Intelligent Human Systems Integration (IHSI), Vol. 200, 2026, 920–928
https://doi.org/10.54941/ahfe1007153

High-Fidelity Simulation and Digital Twin Framework for Real-Time Adaptive Human–Robot Collaboration

Isaiah Nassiuma, Ella-Mae Hubbard, and Yee Mey Goh

Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, Loughborough, UK

ABSTRACT

Effective real-time adaptation in human–robot collaboration requires significant testing and validation to avoid compromising human safety and improve task efficiency. Developing and validating such systems in real-world industrial environments is resource intensive. This paper presents a high-fidelity simulation and visualisation platform for shared robotic workspaces that supports both synthetic data generation and digital twin operation in Unity 3D. The platform models and visualises human body movement and gaze behaviour, enabling the representation of body movement and visual attention during collaborative manufacturing tasks. Using procedurally generated scenarios, the system supports controlled and repeatable experimentation for the development and evaluation of real-time adaptation strategies. When integrated with real sensor and eye-tracking data, the platform operates as a digital twin, allowing real-time mirroring, monitoring, and analysis of human– robot interactions under operational conditions. By enabling both offline simulation and online system integration, the platform facilitates rapid prototyping and systematic assessment of real-time adaptation approaches, helping bridge the gap between theoretical methods and practical deployment in intelligent human–robot collaborative systems.

Keywords: Human-robot collaboration, Simulation, Digital twin

INTRODUCTION

Human–robot collaboration (HRC) requires continuous mutual adaptation, as both human and robot actions affect task progress, safety, and efficiency. Real-time adaptation is therefore essential: robot motions, interaction timing, and system feedback must adjust dynamically in response to changes in task state and human behaviour (Li et al., 2023). Effective adaptive systems simultaneously optimise robot performance and support human operators, modulating guidance, information delivery, or interventions to maintain coordination, reduce cognitive load, and enhance task fluency. By responding to human behaviour while shaping the collaborative interaction, adaptation mechanisms enable more robust, efficient, and natural shared control, underscoring their central role in intelligent HRC design.

Training and evaluating adaptive collaboration strategies in real-world environments is resource intensive and difficult to scale. Physical experiments require substantial time and cost, including participant recruitment, safety

Received December 30, 2025; Revised January 19, 2026; Accepted January 24, 2025; Available online February 11, 2026

management, and repeated trials to capture variability in human behaviour and system responses. These constraints limit the exploration of alternative strategies, rare events, and edge cases that are critical for robust system design. Simulation therefore plays a central role in the development and validation of real-time adaptive systems prior to deployment. Offline simulation enables controlled evaluation of adaptive strategies without physical hardware, while online simulation supports real-time interaction by synchronising virtual and physical systems.

However, existing simulation environments focus primarily on robot control or human body movement (Grushko et al., 2021), offering limited support for studying adaptive interaction mechanisms that depend on integrated embodied movement and visual attention. This motivates the need for simulation frameworks that explicitly model these human factors while supporting both controlled offline experimentation and real-time integration with physical systems.

This paper presents a high-fidelity simulation and visualisation platform for HRC that supports both offline simulation and real-time digital twin operation. The platform models human body movement and visual attention within shared robotic workspaces using predefined scenarios, enabling realistic representation of human physical state during collaborative tasks.

The remainder of this paper reviews related work in simulation-based HRC, presents the system architecture and digital twin implementation, and details the human behaviour modelling approach and adaptation interface. Representative use case scenarios are then demonstrated, followed by a discussion of limitations and implications. The paper concludes with a summary of contributions and directions for future work.

RELATED WORK

Research in HRC has explored numerous approaches aimed at improving safety, efficiency, and coordination in shared workspaces. Many studies focus on robot-centric adaptation (Zhang et al., 2022; Suresh et al., 2024), including trajectory replanning (Grushko et al., 2021), speed modulation (Askin and Bitsch, 2023), and collision avoidance based on human proximity or predicted motion (Liu and Wang, 2021). While these approaches improve physical safety, they often treat human behaviour as a reactive element, limiting their ability to support context-aware adaptation during collaboration.

Interactive communication and coordination mechanisms in HRC are being investigated through visual, auditory, and haptic modalities to convey robot intent, guide task progress, or improve mutual understanding (Li et al., 2023). Recent work investigates adaptive and learning-based approaches for determining when and how nudges and other notifications that influence human behaviour should be delivered (Yang et al., 2022; Nassiuma et al., 2024). However, the development of these systems would benefit by simulation environments that include realistic models of human response and engagement.

Simulation has emerged as a key tool for studying HRC due to the cost, safety, and scalability limitations of real-world experimentation (Cimino et al., 2024; Baratta et al., 2025). Virtual environments enable controlled evaluation of interaction strategies across a wide range of scenarios (Arntz, Eimler and Hoppe, 2021). Several frameworks leverage physics-based simulation (Weistroffer et al., 2022) and digital twins to model collaborative assembly tasks (Cimino et al., 2024), allowing researchers to test robot behaviours under varying human positions and task conditions.

Existing approaches range from scripted animations and motion capture replay to probabilistic and data-driven human motion models (Xia et al., 2017). These techniques are commonly used to predict human trajectories or estimate occupancy regions for safety planning (Pereira and Althoff, 2017). However, many models focus primarily on gross body motion and /or eye tracking but don't merge the two to form an overarching understanding that could provide more insight for context adaptation.

Some studies incorporate stochastic or performance-based variability to represent differences across users and execution conditions (Xia et al., 2026). While this improves realism, the resulting human models are often decoupled from interaction mechanisms such as notifications, warnings, or guidance systems. As a result, simulations may fail to capture how human behaviour changes in response to system-generated information.

Simulation frameworks for HRC have been showcased to model both humans and robots as autonomous agents operating within a shared environment (Antakli et al., 2019; Buerkle et al., 2021). The approach emphasises task-level coordination and workflow optimization, demonstrating how agent-based modelling can be used to evaluate collaborative assembly processes. While effective for analysing system-level interactions and task allocation, its focus isn't on reactionary aspects of human behaviour related to virtual information.

This paper builds on prior simulation-based HRC research by observing how human behaviour changes in response to adaptively generated information. Changes in body movement and visual attention are analysed and used to inform the adaptive framework with the bid of improving safety and efficiency of the

SYSTEM OVERVIEW

System Architecture and Digital Twin Environment

The system architecture adopts a modular and layered design to support the development, simulation, and validation of real-time adaptive human-robot collaboration. The architecture integrates multimodal human sensing, digital twin simulation, human behaviour modelling, and interaction management within a unified framework implemented in Unity 3D as shown in Figure 1.

At the physical detection layer, human motion is captured using a multi-camera ZED stereo setup that provides real-time full-body skeletal tracking, including joint positions and orientations. Visual attention and head pose are acquired through a HoloLens 2 headset using the Mixed Reality Toolkit (MRTK3), providing gaze direction and fixation estimates within the workspace. These sensing streams are fused into a unified human state representation within Unity, enabling a consistent description of human posture, movement dynamics, and visual attention. While the current implementation focuses on motion and gaze, the architecture provides provisions for integrating additional physiological sensing modalities, such as electromyography (EMG) or electroencephalography (EEG), through standardized data interfaces.

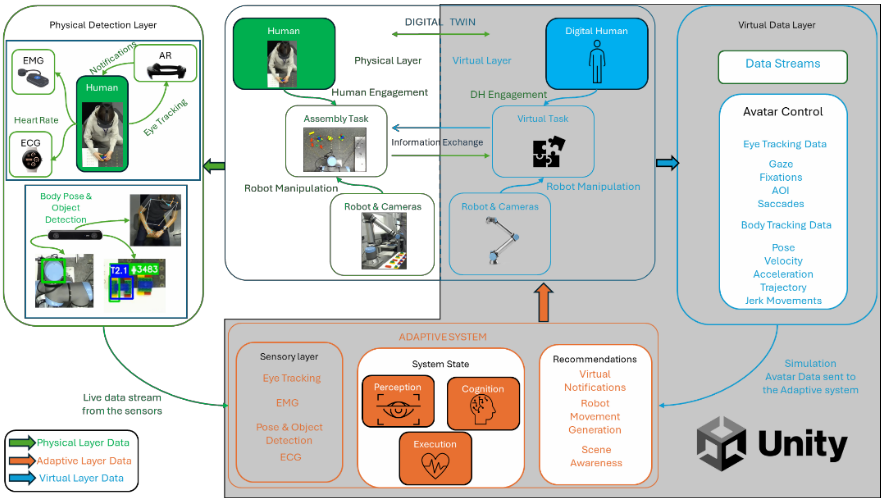

Figure 1: System architecture: a combination of physical sensors monitoring the physical environment (human, objects and robot). A digital twin interface run and adaptation architecture run via unity 3D.

The digital twin layer serves as a central integration hub within the simulation architecture. It contains synchronised and kinematically consistent virtual replicas of the human, the UR10e robot, and the shared workspace within the Unity 3D environment. Human motion data from the tracking pipeline are retargeted to an articulated humanoid avatar via a calibrated skeletal mapping, preserving joint hierarchy, segment lengths, and joint limit constraints to ensure physically plausible pose reconstruction.

The UR10e is modelled as a 6-DoF serial manipulator whose joint states are governed by the virtual controller: joint positions, velocities, and accelerations are generated in simulation from task-level commands using inverse kinematics and time-parameterised trajectory generators. These joint-space profiles are streamed to the physical robot, which also feeds back measured joint states to close the synchronisation loop between the virtual and physical instances. In this way, the robot controller is embedded as one module in a broader interaction pipeline, where human state estimation, environment representation, and adaptation logic are tightly integrated with the digital twin rather than treated as isolated subsystems.

The virtual data layer incorporates avatar control signals derived from sensor inputs and computational models, which are refined to ensure accurate representation of human actions in the virtual environment. Additionally, this layer generates visual attention metrics and body tracking data, enabling precise monitoring of gaze direction and physical posture. These processed outputs are transmitted to the simulated avatar, facilitating realistic execution of virtual tasks within the digital twin framework.

The interaction and adaptation layer implements a bidirectional control and communication loop between the human operator and the HRC system, operating on a fused representation of human and robot state. It treats the HoloLens 2 as both a sensing endpoint and an interaction terminal: head- and eye-gaze measurements are streamed into the state estimation pipeline, while context-dependent feedback is sent back as rendered holographic overlays and spatialised audio cues. The fused human–robot state is streamed to the adaptive modules, which can execute rule- based or learning-based policies to trigger interaction responses or update robot motion parameters.

In simulation, adaptation decisions are computed against procedurally generated or replayed interaction scenarios, whereas during live operation they are computed online from continuously streamed sensor data originating in the physical workspace. This separation of execution context allows the same adaptation logic and interfaces to be verified under controlled, repeatable conditions and then deployed for real-time use on the physical system without modification to the underlying decision mechanisms.

Human Behaviour Model

To support scalable experimentation beyond real-time data capture, the platform incorporates a generative human behaviour model capable of synthesizing realistic body movement and visual attention. The model is grounded in empirical human data collected, ensuring that generated behaviours remain consistent with observed human motion patterns as shown in Figure 2.

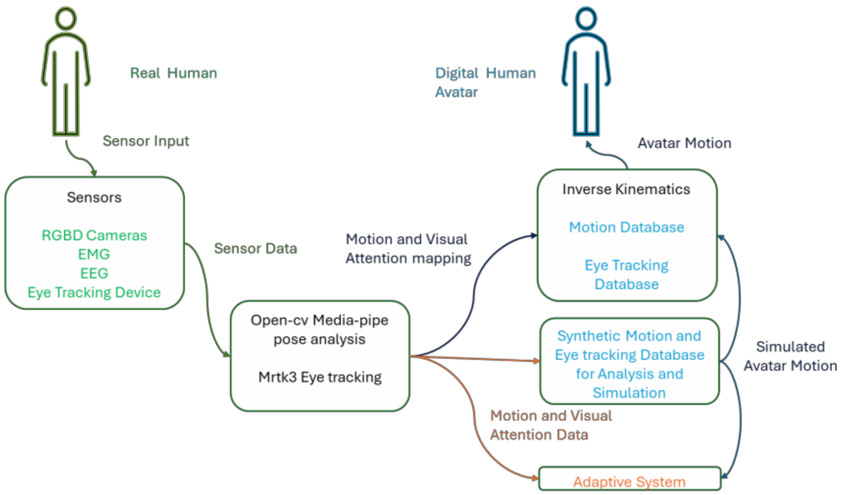

Figure 2: The real time data of the human captured by the devices and analysed and this feeds into the synthetic human database as well as the controlling of the digital human. In offline mode the synthetic database runs the human avatar.

Human behaviour within the platform is represented through a hierarchical model that captures both body movement and visual attention across multiple levels of abstraction. Low-level kinematic variables, including joint positions, velocities, and accelerations, describe physically plausible motion, while higher- level temporal structure encodes task-relevant phases and coordination patterns. This unified representation enables consistent modelling of human behaviour during real-time digital twin operation and offline simulation.

In real-time operation, body movement is captured using a ZED stereo camera setup with Media Pipe pose estimation, extracting 3D skeletal joint positions, velocities, accelerations, and jerk metrics from video streams. These data are filtered and retargeted to a Unity humanoid avatar, preserving kinematic constraints and temporal dynamics. Concurrently, visual attention is acquired via HoloLens 2 with MRTK3 eye-tracking, processing raw gaze data into fixation durations, saccade trajectories, and area of interest dwell time metrics. The fused body-gaze streams drive an articulated digital avatar that mirrors the human operator's posture, motion dynamics, and visual focus within the digital twin workspace

For offline simulation, the platform independently employs two complementary generative methods with distinct roles. First, bootstrap resampling is applied to recorded pose and gaze trajectories to generate controlled variations of observed behaviour. This method preserves task structure and coordination patterns while enabling systematic manipulation of movement speed, trajectory shape, fixation duration, and attention transitions, supporting robustness testing and scenario exploration. Second, a variational autoencoder (VAE) trained on aggregated pose–gaze sequences are used to synthesise novel trajectories by sampling from a learned latent space. This enables generation of behaviours not explicitly observed in the dataset, capturing inter-subject variability in movement style and gaze strategies while maintaining anatomical and task-level plausibility.

In simulation mode without live eye tracking, synthesized gaze vectors are mapped to head orientation using a gaze-to-head coupling model to ensure anatomical consistency between eye position and head pose. Together, the bootstrapped and VAE-generated models provide complementary mechanisms for reproducing observed behaviour and exploring novel interaction patterns, forming a unified behavioural foundation for adaptive HRC experimentation.

Adaptation Framework

The platform exposes a modular adaptation interface that enables external adaptation strategies to observe and influence the collaborative environment in a structured manner. As illustrated in Figure 3, the adaptation process operates on an interaction state that integrates human motion state, visual attention state, and robot execution state. This abstraction decouples adaptation logic from low-level sensing and ensures that identical decision-making mechanisms can be applied in both offline simulation and real-time digital twin operation.

Adaptation strategies are organised into three complementary classes. Rule-based adaptation encodes deterministic constraints and heuristics, such as safety thresholds, temporal conditions, and context-specific rules, providing predictable and verifiable system behaviour. Learning-based adaptation operates in an unsupervised or data-driven manner, leveraging latent state representations to discover behavioural patterns, cluster interaction states, or learn adaptive policies from data. Hybrid strategies combine these approaches by constraining learned behaviours with explicit rules or using learned triggers to activate deterministic responses. This separation allows systematic comparison and integration of different adaptation paradigms within the same framework.

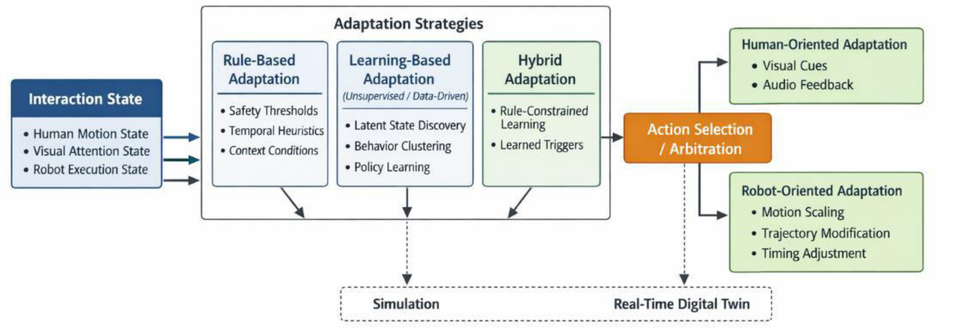

Figure 3: Schematic of the modular adaptation interface, showing state observations, parallel rule-based and unsupervised adaptation paths, and action outputs for robot adjustment and human information delivery.

Outputs from a chosen adaptation strategy determines the active response. Adaptation actions are applied along two primary channels: human-oriented adaptation, such as visual or audio feedback delivered through the HoloLens, and robot-oriented adaptation, including motion scaling, trajectory modification, or timing adjustment. By supporting both simulation-based evaluation and real-time execution using the same interface, the platform enables consistent development, testing, and deployment of adaptive HRC strategies.

Use Case Scenarios

The framework supports application areas that advance both the analysis of human behaviour in HRC and the development of adaptive interaction strategies. One application area focuses on systematically characterising how operators' body movements and visual attention respond to context-adaptive information delivered through the AR interface. The platform enables quantitative evaluation of gaze deviations from task-critical Areas of Interest (AOIs) and the effectiveness of automatically triggered visual or auditory cues. Comparative analysis of cue modalities can reveal their impact on proximity corrections, task resumption latency, and attention allocation. By modelling individual differences in baseline movement and visual patterns, the framework informs the design of personalised adaptation thresholds that optimise cue timing, salience, and modality for diverse operators.

A second application area examines human responses to robot motion anomalies, including path deviations or unexpected hesitations, to inform reassurance strategy design. The system can quantify correlations between error severity, hesitation duration, and gaze reallocation toward the robot end-effector. AR overlays, such as trajectory previews, are evaluated for their effectiveness in reducing operator withdrawal distance and restoring task engagement. Procedural behavioural variation in simulation supports counterfactual analysis, enabling systematic testing of multi-modal reassurance sequences that combine visual, audio, and motion synchronisation cues.

Across both application domains, the platform's integrated simulation and digital twin capabilities establish causal relationships between system actions, human responses, and collaborative performance metrics, providing a robust foundation for iterative refinement of context-aware HRC policies.

DISCUSSION AND CONCLUSION

The proposed platform provides a scalable framework for developing and evaluating adaptive HRC strategies. By integrating high-fidelity simulation with real-time digital twin operation, it enables systematic study of human responses to variations in robot motion, interaction cues, and error recovery mechanisms. The platform captures body movement, gaze behaviour, and task engagement, supporting iterative design and testing of adaptation strategies prior to real-world deployment. Its dual simulation–digital twin capability allows offline scenario generation for policy training and online mirroring for real-time adaptation, reducing reliance on resource-intensive physical trials and enhancing safety.

Hierarchical human behaviour modelling abstracts low-level sensory data into structured interaction states suitable for rule-based, learning-based, or hybrid adaptation strategies. These states inform robot motion adjustments or human- directed feedback, enabling flexible and method-agnostic adaptation. The modular design supports diverse robots, sensors, and human–machine interfaces, allowing exploration of application areas such as responses to adaptive cues, reactions to robot motion anomalies, and evaluation of multi-modal strategies. While the current work focuses on physical and perceptual aspects of human behaviour, this provides a foundation for future extensions which may incorporate higher-level task reasoning, intention modelling.

The platform unifies human behaviour modelling, adaptation strategy testing, and digital twin operation, providing a repeatable and controlled environment to refine context-aware HRC policies and improve collaborative efficiency, safety, and fluency in industrial settings.

REFERENCES

Antakli, A. et al. (2019) 'Agent-based web supported simulation of human-robot collaboration', WEBIST 2019 - Proceedings of the 15th International Conference on Web Information Systems and Technologies, pp. 88–99. Available at: https://doi.org/10.5220/0008163000880099.

Arntz, A., Eimler, S.C. and Hoppe, H.U. (2021) 'A Virtual Sandbox Approach to Studying the Effect of Augmented Communication on Human-Robot Collaboration', Frontiers in Robotics and AI, 8, p. 728961. Available at: https://doi.org/10.3389/FROBT.2021.728961/BIBTEX.

Askin, J. and Bitsch, G. (2023) 'Investigating the Influence of a Cobot's Average Tool Center Point Speed on Human Work Behavior in a Cooperative Human-Robot Collaboration Assembly Station', Procedia CIRP, 118, pp. 217–222. Available at: https://doi.org/10.1016/J.PROCIR.2023.06.038.

Baratta, A. et al. (2025) 'Conceptual Modeling for a Simulation-based Digital Twin in Human-Robot Collaboration', Procedia Computer Science, 253, pp. 3247–3256. Available at: https://doi.org/10.1016/J.PROCS.2025.02.049.

Buerkle, A. et al. (2021) 'An adaptive human sensor framework for human–robot collaboration', The International Journal of Advanced Manufacturing Technology 2021 119:1, 119(1), pp. 1233–1248. Available at: https://doi.org/10.1007/S00170-021-08299-2.

Cimino, A. et al. (2024) 'Simulation-based Digital Twin for enhancing human-robot collaboration in assembly systems', Journal of Manufacturing Systems, 77, pp. 903–918. Available at: https://doi.org/10.1016/J.JMSY.2024.10.024.

Grushko, S. et al. (2021) 'Improved Mutual Understanding for Human-Robot Collaboration: Combining Human-Aware Motion Planning with Haptic Feedback Devices for Communicating Planned Trajectory', Sensors 2021, Vol. 21, Page 3673, 21(11), p. 3673. Available at: https://doi.org/10.3390/S21113673.

Li, S. et al. (2023) 'Proactive human–robot collaboration: Mutual-cognitive, predictable, and self-organising perspectives', Robotics and Computer-Integrated Manufacturing, 81, p. 102510. Available at: https://doi.org/10.1016/J.RCIM.2022.102510.

Liu, H. and Wang, L. (2021) 'Collision-free human-robot collaboration based on context awareness', Robotics and Computer-Integrated Manufacturing, 67, p. 101997. Available at: https://doi.org/10.1016/J.RCIM.2020.101997.

Nassiuma, I., Goh, Y.M. and Hubbard, E.M. (2024) 'Positive and Negative Reinforcement Nudges for Human-Robot Collaboration using a Mixed Reality Interface', ACM International Conference Proceeding Series [Preprint]. Available at: https://doi.org/10.1145/3673805.3673844.

Pereira, A. and Althoff, M. (2017) 'Calculating human reachable occupancy for guaranteed collision-free planning', IEEE International Conference on Intelligent Robots and Systems, 2017-September, pp. 4473–4480. Available at: https://doi.org/10.1109/IROS.2017.8206314.

Suresh, P.; et al. (2024) 'Open Human-Robot Collaboration using Decentralized Inverse Reinforcement Learning'. Available at: https://www.merl.com (Accessed: 15 December 2025).

Weistroffer, V. et al. (2022) 'Using Physics-Based Digital Twins and Extended Reality for the Safety and Ergonomics Evaluation of Cobotic Workstations', Frontiers in Virtual Reality, 3, p. 781830. Available at: https://doi.org/10.3389/FRVIR.2022.781830/BIBTEX.

Xia, G. et al. (2026) 'A hierarchical human behavior modeling framework for safe and efficient human-robot collaborative assembly', Robotics and Computer-Integrated Manufacturing, 99, p. 103202. Available at: https://doi.org/10.1016/J.RCIM.2025.103202.

Xia, S. et al. (2017) 'A Survey on Human Performance Capture and Animation', Journal of Computer Science And Technology, 32(3), pp. 536– 554. Available at: https://doi.org/10.1007/s11390-017-1742-y.

Yang, X. et al. (2022) 'Towards the Understanding of Nudging Strategies in Cyber-Physical-Social System In Manufacturing Environments', Proceedings of the ASME Design Engineering Technical Conference, 3-B. Available at: https://doi.org/10.1115/DETC2022-90863.

Zhang, Z. et al. (2022) 'Prediction-Based Human-Robot Collaboration in Assembly Tasks Using a Learning from Demonstration Model', Sensors (Basel, Switzerland), 22(11), p. 4279. Available at: https://doi.org/10.3390/S22114279.