Editorial: Beyond the overlay: generative AI, vision language models and skills transfer in maintenance

The translation gap

Ask a maintenance engineer to "replace the seal." A quiet chain of interpretation begins. Which seal? In which orientation? To what torque? In what order? For decades, experience filled these gaps. Paper documented the rest. Augmented reality (AR) promised to move that information into the technician's line of sight. In several well-defined maintenance and repair settings, it has delivered measurable benefits.

The harder question for this journal is now different. What does AR become when generative artificial intelligence can author, interpret and reason about the work itself?

This editorial argues that the next research frontier is not AR display technology alone, but the generation, grounding, verification, governance and learning effects of AI-mediated maintenance guidance. The central question is whether AR systems can produce correct guidance, connect it to the physical asset, verify it before action and leave the technician more capable over time.

The maturing baseline

AR in maintenance has moved beyond speculation in several task settings. Overlays can register instructions onto the physical asset. Headsets can keep the hands free. Remote experts can see what the maintenance worker sees. Guided procedures can generate digital records for audit.

The evidence base has also matured. Eversberg and Lambrecht (2023) studied turbine blade repair with ten experienced metalworkers. Using an AR assistance system, the workers completed a familiar repair task 21% faster. Their perceived workload fell by 26%. Every participant preferred the system to paper. Earlier comparative work reached compatible conclusions across different media (Funk *et al.*, 2016).

These results are important, but they should be placed in perspective. The demonstration that AR can help under suitable task and deployment conditions should no longer be treated as the central research question. The frontier has shifted toward scalable authoring, contextual interpretation, verification, human learning and organizational integration. This is consistent with systematic review evidence showing that AR in maintenance has progressed substantially, but that industrial deployment remains constrained by fragmentation in hardware, software, tracking, interaction design and evaluation methods (Palmarini *et al.*, 2018).

The authoring bottleneck

The constraint that has limited AR maintenance at scale is content. Every asset needs spatially registered instructions. Every engineering change demands new ones. When the underlying model changes, the AR content quietly goes stale. The dominant cost was never only the headset. It was the authoring labor. It was also the burden of keeping thousands of procedures aligned with engineering reality.

Generative AI addresses this problem first. This is its least speculative contribution. Emerging commercial platforms are beginning to generate AR work instructions from CAD models and process definitions (Druet, 2025). However, independent evidence on accuracy, maintainability and lifecycle cost remains limited. Computer vision pushes this further.

Journal of Quality in Maintenance Engineering
Vol. 32 No. 3, 2026
pp. 629-633

e-ISSN: 1758-7832
p-ISSN: 1355-2511
DOI 10.1108/JQME-08-2026-165

Systems that recognize objects and assembly states can detect what the technician has actually done (Malta *et al.*, 2023). They can advance steps automatically. They can flag an action that did not produce the expected result.

Two distinct forms of generative capability should be separated. The first is offline authoring, where AI converts engineering documents, CAD models and static process definitions into AR-ready instructions. This supports maintainability and couples the content lifecycle directly to the engineering lifecycle. The second is online guidance, where AI dynamically interprets the live scene and adapts instructions during execution. This form is not only more powerful but also more safety critical.

This structural shift matters. AR moves from a bespoke deployment to a maintainable system. For a maintenance organization, this is decisive. It is the difference between a pilot that impresses on one asset and a program that survives a large equipment register. The research question changes accordingly. It is no longer whether we can overlay a step. It is whether we can generate, validate and version that step automatically.

Closing the translation gap

An overlay is not understanding. The deeper frontier is interpretation. The system must read an abstract instruction and ground it in the scene in front of the worker. It must resolve "tighten the bolt" into which bolt, in what sequence, on this assembly.

Early work paired large language models with AR for exactly this purpose. Xu *et al.* (2024) converted textual maintenance instructions into executable, context-specific actions. They targeted the ambiguity that arises when abstract text must become precise motion. More recent systems generate the visual guidance itself. Zhao *et al.* (2025) combined language models with vision models to produce situated, visually rich guidance on demand. No author needs to place each cue in advance. Vision language models enable content that adapts to the scene in real time (Behravan and Gračanin, 2024). The broader trajectory points the same way. Generative AI is meeting extended reality. The result is interfaces that interpret ambiguous instructions, parse physical scenes and generate three-dimensional content (Zhu *et al.*, 2026).

For maintenance, the implication is concrete. In such systems, the AR interface becomes the visible layer of a broader reasoning pipeline. Language models interpret instructions. Vision models ground them in the physical scene. Verification modules assess whether proposed actions are safe and correct. AR therefore begins to support task reasoning rather than merely task display. It can recognize an unexpected condition. It can suggest a revised sequence. It can explain the basis for that suggestion. This is also where AR meets the wider shift in maintenance toward agentic and foundation models. Those systems increasingly act as copilots to human experts (Lukens *et al.*, 2024).

Verification, provenance and trust

The move from display to reasoning changes the risk profile. In safety-critical maintenance, an incorrect instruction is not merely a usability defect. It can create asset damage, production loss, environmental risk or injury. A vision language model that confidently misidentifies a component in a safety-critical task is worse than no guidance.

AI-generated guidance therefore requires an independent verification layer. That verifier could draw on a knowledge graph, a numerical solver, a rules engine, a validated procedure library or live digital twin state. The important point is independence. The model that generates a step should not be the only authority that approves it.

Provenance is equally important. Procedures authored or adapted by AI need traceability to an engineering source of truth. They need model version records, approval workflows, technician override mechanisms and audit trails. In regulated assets, the key question is not only whether the instruction is useful. It is who approved it, based on what evidence, under which asset conditions and with what accountability.

Trust must also be calibrated. Technicians should neither ignore useful guidance nor comply blindly with uncertain guidance. Systems should disclose uncertainty, identify the source of recommendations and make clear when an instruction is generated, retrieved, modified or verified. Trust in AI-mediated AR should be earned through evidence, not assumed through interface design.

Skills transfer and the human factor

The most important long-term implication concerns the workforce. The maintenance labor force is aging and thinning. Tacit knowledge leaves faster than it can be captured. This concern recurs across the prognostics literature as much as in AR (Lukens *et al*., 2024). Reframed this way, AR becomes a training instrument. Adaptive AR tutoring can tailor guidance to the individual learner. It can shorten the path to proficiency (Huang *et al*., 2021). In some settings, structured AR tutoring may reduce dependence on prolonged shadowing.

A serious editorial must hold the counterweight. The evidence on cognitive load is mixed and strongly task-dependent. Kolla *et al*. (2021) found AR superior to paper in errors and usability. Yet they found no significant difference in completion time or workload. Benefits often concentrate in complex tasks and can vanish on simple ones.

There is also a subtler risk that should concern anyone serious about skills transfer: deskilling driven by automation complacency. A technician who passively follows step-by-step overlays may never build a durable mental model of the machinery. If guidance routinely substitutes for spatial and mechanical reasoning, the worker becomes highly capable during assisted execution, yet less capable over time when left unaided. This represents a decoupling of immediate task performance from long-term diagnostic capability (Parasuraman and Riley, 1997).

The design imperative follows. AR for maintenance should build understanding, not merely secure compliance. A system that improves immediate task completion while weakening long-term diagnostic judgment is a poor organizational investment. Future studies should therefore measure not only assisted performance but also unaided retention, transfer to unfamiliar faults and the technician's willingness and ability to challenge incomplete or incorrect AI guidance.

The evidence problem

Much of the headline benefit circulates as vendor benchmarks rather than controlled science. The peer-reviewed picture is more nuanced. Some trials show that AR can reduce errors, but benefits depend heavily on display technology, task complexity and operator expertise (Funk *et al*., 2016). These moderators are too often left uncontrolled in empirical designs.

This journal can help raise the standard. We invite properly powered comparative trials. We ask authors to stratify by task complexity and operator expertise and to employ standardized error taxonomies rather than ad hoc error counts. Crucially, the generative era introduces a comparison the field has not yet made rigorously: guidance authored by AI against guidance authored by humans, held to the exact same evaluative standard.

Because skills transfer is the ultimate claim, outcome measures must include long-term retention and transfer. They cannot stop at completion time measured during active assistance. Researchers should ask whether a technician can perform the task later without guidance, diagnose a variant fault, detect an incorrect instruction and explain the engineering rationale behind their maintenance actions.

Open challenges and a research agenda

Transitioning from basic visual overlays to generative reasoning introduces deep socio-technical hurdles. Beyond structural and algorithmic verification, industrial plants impose harsh operating constraints. These include edge inference requirements, latency penalties, unreliable connectivity, device ruggedness and the physical ergonomics of prolonged headset use during long shifts. In such settings, verification cannot be treated as an optional safeguard.

Beneath these constraints sits a fundamental governance question: who validates an instruction authored by AI and against what reference and asset state?

To address these challenges, this journal proposes an organized research agenda for prospective contributors, structured around four core clusters.

(1) Content lifecycle

- Authoring fidelity: Develop methods and benchmarks to automatically generate spatially registered AR work instructions from CAD and PLM models. Validate their correctness against physical ground truth and detect stale content automatically when designs change.
- AI-authored vs human-authored guidance: Run adequately powered trials using standardized error taxonomies to compare AI-generated guidance directly against expert human-generated documentation under identical task and asset environments.

(2) Contextual intelligence

- Grounded text-to-action: Build vision-language pipelines capable of resolving referential ambiguity using real-time scene context, asset models and historic maintenance logs. Quantify grounding accuracy in cluttered, low-light or hazardous industrial settings.
- Digital twin integration: Drive AR overlays dynamically from live asset health data and failure histories. Close the loop from a prognostic alert to an AR-guided intervention, ensuring the system captures outcomes to update the maintenance knowledge base and improve future guidance.

(3) Safety and governance

- Verification-in-the-loop: Couple generative pipelines with independent semantic verifiers, such as knowledge graphs and rules engines. Confirm the technical validity of an instructional step before it is visually displayed to the user and benchmark how effectively these verifiers catch model errors.
- Trust calibration and provenance: Study how technicians calibrate their reliance on AI-generated guidance. Define user interface approaches that transparently disclose uncertainty, track traceability back to an engineering source of truth and formalize clear approval workflows for regulated assets.

(4) Human capability

- The science of skills transfer: Design controlled studies isolating whether AR prompts build durable mental models or foster operational dependence. Measure retention, diagnostic flexibility and error-detection capability days or weeks after assistance is removed.
- Adaptive scaffolding: Engineer AR tutoring systems that progressively withdraw explicit visual assistance and hints as the technician's competence grows.
- The human factors envelope: Systematically map the inflection points where AR assistance harms cognitive engagement, accounting for task complexity, baseline expertise, display modalities and the ergonomic load of extended headset wear.
- Multilingual and equitable guidance: Explore the capacity of generative models to translate and adapt instructions across varied languages and technical literacy levels, mitigating regional labor constraints.

The overlay was the first phase of maintenance AR; the questions ahead are more consequential. Can automated guidance be shown to be correct? Can it be grounded reliably in complex physical environments? Can it be verified before it reaches a worker's line of sight? Most importantly, can it leave that worker more capable, not merely more compliant?

These questions will shape the next decade of industrial AR. This journal welcomes the empirical trials, system architectures, human factors studies and field deployments that will answer them.

Mohamed Ben-Daya

Acknowledgments
LLMs were used for English editing and searching for some suitable references.

References

Behravan, M. and Gračanin, D. (2024), "Generative multi-modal artificial intelligence for dynamic real-time context-aware content creation in augmented reality", *Proceedings of the 30th ACM Symposium on Virtual Reality Software and Technology*, pp. 1-2.

Druet, H. (2025), "Augmented reality will change manufacturing workflows", Automation.com, available at: https://www.automation.com/article/augmented-reality-change-manufacturing-workflows (accessed 15 June 2026).

Eversberg, L. and Lambrecht, J. (2023), "Evaluating digital work instructions with augmented reality versus paper-based documents for manual, object-specific repair tasks in a case study with experienced workers", *The International Journal of Advanced Manufacturing Technology*, Vol. 127 No. 3, pp. 1859-1871, doi: 10.1007/s00170-023-11313-4.

Funk, M., Kosch, T. and Schmidt, A. (2016), "Interactive worker assistance: comparing the effects of in-place projection, head-mounted displays, tablet, and paper instructions", *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*.

Huang, G., Qian, X., Wang, T., Patel, F., Sreeram, M., Cao, Y., Ramani, K. and Quinn, A.J. (2021), "Adaptutar: an adaptive tutoring system for machine tasks in augmented reality", *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1-15.

Kolla, S.S.V.K., Sanchez, A. and Plapper, P. (2021), "Comparing effectiveness of paper based and augmented reality instructions for manual assembly and training tasks", *SSRN Electronic Journal*.

Lukens, S., McCabe, L.H., Gen, J. and Ali, A. (2024), "Large language model agents as prognostics and health management copilots", *Annual Conference of the PHM Society*, Vol. 16 No. 1, doi: 10.36001/phmconf.2024.v16i1.3906.

Malta, A., Farinha, T. and Mendes, M. (2023), "Augmented reality in maintenance—history and perspectives", *Journal of Imaging*, Vol. 9 No. 7, 142, doi: 10.3390/jimaging9070142.

Palmarini, R., Erkoyuncu, J.A., Roy, R. and Torabmostaedi, H. (2018), "A systematic review of augmented reality applications in maintenance", *Robotics and Computer-Integrated Manufacturing*, Vol. 49, pp. 215-228, doi: 10.1016/j.rcim.2017.06.002.

Parasuraman, R. and Riley, V. (1997), "Humans and automation: use, misuse, disuse, abuse", *Human Factors*, Vol. 39 No. 2, pp. 230-253, doi: 10.1518/001872097778543886.

Xu, F., Nguyen, T. and Du, J. (2024), "Augmented reality for maintenance tasks with ChatGPT for automated text-to-action", *Journal of Construction Engineering and Management*, Vol. 150 No. 4, 04024015, doi: 10.1061/jcemd4.coeng-14142.

Zhao, A.Y., Gunturu, A., Do, E.Y.L. and Suzuki, R. (2025), "Guided reality: generating visually-enriched AR task guidance with LLMs and vision models", *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*, pp. 1-15.

Zhu, M., Chen, J. and Li, B. (2026), "When generative AI meets extended reality: enabling scalable and natural interactions", *IEEE Internet Computing*, pp. 15-24.