guide to AI in Robotics - Encord, accessed July 3, 2025, https://encord.com/blog/embodied-ai/
55. Key Differences Between Embodied AI and Traditional AI - Vertu, accessed July 3, 2025, https://vertu.com/ai-tools/embodied-ai-vs-traditional-ai/
56. How to navigate the ethical dilemmas posed by the future of digital identity, accessed July 3, 2025, https://www.weforum.org/stories/2024/03/navigate-ethical-dilemmas-future-digital-identity/
57. Embodiment, AI, and the Human Question — A Conversation on Technology and Theology with Jared Hayden - Hungarian Conservative, accessed July 3, 2025, https://www.hungarianconservative.com/articles/interview/embodiment-ai-digital-age-mcc-summit-interview-jared-hayden/
58. Toward Embodied AGI: A Review of Embodied AI and the Road Ahead - arXiv, accessed July 3, 2025, https://arxiv.org/html/2505.14235v1