

Original Article

AI Copilots for Warehouse Supervisors: Human-in-the-Loop Decisioning in SAP EWM

Mahir Anwar

University of Southern California, Los Angeles, CA

Received Date: 07 September 2026

Revised Date: 12 September 2026

Accepted Date: 18 September 2026

Abstract: AI copilots may enhance SAP Extended Warehouse Management beyond mere transactions, enabling warehouse supervisors to make better decisions on warehouse-related issues, such as wave release, workload balancing and reallocation, replenishment escalation, and exception resolution. Unlike fully automated control, the human-in-the-loop use case considered here keeps recommendations subject to human judgment, accountability, and intervention. This review examines the conceptual and empirical foundations of AI-supported warehouse supervision, drawing on warehouse automation, logistics digitalisation, human factors, algorithmic control, and hybrid-intelligence research. When task boundaries are clear, confidence information is communicated, and override rights are available, the literature suggests that AI support can enhance information processing and decision consistency. The reviewed literature also identifies risks including automation bias, algorithm aversion, opaque recommendations, skill erosion, and conflicts between local and system-level objectives. There is significant uncertainty about the effectiveness, safety, governance, and user acceptance of an embedded SAP EWM supervisor copilot as no peer-reviewed journal study identified in this review directly assesses such an implementation. To progress, then, it will be necessary to rely on field experiments, decision architectures that can be tracked, calibrated explanations, and an evaluation against combined operational and human-centred outcomes.

Keywords: Algorithmic Decision Support, Human-In-The-Loop, SAP EWM, Warehouse Copilots, Warehouse Supervision

I. INTRODUCTION

Warehouse supervision requires rapid decision-making under dynamic workload, inventory, equipment, labour, and service-level constraints. Relevant decisions in SAP Extended Warehouse Management (EWM) can include releasing or splitting waves, reassigning warehouse orders, changing the priority of the queues, escalating replenishment shortages, managing resource shortages and resolving execution exceptions. An AI copilot can monitor these operational indicators, predict outcomes, rate interventions and generate a rationale and supporting evidence for supervisory review. This should be treated as a structured, bidirectional collaboration between computational recommendation and accountable human judgement. This collaboration between AI and humans has been shown to lead to more comprehensive decision-making, with AI's ability to detect patterns at machine speed and humans' capacity to interpret them in context, especially when the operational environment is beyond the stability limits of fixed rules [1].

A warehouse copilot should be evaluated not only on predictive accuracy but also on the decision classes for which it may recommend, initiate, or execute actions. AI can provide recommendations; a supervisor can approve, modify, defer or reject an intervention. Decision authority and approval thresholds should vary according to reversibility, financial exposure, safety implications, and service-level risk. Shrestha et al. [2] distinguish three AI-supported organizational decision structures: full delegation, sequential decision-making, and aggregated human-machine judgment. In SAP EWM, these are distinct designs of the same structure. If the queue is in a low risk state, the reprioritization can be automatically executed within specified limits, but if a decision concerns blocked stock, a labour redeployment or a major wave cancellation might need explicit permission.

The subject has an important meaning in the fields of information systems, operations management, ergonomics and organizational research. The automation-augmentation paradox describes how automation can both enhance human capability and reduce expertise, situational awareness, or meaningful control [3]. In addition to accuracy, there are aspects of trust in research that affect its acceptance, such as perceived reliability, transparency, task type, embodiment, and institutional safeguards [4]. The situation is exacerbated in warehouse supervision as the activities are time coupled. A locally rational recommendation can lead to downstream congestion, starvation, travel, overtime or service failure. The ability to support system level reasoning, yet not take over operational responsibility, is thus crucial for an effective copilot.

The research base is still not well-connected. Warehouse studies offer rich models of picking, automation, routing, storage, and robot control, but do not commonly include assessments of conversational/generative AI interfaces for supervisors. Human-AI research focuses on delegation, trust, opacity, and collaboration, but typically treats tasks in



laboratories or external workplaces not related to logistics. No peer-reviewed research was found that specifically studies the use of an AI copilot within SAP EWM. For this reason, the intent of this review is to evaluate the transferability of these findings to the context of SAP EWM supervision, identify the context where transfer is still speculation, and provide a bounded framework for human in the loop decisioning. The analysis focuses on decision architecture, the quality of evidence, operational outcomes, human-centred risks, and research designs which can add causal value.

II. LITERATURE REVIEW

Human-factor research on order picking shows that warehouse performance is not adequately captured by the travel distance, throughput or utilization. Even though fatigue, learning, cognitive load, physical strain, error propensity and heterogeneity of workers have direct effects on productivity and quality, they are often underrepresented in planning models [5]. The human aspect was not consistently represented in all order picking studies, as revealed by a subsequent content analysis, where the physical factors of order picking are more evident than cognition and social factors [6]. This discovery restricts the recommendations that can be made solely based on execution telemetry for supervisor copilots. A recommendation that balances the labour may reduce the short-horizon throughput and increase inequitable task allocation, cognitive switching, or error due to fatigue. Literature therefore demonstrates the capabilities of multi-criteria decision making support but does not provide strong evidence about the way human-state information should be inferred, governed or presented within an enterprise warehouse system.

Research on logistics automation-human reaction is more directly related to the copilot concept. Klumpp [7] argues that adaptation, acceptance, competence, and collaboration by humans is a crucial issue in logistics automation and should be captured explicitly and not assumed to be frictionless technological replacement. This is in contrast to the literature on warehouse design that focuses on equipment configurations and control logic. Ecommerce warehousing surveys reveal increased demands for responsiveness, high variety, small orders, and dynamic demand, which are increasingly driving requirements for adaptive decision support [8]. In reviews of robotized warehouse systems, there is also a general trend of growing complexity of control over various automated storage, shuttle, robot, and picking configurations [9]. However, these reviews focus on system design and optimization of operation. The other four concerns supervisory cognition, explanation requirements, accountability, and override behaviour remain peripheral.

Table 1: Summary of key Findings

Ref	Focus	Key Findings
[5]	Human factors in order-picking models	Operational models underrepresent fatigue, cognition, learning, and worker heterogeneity; human-centred criteria should enter planning objectives.
[6]	Content analysis of human factors in picking	Physical factors receive more attention than cognitive and psychosocial factors; evidence is fragmented across methods.
[7]	Human reaction to logistics automation and AI	Collaboration, competence, acceptance, and task redesign mediate the value of automation in logistics settings.
[8]	E-commerce warehouse design and control	Volatility, assortment growth, and small-order profiles increase the need for responsive, integrated control decisions.
[9]	Robotized and automated warehouse systems	Automation creates interconnected design and control choices; performance depends on coordination across subsystems.
[10]	Logistics 4.0 research	Digital integration enables adaptive logistics but lacks consistent reference architectures and mature implementation evidence.
[11]	Autonomous mobile robot planning and control	Assignment, routing, congestion, charging, and human interaction require coordinated control rather than isolated optimization.
[12]	AI in supply-chain management	Research favours predictive and optimization applications; organizational deployment and causal performance evidence remain limited.

Logistics 4.0 research extends the analytical level and involves the integration of data, automation, and decentralised control, as well as cyber-physical systems [10]. Intralogistics performance is influenced by task allocation, routing, congestion management, charging, fleet coordination and interaction with the human-operated process, as shown in planning and control studies for autonomous mobile robots [11]. The findings are relevant to SAP EWM as an AI copilot can suggest resource or queue interventions in tightly coupled execution processes. However, the objective functions and state information used in technical optimization models often are more explicit than those in real warehouses. Sensor gaps, delayed confirmations, informal workarounds, master-data inaccuracies and unrecorded operational knowledge can negate the validity of

recommendations. Improvements in reported optimization, therefore, cannot be readily applied to a copilot visible to a supervisor without testing in the field.

In supply-chain management, other areas of AI research are largely interested in forecasting, optimization, risk management, and operational decision support, while there is limited support for the deployment maturity and organizational impact of AI. Algorithmic control also signals that autonomy, surveillance, evaluation and contestability can be modified by algorithmic systems [13]. Experiments with human-AI delegation suggest that the combined team's effectiveness is not just about having a human follow an algorithm, but rather about aligning the specific task components with comparable capabilities [14]. Research on opaque algorithms reveals that users can switch off, defer without evaluation and develop local practices to evaluate the credibility of recommendations [15]. Research on hybrid-intelligence, therefore, focuses on the outcome of complementary capabilities and learning loops, and research on machines as teammates focuses on coordination, communication, shared mental models, and role clarity. In all of these clusters, the central gap is clear: In every cluster there are some principles that are relevant, but there is no direct evidence of principles for SAP EWM supervisor copilots.

As presented in Table 1, the most related literature in the warehouse focuses on complexity and human factor exposure without assessing conversational decision support in SAP EWM. The best transferable finding is that evaluation of copilot performance needs to focus on system performance, decision quality, operator workload, and governance—not just on the accuracy of the recommendations.

III. METHODOLOGY

From the literature reviewed, four general methodological approaches are used. First, warehouse engineering studies include analytical models, simulation, classification, and structured reviews of the routing, picking, storage, automation, and robot coordination issue [5, 8, 9, 11]. These methods give clear goals and allow for the comparison of operational policies within a predetermined scope. One problem for the SAP EWM copilots is ecological distance simulation studies typically use cleaner, more complete state variables, more stable objectives, and more explicit constraints than those encountered in live warehouse operations. A recommendation that performs well in simulation may perform poorly in practice when confirmation delays, inconsistent exception codes, or undocumented local priorities are present.

Table 2. Method Comparison

Ref	Method	Strengths	Limitations
[5]	Conceptual framework grounded in order-picking models	Connects human factors with formal planning objectives	Limited validation of cognitive variables in live warehouse control
[6]	Structured content analysis	Identifies systematic omissions across the picking literature	Publication coding cannot establish operational effect sizes
[8]	State-of-the-art survey and problem classification	Links warehouse design pressures to control requirements	Limited evidence on supervisor interaction with adaptive systems
[9]	Review and taxonomy of automated warehouse systems	Clarifies subsystem interdependence and control complexity	Human authorization and explanation receive little treatment
[11]	Literature review of AMR planning and control	Identifies coupled assignment, routing, congestion, and charging decisions	Model assumptions often omit enterprise-data defects and informal work
[12]	Systematic literature review	Maps AI methods and supply-chain application areas	Broad scope weakens product-specific and role-specific conclusions
[14]	Controlled human-AI delegation experiments	Isolates cognitive effects of task allocation	Simplified tasks limit external validity for warehouse supervision
[15]	Qualitative field study of opaque AI use	Reveals practices for judging uncertain recommendations	Medical context differs from warehouse tempo and system architecture
[16]	Conceptual hybrid-intelligence framework	Models' complementary capability and reciprocal learning	Lacks SAP EWM implementation evidence
[17]	Multidisciplinary research-agenda development	Highlights communication, role clarity, and team coordination	Agenda propositions require operational field testing

Second, human-factor and organizational studies rely on literature coding, surveys, experiments, qualitative fieldwork, and conceptual theory to focus on the trust, delegation, opacity, control, and collaboration [6], [7], [13]–[17]. These methods reveal behavioural mechanisms that are not considered by operational optimization. Experimental work can help account for the delegation effect and qualitative work can help account for how professionals challenge the uncertain recommendation. Because of the time pressure, physical execution, constraints from enterprise system and cascading operational consequences, transfer remains limited. A laboratory classification job will not simulate a wave release decision relating to labour/staging, replenishment, transport cut-off, or congestion.

Third, AI and digital-logistics reviews trace the application areas, technologies and research gaps [10, 12]. These studies provide breadth but often combine forecasting, robotics, planning, risk and decision support. Key differences between autonomous optimization and human-in-the-loop recommendation can be hidden. A methodologically clear evaluation of an SAP EWM copilot should define the decision object, action authority, confidence representation, explanation format, time horizon, and rollback mechanism.

Fourth, hybrid-intelligence and team-based approaches view the human and AI elements as a "learning unit" able to adaptively handle decisions [16, 17]. This is best suited when the copilot is able to document responses of the supervisor and enhance routing/recommendation policies without training every override as truth. An override might not be a mistake in the algorithm, but a result of missing context, priority or a lack of knowledge in the short term. This distinction is formalized in the control architecture of Figure 1, where recommendation, explanation, authorization, execution and outcome monitoring are all separately auditable.

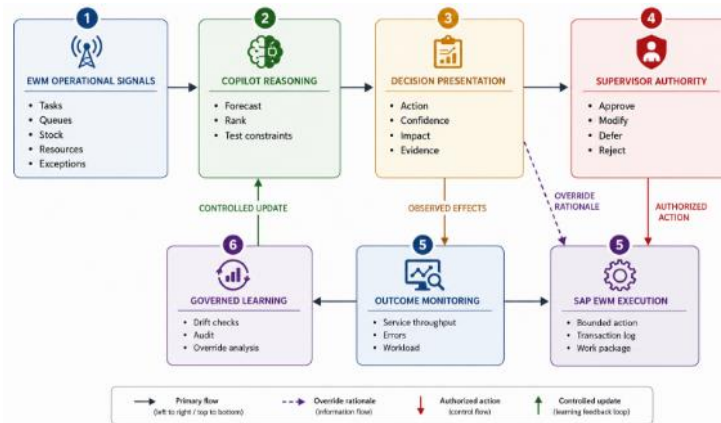


Figure 1: Human-In-The-Loop Decision Architecture for an SAP EWM Supervisor Copilot

The architecture separates data acquisition, recommendation generation, explanation, authorization, execution, and outcome monitoring. Before an approved action is executed in SAP EWM, the supervisor evaluates its contextual appropriateness. The conceptual model was developed based on the results of the decision-allocation, warehouse-control, hybrid-intelligence, and AI-teamwork models observed in [2] and [11] [14] [16] [17].

IV. DISCUSSION

The first major finding is that technical AI capability alone does not establish operational value. Warehouse-automation research demonstrates substantial optimisation potential, but it does not directly validate supervisor-facing AI copilots [8, 9, 11]. On the other hand, an AI Copilot is not a layer of an optimization engine. The copilot needs to identify when model output is relevant for making a decision, translate model outputs into warehouse-relevant concepts, highlight constraints, and assist with action in SAP EWM. This translation layer introduces failure modes which are not represented by routing/scheduling performance. A forecast can be used to justify an inappropriate action where the action is in conflict with a transport cut-off, labour agreement, quality hold or equipment restriction or undocumented priority. This makes it difficult to distinguish between predictive accuracy, recommendation quality, supervisory acceptance, execution correctness and downstream operational effect during the evaluation.

The second relates to complementarity. Human-AI research is used to aid in delegation of task parts based on comparative capability [14,16]. The following are computational benefits of SAP EWM: Continuous monitoring, cross-object aggregation, fast scenario comparison, anomaly detection, consistency checking. The strengths of supervision involve interpreting an unusual context, recognizing incomplete records, negotiating across functions, being accountable, and evaluating consequences that are not explicitly expressed in the objective function. Human review should not function as a universal approval gate. Excessive confirmation requirements may create alert fatigue and process delays, whereas excessive

automatic execution may increase automation bias and reduce meaningful control. Decision rights should be different depending on risk classification. Changes to the queue that can be reverted and have a bounded impact can be supported by automatic execution and retrospective review. Exceptional actions – such as a wave cancellation, blocked-stock disposition, heavy labour reallocation, or safety or contractual priorities – necessitate explicit approval and more evidence.

This allocation is complicated by the trust findings. Calibrated reliance doesn't always result from reliable recommendations [4]. A recommendation can be refused by a supervisor after a visible error or accepted by a supervisor as a weak recommendation if the interface is perceived as authoritative. According to qualitative evidence for opaque AI, professionals develop local strategies for determining the credibility of algorithms, such as comparing cases, cross-checking, and selectively consulting [15]. The warehouse copilot should therefore provide operationally relevant evidence such as affected warehouse orders, binding constraints, estimated downstream effects, confidence intervals or ranges, data freshness, and plausible alternatives. Sometimes generic natural-language explanations sound coherent without providing any information about causal/optimization logic. Explanation quality should be evaluated through fidelity, calibration, comprehensibility, decision usefulness, response time, and the ability to detect or prevent unsafe recommendations.

The third one is that the evidence is still not equally distributed. Figure 2 represents the literature that has been retained after 2015 by primary contribution and transfer relevance. Direct evidence for SAP EWM supervisor copilots is missing, while warehouse-operations and human-AI studies provide only indirect support. This distribution is important because any possible research programme needs to link operational models and their evaluation with behavioural and organizational evaluation, with neither stream alone being sufficient to prove implementation value. The bars represent the number of journal articles that were retained for each primary evidence domain, and the diamond markers represent an analytical transfer-relevance score ranging from one to five. It's important to note that the zero directly reported SAP EWM copilot trials is not a statement that there are no unpublished or non-journal implementations, but rather an evidence-gap indicator. Authors referenced [5]–[17] by contribution to primary research and context proximity to supervisor facing decision support in SAP EWM.

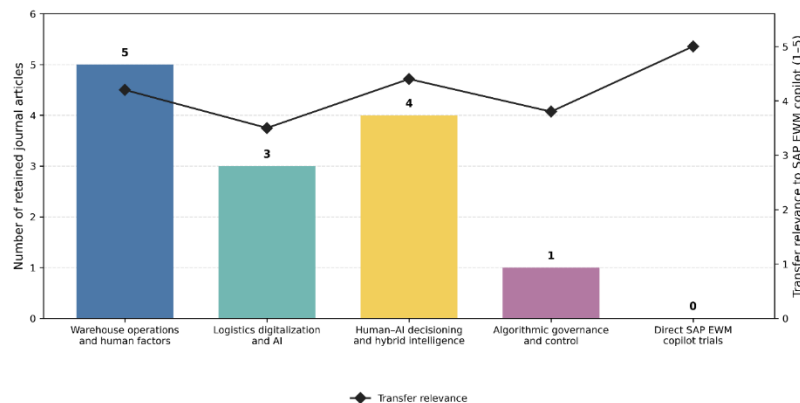


Figure 2: Evidence Volume and Transfer Relevance for SAP EWM Supervisor Copilots

The fourth finding concerns the tension between local and system-level optimization. In the study of warehouse control, the operations can be divided into assignment, routing, storage, batching, congestion and charging problems [8, 9, 11]. However, SAP EWM supervisors are faced with interactions between these decisions. Staging may be overloaded due to releasing additional work, which may increase short term resource utilization. Reassigning a resource may reduce work in one queue while depriving replenishment activities of capacity. Expediting a high-priority order may increase travel time or delay a larger group of nearly due orders. A copilot is thus required to highlight objective trade-offs and foresee possible spillovers. Single score recommendation interfaces mask value judgements on throughput, service, labour, energy, safety and fairness.

This worry is further supported by human-factors data. There are models that do not adequately represent worker characteristics and cognitive or physical requirements in order picking research [5] [6]. If an SAP EWM copilot has been trained on past confirmations, it can replicate allocation patterns that are based on uneven workload or on the dependency of experienced workers. Historical speed may also represent familiarity of the task, availability of equipment, environment of the zone, or assistance that may not have been documented. So, constraints on workload distribution, qualification, accessibility, fatigue exposure and contestability must be included in labour recommendations. The goal of performance monitoring should never be to create a black box ranking system for the workers.

A fifth effect is that organizational control is a component of system performance. Algorithmic-control research points to changes in monitoring, directing, assessing, disciplining when computational systems intervene in work [13]. A copilot can expand the capacity for decision-making or restrict it by default recommendations and compliance checks for the warehouse

supervisor. Interface design has an impact on this balance. The 'override' field with a strong justification requirement may deter valid dissent. On the other hand, not having an override process could cause learning to be lost or make model failure systematic. Governance must document the context of the recommendation, confidence, action, modification, reason category and outcome and must avoid over-surveillance and suppress operational candour.

Table 3 highlights the most salient reported outcomes and bases for interpreting them. The evidence confirms conditional value of AI-supported decisioning but cannot provide an established effect size for SAP EWM. Studies rarely report gains in warehouse-copilot throughput, but rather frameworks, taxonomies, experimental mechanisms, or qualitative patterns. These studies therefore do not support the assumption that a supervisor copilot will automatically increase throughput.

Table 3: Results Comparison

Ref	System / Context	Metric / Basis of Comparison	Outcome
[5]	Manual order picking	Inclusion of physical and cognitive variables in planning models	Human-factor omission can produce operationally incomplete policies despite efficient travel or assignment results.
[6]	Order-picking research corpus	Distribution of human-factor topics and methods	Cognitive and psychosocial evidence is less developed than physical-factor evidence.
[7]	Business logistics automation	Human reaction, competence, acceptance, and collaboration	Technical automation value depends on task redesign and human adaptation.
[8]	E-commerce warehousing	Design and control problem classes under volatile demand	Responsiveness requires coordinated decisions across storage, picking, batching, and fulfilment.
[9]	Robotized warehouse systems	System configurations and control dependencies	Automation raises coordination requirements and does not remove managerial control problems.
[11]	AMR-based intralogistics	Assignment, routing, congestion, charging, and interaction criteria	Isolated optimization can displace congestion or delay across the warehouse system.
[12]	AI in supply-chain management	Application distribution and evidence maturity	Predictive and optimization research exceeds causal evidence from deployed organizational use.
[13]	Algorithmic work control	Direction, evaluation, monitoring, and worker contestation	Algorithmic systems can increase consistency while reducing transparency and discretion.
[14]	Human-AI task delegation	Joint task performance under alternative delegation patterns	Productive collaboration depends on capability-aligned task allocation and user learning.
[15]	Professional judgment with opaque AI	Engagement practices and perceived credibility	Users develop local validation practices when model logic is difficult to inspect.
[16]	Hybrid-intelligence systems	Complementarity and reciprocal learning	Joint capability requires designed interaction rather than simple human approval after prediction.
[17]	AI as a team participant	Coordination, communication, roles, and shared understanding	Effective teaming requires explicit role boundaries and interaction protocols.

The sixth is that decision classes need to have different copilot policies. The automation suitability vs oversight requirement, cost of error, and data observability are plotted against each of five representative SAP EWM supervisory decisions in Figure 3. When downstream constraints are not known, there can be a lot of room for observation of the effects of wave prioritization and task reassignment, but still a need to be watched. Quality-hold or blocked-stock decisions have a relatively lower automation suitability due to legal, quality, and ownership considerations that may go beyond those of warehouse telemetry. The horizontal position indicates the level of automation suitability; the vertical position indicates the level of human supervision needed; the marker area indicates the relative cost of errors; the colour indicates the level of observability of operational data. The configuration shows why there is a need to have different policies for different classes of

supervisor decisions. Analytical coding using results from human-factor, warehouse-control, delegation and opacity references [5] and [11] and algorithmic-governance references [13]–[16].

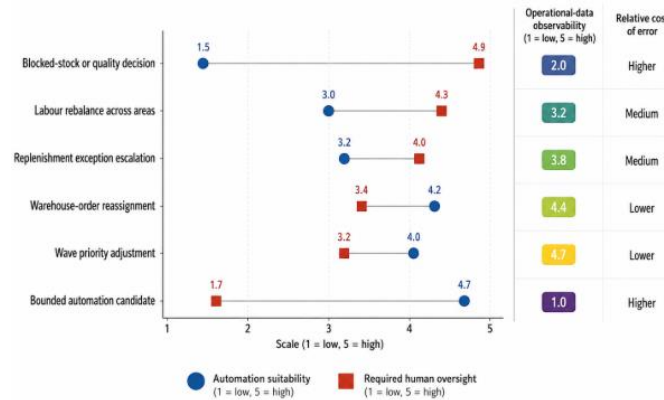


Figure 3: Comparative Decision Profile for Human-In-The-Loop Control in SAP EWM

The combined evidence supports five design implications. First, recommendations should be attached to explicit decision objects in SAP EWM rather than presented as general advice. Second, each recommendation should expose confidence, data age, constraints, alternatives, and expected operational impact. Third, authorization thresholds should depend on reversibility, cost of error, and cross-process consequences. Fourth, outcome monitoring should measure service, throughput, error, workload, override quality, and decision latency. Fifth, learning from supervisor action should distinguish disagreement, missing context, policy exception, and model error.

Important contradictions remain. Automation research often treats reduced human intervention as an efficiency gain, whereas human-AI research treats meaningful involvement as necessary for contextual judgment and accountability. Trust research promotes transparency, while time-critical warehouse operations impose strict limits on explanation length. Adaptive learning promises contextual improvement, yet frequent model change can weaken validation and reproducibility. Historical EWM data offer high operational relevance but may encode outdated layouts, temporary disruptions, workaround behaviour, or inequitable allocation. These contradictions indicate that the central research question is not whether a copilot should automate or augment supervision. The more precise question concerns which decision rights, evidence displays, learning rules, and safeguards produce superior joint performance for each warehouse decision class.

V. FUTURE DIRECTIONS

Research should start with prospective studies in a live SAP EWM environment, and not with the demonstration of previously exported transaction data. The appropriate design includes stepped-wedge deployments in different warehouse areas, randomized recommendation visibility at decision-event level and controlled crossover studies between rule-based alerts, AI-ranked recommendations, and AI recommendations with explanation. Primary outcomes should be Service Attainment, Decision Latency, Travel, Congestion, Replenishment Delay, Execution Error, Workload, Override Rate, Downstream instability. Recommendation acceptance should be distinguished from correctness in order to determine if there is value or automation bias.

Decision-specific uncertainty calibration is also needed for future copilots. Confidence must be judged at the action level and verified during distribution shift (due to seasonality, layout change, product introduction, equipment failure, or master-data deterioration). Explanation formats differing in terms of counterfactual actions, binding constraints, precedent cases, and projected process effects should be compared in research. Evaluation should determine whether an explanation helps a supervisor identify, reject, or correct an unsafe or suboptimal recommendation, rather than merely making the recommendation appear clearer [4, 15].

The interdisciplinary work should be linked to the fields of operations research, information systems, ergonomics, industrial psychology, labour governance and enterprise architecture. Human-related data collection should be proportionate, privacy-preserving, and limited to operationally justified purposes, particularly when fatigue or productivity indicators influence work allocation. Research should examine whether supervisors retain meaningful authority when compliance with or deviation from default recommendations is systematically tracked [13]. Technical studies should be made to develop secure access to SAP EWM objects, transaction safe execution, rollback, model drift and audit preservation, but not assume a language model is supposed to directly write operational transactions.

Clearly bounded decision classes should be a high-priority research focus including wave release, queue prioritization, warehouse-order reassignment, replenishment escalation and congestion response. A preregistered decision policy, basic

comparator, failure taxonomy, human-factor protocol and economic evaluation are required per class. Multi-site replication is necessary because the warehouse layout, automation, the labour model, product profile and EWM configuration play a significant part in the results.

VI. CONCLUSION

While a possible extension of SAP EWM decision support is the AI copilot for warehouse supervisors, the academic support is still indirect. The operational decisions are linked throughout the picking, replenishment, routing, storage, labour and automated resources in the warehouse. The human-AI research literature reveals that the value of computational recommendations relies on intentional design of task allocation, trust calibration, explanation, accountability, and override authority.

The most defensible architecture is one where there is a distinction between recommendation and authorization/execution. Bounded automation can be achieved by low-risk, reversible actions, whereas decisions related to safety, quality, labour, significant exposure of the service, or weakly observed context must be controlled explicitly. Evaluation should not be limited to model accuracy, but also consider downstream operational impacts, workload, decision latency, error detection, quality of human-AI reliance, fairness, and organizational control.

Substantial evidence gaps remain. There is no existing peer-reviewed journal article that directly compares an AI copilot in an SAP EWM application with a live human supervisor, and previous studies of the warehouse system do not reflect the contextual knowledge, data inaccuracies, or cross-process implications of real-time supervision. The effectiveness of the warehouse cannot be determined by evidence from delegation tasks in the laboratory or non-logistics professions. The field should progress in the following steps: decision specific field experiments, calibrated uncertainty, auditable explanations, transaction safe integration and multi-site replication. SAP EWM copilots should be considered decision-support systems, not autonomous supervisory agents, until there is proof of their capabilities.

VII. REFERENCES

- [1] Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577-586.
- [2] Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review*, 61(4), 66-83.
- [3] Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation-augmentation paradox. *Academy of Management Review*, 46(1), 192-210.
- [4] Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627-660.
- [5] Grosse, E. H., Glock, C. H., Jaber, M. Y., & Neumann, W. P. (2015). Incorporating human factors in order picking planning models: Framework and research opportunities. *International Journal of Production Research*, 53(3), 695-717.
- [6] Grosse, E. H., Glock, C. H., & Neumann, W. P. (2017). Human factors in order picking: A content analysis of the literature. *International Journal of Production Research*, 55(5), 1260-1276.
- [7] Klumpp, M. (2018). Automation and artificial intelligence in business logistics systems: Human reactions and collaboration requirements. *International Journal of Logistics Research and Applications*, 21(3), 224-242.
- [8] Boysen, N., de Koster, R., & Weidinger, F. (2019). Warehousing in the e-commerce era: A survey. *European Journal of Operational Research*, 277(2), 396-411.
- [9] Azadeh, K., de Koster, R., & Roy, D. (2019). Robotized and automated warehouse systems: Review and recent developments. *Transportation Science*, 53(4), 917-945.
- [10] Winkelhaus, S., & Grosse, E. H. (2020). Logistics 4.0: A systematic review towards a new logistics system. *International Journal of Production Research*, 58(1), 18-43.
- [11] Fragapane, G., de Koster, R., Sgarbossa, F., & Strandhagen, J. O. (2021). Planning and control of autonomous mobile robots for intralogistics: Literature review and research agenda. *European Journal of Operational Research*, 294(2), 405-426.
- [12] Toorajipour, R., Sohrabpour, V., Nazarpour, A., Oghazi, P., & Fischl, M. (2021). Artificial intelligence in supply chain management: A systematic literature review. *Journal of Business Research*, 122, 502-517.
- [13] Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, 14(1), 366-410.
- [14] Fügner, A., Grahl, J., Gupta, A., & Ketter, W. (2022). Cognitive challenges in human-artificial intelligence collaboration: Investigating the path toward productive delegation. *Information Systems Research*, 33(2), 678-696.
- [15] Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. *Organization Science*, 33(1), 126-148.
- [16] Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering*, 61(5), 637-643.
- [17] Seeber, I., Bittner, E., Briggs, R. O., de Vreede, G.-J., de Vreede, T., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, 57(2), 103174.