Journal of Inclusive Methodology and Technology in Learning and Teaching
ISSN 2785-5104
Anno 6 n. 3 (2026)

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

Learner and Instructor Perceptions of Teaching Industrial Robotics with Artificial Intelligence and Extended Reality

Davide Ermellino [1], Mirko Marras [1,*], Andrea Giovanni Martis [1], Diego Reforgiato Recupero [1], Daniele Riboni [1], Marco Sau [1] and Giuseppe Scarpi [1,2]

[1] University of Cagliari; d.ermellino1@studenti.unica.it, mirko.marras@unica.it, a.martis12@studenti.unica.it, diego.reforgiato@unica.it, daniele.riboni@unica.it, m.sau11@studenti.unica.it, giuseppe.scarpi@unica.it
[2] R2M Solution srl; giuseppe.scarpi@r2msolution.com
* Correspondence: mirko.marras@unica.it

Abstract: Extended reality offers opportunities for technical education, but its educational value depends on instructional design and usability rather than immersion alone. This study presents an approach to industrial robotics training that combines structured educational specifications, intelligence-assisted scenario authoring, and immersive practice. Two scenarios, addressing waste automotive battery sorting and refrigerator dismantling, were developed using a common five-module progression from guided familiarization to autonomous decision-making. A formative evaluation involved 16 learners and 4 instructors. Data were collected through questionnaires, anonymized interaction logs, and qualitative feedback. Learners reported positive perceived learning effectiveness and overall satisfaction, while instructors positively assessed the approach's suitability for teaching and training. Interaction logs indicated high completion across the guided modules but greater difficulty during autonomous decisions. Qualitative feedback identified opportunities to improve instructional clarity, navigation, readability, and interaction consistency. These findings support feasibility and perceived educational relevance of pedagogically structured immersive scenarios while identifying the need for improved scaffolding, accessibility-oriented design, and further evaluation using direct measures of learning outcomes.

Keywords: Extended Reality; Immersive Learning; Industrial Robotics Education; Instructional Design; Generative Artificial Intelligence.

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Extended reality (XR) is being explored in education and professional training because it allows learners to easily interact with simulated environments, manipulate virtual objects, and practice procedures that may be otherwise difficult to reproduce in conventional classrooms. These capabilities are particularly relevant to technical disciplines, where learning frequently involves spatial understanding, procedural knowledge, situational awareness, and decision-making. Research on immersive virtual reality has documented applications across higher education and identified opportunities for experiential and practice-oriented learning (Radianti et al., 2020; Stracke et al., 2025). Industrial robotics represents a relevant application domain among them, particularly as digital transformation is reshaping manufacturing processes and the competencies required to work with increasingly automated

production systems (Chirumalla et al., 2025; Milošević et al., 2024). Training in human–robot collaboration in manufacturing settings requires learners to understand operational sequences, recognize safety constraints, coordinate actions with automated equipment, and respond to changing task conditions. Indeed, physical training environments provide authentic experiences but may involve substantial costs, restricted equipment availability, and safety limitations. Immersive simulations can complement these environments by providing opportunities for repeated practice without requiring continuous access to industrial infrastructure.

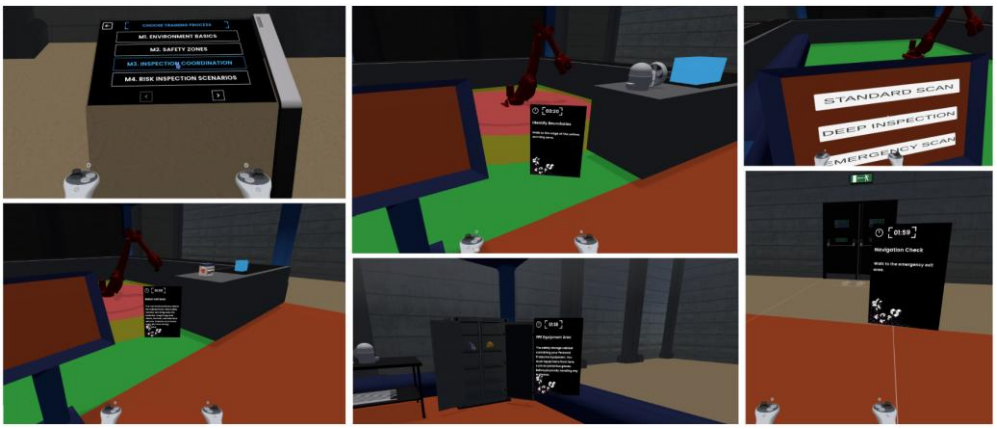
However, the relationship between immersion and learning remains under-explored. Although immersive environments may increase engagement and the perceived realism of educational activities, these characteristics do not necessarily translate into improved learning outcomes. Makransky et al. (2019), for example, found that an immersive science simulation increased perceived presence but produced lower learning outcomes and higher cognitive load than a desktop alternative. Similarly, Mayer et al. (2023) argue that immersive environments can promote meaningful cognitive engagement while also introducing distractions. These emerging findings challenge the assumptions that technological immersion is inherently educationally beneficial. Instead, they suggest that the effectiveness of XR depends on how learning activities are structured, how information is presented, and how learners are supported throughout the experience. Specific reviews targeting educational XR emphasize the importance of instructional design, appropriate interaction mechanisms, and alignment between learning objectives and the characteristics of immersive environments (Radianti et al., 2020; Stracke et al., 2025).

One important consideration under this context is the relationship between instructional guidance and learner autonomy. Technical training often requires learners to progress from understanding an environment and following prescribed procedures toward interpreting operational situations and making independent decisions. Excessive guidance may limit opportunities for autonomous application, whereas insufficient guidance may increase uncertainty and cognitive demands, particularly when learners must simultaneously process instructional content and interact with an immersive environment (Mayer et al., 2023). In immersive environments, this challenge is compounded by the need to learn navigation and interaction mechanisms alongside the target educational content. Consequently, the design of XR learning experiences should distinguish difficulties associated with the subject from those introduced by the interface. For example, an incorrect inspection decision may reflect inadequate procedural understanding, ambiguous visual information, or uncertainty about how to interact with an object. These explanations have different instructional implications and cannot be reliably distinguished through satisfaction scores alone. A further consideration concerns the development and maintenance of immersive educational content. Creating XR scenarios typically requires coordination between instructional designers, domain experts, software developers, and 3D creators. Recent advances in generative and agentic artificial intelligence have introduced architectures in which multiple agents can coordinate planning, tool use, content generation, and iterative task execution (Park et al., 2023; Wu et al., 2024). Nevertheless, automated content production raises questions about the preservation of educational intentions, the consistency and controllability of generated interactions, and the ability of human experts to inspect, correct, and supervise system outputs (Amershi et al., 2019).

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

In this paper, we address these abovementioned challenges through a design process that places an explicit pedagogical specification layer before technical implementation. First, the learning objectives, competence references, operational parameters, instructional sequences, and monitoring indicators are documented in a structured worksheet. Then, a multi-agent framework supports the translation of this specification into executable immersive environments. The resulting training format organizes activities into modules that move from familiarization and guided practice toward autonomous operations (Figure 1). In the present study, we evaluate this approach through two industrial training scenarios addressing waste automotive battery sorting and refrigerator dismantling. Unlike evaluations focused exclusively on technical functionality, the study combines quantitative and qualitative learner perceptions, behavioural interaction logs, and instructor assessments. This multi-perspective design allows the scenario format to be examined both as an interactive learning experience and as a potentially reusable educational resource. The main contribution is an empirical examination of the relationship between pedagogical structure, learner interaction, and stakeholder perceptions in intelligence-assisted immersive training. The findings indicate positive perceived educational value and reliable progression via guided activities, while identifying autonomous decision-making and interaction clarity as priorities for refinement. Rather than showing learning gains, the present study provides formative evidence to guide the further development and evaluation of pedagogically grounded XR training environments.

Figure 1. Representative views of the used immersive training environment. The panels illustrate key components of the learner experience, including module selection, navigation and safety guidance, interaction with the industrial workspace and robotic system, inspection options, personal protective equipment preparation, and task-specific prompts supporting progression through the training workflow.



2. Materials and Methods

2.1. Research Design

The present study adopted a formative, descriptive evaluation design combining questionnaire responses, application-generated interaction logs, and qualitative feedback. The objective was to assess the technical executability, usability, perceived educational value, and instructional suitability of the immersive scenario format. The evaluation involved two main participant categories: learners and instructors. Each category contributed evidence relevant to its relationship with the environment. Learners directly experienced the operational workflow, while instructors assessed

its suitability for teaching and training. The study did not employ a randomized experimental design or a control group for learners to measure learning gains. Accordingly, the evaluation was intended to characterize the developed experience and identify opportunities for improvement rather than establish learning outcomes.

2.2. Study Context and Technical Environment

The present study was conducted in the TI-RAX project (http://master-ti-rax.github.io) under the MASTER ecosystem (https://www.master-xr.eu/), with the aim of supporting industrial robotics education through immersive training. Two industrial scenarios were implemented: waste automotive battery sorting and refrigerator dismantling. Both scenarios were designed using a common educational structure and deployed through VIROO (https://viroo.virtualwareco.com/). The scenarios combine virtual industrial environments, interactive objects, instructional panels, task progression mechanisms, and decision points associated with safety, inspection, and robot-assisted operations. Learners are required to navigate the environment, interpret the provided instructions, interact with virtual objects, and complete progressively more demanding activities. Although the scenarios represent different industrial processes, their common instructional structure allows the format to be examined beyond the characteristics of an individual application.

2.3. Educational Design for Immersive Scenarios Generation

The educational design process begins with a structured natural-language educational worksheet. This document specifies the intended audience, learning objectives, competence references, operational parameters, learning activities, expected outcomes, monitoring indicators, and instructional progression. The worksheet serves as a shared representation of educational intentions for instructors, domain experts, and developers. It also provides the main basis for translating pedagogical requirements into executable interactions. The learning experience supported by the worksheet is organized into five sequential modules, summarized in Table 1. The sequence is designed to reduce instructional support progressively. Initial modules introduce navigation, safety, and procedural knowledge, while later modules require learners to apply encountered information with greater autonomy.

Table 1. Instructional organization of the training format. The modules describe the evolution of the learner experience from environment familiarization and safety awareness to guided inspection, scenario-based practice, and increasingly autonomous decision-making activities.

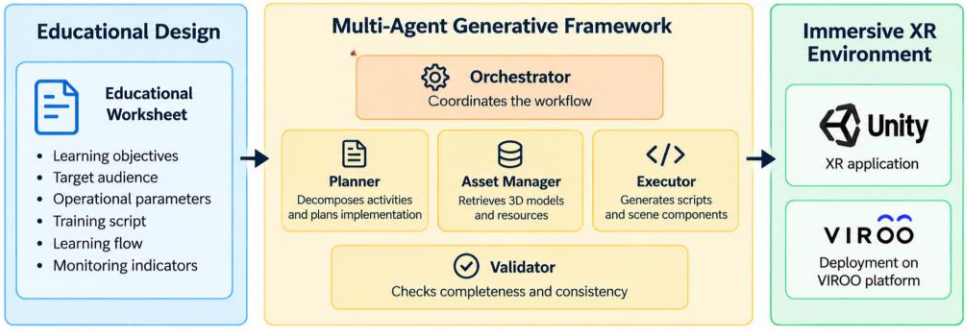
Module	Educational purpose	Main activities	Guidance
M1. Environment Basics	Familiarization with the environment	Orientation, movement, safety checks, robot awareness	High
M2. Safety Zones	Recognition of safety constraints	Safety-zone identification, restricted-area behaviour, emergency procedures	High
M3. Inspection Coordination	Highly guided procedural practice	Inspection tasks and coordination with robotic operations	High
M4. Scenario Inspections	Core application of inspection procedures	Assessment of intact, deformed, leaking, and challenging cases	Moderate
M5. Autonomous Operations	Independent procedural tasks	Safety preparation and routing decisions conducted alone	Low

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

2.4. Immersive Scenario Generation with Artificial Intelligence

The technical authoring process developed on top of the described instructional design is supported by a multi-agent generative workflow (Figure 2). The framework translates the structured educational specification into executable XR components while preserving traceability between pedagogical requirements and implementation. An *Orchestrator* manages the workflow and coordinates the interaction among specialized agents. The *Planner* converts the educational worksheet into an ordered set of implementation tasks, identifying the scene elements, interactions, and dependencies required to realize each activity. The *Asset Manager* retrieves the 3D models, scripts, and project resources needed for those tasks. The *Executor* uses the resulting plan and assets to generate or assemble Unity scripts and scene components. The *Validator* then checks the generated artifacts for completeness, consistency, and alignment with the requested specifications, providing feedback when further refinement is required. This creates an iterative generation-and-validation loop. Human experts remain involved throughout the workflow and can inspect, modify, or approve intermediate developed outputs, ensuring that automation supports - rather than replaces - pedagogical and technical oversight.

Figure 2. Multi-agent framework for generating immersive training scenarios. The framework translates a structured educational worksheet into an XR training application through coordinated planning, asset retrieval, content generation, and validation, while preserving human oversight and enabling subsequent evaluation.



2.5. Participant Recruitment for Validation

The validation involved 20 participants: 16 learners and 4 instructors. Learners were university-level students aged 19–30 years with general or entry-level technical backgrounds. Instructors were university-level teaching or evaluation personnel, aged 30–60 years. Participants were recruited through a snowballing strategy. The validation sessions could be attended in person or remotely where compatible equipment was available. Laboratory equipment was provided for participants who lacked the necessary devices or operating systems. Seventeen learners initially registered, of whom 16 completed the questionnaire. Fifteen valid interaction logs were available for behavioural analysis.

2.6. Validation of the Immersive Scenarios with Participants

The validation followed five phases. First, participants received an introduction to the project, the scenarios, the intended learning objectives, and the immersive environment. Second, the instructional structure was presented, including the role of guidance, safety-related information, checkpoints, and autonomous decision-

making. Third, participants accessed or reviewed the application according to their role. Learners completed the operational workflow, proceeding sequentially through the five modules. Instructors examined the experience from the perspective of educational suitability. The learner workflow was configured for a validation session of approximately 30 minutes. Parameterized activities, including routing decisions in the autonomous module, were reduced to make completion feasible within the session time. Fourth, learner interaction logs were automatically generated and exported. Participants were instructed to save and send these log files before closing the application. Finally, questionnaires and qualitative feedback were collected.

2.7. Analysis of the Collected Quantitative and Qualitative Feedback

Closed-ended questionnaire responses were mainly summarized via descriptive statistics, including means, medians, and response distributions. Interaction logs were analysed at module and task level. Module completion was classified as full when all expected activities were completed, partial when at least one activity was performed without reaching full progress, and not completed when no meaningful progress was recorded. Task outcomes were classified as correct, incorrect, or not evaluated according to the application-generated records. Durations were examined to characterize differences in execution time across modules. Qualitative comments and discussion notes were examined inductively to identify recurring concerns and observations. Themes included instructional guidance, readability, navigation, object interaction, decision support, and potential scenario reuse. The evidence sources were interpreted to distinguish perceived experience, observable task execution, and stakeholder assessments. No inferential statistical tests were performed.

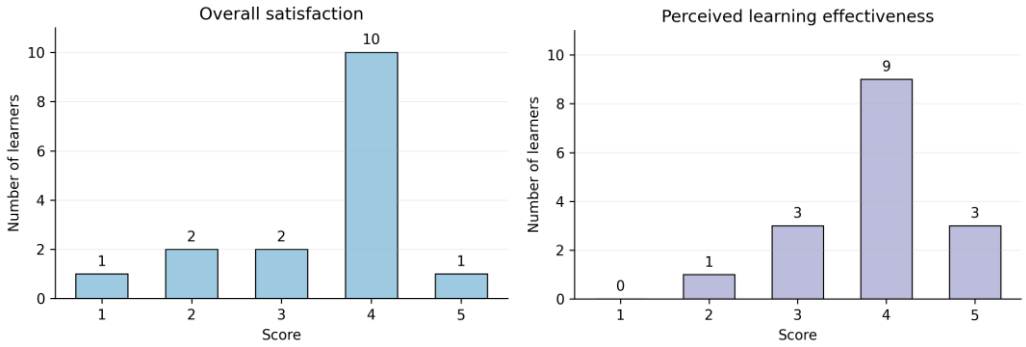
3. Experimental Results

3.1. Learner Perceptions

Sixteen learners completed the post-experience questionnaire (Figure 3). Their satisfaction reached a mean of 3.50/5 and a median of 4.00. Eleven of the 16 learners selected a score of 4 or 5. Perceived effectiveness in supporting learning was higher, with a mean of 3.88/5 and a median of 4.00. Twelve of the 16 learners selected a score of 4 or 5. The difference between these two indicators is relevant because it suggests that learners could recognize educational usefulness even when their overall experience was affected by aspects of navigation or interaction. Additional feedback was consistent with this interpretation. Visual quality was generally perceived positively and interaction mechanisms were accepted, although some responses remained neutral. Navigation and locomotion generated the most heterogeneous perceptions. Learners also evaluated the use of immersive training positively in relation to safety, an important consideration in industrial robotics contexts in which physical practice involves hazards or limited access to equipment. Open-ended comments showed several recurring areas for improvement. Learners requested clearer instructions and more visible checkpoints, particularly when moving from one task state to another. Some reported insufficient readability of text embedded in the environment and suggested larger text, stronger visual contrast, or more persistent presentation. Movement was occasionally perceived as slow or rigid. Other comments concerned the consistency of object-interaction cues and the amount of information available when choosing among alternative inspection or

routing actions. These observations indicate that the main perceived problems were not related to the general rationale of immersive training but to the interaction layer.

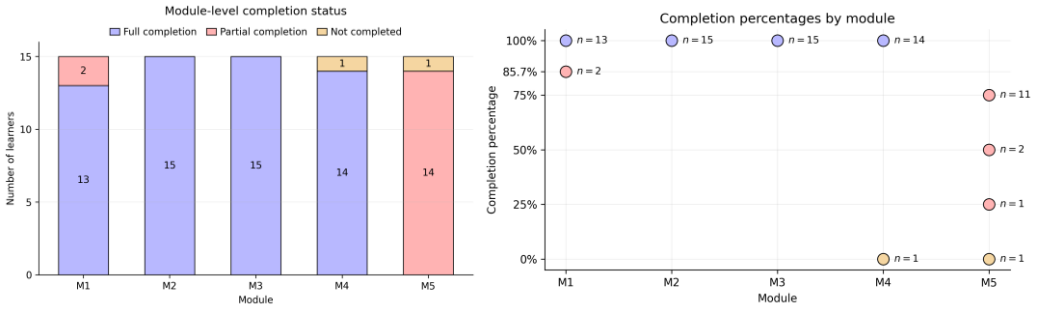
Figure 3. Explicit learner evaluations of the training experience. Distribution of responses to the two closed-ended questionnaire items assessing (a) overall satisfaction and (b) perceived effectiveness in supporting learning. Responses were provided on a five-point scale, with higher scores indicating more positive evaluations; values above the bars report the number of learners selecting each score.



(**a**) (**b**)

Fifteen valid learner logs were available, producing learner-module records across the five training modules. Figure 4 (a) summarizes their completion status. *Environment Basics* (M1) was fully completed by 13 learners and partially completed by two. *Safety Zones* (M2) and *Inspection Coordination* (M3) were fully completed by all 15 learners. *Scenario Inspections* (M4) was fully completed by 14 learners, while one did not complete the module. A substantially different pattern emerged for *Autonomous Operations* (M5), where no learner reached full completion: 14 learners partially completed the module and one did not complete it. Figure 4 (b) provides a more fine-grained view by showing the completion percentages reached within each module. In M1, 13 learners reached 100% completion and two reached 85.7%, whereas all learners reached 100% in both M2 and M3. In M4, 14 learners completed the entire module and one remained at 0%. M5 showed a broader distribution: 11 learners reached 75% completion, two reached 50%, one reached 25%, and one remained at 0%. Taken together, the results show that incomplete progression was not distributed uniformly across the training experience. Learners progressed consistently through the first four modules, particularly the highly guided M2 and M3 activities, whereas the main completion bottleneck was concentrated in the autonomous module. Additional log-level evidence supports this interpretation. In M1, the longest activities were *Movement* (261.4 s) and *Orientation* (126.5 s), although both were completed correctly by all learners, suggesting that the additional time was primarily associated with familiarization rather than task failure. By contrast, the final module showed a deterioration in decision performance: across the 45 routing-decision records, 24 were correct, 13 incorrect, and 8 not evaluated, with correct responses decreasing from 12 in the first routing decision to 11 in the second and only 1 in the third. These observations suggest that the transition from predominantly full completion in M1–M4 to partial completion in M5 was accompanied by a shift from interaction familiarization toward greater procedural and decision-making difficulty. However, the log data do not allow us to determine whether this emerging pattern was mainly driven by increased conceptual demands, reduced instructional support, or interaction-related difficulties.

Figure 4. Implicit learner evaluations of the training experience. Distribution of learner completion outcomes across the five training modules, showing (a) the number of learners achieving full, partial, or no completion and (b) the completion percentages reached within each module. Values reported in the plots indicate the number of learners associated with each completion outcome or percentage.



(**a**) (**b**)

3.2. Instructor Perceptions

Four instructors evaluated the usefulness of the experience for teaching or training activities. The mean response was 4.25/5, with a median of 4.50. Two instructors selected the maximum score of 5, one selected 4, and one selected 3; no instructor selected either of the two lowest response categories. Supplementary usability evidence was consistent with this result. The instructors obtained an aggregate System Usability Scale score of 87/100, indicating a very positive usability assessment within this group. Instructor feedback was particularly useful for interpreting the learner difficulty observed in the final module. The transition from highly guided activities to autonomous decision-making was not considered undesirable in itself. Instead, instructors viewed this transition as educationally meaningful because it moves learners from recognition and procedural reproduction toward independent application. From this perspective, the low completion observed in the final module should not necessarily be addressed by converting it into another fully guided activity. A more appropriate response is to improve the timing and quality of scaffolding. Learners should remain responsible for making decisions while receiving enough information to interpret the task state and understand the consequences of an incorrect choice. Instructors also emphasized the value of making industrial procedures observable and repeatable. Unlike a textual explanation, the scenario requires learners to move within an operational space, identify objects and safety zones, perform actions, and observe consequences of decisions. The same environment can therefore support instructor-led discussion and debriefing. At the same time, instructors stressed that this function depends on an unambiguous instructional layer. Task objectives, prompts, feedback, and interaction cues need to be sufficiently clear to ensure that learners are challenged by the decision rather than by uncertainty about how the interface works.

4. Discussion

The present study provides a formative evaluation of a pedagogically structured immersive training approach for industrial robotics, combining learner perceptions, behavioural evidence, and instructor feedback. Rather than assessing immersion as a technology in isolation, the analysis focuses on how instructional structure, learner

autonomy, and intelligence-assisted authoring interact within the training experience. Three main considerations emerge from the results.

A first takeaway is that the educational value of immersive training depends on the pedagogical structure of the experience rather than on immersion alone. This is consistent with prior research showing that XR can support experiential and practice-oriented learning while also introducing cognitive and interaction-related demands when instructional design is insufficiently calibrated (Mayer et al., 2023; Radianti et al., 2020; Stracke et al., 2025). In the present study, learners progressed reliably through the guided safety and inspection activities, whereas the main difficulties emerged in the final autonomous module. The log data further show why aggregate indicators should be interpreted cautiously: longer execution times during initial orientation were associated with correct performance and mainly reflected familiarization with the environment, while shorter routing decisions in the autonomous phase produced substantially more errors. These findings suggest that completion time, correctness, and progress are most informative when interpreted together and in relation to the instructional function of each activity.

A second takeaway concerns the balance between scaffolding and autonomy. The lower completion observed in the final module should not necessarily be interpreted as evidence that learners require more guidance throughout the entire training experience. The purpose of this phase was precisely to shift from structured practice toward more independent application of previously introduced procedures. At the same time, the results indicate that reduced guidance may expose uncertainty related to task state, decision requirements, or interaction cues, in line with previous work highlighting the cognitive demands that immersive environments can impose on learners (Mayer et al., 2023). The design challenge is therefore to preserve meaningful learner responsibility while reducing avoidable ambiguity. More visible checkpoints, concise recap prompts, clearer task-state information, and explanatory or adaptive feedback could provide such desired support without transforming the autonomous phase into another fully guided activity.

A third takeaway concerns the role of instructional design and human oversight in the intelligence-assisted authoring. The presented workflow maintains an explicit educational specification layer before technical implementation, making learning objectives, task sequencing, expected outcomes, and monitoring criteria inspectable independently of the generated XR components. This helps prevent the authoring process from becoming purely technology-driven and is consistent with broader recommendations for maintaining transparency and meaningful human control in intelligence-supported systems (Amershi et al., 2019). The same principle applies to usability and accessibility: learner feedback on readability, navigation, and interaction consistency shows that interface characteristics can directly affect engagement with the educational content and should therefore be treated as pedagogical concerns rather than secondary technical details. Overall, the results support a hybrid model in which artificial intelligence assists implementation, while instructors and domain experts retain responsibility for defining, inspecting, and refining the experience.

5. Conclusions

In this study, we presented and evaluated a pedagogically structured approach to immersive industrial robotics training that combines explicit educational specifications, multi-agent intelligence-assisted authoring, and immersive practice.

The findings provide positive formative evidence regarding perceived educational value, instructional suitability, and learner progression through guided activities, while identifying autonomous decision-making as the main point of difficulty. Taken together, our findings suggest that effective immersive training depends on the alignment between instructional objectives, interaction design, scaffolding, and technical implementation. Artificial intelligence can support the realization of these environments, but its contribution is most valuable when embedded within a transparent workflow that preserves pedagogical intent and human oversight.

The study also presents limitations that directly motivate future work. The relatively small and non-representative sample limits generalizability and should be addressed through larger studies involving vocational learners, industrial trainees, and professionals engaged in upskilling or reskilling. The evaluation covered only two scenarios built on the same instructional format, calling for validation across more diverse industrial processes and task requirements. Moreover, the current evidence concerns satisfaction, perceived usefulness, instructor judgments, and behavioural execution rather than direct learning gains; future studies should hence include pre/post assessments, retention measures, transfer tasks, and controlled comparisons with non-immersive or conventional training approaches. Finally, the short first-use setting and the limited explanatory power of interaction logs motivate longitudinal evaluations combining behavioural traces with debriefings, qualitative explanations, and accessibility-oriented assessments to better understand the extent to which learners develop competence and autonomy over repeated use.

References

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human–AI interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Paper 3, pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300233

Chirumalla, K., Oghazi, P., Nnewuku, R. E., Tuncay, H., & Yahyapour, N. (2025). Critical factors affecting digital transformation in manufacturing companies. International Entrepreneurship and Management Journal, 21, Article 54. https://doi.org/10.1007/s11365-024-01056-3

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. Learning and Instruction, 60, 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007

Mayer, R. E., Makransky, G., & Parong, J. (2023). The promise and pitfalls of learning in immersive virtual reality. International Journal of Human–Computer Interaction, 39(11), 2229–2238. https://doi.org/10.1080/10447318.2022.2108563

Milošević, I. M., Plotnic, O., Tick, A., Stanković, Z., & Buzdugan, A. (2024). Digital transformation in manufacturing: Enhancing competitiveness through Industry 4.0 technologies. Precision Mechanics & Digital Fabrication, 1(1), 31–40. https://doi.org/10.56578/pmdf010104

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (Article 2, pp. 1–22). Association for Computing Machinery. https://doi.org/10.1145/3586183.3606763

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. Computers & Education, 147, Article 103778. https://doi.org/10.1016/j.compedu.2019.103778

Stracke, C. M., Bothe, P., Adler, S., Heller, E. S., Deuchler, J., Pomino, J., & Wölfel, M. (2025). Immersive virtual reality in higher education: A systematic review of the scientific literature. Virtual Reality, 29, Article 64. https://doi.org/10.1007/s10055-025-01136-x

Wu, Q., Bansal, G., Zhang, J., Wu, Y., Li, B., Zhu, E., Jiang, L., Zhang, X., Zhang, S., Liu, J., Awadallah, A. H., White, R. W., Burger, D., & Wang, C. (2024). AutoGen: Enabling next-gen LLM applications via multi-agent conversations. In First Conference on Language Modeling (COLM).