

Review

Open Access



Digital twins to embodied artificial intelligence: review and perspective

Junfei Li, Simon X. Yang

School of Engineering, University of Guelph, Guelph ON N1G 2W1, Canada.

Correspondence to: Prof. Simon X. Yang, School of Engineering, University of Guelph, 50 Stone Road East, Guelph ON N1G 2W1, Canada. E-mail: syang@uoguelph.ca

How to cite this article: Li, J.; Yang, S. X. Digital twins to embodied artificial intelligence: review and perspective. *Intell. Robot.* 2025, 5(1), 202-27. <http://dx.doi.org/10.20517/ir.2025.11>

Received: 13 Sep 2024 **First Decision:** 11 Nov 2024 **Revised:** 17 Feb 2025 **Accepted:** 17 Feb 2025 **Published:** 18 Mar 2025

Academic Editor: Hang Su **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

Embodied artificial intelligence (AI) is reshaping the landscape of intelligent robotic systems, particularly by providing many realistic solutions to execute actions in complex and dynamic environments. However, Embodied AI requires a huge data generation for training and evaluation to ensure safe interaction with physical environments. Therefore, it is necessary to build a cost-effective simulated environment that can provide enough data for training and optimization from the physical characteristics, object properties, and interactions. Digital twins (DTs) are vital issues in Industry 5.0, which enable real-time monitoring, simulation, and optimization of physical processes by mirroring the state and action of their real-world counterparts. This review explores how integrating DTs with Embodied AI can bridge the sim-to-real gap by transforming virtual environments into dynamic and data-rich platforms. The integration of DTs offers real-time monitoring and virtual simulations, enabling Embodied AI agents to train and adapt in virtual environments before deployment in real-world scenarios. In this review, the main challenges and the novel perspective of the future development of integrating DTs and Embodied AI are discussed. To the best of our knowledge, this is the first work to comprehensively review the synergies between DTs and Embodied AI.

Keywords: Digital twins, Embodied AI, robotic digital twins, human digital twins, human-robot collaboration

1. INTRODUCTION

Embodied artificial intelligence (AI) represents a general movement toward realistic systems that require the ability to adapt to dynamic and uncertain situations in the real world^[1]. Unlike traditional disembodied AI,



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



which often relies on abstract data processing in cyberspace, Embodied AI emphasizes the importance of physical interaction, perception, and motion in accomplishing multi-objective tasks, including sensing and multi-sensor fusion for Embodied robots, safety navigation and interaction subject to uncertainties and disruptions, and self-supervised learning and reasoning to understand human behaviors^[2–4]. Embodied AI requires a huge amount of data generation for training and evaluation in their surrounding environments. In addition, the potential harm to robots in hazardous scenarios and the possible injuries to humans during interactions have led to the current Embodied AI being unable to move beyond simulation environments. Therefore, it is necessary to build a cost-effective simulated environment for the training and optimization of Embodied AI.

Many Embodied simulators have been developed to provide suitable platforms for testing the Embodied AI. Duan *et al.* provided a comparison review for the current Embodied AI simulators^[5]. The comparison study indicated that AI2-THOR^[6], Gibson^[7], and Habitat-Sim^[8] are highly respected and widely adopted in Embodied AI research due to the evaluation features of realism, scalability, and interactivity. Liu *et al.* provided an overview of various simulators used in Embodied AI research^[9]. They mentioned that the growing development of simulators (e.g., RoboGen^[10], HOLODECK^[11], PhyScene^[12], and ProcTHOR^[13]) is valuable in Embodied AI research to enhance the generation of interactive, diverse, and physically consistent simulation environments. However, most simulators have limitations. For example, AI2-THOR offers many interactive scenes and its script-based interactions lack physical realism that is only suitable for tasks that do not require highly accurate interactions. Matterport3D^[14] is recognized for its foundational role in Embodied AI benchmarks despite its limitations in agent interaction.

There is a significant challenge to achieving Embodied AI, known as the sim-to-real gap, which refers to challenges faced when transferring a model, algorithm, or control strategy that was trained or developed in a simulated environment to a real-world system. Although simulations provide a cost-effective and safe way to test and train intelligent systems, they often fail to capture the full complexity, variability, and unpredictability of real-world environments due to the high cost, risk, and complexity of real-world experimentation. Digital twins (DTs) are important issues in Industry 5.0^[15] and smart manufacturing^[16]. DTs enable real-time monitoring, simulation, and optimization of physical processes by mirroring the state and action of their real-world counterparts. Recently, they have become critical tools for improving the efficiency and reliability of robotic systems^[17–20]. DTs offer several significant benefits in robotic applications. For instance, DTs can create a virtual representation of physical robots and their environments, which allows for real-time simulation and testing of different scenarios to improve the efficiency and safety of collaborative robotics. In addition, DTs provide a shared visual representation of the physical system, which can enhance the understanding of the system and its behavior among human-robot collaborative tasks.

Recent developments in DTs technology have the potential to fulfill the requirements of Embodied AI for training and optimization in virtual environments while also enhancing real-time response capabilities. Therefore, this review provides an extensive overview of both DTs and Embodied AI and offers new perspectives on how their integration may shape future research. The integration of DTs and Embodied AI presents a promising avenue for advancing intelligent systems, as DTs enable AI agents to test and refine behaviors in digital virtual environments before physical deployment. In addition, human and environmental DTs can broaden the scope of Embodied AI to enhance the human–robot collaboration (HRC) and operational efficiency in large-scale environments. In this review, a comprehensive analysis of DTs is proposed from their conceptual origins to current applications in robotics and beyond. In addition, this review investigates the core components of state-of-the-art Embodied AI. To the best of our knowledge, this is the first comprehensive review of the synergies between DTs and Embodied AI, underscoring how DTs can deliver highly accurate real-time virtual replicas that enable Embodied agents to learn, adapt, and operate effectively in complex environments.

This paper is structured as follows. Section 2 presents the overview of this paper. Section 3 provides the current

applications of DTs in robotic systems. Section 4 discusses the key technologies of Embodied AI. Section 5 explores the potential connections between DTs and Embodied AI, examining how they can complement and enhance each other. Section 6 concludes with a perspective on future research directions.

2. OVERVIEW

In this section, the background of DTs is first introduced, the literature review process is then outlined, and the preliminary results are finally presented.

2.1. Backgrounds of DTs

The concept of DTs can be traced back to the 1960s when National Aeronautics and Space Administration (NASA) used virtual simulation technology for testing and simulating spacecraft during the Apollo program [21]. The developed technologies allowed engineers to identify and resolve potential issues before launch, thereby increasing mission success rates. Although the term “Digital Twin” was not used at that time, the tight integration between virtual models and physical objects was already evident [22]. In 2002, Grieves [23] first introduced the concept of the “Digital Twin” to Product Lifecycle Management. DTs technology involves creating a virtual model of a product that spans its entire lifecycle, from design and manufacturing to maintenance. In the 2010s, the rapid development of the Internet of Things (IoT) and Big Data technologies provided powerful support for the application of DTs [24–26]. IoT devices can collect real-time data from physical objects and transmit it to virtual models, enabling real-time updates and simulations. In addition, DTs have become a key technology in Industry 4.0, widely used in smart manufacturing, predictive maintenance, and production optimization [27–29]. Based on real-time monitoring and data analysis, DTs technology significantly improved production efficiency and product quality. In 2012, NASA redefined the concept of DTs as highly accurate simulations that integrate multiphysics and multiscale modeling with probabilistic methods [30]. These DTs provide real-time reflections of their physical counterparts by using historical data, real-time sensor inputs, and detailed physical models to ensure high fidelity and timely updates.

Recently, DTs have become a popular research issue in various industry applications with many novel technologies (e.g., AI, 5G, and blockchain). Huang *et al.* integrated large language models (LLMs) into DTs to simulate complex scenarios and accurately forecast future conditions with remarkable precision [31]. Rodrigo *et al.* developed a DT mobile network for 5G and 6G environments based on blockchain technology to consider cybersecurity and Industry 4.0 scenarios [32]. In their network, DTs offer a cost-effective solution to assess performance and make informed decisions. Hasan *et al.* proposed a blockchain-based process to create DT to ensure the provenance of secure and trusted data [33]. Feng *et al.* proposed a DT-based intelligent gear health management system to assess the progression of gear degradation with a novel transfer learning algorithm [34,35]. Zhang *et al.* proposed a DT-based intelligent diagnosis method for the assessment of bearing conditions with a partial domain adaptation algorithm [36]. Feng *et al.* proposed a novel DT-enabled domain adversarial graph network to provide effective fault diagnostics of bearings with limited knowledge [37]. Beyond industrial applications, DTs are beginning to demonstrate potential in various other scenarios, including healthcare [38,39], agriculture [40,41], and transportation [42,43]. Table 1 summarizes the highly cited DT papers from the Web of Science (WOS) database over the past five years.

As shown in Table 1, DTs are applied in various domains, including power systems, manufacturing, healthcare, smart cities, transportation, medical imaging, sensor technology, HRC, and smart batteries. The key results highlight significant improvements in system efficiency, predictive capabilities, and operational optimization, underscoring the critical role of DTs in enhancing performance and decision-making across multiple industries. However, the data sources in many studies are still primarily based on simulated data rather than real-time data collected from actual environments.

Table 1. Highly cited DT papers from the WOS database

Ref.	Year	Application field	Core technologies	Data sources	Results
Arraño-Vargas and Konstantinou [44]	2023	Power systems	Modular framework, real-time model, EMT	Australian National Electricity Market	Developed flexible, robust, cost-effective PSDT framework
Lv et al. [45]	2023	Manufacturing	Active learning-DNN, domain adversarial neural networks	experimental data	Designed DT-based workshop system for fault diagnosis, trend prediction
Mo et al. [46]	2023	Manufacturing	Modular AI algorithms, knowledge graph	Simulation data	Enabled dynamic reconfiguration of manufacturing systems
Kapteyn et al. [47]	2022	UAVs	Reduced-order models, Bayesian state estimation	Structural sensor data from UAV	Developed physics-based DT and library of damaged/pristine UAV components
Lv et al. [48]	2022	Healthcare	DL, improved AlexNet	Simulation data	Improved transmission delays, energy use, task completion, resource utilization
Li et al. [49]	2022	Smart cities	Big data, DL, CNN, multi-hop transmission	Wireless sensor network	Provided experimental references for smart city digital development
Lv et al. [50]	2022	Transportation	DL, CNN, support vector regression	Simulation data	Maintained high data transmission, optimal vehicle path planning
Wan et al. [51]	2021	Medical imaging	Semi-supervised SVM, improved AlexNet	MRI data from brain tumor dept	Improved model performance compared to other methods
Darvishi et al. [52]	2021	Sensor technology	Machine learning, fault detection	Air quality, wireless sensor network data	Developed machine-learning-based architecture for sensor validation
Malik and Brem [53]	2021	HRC	Virtual counterpart, validation	Simulation data	Explored DT opportunities in collaborative production systems
Dembski et al. [54]	2020	Smart cities	3D modeling, urban mobility simulation	Sensor network, mobile app data	Enhanced urban planning through participatory and collaborative processes
Wu et al. [55]	2020	Smart battery	Battery modelling, machine learning	Onboard sensing, diagnostics	Framework for smarter battery control and longer lifespan

DT: Digital twin; WOS: Web of Science; EMT: electromagnetic transient; PSDT: power system digital twin; DNN: deep neural network; AI: artificial intelligence; UAVs: unmanned aerial vehicles; DL: deep learning; CNN: convolutional neural network; SVM: support vector machine; MRI: magnetic resonance imaging; HRC: human-robot collaboration.

2.2. Literature review process

To facilitate efficient search and collection of research papers, WOS (<https://webofscience.com>) and Engineering Village (EI) (<http://www.engineeringvillage.com/>) were selected due to their comprehensive representation of high-quality peer-reviewed publications in the engineering field. In addition, Google Scholar was selected to collect additional highly cited papers, especially unpublished papers from arXiv (<https://arxiv.org/>).

The journal publications are collected on the basis of their relevance to DTs and Embodied AI and the Journal Citation Reports in WOS. These categories cover diverse research areas, including information science, automation, bioinformatics, imaging science, industrial manufacturing, electrical and electronic engineering, and robotics. In addition, top-tier AI conferences [e.g., The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) and IEEE/CVF International Conference on Computer Vision (ICCV)] and robotics conferences [e.g., IEEE International Conference on Robotics and Automation (ICRA) and the International Conference on Intelligent Robots and Systems (IROS)] are included.

In this review, two aspects are covered: one for DTs and one for Embodied AI. The search sentence (WOS version) for DTs AND HRC includes (TS = ((digital twin* OR “digital twins” OR “digital-twin” OR “robotic digital twin*” OR “human digital twin*” OR “environmental digital twin*” OR “industrial digital twin*” OR “manufacturing digital twin*” OR “agricultural digital twin*” OR “environmental digital twin*” OR “urban digital twin*”) AND (human-robot OR “human robot” OR “human machine” OR interaction OR interactive OR collaboration OR collaborative) AND (industrial OR industry OR manufacturing OR production))) AND

(PY = 2020 OR 2021 OR 2022 OR 2023 OR 2024), while the search sentence for embodied AI includes (TS = (“embodied AI” OR “embodied artificial intelligence”) AND (sensing OR perception OR “motion planning” OR “motion control” OR learning OR reasoning OR interaction OR collaboration OR train OR “real-world environment”))) AND (PY = 2020 OR 2021 OR 2022 OR 2023 OR 2024).

As a result, 107 relevant papers were selected after filtering out papers not related to the survey topic based on title, keywords, and abstract. In addition, relevant papers are included from the references and support materials have been selected. Thus, a total of 138 papers form the basis of this review paper.

2.3. Preliminary results

During the literature review, several preliminary observations emerged. A major challenge in Embodied AI research is the sim-to-real gap. Many simulation algorithms rely on simplified or predefined models, which often fail to perform adequately in real-world settings. This challenge becomes more pronounced when managing tasks that require large datasets in large-scale cooperative control of robots. In addition, current research on DTs often emphasizes the creation of virtual environments; however, most studies do not extend these virtual spaces for learning, prediction, or inference. As a result, practical strategies or recommendations derived from such DTs have limited impact on real-world scenarios.

This review suggests that an integrated framework that combines DT and Embodied AI could address the limitations observed in both fields. By offering high-fidelity models of real-world environments, DTs can reduce the sim-to-real gap common in Embodied AI. Meanwhile, AI-driven learning and inference mechanisms within DTs can transform virtual environments from passive simulations into data-rich platforms for ongoing training and validation. This synergistic relationship thereby enables continuous refinement of both the physical and virtual components, ultimately fostering adaptive strategies that are more robust and seamlessly transferable to real-world scenarios.

However, despite the promising potential of combining DTs and Embodied AI, this review also identifies several challenges in their integration, which are summarized as follows:

- **Data latency and transfer:** Large data volumes must travel between physical systems and digital models, causing delays that hinder real-time feedback. Such latency is especially problematic in contexts that require quick responses, such as healthcare or industrial robotics.
- **Synchronization issues:** Delays or mismatches in updates can result in decisions based on outdated or incorrect data, particularly in fast-changing settings, where continuous synchronization is essential.
- **Data privacy:** Human-centric applications often rely on personal data such as health metrics or behavioral patterns, introducing privacy risks that require robust security measures without compromising system functionality.
- **Scalability of DTs:** As systems expand, managing interoperability and resilience becomes increasingly complex. This issue is significant in domains such as smart cities and large industrial networks, which require efficient, adaptable, and secure DT operations.

3. DTS IN ROBOTIC SYSTEMS

This section explores the integration of DTs in robotic systems, covering their historical development, their role in robotic digital twins (RDTs), and applications in industry and HRC. It highlights the key aspects of DTs that are of primary concern: the physical components involved, the simulation platforms used, the methods of data collection, and the services provided by these systems. These factors are crucial in determining the performance and applicability of DTs, making their thorough understanding essential for the advancement of research and practical implementation in this field.

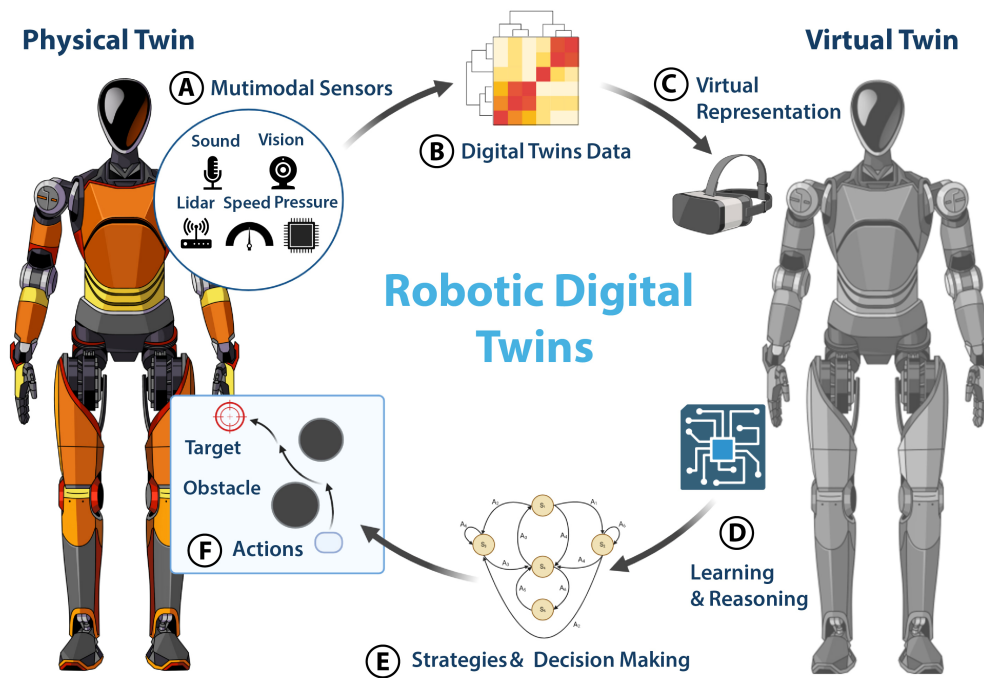


Figure 1. Example of RDT. (A) Multi-modal sensors combine data from multiple sensor types, such as visual, auditory, and tactile inputs, to provide comprehensive and robust environmental perception and enhance the accuracy and reliability of intelligent systems; (B) DTs data include real-time data representing the physical and functional characteristics of a physical robot; (C) DTs create a virtual representation of a physical robot and enable real-time simulation of its physical counterpart; (D) DTs simulate the physical characteristics, environment, and operating conditions of robots to enable robots to train and learn in a virtual environment; (E) DTs provide real-time data and insights, which are crucial for making informed decisions in the complex environment; (F) DTs help optimize the execution of complex tasks by identifying the most effective sequences and methods. RDT: Robotic digital twin; DTs: digital twins.

3.1. RDTs

Robots play an increasingly important role in industrial production and manufacturing processes^[56]. The integration of robotic systems into industrial processes enables human operators to focus on more strategic and creative functions to deal with complex tasks^[57]. RDTs represent a comprehensive convergence of robotics, digital modeling, and real-time data analytics. **Figure 1** shows a general example of RDTs. The virtual counterparts of physical robots enable enhanced interaction, monitoring, and control by providing a high-fidelity simulation of the robot's dynamics, kinematics, and operational environment. Based on real-time multi-modal sensor data, RDTs continuously update to mirror the state of their physical counterpart, which allows for predictive maintenance, optimization of tasks, and adaptability to changing conditions. By enabling simulation-driven design, testing, and deployment, RDTs can significantly accelerate the training processes in robotics, which offers a robust platform to validate AI-driven control strategies and ensure operational reliability in real-world applications.

An important characteristic of RDTs lies in their capacity to accurately simulate the dynamics of robotic systems. Achieving this requires high-precision modeling of the robotic system, often utilizing industrial software such as CAD^[58] or game engines such as Unreal^[59]. **Table 2** summarizes current RDTs research focusing on their physical and virtual components, data collection methods, and services. Sun *et al.* developed a learning-based framework for complex assembly tasks based on a multi-modal DT environment that integrates visual, tactile force, and proprioception data^[60]. By collecting VR-based demonstration data and constructing a skill knowledge base through multi-modal analysis, they enable the transfer of learned assembly skills from the digital realm to real-world settings. Xu *et al.* proposed a cognitive DTs framework for multi-robot collaborative

manufacturing, where an incomplete multi-modal Transformer with deep autoencoder enhances cognition of the system by efficiently handling incomplete data^[61]. This multi-modal approach improves robustness and adaptability in DTs environments. Liu *et al.* proposed a high-fidelity modeling approach and architecture for RDT to enhance real-time monitoring and intelligent applications^[62]. By closely aligning virtual and physical systems, this framework offers advantages in transparent management, comprehensive simulation, and data visualization. However, potential drawbacks include high computational demands, scalability challenges, and complexities in adapting the model to diverse industrial environments. Li *et al.* proposed a DTs-based synchronization architecture for inspection robots^[63]. The robotic system is based on SLAM for constructing real-world maps, while wireless communication ensures smooth data transmission and robot synchronization, allowing for real-time movements and algorithm training in the virtual environment. Liang *et al.* developed a real-time DT system to provide high accuracy in trajectory planning and execution with real-time synchronization between the simulated and physical robots^[64]. While the system successfully bridges virtual and physical environments in real time, challenges related to system complexity and potential latency in data transmission could restrict scalability and performance in more dynamic and large-scale construction settings. Zhang *et al.* presented a novel high-fidelity simulation platform for robot dynamics in a DTs environment^[65]. This approach provides accurate control logic mirroring real systems, simplifies algorithm development, and supports simultaneous testing of real and simulated environments. However, implementing detailed dynamic models may lead to increased computational costs and complexity, particularly when scaling to larger or more varied industrial settings. Liu *et al.* developed a DT-enabled approach for transferring deep reinforcement learning (DRL) algorithms from simulation to physical robots^[66]. This approach uses DTs to correct real-world grasping points, enhancing the accuracy of robotic grasping and validating the effectiveness of sim-to-real transfer mechanisms. Mo *et al.* proposed Terra, a smart DT framework that employs a multi-modal perception module to capture and continuously update real-time environmental and robot states^[67]. This multi-modal approach enriches the DT representation, enabling on-the-fly policy updates and real-time feedback loops between virtual and physical spaces. However, while Terra demonstrates efficacy in an obstacle-rich environment with a simple robot, its scalability may pose limitations when adapting to more complex or resource-constrained scenarios. Cascone *et al.* created VPepper, a virtual replica of the Pepper humanoid robot, to improve interaction with smart objects in a smart home environment, allowing for intensive machine learning training without risking physical robot deterioration^[68]. In addition, Xu *et al.* introduced a DT-based industrial cloud robotics framework that integrates high-fidelity digital models with sensory data, encapsulating robotic control capabilities as robot control as-a-service to ensure accurate synchronization and interaction between digital and physical robots^[69].

As shown in Table 2, the results show that platforms such as Unity 3D are widely used due to their versatility in supporting real-time monitoring, object recognition, and policy updates. Other platforms, such as V-REP and PhysX physics engines, focus on precise tasks such as grasp correction and simulation-to-reality transfer. Physical components such as KUKA industrial robots are often paired with advanced platforms such as ROS and Tecnomatix, enabling motion optimization and decision making in industrial settings. Meanwhile, Pepper humanoid robots and custom-built robots are used with Unity 3D for intelligent perception and emergency intervention in HRC scenarios. This analysis highlights the adaptability of Unity 3D, making it a preferred choice across various applications.

3.2. RDTs in industrial applications

In industrial applications, RDTs play a crucial role in optimizing manufacturing processes, improving efficiency, and reducing costs. The robot modeling involves managing the robot system in the 3D environment^[71]. However, a simple virtual visualization does not suffice to define the RDTs. Crucially, RDTs should have the ability to learn and optimize in response to varying tasks and environmental uncertainties, thereby allowing dynamic adjustments to the physical robotic system. Zhang *et al.* also proposed a novel DT modeling method for large-scale components in robotic systems, using a knowledge graph for knowledge representation and

Table 2. Summary of recent RDTs

Ref.	Year	Physical part	Virtual platform	Data collection	Service
Sun <i>et al.</i> [60]	2024	Robotic arm	Unity 3D	Visual, tactile force, and proprioception data	Transfer knowledge to the robot based on the multi-modal DT environment
Xu <i>et al.</i> [61]	2024	UR5	ROS and URScript	8-modality data from the system	Improve the robustness based on the incomplete multi-modal dataset
Liu <i>et al.</i> [62]	2023	SD3/500 robotic arm	Unity 3D, Pixyz plugin and MATLAB	Robot operational data	High-fidelity modeling approach and architecture for DTs of industrial robots
Li <i>et al.</i> [63]	2022	Car-like robot	Unity 3D and ROS	LiDAR data and 3D point cloud	Enable simultaneous mapping, data interaction, and motion control between the physical and virtual robots
Liang <i>et al.</i> [64]	2022	KUKA KR120 robotic arm	Gazebo	Transmission times, joint angles, and end-effector poses	Ensure minimal mean square errors between the virtual and physical robots
Zhang <i>et al.</i> [65]	2022	Kinova Gen3 robot and the MPS station	Festo CIROS and MATLAB	Optic, magazine and terminal sensors	High-fidelity representations of robot dynamics
Liu <i>et al.</i> [66]	2022	Dobot Magician (robotic arm)	V-REP PRO EDU 3.6.2	RGB images, depth images	Real-time perception, grasp correction, sim-to-real transfer
Mo <i>et al.</i> [67]	2021	Custom-built robot	Unity 3D, PhysX physics engine	Multiview cameras	Real-time monitoring, intelligent perception, feedback loops for policy updates
Cascone <i>et al.</i> [68]	2021	Pepper (humanoid robot)	Unity 3D	Cameras, infrared, laser, sonar sensors	Real-time monitoring, object recognition, safe object handling, emergency intervention
Xu <i>et al.</i> [69]	2021	KUKA KR3 R540 and KR6 R700	JavaScript and C-based simulation platform	Speed, acceleration, position	Real-time monitoring, remote control, robotic disassembly operations
Kaigom and Roßmann [70]	2021	KUKA LWR, IIWA, and KUKA KR 16	Tecnomatix, ROS, Unity 3D and CIROS	Joint torques, velocities, and other sensor data	Motion optimization, decision-making

RDTs: Robotic digital twins; DT: digital twins; ROS: robot operating system; RGB: red-green-blue.

function blocks for multi-source data integration [72]. Zhang *et al.* developed a generic architecture for constructing comprehensive industrial robot DT based on an ontology information model to address scalability and reusability challenges [73]. The ontology information model is a structured framework that represents knowledge about a particular domain using a formal ontology. In manufacturing, Chen *et al.* introduced a multi-sensor fusion-based DT for quality prediction in robotic laser-directed energy deposition additive manufacturing [74]. This method synchronizes multi-sensor features with real-time robot motion data to predict defects such as cracks and keyhole pores. Furthermore, Li *et al.* developed a DT model to automate the assembly process of the PCB kit box build, incorporating a symmetry-driven method to optimize pose matching and a quality prediction model using small displacement torsor (SDT) and Monte Carlo methods [75]. Liu *et al.* used DTs and a deep Q-learning network to dynamically optimize robotic disassembly sequence planning under uncertain missing conditions [76]. Wenna *et al.* proposed a DT framework for 3D path planning of large-span curved-arm gantry robots, integrating bilateral control, 3D mesh construction, and NavMesh-based multi-objective path planning to enhance safety and real-time performance [77]. Hu *et al.* proposed a grasps-generation-and-selection convolutional neural network trained and implemented in a DT of intelligent robotic grasping to identify optimal grasping positions [78]. Tipary and Erdős facilitated the iterative refinement of work cells in both digital and physical spaces, accelerating the commissioning and reconfiguration processes while minimizing physical work, demonstrated through a pick-and-place task and reconfiguration scenario [79].

As shown in Table 3, Unity 3D emerges as a widely used platform, providing capabilities such as 3D path planning, motion analysis, and dynamic sequence planning, which are critical for industrial tasks. Platforms such as MATLAB and MWorks are employed for predictive maintenance and lifecycle management, while

Table 3. Summary of recent RDTs in industrial applications

Ref.	Year	Physical part	Virtual platform	Data collection	Service
Zhang <i>et al.</i> [72]	2024	Stäubli TX200	Neo4j	Leica AT960 LT for large-scale measurements	Data exchange, assembly simulation, accuracy analysis
Zhang <i>et al.</i> [73]	2023	KUKA R2700	KR210 MATLAB, MWorks, Unity 3D	Torque, velocity, positional data	Real-time monitoring, control, predictive maintenance, lifecycle management
Chen <i>et al.</i> [74]	2023	KUKA robot	6-axis ROS	Infrared camera, CCD, microphone, laser sensor	Real-time monitoring, defect prediction, adaptive control
Li <i>et al.</i> [75]	2023	6-DOF manipulator	Rviz, Gazebo, OpenCASC-CADE	RealSense RGB-D, 3D sensors	Data exchange, assembly simulation, accuracy analysis
Liu <i>et al.</i> [76]	2023	KUKA LBR iiwa 14 R820	Unity 3D	3D point clouds, multi-perspective depth images	Real-time monitoring, dynamic sequence planning, disassembly
Wenna <i>et al.</i> [77]	2022	Gantry robot	Unity 3D	3D position data	3D path planning, motion analysis, bilateral control
Hu <i>et al.</i> [78]	2022	UR5E robotic arm	Bullet physics engine	RGB-D images	Grasp planning, quality evaluation, execution
Tipary and Er-dos [79]	2021	UR5 robotic arm, Robotiq 2F-85 gripper	LinkageDesigner, PQP, V-Collide	2D camera	Path planning, collision detection, task-specific uncertainty resolution

RDTs: Robotic digital twins; ROS: robot operating system; CCD: charge-coupled device; RGB: red-green-blue-depth.

ROS and OpenCASCADe focus on real-time monitoring and adaptive control. Physical components such as KUKA and UR5 robotic arms are commonly used, supporting tasks such as grasp planning, path optimization, and collision detection. Advanced data collection methods, including RGB-D sensors, infrared cameras, and 3D position data, enhance the accuracy and efficiency of these applications. Additionally, Neo4j and Leica AT960 LT are used for large-scale measurements and assembly simulations, emphasizing data exchange and analysis.

3.3. RDTs in HRCs

HRC is increasingly driven by the development of collaborative robots designed specifically to work alongside human operators [80]. Unlike traditional robots, which typically function in isolated environments, cobots are equipped with advanced sensors and AI capabilities that enable them to interact safely and effectively with human partners. These robots are frequently used to perform repetitive, hazardous, and physically demanding tasks, improving both safety and productivity in the workplace [81–83]. In the broader context of industrial evolution, Industry 5.0 is emerging as a new paradigm that emphasizes the central role of humans in industrial and manufacturing processes [18]. Unlike traditional industrial robots, which focus primarily on optimizing system performance by improving output and reducing costs [84,85], Industry 5.0 advocates the integration of human skills, creativity, and decision-making abilities with advanced technologies. This human-centric approach fosters innovation by combining the strengths of humans and robots in collaborative settings [86–88].

In the context of HRC, DTs are increasingly being used to enhance safety, efficiency, and overall user experience, as shown in Table 4. Park *et al.* proposed a DTs-based framework that used an exoskeleton-type robotic system to improve manufacturing processes [89]. This framework provides an efficient method by which robotic systems can learn knowledge and skills from the virtual representation of collaborative robots and operators. In the real HRC environment, the human operator wears the VR device and the exoskeleton robot to work with the virtual robot within the DTs. At the same time, the real robot performs the given task and transmits the sensor information to the virtual representation. Wang *et al.* introduced a deep learning (DL)-enhanced DTs framework to improve safety and reliability in HRC [90]. This framework enables intelligent decision-making by accurately detecting and classifying human and robot actions. In their real-world experiments, the robotic system was trained to detect human collisions based on information from virtual DTs, which might avoid

Table 4. Summary of the recent RDTs in HRC

Ref.	Year	Physical part	Virtual platform	Data collection	Service
Park <i>et al.</i> [89]	2024	Doosan Robotics (6-DOF arm)	Unity 3D	Force, torque sensor	Real-time monitoring, path planning, collision detection, task execution
Wang <i>et al.</i> [90]	2024	Universal Robots UR10	Unreal Engine 4	Kinect V2 for images, depth data	Real-time monitoring, safety decision-making, task execution
Kim <i>et al.</i> [17]	2024	Universal Robots UR3	RoboDK, machine learning toolkit	3D cameras (Intel RealSense D435)	Object detection, gripping action optimization (RL), path planning
Choi <i>et al.</i> [91]	2022	Universal Robots UR3	Unity 3D	Azure Kinect depth sensors, RGB-D sensors	Real-time monitoring, safety distance calculation, task assistance
Malik and Brem [53]	2021	Universal Robot UR5	Tecnomatix Process Simulate	Proximity, photoelectric, property, joint value sensors	Real-time monitoring, task allocation, collision detection, path optimization

RDTs: Robotic digital twins; HRC: human-robot collaboration; RL: reinforcement learning; RGB-D: red-green-blue-depth.

potential dangers to the human operator. Choi *et al.* developed an integrated mixed reality (MR) system for HRC that presents a fast and accurate 3D offset-based safety distance calculation [91]. In their test, the safety distance can be measured based on the RDT and the human skeleton instead of using 3D point cloud data. Furthermore, Kim *et al.* proposed a DT-based method for training cobots in virtual environments using synthetic data and the point cloud framework, addressing the challenges of time and cost constraints in real-world training [17]. Lastly, Malik and Brem explored the potential of DTs to enhance HRC in assembly processes by developing a DTs-based virtual counterpart for validation and control of a physical HRC system throughout its life cycle, discussing the building blocks and advantages of this approach [53].

As shown in Table 4, platforms such as Tecnomatix Process Simulate emphasize task allocation and path optimization, while Unreal Engine 4 is utilized for safety decision-making and task execution. RoboDK, combined with machine learning toolkits, focuses on object detection and action optimization based on reinforcement learning. Physical components, including Universal Robots (UR3, UR5, UR10) and Doosan Robotics arms, are equipped with advanced data collection systems such as Azure Kinect depth sensors, RGB-D sensors and 3D cameras. These systems enable precise force measurement, proximity sensing, and joint value monitoring, enhancing real-time task execution and safety in HRC scenarios.

4. KEY TECHNOLOGIES IN EMBODIED AI

Embodied AI enables intelligent agents (e.g., robots) to perceive, act, and learn in real-world environments. The successful deployment of Embodied AI systems depends on several key technologies, including perception, interaction, learning, and motion control. Although these technologies have significantly advanced embodied intelligence, they also face notable limitations, such as the sim-to-real gap, sensor inaccuracies, and challenges in real-time adaptation. This section provides a comprehensive review of the strengths and weaknesses of current Embodied AI technologies, identifying critical gaps that hinder their real-world implementation. In Section 5, it will explore how DTs can address these limitations by providing real-time virtual environments, improving data fidelity, and improving system adaptability for robust and scalable Embodied AI applications.

4.1. Sensing and perception

Sensing and perception are fundamental to Embodied AI, which is a prerequisite for allowing agents to understand and interpret the work environment. Embodied AI systems typically employ multi-modal sensors, such as cameras, LiDAR, sonar, and tactile sensors, to capture diverse data streams. By integrating information from multiple sensor modalities, Embodied AI agents can build a comprehensive representation of the environment to provide accurate decision-making and action, as shown in Figure 2.

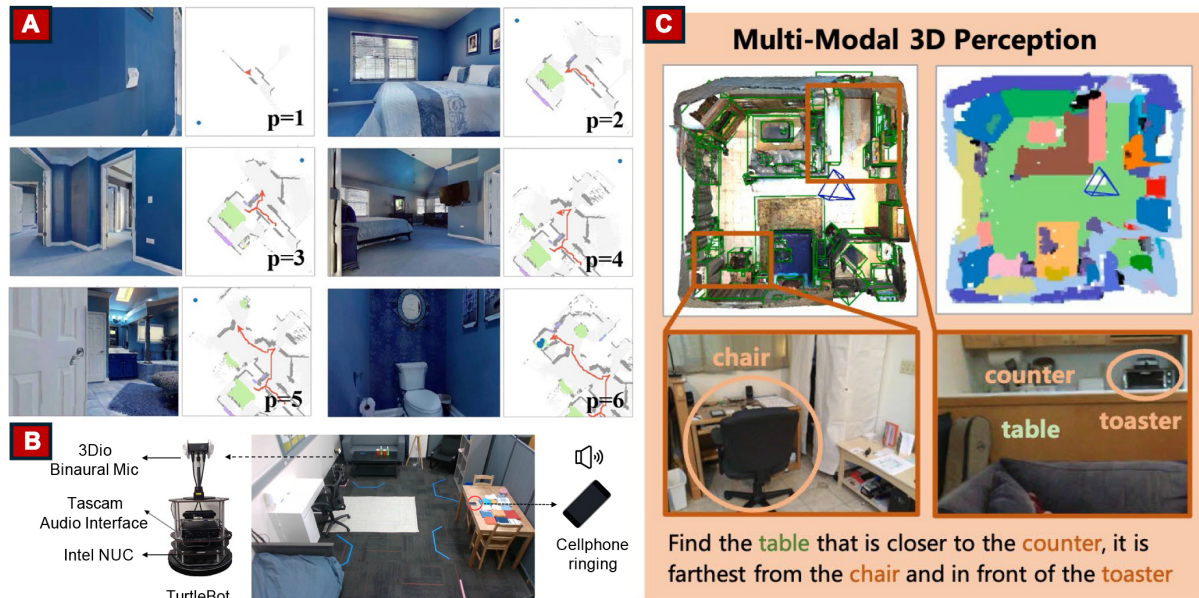


Figure 2. Examples of embodied perception. (A) The first perspective views of perception and bird-eye views of the robot when detecting the environment [92]; (B) TurtleBot platform that can collect both image and voice information [93]; (C) EmbodiedScan that benchmarks language-grounded holistic 3D scene understanding capabilities [94].

In the realm of depth perception and environmental understanding, He *et al.* focused on enhancing depth perception in dynamic physical environments by proposing a time-aware contrastive pre-training method [95]. This approach utilizes temporal similarities between frames to generate robust depth representations, allowing agents to better understand and interact with their surroundings across various domains. However, the significant computational resources required for pre-training, especially with large datasets such as UniRGBD, may limit its applicability in resource-constrained scenarios. Similarly, Jing and Kong introduced a dynamic exploration policy for embodied perception, which guides agents to efficiently gather training samples in new environments [96]. This method, while effective, faces challenges related to the computational cost of updating 3D semantic distribution maps and the reliance on depth sensors, which may limit its performance in environments with poor depth data or complex surfaces.

Focusing on communication and interpretability, Patel *et al.* explored the interpretability of communication between embodied agents and their underlying perception of the physical space around them [97]. They developed communication strategies tightly linked to the perceptions of agents, improving the understanding of the environment. However, the unstructured communication approach, while flexible, often results in less efficient and harder-to-interpret messages, reducing overall performance. The study highlights a gap between the performance of structured communication and the ideal scenario, indicating room for improvement in optimizing communication strategies for complex real-world tasks. In the area of object detection and forecasting, Peri *et al.* addressed the often-isolated study of these problems by proposing the FutureDet method, which reframes trajectory forecasting as future object detection based on LiDAR data [98]. While effective in handling linear and stationary objects, the method shows limitations in dealing with non-linear trajectories, potentially leading to overfitting in constant velocity scenarios. This limitation highlights the need for further refinement to enhance the adaptability to more dynamic and unpredictable environments. Guo *et al.* proposed an object-driven navigation strategy based on active perception and a semantic association model using RGB-D sensors, achieving improved navigation success and efficiency in dynamic environments [92].

When it comes to multi-modal perception and simulation, Gao *et al.* introduced the Sonicverse platform,



Figure 3. Examples of embodied planning and control. (A) Safe collision-avoidance embodied navigation^[99]; (B) Embodied motion control for the dual-robot system of the bionic robot controller^[100]; (C) Example of motion planning for response to the operator's request.^[101]; (D) Skill Transformer is a unified policy model that directly maps visual observations (pixels) to actions, enabling a robot to autonomously perform complex, multi-step tasks such as mobile manipulation by inferring the required skills and executing precise control actions^[102].

which allows agents to perceive their surroundings using both sight and sound within a simulated environment^[93]. Despite its potential, the high computational intensity required to render high-fidelity audio and visual simulations in real time may limit its accessibility for some applications. In addition, the platform excels in static environments, but dynamic changes present challenges, particularly in simulating complex interactions between moving objects and varying sound sources. Complementing this, Wang *et al.* presented the EmbodiedScan dataset, a comprehensive multi-modal 3D perception dataset that integrates visual, depth, and language data to facilitate holistic 3D scene understanding^[94]. Although the dataset improves the ability of embodied agents to interpret and interact with complex indoor environments, the computational complexity of processing and integrating the diverse data streams poses challenges, especially in real-time applications.

4.2. Motion planning and control

Motion planning and control are fundamental components of Embodied AI, enabling robotic systems to navigate through and interact with their environments effectively. These processes can be broadly categorized into hierarchical and optimization-based planning, morphology-based control, and vision or transformer-based planning approaches, as shown in Figure 3.

Hierarchical structures and optimization techniques are used to achieve effective motion planning and control. These approaches often involve breaking down complex tasks into manageable sub-tasks or optimizing specific criteria to ensure efficient navigation and interaction with the environment. Gan *et al.* proposed a hierarchical planning and control method utilizing RGB-D sensors for task and motion planning in dynamic environments^[99]. Although ThreeDWorld provides high-fidelity physical environments, the paper does not address limitations related to computational resources and real-time performance, which may affect its feasi-

bility for large-scale applications. In addition, experiments also do not fully explore the diversity of physical properties in the environment, such as friction and gravity, and their potential impact on task performance. Li *et al.* introduced an optimization-based trajectory planning method based on vehicle kinematics and geometry^[103]. The model achieves a good balance between time optimality and safety, with a denser distribution of discrete points in bottleneck areas. However, the simulations mainly focus on simple static obstacle environments, without considering dynamic obstacles or larger and more complex real-world traffic scenarios, such as urban driving. Furthermore, although field tests were conducted, the testing scale was small (1.75 m × 1.20 m indoor environment), and its applicability to large-scale real-world vehicles has not been verified.

Some researchers used embedding control mechanisms directly into the physical design or morphology of the robot. The morphology control methods use the structure of the robot to minimize computational requirements and enhance task performance, although they may face limitations in adaptability and sensor precision. Pervan and Murphey presented a novel algorithmic design method to transfer task information from centralized control to the physical structure of robots, enabling task embodiment through physical morphology^[104]. However, the study does not address dynamic obstacles or complex external environments, which may limit its applicability. In addition, the simulation environment is idealized, and the method's robustness in real-world physical systems remains insufficiently validated. Gan *et al.* explored a morphology-based approach with a bionic robot controller that integrates vision and motion planning, showing promise in dual-robot collaborations^[100]. Although a solution for dynamic environments is proposed, the experimental setup is relatively idealized and does not sufficiently validate the performance in more complex and unpredictable scenarios. In addition, detailed analysis and testing of dynamic obstacles and multi-agent interaction are lacking.

Vision-based perception and advanced machine learning techniques, such as transformers, are also used to enable robust motion planning and control in complex environments. These approaches often aim to improve the robot's ability to understand and react to dynamic and uncertain environments. Behrens *et al.* developed an embodied reasoning system that integrates vision-based scene understanding with action planning, achieving significant success in dynamic environments but requiring further refinement for complex real-world applications^[101]. Huang *et al.* introduced the Skill Transformer, a transformer-based policy that integrates visual and proprioceptive inputs, excelling in long-horizon tasks but challenged by computational efficiency^[102]. The Skill Transformer takes as input a sequence of tokens that encode the observations over time. Each token represents its state and is made up of visual observations (depth-only), non-visual observations (joint angles and the holding state of the end effector) and positional embedding. The Skill Inference module was then used to process tokens using a causal transformer. By applying a causal mask, the transformer outputs a one-hot skill vector that represents the predicted skill to execute. As the final step, the Action Inference module utilizes inferred skill, observational data, and previous action information to determine the low-level action that the robot will execute.

4.3. Learning and reasoning

Learning and adaptation are central to the development of intelligent Embodied AI agents, enabling them to optimize their behavior through experience and adapt to new tasks, environments and situations, as shown in [Figure 4](#).

Reinforcement learning has become a vital tool in embodied intelligence. It allows agents to interact with physical or simulated environments and progressively learn to handle complex conditions and tasks. A significant strength of reinforcement learning in this context is its ability to adapt and learn online. By trying out actions and observing their consequences, reinforcement learning algorithms continually refine their policies in response to changing circumstances. This adaptability is especially beneficial in dynamic or uncertain environments, where predefined rules may not be sufficient. Another key advantage of reinforcement learning is that it does not require a precise model of the environment or the agent's dynamics. Many real-world

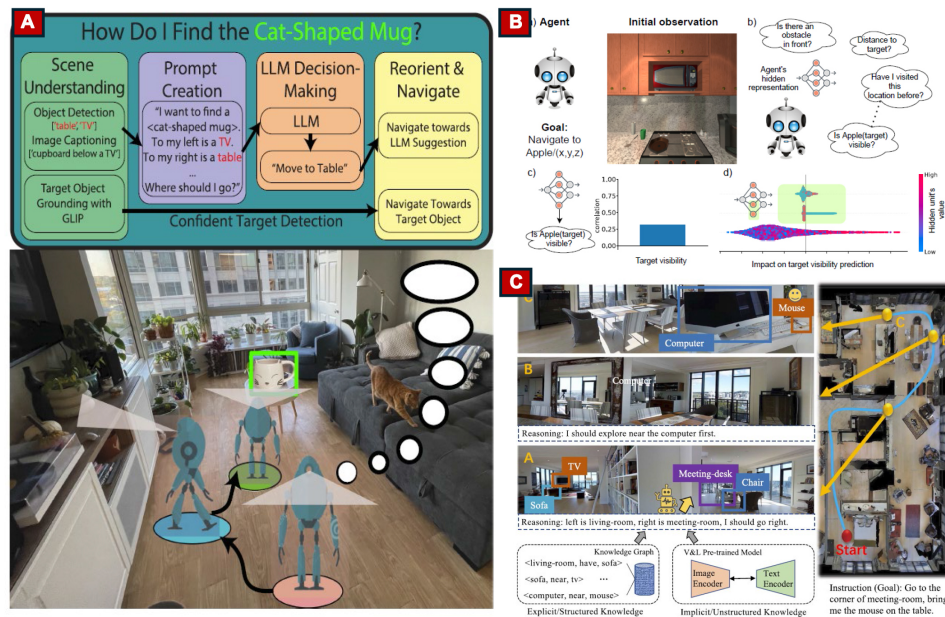


Figure 4. Examples of embodied learning and reasoning. (A) Example of navigation and detect the object based on LLM model in unknown environments [105]; (B) The iSEE framework that can learn the environmental information when the robot travels in the environment [106]; (C) Example of reasoning and producing navigation decisions based on prior knowledge from the knowledge graph and the pre-trained model [107]. LLM: Large language model; iSEE: interpretability system for embodied agents.

systems, such as those involving complex contact forces or multi-body interactions, are difficult to describe with accurate mathematical models. Reinforcement learning circumvents this challenge by using feedback directly from the environment to improve the agent's behavior. As a result, it can be applied to a broader range of problems where modeling is either infeasible or extremely costly. Zhao *et al.* introduced the embodied representation and reasoning architecture, which combines LLMs with reinforcement learning, depth, and tactile sensors to enhance long-horizon manipulation tasks [108]. Despite its success, the system is constrained by its reliance on simulated training data and hardware limitations. Liu *et al.* proposed a hierarchical decision framework for multi-agent embodied visual semantic navigation, which uses reinforcement learning and scene prior knowledge, along with RGB and depth sensors [109]. This framework improves collaboration and efficiency in navigation tasks, though it faces challenges related to communication bandwidth limitations and sensor accuracy.

However, reinforcement learning models provide minimal to no inherent interpretability regarding the concepts and skills they have learned or the reasoning behind their actions. Dwivedi *et al.* investigated the environmental information learned in the internal representations of Embodied agents trained using reinforcement learning and supervised learning [106]. The study introduces an interpretability framework to analyze the information AI agents encode about their surroundings. These include obstacle locations, target visibility, and reachable areas, which enable effective movement and task completion. In addition, one major challenge in applying reinforcement learning to real-world robotic tasks is its low sample efficiency. Most reinforcement learning algorithms learn a policy by interacting with the environment many times, which can be costly and risky in real robot systems. Each interaction involves physical motions that cause wear to the hardware and may pose safety hazards. For instance, a small mistake during training can damage robot parts or endanger nearby people. To reduce these risks, researchers commonly train agents in simulation first, where mistakes are safer and cheaper. However, this approach leads to the sim-to-real gap. Even the best simulations cannot perfectly capture every detail of the physical world, such as slight differences in friction, vibrations in the joints, or unexpected sensor noise. Because of these discrepancies, a policy that performs well in simulation may fail

when transferred to a real robot.

Incorporating language and vision models has been another focus area, with the aim of improving the interaction with its environment. Dorbala *et al.* developed a language-guided exploration method that uses LLMs and vision-language models using RGB and depth sensors, achieving superior zero-shot object navigation performance^[105]. However, this method is computationally intensive and heavily dependent on high-quality sensor data. Similarly, Sermanet *et al.* introduced the RoboVQA framework, which utilizes video-conditioned contrastive captioning with RGB and depth sensors for multi-modal long-horizon reasoning in robotics^[110]. Although this framework significantly improves task execution, it requires high computational resources and quality sensor data. Another category includes approaches that focus on using 3D data for visual learning and navigation tasks. Zhao *et al.* proposed a 3D divergency policy based on real 3D point cloud data, effectively improving robotic active visual learning in novel environments^[111]. However, this method introduces challenges related to handling sensor noise and the high resource demands associated with processing 3D data.

Finally, the integration of cross-modal reasoning and transformer-based architectures has been explored to enhance adaptability in complex environments. Cross-modal reasoning refers to the ability of an AI system to integrate and reason over information from multiple sensory or data modalities, such as vision, language, audio, and tactile inputs. Gao *et al.* introduced the cross-modal knowledge reasoning framework, which utilizes the transformer architecture along with RGB sensors and object detectors^[107]. This framework effectively enhances navigation and object localization in embodied referring expression tasks. However, it is challenged by computational complexity and a heavy reliance on pre-trained models.

4.4. Interaction and collaboration

Interaction and collaboration are essential for Embodied AI systems to function effectively in environments where they must work alongside humans, other agents, or complex systems. Embodied AI agents are designed to interact with their surroundings, communicate with other agents, and collaborate with humans to achieve shared goals, as shown in [Figure 5](#).

In the domain of real-time interaction and sensory integration, Sagar *et al.* proposed the BabyX platform, which employs visual and touch sensors for real-time interaction and event processing in a virtual infant model^[114]. This platform achieves realistic caregiver-infant interactions, demonstrating the potential of Embodied AI in simulating complex human-like interactions. However, the platform faces challenges with sensor sensitivity and the complexity of motor execution, which can influence overall performance in dynamic scenarios. Similarly, Legrand *et al.* developed a variable stiffness anthropomorphic finger that uses pneumatic pressure sensors to dynamically adjust joint stiffness^[115]. This innovation enables real-time adaptability and enhanced interaction stability, although it requires complex control systems and may experience delays during rapid force transitions, which presents challenges in fast-paced environments.

For cognitive modeling and perspective-taking, Fischer and Demiris introduced a computational model for visual perspective-taking^[116]. This model uses simulated visual sensors and action primitives to enable efficient cognitive modeling and rapid response times, making it suitable for tasks requiring quick mental adaptations. However, its reliance on a simulated environment and task-specific optimizations limits its applicability to more generalized real-world settings, where unexpected variables and broader contexts may come into play. Focusing on human-in-the-loop learning and simulation, Long *et al.* proposed an interactive simulation platform designed for surgical robot learning^[112]. This platform leverages haptic devices and visual sensors to enhance learning efficiency through embodied intelligence, where human input is directly integrated into the learning process. The platform achieves realistic interactions and improves the training of surgical robots, though it faces challenges related to the complexity of the simulation and its heavy dependence on expert demonstrations, which can limit scalability and broader application.

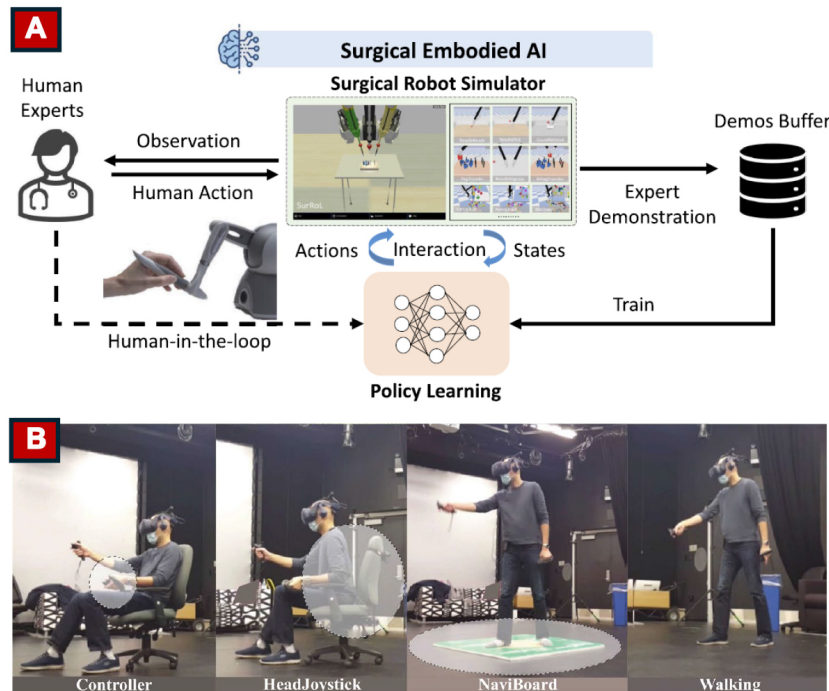


Figure 5. Examples of embodied learning and reasoning. (A) Example of human-in-the-loop demonstrations for robot learning^[112]; (B) An example of VR-based interaction affects user experience, usability, and effectiveness^[113].

In the area of VR interfaces and interaction, Hashemian *et al.* developed learning-based interfaces that use head movement sensors in head-mounted displays to improve VR locomotion and object interaction^[113]. These interfaces improve spatial presence and multitasking performance in VR environments, offering a more immersive experience. However, the interfaces exhibit lower precision compared to physical walking and impose increased physical demands on users, which could impact usability and comfort during extended VR sessions.

5. PERSPECTIVE OF SYNERGIES BETWEEN DTS AND EMBODIED AI

The intersection of DTs and Embodied AI presents an unprecedented opportunity to refine intelligent systems for dynamic and uncertain environments. While DTs create real-time, high-fidelity virtual representations of the physical world, Embodied AI uses the virtual space for perception, learning, and interaction. This synergy enables embodied agents to train in digital replicas of real-world environments, reducing training costs, minimizing safety risks, and improving real-time adaptability. In this section, novel perspectives of the synergy of DTs and Embodied AI are discussed to anticipate the future development of DTs-based Embodied AI systems.

5.1. Embodied AI training based on DTs virtual space

DTs technologies play a central role by creating a real-time virtual environment. As shown in Tables 2-4, DTs are designed to mirror physical systems in real time. This real-time synchronization offers significant advantages for Embodied AI training, enabling more efficient learning, perception, and decision-making in complex environments. DTs continuously collect and process data from physical systems using sensors such as depth cameras, RGB-D sensors^[75,91], and force-torque sensors^[89]. The collected real-time data are then mapped into a virtual space to ensure real-time reflection of physical states. Real-time feedback is essential for Embodied AI, as it allows AI models to adjust their decisions based on current environmental changes.

Table 5. Comparison of traditional simulations and DTs

Aspect	Traditional simulations	DTs
Connection with physical systems	Often isolated virtual models, approximate real-world conditions	Maintains continuous, real-time synchronization with physical systems
Real-time feedback	Limited or no real-time feedback, usually operates on predefined scenarios	Provides real-time feedback based on sensor data, enabling dynamic decision-making
Data source	Typically uses predefined models or simulated data	Utilizes real-world sensor data for accurate reflection of physical states
Interactivity	Limited or static interaction with the system, lacks two-way communication	Two-way interaction with physical systems, virtual models directly influence physical systems
Flexibility in environment simulation	Limited flexibility in simulating complex, changing environments	High flexibility, easily simulates dynamic, complex environments and adapts to real-time changes
Application to embodied AI training	Often struggles to train AI models that adapt well to real-world environments	Enables Embodied AI to refine decision-making and adapt to dynamic real-world environments, reducing the "sim-to-real gap"

DTs: Digital twins; AI: artificial intelligence.

In comparison to traditional simulations, DTs maintain a continuous, real-time connection with physical systems. Although traditional simulations often create isolated virtual models that approximate real-world conditions, DTs enable seamless synchronization with the actual physical environment. The close coupling allows the embodied agent to operate within a virtual space that more accurately mirrors real-world conditions. As a result, embodied agents can develop and refine their decision-making processes based on real-time physical data, significantly reducing the sim-to-real gap. In addition, DTs emphasize real-time interactive ability, which is dynamic and often connected to the ongoing processes of physical systems. This two-way interaction means that the virtual model is continuously updated with data from the physical system, while decisions made within the virtual environment can directly influence the physical system. Finally, DTs are based on actual sensor data from physical systems, providing a more realistic and precise training environment for embodied AI. Traditional simulations often depend on predefined models that may not fully capture the complexities or variabilities of real-world scenarios. In contrast, DTs feed real-world data, such as force measurements, temperature readings, or positional data, into the virtual environment, ensuring that embodied AI is exposed to more accurate and varied conditions during its training. This data-driven approach enables embodied AI to better anticipate and respond to the unpredictable and dynamic nature of real-world situations. Table 5 summarizes the comparison of traditional simulations and DTs. DTs offer better flexibility by simulating complex environments to enable AI to train in a wide range of scenarios and improve its ability to adapt to real-world challenges.

However, using DTs as virtual training environments also faces challenges, such as data latency and excessive data transfer. To address these challenges, several practical solutions and technologies can be implemented: (1) Data Compression and Optimization: Data compression techniques can be applied to reduce the size of data before transmission while retaining essential information. This can be achieved through methods such as principal component analysis (PCA) for dimensionality reduction^[117], neural network-based compression techniques such as auto-encoders to extract compact data representations^[118], and video or image compression using efficient formats; (2) Decentralized Architectures: Distributing data processing and storage across multiple nodes is an effective way to avoid bottlenecks in centralized systems. This can be implemented by using Apache Kafka^[119] for real-time data streaming and communication between distributed nodes and employing distributed databases such as MongoDB or Cassandra^[120] to ensure scalable and efficient data storage; (3) Hybrid and Multi-Cloud Architectures: Adopting hybrid and multi-cloud architectures can provide flexible and scalable solutions to manage data processing and storage efficiently. Multi-cloud architectures allow organizations to distribute workloads across different cloud providers, reducing dependency on a single platform and improving system resilience. Implementing container orchestration tools such as Kubernetes ensures seamless integration between cloud environments, while data transfer optimization tools such as AWS DataSync

or Google Transfer Appliance streamline data movement^[121]; (4) 5G/6G communication network: Using advanced network technologies such as 5G and emerging 6G standards can provide high-bandwidth, low-latency communication, enabling real-time data transmission and reducing delays in critical operations^[122]. These networks are particularly effective for supporting massive IoT deployments and high-speed data exchange in dynamic environments^[123].

5.2. Embodied AI with human DTs

As shown in Sections 3.3 and 4.4, DTs and Embodied AI are widely applied in HRC. However, only a few studies consider the human aspect or merely treat humans as simple obstacles. If human behavior and status become more predictable, robots can provide more efficient strategies for human operators and generate data-informed decisions to assist human operators. However, modeling and simulation of human status and behaviors is difficult in HRC. Firstly, compared to robotic systems, human modeling requires a larger dataset, covering more complex geometric configurations, the acquisition of biological signals, and significant variations among different individuals. Secondly, human behavior modeling demands higher real-time performance due to the dynamic and unpredictable nature of human actions. This involves processing a continuous stream of data from multiple sensors to accurately track and respond to human movements and states. In addition, human performance and behavior are also subject to subjective influences, such as psychological stress and fatigue levels, which can significantly affect both physical and cognitive functions.

The development prospects of human digital twins (HDTs) are promising, particularly in enhancing HRC^[124]. The HDTs hold significant potential for improving the interaction between embodied agents and humans^[125]. Despite ongoing debates on the definition of HDTs, it is generally agreed that HDTs are aimed at collecting comprehensive information to create a virtual digital human^[126]. As shown in Figure 6, HDTs collect both physical and mental data from humans, allowing AI to adjust its behavior in real time during virtual training. Instead of treating humans as simple agents, HDTs enable AI to recognize and adapt to the physical and mental states of humans, such as fatigue, stress, or movement patterns. This adaptive capability is crucial in building human-centric systems, where AI can respond more intelligently and sensitively to human needs and conditions.

In addition, HDTs have widely potential applications in the field of human-robot interaction. They can support real-time monitoring of human body state within human-robot interactions. Through the use of AI technologies, the robotic system simultaneously tracks 3D posture, action intention, and ergonomic risk, which is a more holistic view of the human operator than traditional methods^[127]. Beyond the body information, HDTs can continuously monitor human fatigue state during human-robot collaborative assembly. They can assess worker real-time fatigue in ongoing tasks to reallocate tasks between humans and robots to reduce physical fatigue, thus optimizing the overall HRC in manufacturing^[128]. Moreover, training robots to interact with humans in real time can carry inherent risks, especially in scenarios that require immediate responses, such as emergency interventions or collaborative tasks. By simulating human responses and conditions within a virtual space, HDTs allow robots to learn without posing a direct threat to human safety. This approach ensures that robots can develop the necessary skills and reaction times without the risk of causing physical harm during the training process^[91].

The future of HDTs in practical applications of Embodied AI holds significant promise. HDTs can enable embodied agents to work more effectively in dynamic and human-centric environments by integrating real-time data on human physical and mental states. For example, in healthcare, HDTs can simulate patient conditions, allowing embodied agents to provide personalized care, such as monitoring fatigue or stress levels and adapting their actions to meet individual needs^[129]. In manufacturing, HDTs can help improve worker safety and efficiency by analyzing operator movements and suggesting optimized workflows to reduce strain or prevent injuries^[130]. Similarly, in autonomous systems, HDTs can create realistic simulations of human behavior,

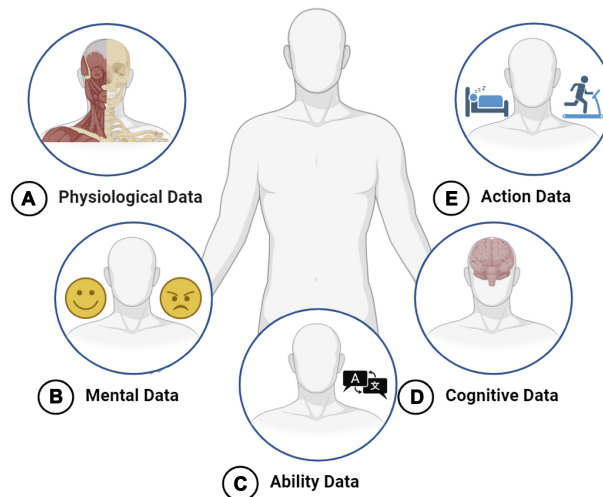


Figure 6. Common types of information collected include: (A) physiological data refers to the biological information that reflects the functions and processes of the human body, such as heart rate, blood pressure, body temperature, respiratory rate, and ECG readings; (B) mental data refers to information related to an individual's emotional states, encompassing aspects of mental health, mood, and psychological well-being; (C) ability data refers to information that captures an individual's capabilities, reflecting how well a person can execute certain functions, which can be physical, such as motor skills, or cognitive, such as problem-solving abilities; (D) cognitive data refers to information related to an individual's reason processes involved in acquiring knowledge, problem-solving, memory, attention, and decision-making; and (E) action data refers to information that captures an individual's physical movements, behaviors, or actions in response to stimuli or within a specific environment. ECG: Electrocardiogram.

enabling embodied agents to predict and respond to human actions more intelligently during tasks such as emergency evacuations or collaborative work^[131]. These advances can lead to safer and more efficient interactions between humans and embodied agents, supporting a wide range of industries that require seamless and adaptive collaboration.

However, constructing HDTs is highly challenging, particularly when considering the complexities of human behavior and environmental interactions in large-scale implementations. To address this challenge with real-time adaptability, the following technologies can be applied: (1) Multi-modal Data Collection and Fusion: Use biosensors [e.g., electroencephalography (EEG), electrocardiogram (ECG)], environmental sensors (e.g., temperature, humidity, gas sensors), and visual sensors (e.g., RGB cameras, depth sensors)^[132]. Employ data fusion methods such as Kalman filters, Bayesian networks, or DL to integrate multiple data streams to create a comprehensive and accurate representation of human and environmental states; (2) Real-Time Data Processing: Implement edge computing using devices such as NVIDIA Jetson^[133] or Intel Movidius^[134] for local data processing to reduce latency. Using real-time data streaming frameworks such as Apache Kafka or Apache Flink to ensure fast responses to dynamic environmental changes; (3) Dynamic Modeling and Learning: Use dynamic neural networks such as graph neural networks (GNNs) or recurrent neural networks (RNNs) to model complex relationships in dynamic environments. Apply reinforcement learning algorithms to optimize adaptation and decision-making; (4) DTs Simulation Platforms: Leverage simulation tools such as NVIDIA Isaac Sim, Unity, or Gazebo to create realistic virtual environments. Integrate data-driven modeling using Python or MATLAB to update simulations with real-time input. These technologies work together to enable HDT systems to operate effectively in complex dynamic environments while ensuring reliability, scalability, and ethical compliance.

5.3. Embodied AI in large-scale virtual environments

Currently, most Embodied AI systems are limited to small, indoor environments due to the substantial time and economic costs associated with scaling up to larger, more complex settings such as entire cities. The development prospects of environmental digital twins (EDTs) are especially in advancing AI applications beyond

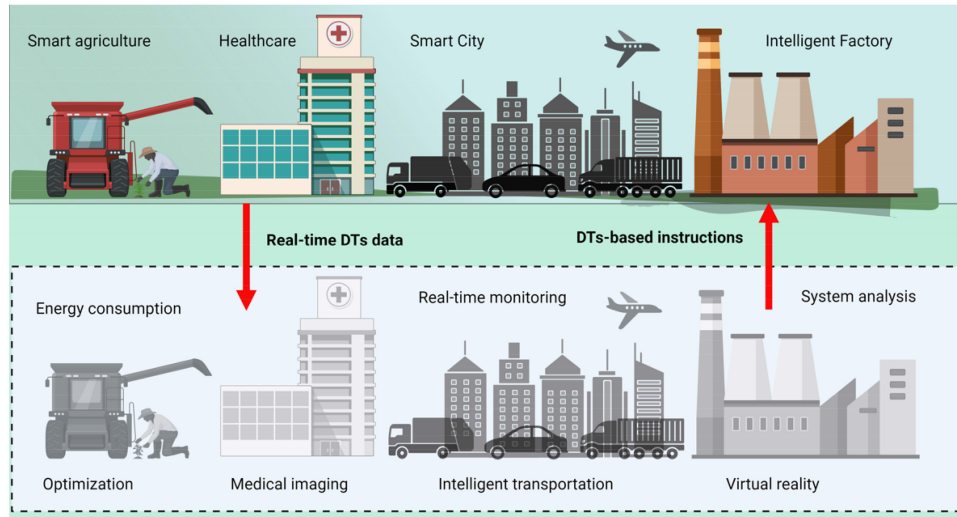


Figure 7. Example of EDTs. DTs technology in both urban and agricultural contexts offers significant advantages by enabling data-driven, real-time simulations. These applications not only optimize resource allocation but also reduce environmental risks, contributing to the sustainable development of cities and agricultural systems. EDTs: Environmental digital twins; DTs: digital twins.

confined indoor environments. EDTs focus on creating digital representations of environments, such as urban areas, providing a comprehensive and dynamic virtual model of complex settings, as shown in [Figure 7](#).

Urban digital twins (UDTs) are increasingly utilized in smart city applications to create highly detailed virtual models of urban environments^[135]. UDTs replicate various aspects of a city, such as infrastructure, traffic patterns, and even environmental factors such as air quality and energy consumption. For Embodied AI, UDTs offer a controlled yet expansive environment for training AI systems designed for tasks such as autonomous driving, public safety monitoring, and traffic management^[136]. By allowing Embodied agents to navigate and interact with complex city environments, UDTs help train Embodied AI models at a large scale with effective operation in real-world conditions. Moreover, training AI in virtual smart city environments can significantly reduce the costs and risks associated with real-world experimentation. Since these DTs provide highly accurate simulations, AI can be trained to respond to traffic incidents, public emergencies, or urban planning scenarios without causing harm to the physical environment or human users^[137].

DTs in agriculture are transforming the modern farming industry. They create virtual representations of agricultural fields, crops, and farming machinery^[138]. For future Embodied AI systems involved in precision agriculture, such as autonomous tractors, drones, or robotic harvesters, DTs allow for safe and cost-effective training in large-scale farming environments. These virtual environments enable Embodied AI to learn how to handle tasks such as soil monitoring, crop management, and resource optimization without the risk of damaging valuable equipment or harming the environment. Using DTs reduces the need for expensive real-world trials, as Embodied AI systems can be fully trained in virtual environments before being deployed in actual fields. In addition, DTs allow for continuous optimization by simulating a range of environmental conditions, such as varying weather patterns or soil types, helping AI adapt to a wide range of agricultural scenarios^[40].

6. CONCLUSIONS

This paper provides a comprehensive review of DTs and Embodied AI, including the foundational principles of DTs and Embodied AI, their ability to synchronize with physical systems in real time, and their applications in automation, HRC, and large-scale environments. In addition, this paper provides a novel perspective on the potential integration of DTs and Embodied AI in future developments. The integration of DTs and Embodied

AI offers a transformative pathway for advancing the capabilities of AI across various domains. DTs provide highly accurate, real-time virtual replicas of physical systems, enabling AI to train, adapt, and perform in environments that closely mirror real-world conditions. This synergy significantly improves AI decision-making, perception, and adaptability, whether in human-centric applications, large-scale urban or agricultural environments, or real-time industrial settings. By leveraging the continuous synchronization between physical and virtual systems, DTs reduce the “sim-to-real gap” and offer more dynamic, interactive, and cost-effective solutions for AI development. The combination of DTs and Embodied AI paves the way for safer, more efficient, and scalable intelligent systems.

Future work in this field should consider the development of training and learning algorithms that use DTs for system and application optimization. While current DTs research mainly focuses on the simulation aspect, it often overlooks the critical potential of DTs to enhance system-level and application-specific performance. This area requires integrating real-time data-driven insights from DTs into learning processes. Another critical direction for future work is to address the limitations associated with the precision of data, the responsiveness in real time, and the high computational demands of DTs in large-scale applications. These challenges are often underexplored, yet they significantly impact the scalability and effectiveness of DTs in practical scenarios. Research should focus on developing techniques to enhance the fidelity of data, optimize computational efficiency, and improve the real-time capabilities of DTs, particularly in complex environments.

DECLARATIONS

Authors' contributions

Made substantial contributions to the research and investigation process, reviewed and summarized the literature, wrote and edited the original draft: Li, L.; Yang, S. X.

Made substantial contributions to review and summarize the literature: Li, L.

Performed oversight and leadership responsibility for the research activity planning and execution, developed ideas and provided critical review, commentary, and revision: Yang, S. X.

Availability of data and materials

Not applicable.

Financial support and sponsorship

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Conflicts of interest

Li, J. is a Junior Editorial Board Member of the journal *Intelligence & Robotics*, and Yang, S. X. is the Editor-in-Chief of the journal *Intelligence & Robotics*. Li, J. and Yang, S. X. were not involved in any steps of editorial processing, notably including reviewers' selection, manuscript handling and decision making.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2025.

REFERENCES

1. Gupta, A.; Savarese, S.; Ganguli, S.; Fei-Fei, L. Embodied intelligence via learning and evolution. *Nat. Commun.* **2021**, *12*, 5721. DOI
2. Nygaard, T. F.; Martin, C. P.; Torresen, J.; Glette, K.; Howard, D. Real-world embodied AI through a morphologically adaptive quadruped robot. *Nat. Mach. Intell.* **2021**, *3*, 410-9. DOI
3. Li, C.; Zhang, R.; Wong, J.; et al. BEHAVIOR-1K: a benchmark for embodied AI with 1,000 everyday activities and realistic simulation. In: *Proceedings of The 6th Conference on Robot Learning, PMLR*; 2023. pp. 80–93. <https://proceedings.mlr.press/v205/li23a.html>. (accessed 2025-02-28)
4. Driess, D.; Xia, F.; Sajjadi, M. S.; et al. PaLM-E: an embodied multimodal language model. *arXiv* **2023**, arXiv:2303.03378. Available online: <https://doi.org/10.48550/arXiv.2303.03378>. (accessed on 28 Feb 2025)
5. Duan, J.; Yu, S.; Tan, H. L.; Zhu, H.; Tan, C. A survey of embodied AI: from simulators to research tasks. *IEEE. Trans. Emerg. Top. Comput. Intell.* **2022**, *6*, 230–44. DOI
6. Kolve, E.; Mottaghi, R.; Han, W.; et al. AI2-THOR: an interactive 3D environment for visual AI. *arXiv* **2017**, arXiv:1712.05474. Available online: <https://doi.org/10.48550/arXiv.1712.05474>. (accessed on 28 Feb 2025)
7. Shen, B.; Xia, F.; Li, C.; et al. iGibson 1.0: a simulation environment for interactive tasks in large realistic scenes. In: *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Prague, Czech Republic, 27 Sep - 01 Oct, 2021. IEEE, 2021; pp. 7520–7. DOI
8. Savva, M.; Kadian, A.; Maksymets, O.; et al. Habitat: a platform for embodied AI research. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea, 27 Oct - 02 Nov, 2019. IEEE, 2019; pp. 9339–47. DOI
9. Liu, Y.; Chen, W.; Bai, Y.; et al. Aligning cyber space with physical world: a comprehensive survey on Embodied AI. *arXiv* **2024**, arXiv:2407.06886. Available online: <https://doi.org/10.48550/arXiv.2407.06886>. (accessed 28 Feb 2025)
10. Auerbach, J.; Aydin, D.; Maesani, A.; et al. RoboGen: robot generation through artificial evolution. In: *ALIFE 14: The Fourteenth International Conference on the Synthesis and Simulation of Living Systems*; 2014. pp. 136–7. DOI
11. Yang, Y.; Sun, F. Y.; Weihs, L.; et al. Holodeck: Language guided generation of 3D Embodied AI environments. In: *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, USA, 16-22 Jun, 2024. IEEE, 2024; pp. 16227–37. DOI
12. Yang, Y.; Jia, B.; Zhi, P.; Huang, S. PhysScene: physically interactable 3D scene synthesis for Embodied AI. In: *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, USA, 16-22 Jun, 2024. IEEE, 2024; pp. 16262–72. DOI
13. Deitke, M.; VanderBilt, E.; Herrasti, A.; et al. ProcTHOR: large-scale Embodied AI using procedural generation. *arXiv* **2022**, arXiv:2206.06994. Available online: <https://doi.org/10.48550/arXiv.2206.06994>. (accessed 28 Feb 2025)
14. Chang, A.; Dai, A.; Funkhouser, T.; et al. Matterport3D: learning from RGB-D data in indoor environments. *arXiv* **2017**, arXiv:1709.06158. Available online: <https://doi.org/10.48550/arXiv.1709.06158>. (accessed 28 Feb 2025)
15. Wang, H.; Lv, L.; Li, X.; et al. A safety management approach for Industry 5.0's human-centered manufacturing based on digital twin. *J. Manuf. Syst.* **2023**, *66*, 1–12. DOI
16. Lu, Y.; Liu, C.; Wang, K. I. K.; Huang, H.; Xu, X. Digital twin-driven smart manufacturing: connotation, reference model, applications and research issues. *Robot. Comput. Integr. Manuf.* **2020**, *61*, 101837. DOI
17. Kim, D.; Choi, M.; Um, J. Digital twin for autonomous collaborative robot by using synthetic data and reinforcement learning. *Robot. Comput. Integr. Manuf.* **2024**, *85*, 102632. DOI
18. Lu, Y.; Zheng, H.; Chand, S.; et al. Outlook on human-centric manufacturing towards Industry 5.0. *J. Manuf. Syst.* **2022**, *62*, 612–27. DOI
19. Mazumder, A.; Sahed, M. F.; Tasneem, Z.; et al. Towards next generation digital twin in robotics: trends, scopes, challenges, and future. *Heliyon*. **2023**, *9*, e13359. DOI
20. Zafar, M. H.; Langås, E. F.; Sanfilippo, F. Exploring the synergies between collaborative robotics, digital twins, augmentation, and industry 5.0 for smart manufacturing: a state-of-the-art review. *Robot. Comput. Integr. Manuf.* **2024**, *89*, 102769. DOI
21. Allen, B. D. Digital twins and living models at NASA. In: *Digital Twin Summit*; 2021. <https://ntrs.nasa.gov/citations/20210023699>. (accessed 2025-02-28)
22. Jarvis, C. R. An overview of NASA's digital fly-by-wire technology development program. 1976. <https://ntrs.nasa.gov/citations/19760024052>. (accessed 2025-02-28)
23. Grieves, M. Digital twin: manufacturing excellence through virtual factory replication. <https://www.3ds.com/fileadmin/PRODUCTS-SERVICES/DELMIA/PDF/Whitepaper/DELMIA-APRISO-Digital-Twin-Whitepaper.pdf>. (accessed 2025-02-28)
24. Jacoby, M.; Usländer, T. Digital twin and internet of things - current standards landscape. *Appl. Sci.* **2020**, *10*, 6519. DOI
25. Suhail, S.; Hussain, R.; Jurdak, R.; Hong, C. S. Trustworthy digital twins in the industrial internet of things with blockchain. *IEEE. Int. Comput.* **2021**, *26*, 58–67. DOI
26. De Benedictis, A.; Flammini, F.; Mazzocca, N.; Somma, A.; Vitale, F. Digital twins for anomaly detection in the industrial Internet of Things: conceptual architecture and proof-of-concept. *IEEE. Trans. Ind. Inform.* **2023**, *19*, 11553–63. DOI
27. Leng, J.; Wang, D.; Shen, W.; Li, X.; Liu, Q.; Chen, X. Digital twins-based smart manufacturing system design in Industry 4.0: a review. *J. Manuf. Syst.* **2021**, *60*, 119–37. DOI
28. Stavropoulos, P.; Mourtzis, D. Chapter 10 - Digital twins in Industry 4.0. In: *Design and operation of production networks for mass personalization in the era of cloud technology*. Elsevier; 2022. pp. 277–316. DOI
29. Sharma, A.; Kosasih, E.; Zhang, J.; Brintrup, A.; Calinescu, A. Digital twins: state of the art theory and practice, challenges, and open research questions. *J. Ind. Inf. Integr.* **2022**, *30*, 100383. DOI
30. Glaessgen, E.; Stargel, D. The digital twin paradigm for future NASA and U.S. Air Force vehicles. In: *53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA*; 2012. pp.

1818. DOI
31. Huang, Y.; Zhang, J.; Chen, X.; Lam, A. H. F.; Chen, B. M. From simulation to prediction: enhancing digital twins with advanced generative AI technologies. In: *2024 IEEE 18th International Conference on Control & Automation (ICCA)*, Reykjavik, Iceland, 18–21 Jun, 2024. IEEE, 2024; pp. 490–5. DOI
 32. Rodrigo, M. S.; Rivera, D.; Moreno, J. I.; Álvarez-Campana, M.; López, D. R. Digital twins for 5G networks: a modeling and deployment methodology. *IEEE Access*. **2023**, *11*, 38112–26. DOI
 33. Hasan, H. R.; Salah, K.; Jayaraman, R.; et al. A blockchain-based approach for the creation of digital twins. *IEEE Access*. **2020**, *8*, 34113–26. DOI
 34. Feng, K.; Borghesani, P.; Smith, W. A.; et al. Vibration-based updating of wear prediction for spur gears. *Wear*. **2019**, *426*, 1410–5. DOI
 35. Feng, K.; Ji, J. C.; Zhang, Y.; Ni, Q.; Liu, Z.; Beer, M. Digital twin-driven intelligent assessment of gear surface degradation. *Mech. Syst. Signal. Process.* **2023**, *186*, 109896. DOI
 36. Zhang, Y.; Ji, J.; Ren, Z.; et al. Digital twin-driven partial domain adaptation network for intelligent fault diagnosis of rolling bearing. *Reliab. Eng. Syst. Saf.* **2023**, *234*, 109186. DOI
 37. Feng, K.; Xu, Y.; Wang, Y.; et al. Digital twin enabled domain adversarial graph networks for bearing fault diagnosis. *IEEE Trans. Ind. Cyber. Phys. Syst.* **2023**, *1*, 113–22. DOI
 38. Kamel Boulos, M. N.; Zhang, P. Digital twins: from personalised medicine to precision public health. *J. Pers. Med.* **2021**, *11*, 745. DOI
 39. Alazab, M.; Khan, L. U.; Koppu, S.; et al. Digital twins for healthcare 4.0 - recent advances, architecture, and open challenges. *IEEE Consum. Electron. Mag.* **2022**, *12*, 29–37. DOI
 40. Pylidianis, C.; Osinga, S.; Athanasiadis, I. N. Introducing digital twins to agriculture. *Comput. Electron. Agric.* **2021**, *184*, 105942. DOI
 41. Purcell, W.; Neubauer, T. Digital twins in agriculture: a state-of-the-art review. *Smart. Agric. Technol.* **2023**, *3*, 100094. DOI
 42. Wang, Z.; Lv, C.; Wang, F. Y. A new era of intelligent vehicles and intelligent transportation systems: digital twins and parallel intelligence. *IEEE Trans. Intell. Veh.* **2023**, *8*, 2619–27. DOI
 43. Hu, X.; Li, S.; Huang, T.; Tang, B.; Huai, R.; Chen, L. How simulation helps autonomous driving: a survey of sim2real, digital twins, and parallel intelligence. *IEEE Trans. Intell. Veh.* **2024**, *9*, 593–612. DOI
 44. Arraño-Vargas, F.; Konstantinou, G. Modular design and real-time simulators toward power system digital twins implementation. *IEEE Trans. Ind. Inform.* **2022**, *19*, 52–61. DOI
 45. Lv, Z.; Guo, J.; Lv, H. Safety poka yoke in zero-defect manufacturing based on digital twins. *IEEE Trans. Ind. Inform.* **2022**, *19*, 1176–84. DOI
 46. Mo, F.; Rehman, H. U.; Monetti, F. M.; et al. A framework for manufacturing system reconfiguration and optimisation utilising digital twins and modular artificial intelligence. *Robot. Comput. Integr. Manuf.* **2023**, *82*, 102524. DOI
 47. Kapteyn, M. G.; Knezevic, D. J.; Huynh, D. B. P.; Tran, M.; Willcox, K. E. Data-driven physics-based digital twins via a library of component-based reduced-order models. *Int. J. Numer. Method. Eng.* **2022**, *123*, 2986–3003. DOI
 48. Lv, Z.; Chen, D.; Feng, H.; Zhu, H.; Lv, H. Digital twins in unmanned aerial vehicles for rapid medical resource delivery in epidemics. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 25106–14. DOI
 49. Li, X.; Liu, H.; Wang, W.; Zheng, Y.; Lv, H.; Lv, Z. Big data analysis of the Internet of Things in the digital twins of smart city based on deep learning. *Future. Gener. Comput. Syst.* **2022**, *128*, 167–77. DOI
 50. Lv, Z.; Li, Y.; Feng, H.; Lv, H. Deep learning for security in digital twins of cooperative intelligent transportation systems. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 16666–75. DOI
 51. Wan, Z.; Dong, Y.; Yu, Z.; Lv, H.; Lv, Z. Semi-supervised support vector machine for digital twins based brain image fusion. *Front. Neurosci.* **2021**, *15*, 705323. DOI
 52. Darvishi, H.; Ciuonzo, D.; Eide, E. R.; Rossi, P. S. Sensor-fault detection, isolation and accommodation for digital twins via modular data-driven architecture. *IEEE Sensors J.* **2020**, *21*, 4827–38. DOI
 53. Malik, A. A.; Brem, A. Digital twins for collaborative robots: a case study in human-robot interaction. *Robot. Comput. Integr. Manuf.* **2021**, *68*, 102092. DOI
 54. Dembski, F.; Wössner, U.; Letzgus, M.; Ruddat, M.; Yamu, C. Urban digital twins for smart cities and citizens: the case study of Herrenberg, Germany. *Sustainability* **2020**, *12*, 2307. DOI
 55. Wu, B.; Widanage, W. D.; Yang, S.; Liu, X. Battery digital twins: perspectives on the fusion of models, data and artificial intelligence for smart battery management systems. *Energy AI* **2020**, *1*, 100016. DOI
 56. Keshvarparast, A.; Battini, D.; Battaia, O.; Pirayesh, A. Collaborative robots in manufacturing and assembly systems: literature review and future research agenda. *J. Intell. Manuf.* **2024**, *35*, 2065–118. DOI
 57. Simões, A. C.; Pinto, A.; Santos, J.; Pinheiro, S.; Romero, D. Designing human-robot collaboration (HRC) workspaces in industrial settings: a systematic literature review. *J. Manuf. Syst.* **2022**, *62*, 28–43. DOI
 58. Burghardt, A.; Szybicki, D.; Gierlak, P.; Kurc, K.; Pietrus, P.; Cygan, R. Programming of industrial robots using virtual reality and digital twins. *Appl. Sci.* **2020**, *10*, 486. DOI
 59. Kuts, V.; Otto, T.; Tähemaa, T.; Bondarenko, Y. Digital twin based synchronised control and simulation of the industrial robotic cell using virtual reality. *J. Mach. Eng.* **2019**, *19*, 128–45. DOI
 60. Sun, F.; Liu, N.; Wang, X.; et al. Digital-twin-assisted skill learning for 3C assembly tasks. *IEEE Trans. Cybern.* **2024**. DOI
 61. Xu, W.; Yang, H.; Ji, Z.; Ba, M. Cognitive digital twin-enabled multi-robot collaborative manufacturing: framework and approaches. *Comput. Ind. Eng.* **2024**, *194*, 110418. DOI
 62. Liu, Y.; Wang, X.; Yang, K.; Pan, Y.; Wang, Q. Architecture and implementation of high-fidelity digital twins for industrial robots. In:

- 2023 International Conference on Frontiers of Robotics and Software Engineering (FRSE); IEEE, 2023; pp. 207–14. [DOI](#)
63. Li, J.; Liu, M.; Wang, W.; Hu, C. Inspection robot based on offline digital twin synchronization architecture. *IEEE. J. Radio. Freq. Identif.* **2022**, *6*, 943–47. [DOI](#)
 64. Liang, C. J.; McGee, W.; Menassa, C. C.; Kamat, V. R. Real-time state synchronization between physical construction robots and process-level digital twins. *Constr. Robot.* **2022**, *6*, 57–73. [DOI](#)
 65. Zhang, Z.; Dershan, R.; Enayati, A. M. S.; et al. A high-fidelity simulation platform for industrial manufacturing by incorporating robotic dynamics into an industrial simulation tool. *IEEE. Robot. Autom. Lett.* **2022**, *7*, 9123–8. [DOI](#)
 66. Liu, Y.; Xu, H.; Liu, D.; Wang, L. A digital twin-based sim-to-real transfer for deep reinforcement learning-enabled industrial robot grasping. *Robot. Comput. Integr. Manuf.* **2022**, *78*, 102365. [DOI](#)
 67. Mo, Y.; Ma, S.; Gong, H.; Chen, Z.; Zhang, J.; Tao, D. Terra: A smart and sensible digital twin framework for robust robot deployment in challenging environments. *IEEE Internet of Things J.* **2021**, *8*, 14039–50. [DOI](#)
 68. Cascone, L.; Nappi, M.; Narducci, F.; Passero, I. DTPAAL: digital twinning pepper and ambient assisted living. *IEEE Trans. Ind. Inform.* **2021**, *18*, 1397–404. [DOI](#)
 69. Xu, W.; Cui, J.; Li, L.; Yao, B.; Tian, S.; Zhou, Z. Digital twin-based industrial cloud robotics: framework, control approach and implementation. *J. Manuf. Syst.* **2021**, *58*, 196–209. [DOI](#)
 70. Kaigom, E. G.; Roßmann, J. Value-driven robotic digital twins in cyber - physical applications. *IEEE Trans. Ind. Inform.* **2020**, *17*, 3609–19. [DOI](#)
 71. Petersen, K.; Nagpal, R.; Werfel, J. TERMES: an autonomous robotic system for three-dimensional collective construction. In: *Robotics: science and systems*. Los Angeles, CA, USA; 2011. pp. 257–64. [DOI](#)
 72. Zhang, X.; Zheng, L.; Fan, W.; Ji, W.; Mao, L.; Wang, L. Knowledge graph and function block based Digital Twin modeling for robotic machining of large-scale components. *Robot. Comput. Integr. Manuf.* **2024**, *85*, 102609. [DOI](#)
 73. Zhang, X.; Wu, B.; Zhang, X.; Duan, J.; Wan, C.; Hu, Y. An effective MBSE approach for constructing industrial robot digital twin system. *Robot. Comput. Integr. Manuf.* **2023**, *80*, 102455. [DOI](#)
 74. Chen, L.; Bi, G.; Yao, X.; et al. Multisensor fusion-based digital twin for localized quality prediction in robotic laser-directed energy deposition. *Robot. Comput. Integr. Manuf.* **2023**, *84*, 102581. [DOI](#)
 75. Li, X.; Liu, G.; Sun, S.; Yi, W.; Li, B. Digital twin model-based smart assembly strategy design and precision evaluation for PCB kit-box build. *J. Manuf. Syst.* **2023**, *71*, 206–23. [DOI](#)
 76. Liu, J.; Xu, Z.; Xiong, H.; Lin, Q.; Xu, W.; Zhou, Z. Digital twin-driven robotic disassembly sequence dynamic planning under uncertain missing condition. *IEEE. Trans. Ind. Inform.* **2023**, *19*, 11846–55. [DOI](#)
 77. Wenna, W.; Weili, D.; Changchun, H.; Heng, Z.; Haibing, F.; Yao, Y. A digital twin for 3D path planning of large-span curved-arm gantry robot. *Robot. Comput. Integr. Manuf.* **2022**, *76*, 102330. [DOI](#)
 78. Hu, W.; Wang, C.; Liu, F.; Peng, X.; Sun, P.; Tan, J. A grasps-generation-and-selection convolutional neural network for a digital twin of intelligent robotic grasping. *Robot. Comput. Integr. Manuf.* **2022**, *77*, 102371. [DOI](#)
 79. Tipary, B.; Erdős, G. Generic development methodology for flexible robotic pick-and-place workcells based on Digital Twin. *Robot. Comput. Integr. Manuf.* **2021**, *71*, 102140. [DOI](#)
 80. Weiss, A.; Wortmeier, A. K.; Kubicek, B. Cobots in industry 4.0: a roadmap for future practice studies on human - robot collaboration. *IEEE. Trans. Hum. Mach. Syst.* **2021**, *51*, 335–45. [DOI](#)
 81. Djuric, A. M.; Urbanic, R. J.; Rickli, J. L. A framework for collaborative robot (CoBot) integration in advanced manufacturing systems. *SAE. Int. J. Mater. Manuf.* **2016**, *9*, 457–64. [DOI](#)
 82. Javid, M.; Haleem, A.; Singh, R. P.; Rab, S.; Suman, R. Significant applications of Cobots in the field of manufacturing. *Cogn. Robot.* **2022**, *2*, 222–33. [DOI](#)
 83. Pauliková, A.; Gyurák Babel'ová, Z.; Ubárová, M. Analysis of the impact of human-cobot collaborative manufacturing implementation on the occupational health and safety and the quality requirements. *Int. J. Environ. Res. Public Health.* **2021**, *18*, 1927. [DOI](#)
 84. Lu, Y.; Adrados, J. S.; Chand, S. S.; Wang, L. Humans are not machines - anthropocentric human-machine symbiosis for ultra-flexible smart manufacturing. *Engineering* **2021**, *7*, 734–37. [DOI](#)
 85. Mark, B. G.; Rauch, E.; Matt, D. T. Worker assistance systems in manufacturing: a review of the state of the art and future directions. *J. Manuf. Syst.* **2021**, *59*, 228–50. [DOI](#)
 86. Leng, J.; Sha, W.; Wang, B.; et al. Industry 5.0: prospect and retrospect. *J. Manuf. Syst.* **2022**, *65*, 279–95. [DOI](#)
 87. Huang, S.; Wang, B.; Li, X.; Zheng, P.; Mourtzis, D.; Wang, L. Industry 5.0 and Society 5.0 - Comparison, complementation and co-evolution. *J. Manuf. Syst.* **2022**, *64*, 424–8. [DOI](#)
 88. Sharma, M.; Sehrawat, R.; Luthra, S.; Daim, T.; Bakry, D. Moving towards industry 5.0 in the pharmaceutical manufacturing sector: challenges and solutions for Germany. *IEEE. T. Eng. Manag.* **2022**, *71*, 13757-74. [DOI](#)
 89. Park, H.; Shin, M.; Choi, G.; et al. Integration of an exoskeleton robotic system into a digital twin for industrial manufacturing applications. *Robot. Comput. Integr. Manuf.* **2024**, *89*, 102746. [DOI](#)
 90. Wang, S.; Zhang, J.; Wang, P.; Law, J.; Calinescu, R.; Mihaylova, L. A deep learning-enhanced Digital Twin framework for improving safety and reliability in human-robot collaborative manufacturing. *Robot. Comput. Integr. Manuf.* **2024**, *85*, 102608. [DOI](#)
 91. Choi, S. H.; Park, K. B.; Roh, D. H.; et al. An integrated mixed reality system for safety-aware human-robot collaboration using deep learning and digital twin generation. *Robot. Comput. Integr. Manuf.* **2022**, *73*, 102258. [DOI](#)
 92. Guo, Y.; Sun, J.; Zhang, R.; Jiang, Z.; Mi, Z.; Yao, C. An object-driven navigation strategy based on active perception and semantic association. *IEEE. Robot. Autom. Lett.* **2024**, *9*, 7110-7. [DOI](#)

93. Gao, R.; Li, H.; Dharan, G.; Wang, Z.; Li, C.; Xia, F. Sonicverse: a multisensory simulation platform for embodied household agents that see and hear. In: *2023 IEEE International Conference on Robotics and Automation (ICRA)*, London, UK, 29 May - 02 Jun, 2023. IEEE, 2023; pp. 704–11. DOI
94. Wang, T.; Mao, X.; Zhu, C.; et al. Embodiedscan: a holistic multi-modal 3D perception suite towards embodied AI. *arXiv* **2023**, arXiv:2312.16170. Available online: <https://doi.org/10.48550/arXiv.2312.16170>. (accessed 28 Feb 2025)
95. He, Z.; Wang, L.; Dang, R.; Li, S.; Yan, Q.; Liu, C. Learning depth representation from RGB-D videos by time-aware contrastive pre-training. *IEEE. T. Circ. Syst. Vid.* **2023**, *34*, 4143-58. DOI
96. Jing, Y.; Kong, T. Learning to explore informative trajectories and samples for embodied perception. In: *2023 IEEE International Conference on Robotics and Automation (ICRA)*, London, UK, 29 May - 02 Jun, 2023. IEEE, 2023; pp. 6050–6. DOI
97. Patel, S.; Wani, S.; Jain, U.; Schwing, A. G.; Lazebnik, S.; Savva, M. Interpretation of emergent communication in heterogeneous collaborative embodied agents. In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, Canada, 10-17 Oct, 2021. IEEE, 2021; pp. 15953–63. DOI
98. Peri, N.; Luiten, J.; Li, M.; Ošep, A.; Leal-Taixé, L.; Ramanan, D. Forecasting from LiDAR via future object detection. *arXiv* **2022**, arXiv:2203.16297. Available online: <https://doi.org/10.48550/arXiv.2203.16297>. (accessed 28 Feb 2025)
99. Gan, C.; Zhou, S.; Schwartz, J.; Alter, S.; Bhandwadar, A.; Gutfreund, D. The threedworld transport challenge: a visually guided task-and-motion planning benchmark towards physically realistic embodied AI. In: *2022 International conference on robotics and automation (ICRA)*, Philadelphia, USA, 23-27 May, 2022. IEEE, 2022; pp. 8847–54. DOI
100. Gan, Y.; Zhang, B.; Shao, J.; Han, Z.; Li, A.; Dai, X. Embodied intelligence: bionic robot controller integrating environment perception, autonomous planning, and motion control. *IEEE. Robot. Autom. Lett.* **2024**, *9*, 4559-66. DOI
101. Behrens, J. K.; Nazarczuk, M.; Stepanova, K.; Hoffmann, M.; Demiris, Y.; Mikolajczyk, K. Embodied reasoning for discovering object properties via manipulation. In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*, Xi'an, China, 30 May - 05 Jun, 2021. IEEE, 2021; pp. 10139–45. DOI
102. Huang, X.; Batra, D.; Rai, A.; Szot, A. Skill transformer: a monolithic policy for mobile manipulation. *arXiv* **2023**, arXiv:2308.09873. Available online: <https://doi.org/10.48550/arXiv.2308.09873>. (accessed 28 Feb 2025)
103. Li, B.; Zhang, Y.; Zhang, T.; et al. Embodied footprints: a safety-guaranteed collision-avoidance model for numerical optimization-based trajectory planning. *IEEE. Trans. Intell. Transpor. Syst.* **2023**, *25*, 2046-60. DOI
104. Pervan, A.; Murphey, T. D. Algorithmic design for embodied intelligence in synthetic cells. *IEEE. Trans. Autom. Sci. Eng.* **2020**, *18*, 864–75. DOI
105. Dorbala, V. S.; Mullen, J. F.; Manocha, D. Can an embodied agent find your "cat-shaped mug"? LLM-based zero-shot object navigation. *IEEE. Robot. Autom. Lett.* **2023**, *9*, 4083-90. DOI
106. Dwivedi, K.; Roig, G.; Kembhavi, A.; Mottaghi, R. What do navigation agents learn about their environment? In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, USA, 18-24 Jun, 2022. IEEE, 2022; pp. 10276–85. DOI
107. Gao, C.; Liu, S.; Chen, J.; et al. Room-object entity prompting and reasoning for embodied referring expression. *IEEE. Trans. Pattern. Anal. Mach. Intell.* **2023**, *46*, 994-1010. DOI
108. Zhao, C.; Yuan, S.; Jiang, C.; et al. ERRRA: an embodied representation and reasoning architecture for long-horizon language-conditioned manipulation tasks. *IEEE. Robot. Autom. Lett.* **2023**, *8*, 3230–7. DOI
109. Liu, X.; Guo, D.; Liu, H.; Sun, F. Multi-agent embodied visual semantic navigation with scene prior knowledge. *IEEE. Robot. Autom. Lett.* **2022**, *7*, 3154–61. DOI
110. Sermanet, P.; Ding, T.; Zhao, J.; et al. RoboVQA: multimodal long-horizon reasoning for robotics. In: *2024 IEEE International Conference on Robotics and Automation (ICRA)*, Yokohama, Japan, 13-17 May, 2024. IEEE, 2024; pp. 645–52. DOI
111. Zhao, Q.; Zhang, L.; Wu, L.; Qiao, H.; Liu, Z. A real 3D embodied dataset for robotic active visual learning. *IEEE. Robot. Autom. Lett.* **2022**, *7*, 6646–52. DOI
112. Long, Y.; Wei, W.; Huang, T.; Wang, Y.; Dou, Q. Human-in-the-loop embodied intelligence with interactive simulation environment for surgical robot learning. *IEEE. Robot. Autom. Lett.* **2023**, *8*, 4441–8. DOI
113. Hashemian, A. M.; Adhikari, A.; Aguilar, I. A.; Kruijff, E.; von der Heyde, M.; Riecke, B. E. Learning-based interfaces improve simultaneous locomotion and object interaction in VR compared to the handheld controller. *IEEE. Trans. Vis. Comput. Graph.* **2023**, *30*, 4665-82. DOI
114. Sagar, M.; Moser, A.; Henderson, A. M.; et al. A platform for holistic embodied models of infant cognition, and its use in a model of event processing. *IEEE. Trans. Cogn. Dev. Syst.* **2022**, *15*, 1916–27. DOI
115. Legrand, J.; Wang, H.; Iida, F.; Vanderborght, B. A variable stiffness anthropomorphic finger through embodied intelligence design. *IEEE. Robot. Autom. Lett.* **2023**, *9*, 1580-7. DOI
116. Fischer, T.; Demiris, Y. Computational modeling of embodied visual perspective taking. *IEEE. Trans. Cogn. Dev. Syst.* **2019**, *12*, 723–32. DOI
117. Greenacre, M.; Groenen, P. J.; Hastie, T.; d'Enza, A. I.; Markos, A.; Tuzhilina, E. Principal component analysis. *Nat. Rev. Methods. Primers.* **2022**, *2*, 100. DOI
118. Cancemi, S. A.; Frano, R. L.; Santus, C.; Inoue, T. Unsupervised anomaly detection in pressurized water reactor digital twins using autoencoder neural networks. *Nucl. Eng. Des.* **2023**, *413*, 112502. DOI
119. Raptis, T. P.; Passarella, A. A survey on networked data streaming with apache kafka. *IEEE. Access.* **2023**, *11*, 85333-50. DOI
120. Abramova, V.; Bernardino, J. NoSQL databases: MongoDB vs cassandra. In: *Proceedings of the international C* conference on computer science and software engineering*; 2013. pp. 14–22. DOI

121. Shah, J.; Dubaria, D. Building modern clouds: using docker, kubernetes & Google cloud platform. In: *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, USA, 07-09 Jan, 2019. IEEE, 2019; pp. 0184–9. DOI
122. Khan, L. U.; Saad, W.; Niyato, D.; Han, Z.; Hong, C. S. Digital-twin-enabled 6G: vision, architectural trends, and future directions. *IEEE. Commun. Mag.* **2022**, *60*, 74–80. DOI
123. Vuković, M.; Mazzei, D.; Chessa, S.; Fantoni, G. Digital twins in industrial IoT: a survey of the state of the art and of relevant standards. In: *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*, Montreal, Canada, 14-23 Jun, 2021. IEEE, 2021; pp. 1–6. DOI
124. Liu, X.; Zheng, L.; Wang, Y.; et al. Human-centric collaborative assembly system for large-scale space deployable mechanism driven by Digital Twins and wearable AR devices. *J. Manuf. Syst.* **2022**, *65*, 720–42. DOI
125. Zhou, H.; Wang, L.; Pang, G.; et al. Toward human motion digital twin: a motion capture system for human-centric applications. *IEEE. Trans. Autom. Sci. Eng.* **2024**, *22*, 619-30. DOI
126. Wang, B.; Zhou, H.; Li, X.; et al. Human digital twin in the context of Industry 5.0. *Robot. Comput. Integr. Manuf.* **2024**, *85*, 102626. DOI
127. Fan, J.; Zheng, P.; Lee, C. K. M. A vision-based human digital twin modeling approach for adaptive human–robot collaboration. *J. Manuf. Sci. Eng.* **2023**, *145*, 121002. DOI
128. You, Y.; Cai, B.; Pham, D. T.; Liu, Y.; Ji, Z. A human digital twin approach for fatigue-aware task planning in human-robot collaborative assembly. *Comput. Ind. Eng.* **2025**, *200*, 110774. DOI
129. Okegbile, S. D.; Cai, J.; Niyato, D.; Yi, C. Human digital twin for personalized healthcare: vision, architecture and future directions. *IEEE. Netw.* **2022**, *37*, 262–9. DOI
130. Wang, B.; Zhou, H.; Yang, G.; Li, X.; Yang, H. Human digital twin (HDT) driven human-cyber-physical systems: key technologies and applications. *Chin. J. Mech. Eng.* **2022**, *35*, 11. DOI
131. Almeaid, S.; Al-Rubaye, S.; Tsourdos, A.; Avdelidis, N. P. Digital twin analysis to promote safety and security in autonomous vehicles. *IEEE. Commun. Stands. Mag.* **2021**, *5*, 40–6. DOI
132. Wang, T.; Zheng, P.; Li, S.; Wang, L. Multimodal human-robot interaction for human-centric smart manufacturing: a survey. *Adv. Intell. Syst.* **2024**, *6*, 2300359. DOI
133. Halawa, H.; Abdelhafez, H. A.; Bektor, A.; Ripeanu, M. NVIDIA jetson platform characterization. In: *Euro-Par 2017: Parallel Processing: 23rd International Conference on Parallel and Distributed Computing*, Santiago de Compostela, Spain, 28 Aug - 1 Sep, 2017. Springer, 2017; pp. 92–105. DOI
134. Allison, D.; Smith, P.; McLaughlin, K. Digital twin-enhanced methodology for training edge-based models for cyber security applications. In: *2022 IEEE 20th International Conference on Industrial Informatics (INDIN)*, Perth, Australia, 25-28 Jul, 2022. IEEE, 2022; pp. 226–32. DOI
135. Lei, B.; Janssen, P.; Stoter, J.; Biljecki, F. Challenges of urban digital twins: a systematic review and a Delphi expert survey. *Autom. Constr.* **2023**, *147*, 104716. DOI
136. Lehtola, V. V.; Koeva, M.; Elberink, S. O.; et al. Digital twin of a city: review of technology serving city needs. *Int. J. Appl. Earth. Obs. Geoinf.* **2022**, *114*, 102915. DOI
137. Ye, X.; Du, J.; Han, Y.; et al. Developing human-centered urban digital twins for community infrastructure resilience: a research agenda. *J. Plan. Lit.* **2023**, *38*, 187–99. https://www.researchgate.net/publication/365888617_Developing_Human-Centered_Urban_Digital_Twins_for_Community_Infrastructure_Resilience_A_Research_Agenda. (accessed 2025-03-06)
138. Angin, P.; Anisi, M. H.; Göksel, F.; Gürsoy, C.; Büyükgülcü, A. AgriLoRa: a digital twin framework for smart agriculture. *J. Wirel. Mob. Netw. Ubiquitous. Comput. Dependable. Appl.* **2020**, *11*, 77–96. DOI