International Journal of Mathematical, Engineering and Management Sciences
Vol. 11, No. 4, 1512-1534, 2026
https://doi.org/10.33889/IJMEMS.2026.11.4.062

Improving Efficiency and Effectiveness in Industrial Support Business Processes through Low-Code Conversational AI: Evidence from a Workflow-Embedded Case Study

Paulo Peças
IDMEC, Instituto Superior Técnico,
Universidade de Lisboa, Portugal.
Corresponding author: ppecas@tecnico.ulisboa.pt

Diogo Pires
Instituto Superior Técnico,
Universidade de Lisboa, Portugal.
E-mail: diogotpires9@tecnico.ulisboa.pt

Diogo Jorge
CEO and Head of R&D Department,
Efficiencyrising, Lda. Lisboa, Portugal.
E-mail: diogojorge@erising.pt

(Received on April 4, 2026; Revised on May 15, 2026; Accepted on June 17, 2026)

Abstract
Industrial support business processes often involve work outside core production activities, including record retrieval, spreadsheet checking, supplier communication, and follow-up of operational events. We examine these issues in a maintenance-support case where a low-code conversational Artificial Intelligence (AI) layer was connected to existing information and communication routines. Two agents were configured: ManuBot, for querying and updating maintenance-history data, and MailBot, for recurrent supplier-email handling. The empirical sequence covered baseline diagnosis, prototype testing and implementation-stage evaluation, drawing on workflow observations, user feedback, task comparisons and records from the implemented tools. The clearest measured changes were task-specific. MailBot reduced supplier-email preparation from about 12-15 min to 2-3 min per message. ManuBot reduced maintenance-data retrieval and querying time by approximately 50%. Users also reported easier access to historical malfunction records, better visibility of recurrent events, and more structured email routines. The case remained constrained by incomplete ERP (Enterprise Resources Planning) integration, data-structure quality, platform permissions and differences in user readiness. The evidence points to a task-specific use of low-code conversational AI: gains were observed when the agents were tied to specific records, supplier-email workflows and human validation points.

Keywords- Generative AI, Workflow-embedded AI, Conversational agents, Low-code automation.

1. Introduction

Industrial support business processes often require a large amount of work outside core production activities. In manufacturing environments, engineers, supervisors, and support teams frequently need to retrieve fragmented records, update operational histories, interpret recurrent events, and prepare routine messages for suppliers or internal stakeholders before decisions can be made (John et al., 2021; Peças et al., 2021, 2022; Sampayo and Peças, 2022). These tasks are not part of the physical transformation process, but they affect the speed, consistency, and quality of operational decisions. This type of work has become a concrete target for Generative (Gen)-Artificial Intelligence (AI) applications because it combines repeated language work, fragmented information sources, and human review. Brynjolfsson et al. (2025), for example, showed that Gen-AI assistance can raise worker productivity. Calvino et al. (2025) argued that its effects on productivity and innovation depend less on the mere availability of the technology than on how it is embedded into organizational routines. Corvello (2025) similarly emphasized that organizational value is mediated by appropriation, customization, and the preservation of human agency in use.

This issue is particularly relevant in manufacturing, where AI adoption has expanded from predictive analytics and optimization toward more interactive forms of operational and decision support. Recent reviews indicate that AI is now applied across production planning, quality assurance, maintenance, assembly, and decision support, while Gen-AI and Large Language Models (LLMs) are extending this field by turning natural language into a practical access route to industrial systems (Gao et al., 2024; Ouerghemmi and Ertz, 2025). For support-process work, this shift matters because everyday tasks are often distributed across records, spreadsheets, email threads, and partially connected information systems. Maintenance-support business processes represent one important instance of this broader class, because they combine operational relevance with recurrent data retrieval, record updating, event interpretation, and communication work.

However, the current literature still provides limited field evidence on what changes when conversational AI is inserted into operating workflows. A large share of recent publications consists of reviews, conceptual mappings, technical overviews, or architecture proposals. These contributions frame the field, but are less informative about effects inside ordinary work routines. Ouerghemmi and Ertz (2025) explicitly characterize digital manufacturing research on LLMs as emergent and call for stronger research on implementation and practical integration. Related work in digital improvement and operations management similarly suggests that advanced digital technologies are often discussed at a strategic or enabling level, while task-level and workflow-level empirical evidence remains comparatively scarce. For the present study, the relevant question is therefore not whether Gen-AI is technically capable, but what changes when it is connected to specific tasks, data sources, interaction modes, and human review routines in an operating workflow (Gao et al., 2024; Kim et al., 2024).

The empirical studies available so far are relevant but still concentrated in specific technical settings. In assembly manufacturing, Colabianchi et al. (2024) found that an LLM-based digital intelligent assistant improved operator experience, reduced cognitive workload, and enhanced process performance in a controlled setting. In machine monitoring, Jeon et al. (2025) proposed ChatCNC, a conversational framework integrating LLM with retrieval-augmented generation and real-time manufacturing data, reporting strong reliability in answering complex shop-floor queries. Both studies show that natural-language interfaces can make industrial information easier to access and use. However, they mainly address bounded use cases such as assisted assembly and conversational machine monitoring. They do not yet document what happens in broader support workflows where users alternate between spreadsheet records, email threads, event interpretation, and follow-up communication under existing organizational constraints.

A second issue concerns how the value of Gen-AI support is evaluated. In operations contexts, discussions often emphasize acceleration, response speed, or automation potential. For industrial adoption, however, this is not sufficient. AI support also needs to improve information accessibility, reduce search and transcription effort, support diagnosis and planning, and remain dependable in the conditions under which people actually work. Recent manufacturing research points in this direction. Galdino et al. (2025), in the context of manufacturing quality management systems, argue that LLM-based cognitive assistants must be translated into explicit operational requirements if they are to support decision-making reliably. Tan et al. (2026) similarly show that LLM adoption in digital manufacturing depends on fit with the context, customization, and the economic logic of deployment, not on generic technological enthusiasm. These concerns are reinforced by known limitations of language models regarding hallucination and truthfulness (Ji et al., 2023; Lin et al., 2022). An industrial study of Gen-AI support should therefore measure speed and examine whether the outputs are useful, reliable, and sustainable under the technical and organizational conditions of use (Amaral and Peças, 2021a).

Industrial support workflows provide a suitable setting for examining these issues because they are operationally important but often built around fragmented information work. This paper focuses on a maintenance-support instance of this broader workflow class. The case includes recurrent activities such as retrieving historical records, identifying malfunction patterns, updating operational data, and handling recurring communication with suppliers and internal stakeholders (Amaral and Peças, 2021b; Martinho et al., 2022). These tasks are not directly value-adding in the physical sense, but they affect the speed, consistency, and quality of operational decisions. Because they depend heavily on access to fragmented digital information, they are well suited to a delimited conversational AI intervention, provided that the outputs remain subject to human validation.

The study therefore examines two recurring classes of support work: a data-centered stream and a communication-centered stream. The data-centered stream concerns the retrieval, update, and interpretation of maintenance-history records. The communication-centered stream concerns the handling of recurrent supplier-email exchanges. It asks how conversational AI can improve efficiency and effectiveness in these two streams and which technical and organizational conditions shaped the gains observed. The paper reports what was observed in the case: task-level time reductions, changes in access to maintenance-history information, more structured supplier communication, and limits associated with data quality, system integration, user readiness, and platform permissions. Beyond the specific maintenance case, the paper also shows how low-code conversational AI can be connected to existing information flows, communication routines, and spreadsheet-based support activities without heavy custom software development. This empirical focus responds to calls for more task-level and workflow-based evidence on LLM deployment in digital manufacturing and industrial operations (Ouerghemmi and Ertz, 2025; Tan et al., 2026). The study is guided by the following research question: How does a low-code Gen-AI solution embedded in industrial support workflows and existing digital work routines affect workflow efficiency and effectiveness under real industrial constraints?

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature, Section 3 presents the research design and intervention development, Section 4 reports the empirical results, Section 5 discusses the implications, and Section 6 concludes with contributions, limitations, and directions for future research.

2. Literature Background

2.1 AI-Enabled Support for Repetitive and Knowledge-Intensive Operational Tasks

Gen-AI has brought renewed attention to a practical question in operations management: which parts of work should be automated, which should be supported, and which should remain under human judgement. In organizational settings, Gen-AI is increasingly viewed as both an automation technology and a means of augmenting human work. Holmström and Carroll (2025) argue that firms can innovate with Gen-AI through different combinations of automation and augmentation, with AI framed as more than a substitute for human labor. In a related view, Corvello (2025) stresses that the value of Gen-AI emerges when organizations adapt these systems to their own routines, embed them in situated practices, and keep users actively involved in judgement and control. This distinction is particularly relevant in operational support tasks, where the work is often repetitive but still requires interpretation, exception handling, and contextual understanding. In such cases, Gen-AI is more appropriately framed as a task-limited support mechanism than as a fully autonomous replacement (Calvino et al., 2025; Noy and Zhang, 2023).

Available empirical evidence also suggests that Gen-AI tends to generate clearer benefits in tasks that are language-intensive, recurrent, and based on recognizable work patterns. Brynjolfsson et al. (2025), studying AI assistance in customer support, found significant productivity gains together with strong heterogeneity across users, with greater benefits for less experienced workers. Although that context differs from manufacturing, the underlying work mechanism is relevant for industrial support processes: employees repeatedly convert dispersed information into decisions, answers, records, or written responses. In manufacturing and industrial support environments, this type of work appears in activities such as retrieving maintenance histories, interpreting recurrent fault patterns, consolidating records, and preparing standard communications. These activities are not fully routine, but they are structured enough to benefit from task-bounded LLM assistance (Donauer et al., 2015; Gao et al., 2024; Ouerghemmi and Ertz, 2025; Varriale et al., 2025).

Manufacturing-oriented reviews point in the same direction. Gao et al. (2024) show that AI applications in manufacturing now span production planning, quality assurance, maintenance, assembly, and decision support. Jin et al. (2025) describe AI-driven automation as an evolution from simple rule-based routines toward more adaptive and intelligent architectures. Recent synthesis work further suggests that the industrial value of LLMs depends less on generic text generation and more on their use as interaction layers for data-intensive and communication-intensive work. This is especially relevant when operational routines require people to transform fragmented information into usable outputs (Ma et al., 2025; Ouerghemmi and Ertz, 2025; Zhang et al., 2025).

In industrial support processes, the more realistic opportunity is often narrower than full autonomy. In many support processes, the opportunity is more modest but also more practical: reducing search effort, accelerating access to relevant information, making routine responses more consistent, and allowing skilled workers to spend less time on repetitive information handling. This is particularly important in industrial support work, including maintenance-support business processes, where employees move between records, spreadsheets, and communication channels under time pressure. In these settings, much of the burden lies in accessing, interpreting, and communicating information, while the physical intervention represents only part of the overall work.

2.2 Conversational AI and Low-Code Automation in Industrial Operations

In manufacturing, LLMs have extended the role of AI beyond prediction and optimization by making industrial knowledge, operational records, and support routines accessible through natural language. Recent reviews show that LLMs are being explored for text generation, troubleshooting, operator assistance, knowledge retrieval, process support, and human-machine collaboration. Ma et al. (2025) position LLMs as an important technology for next-generation intelligent manufacturing under Industry 5.0, while Zhang et al. (2025) show that the field is moving toward applications that connect language models with industrial data, contextual reasoning, and workflow support. Ouerghemmi and Ertz (2025) reach a similar conclusion in digital manufacturing, but also stress that the literature is still dominated by reviews, conceptual proposals, and research agendas, with fewer studies documenting embedded operational use. Related LLM surveys further suggest that practical value depends on how language models are combined with tools, memory, orchestration, workflow access, and the tasks performed by users (Ali et al., 2025; Shao et al., 2024).

Conversational AI changes the interaction layer of industrial work. Instead of requiring users to navigate rigid menus, query languages, or fragmented software systems, conversational systems allow operators, engineers, or support staff to express work needs in natural language and receive synthesized outputs. In principle, this reduces interaction cost and lowers access barriers to data and knowledge. The operational value of such systems, however, depends on whether the conversational interface is connected to the actual workflow, data sources, and validation routines used in daily work. Fettke and Di Francescomarino (2025), in their survey of business process management and AI, argue that AI is becoming relevant across process modeling, analysis, redesign, implementation, and monitoring. Ajimati et al. (2025) show that low-code and no-code development has become an important mechanism for broader organizational participation in digital transformation and for faster deployment of process-oriented applications. Human-LLM interaction research also indicates that the user interface, interaction mode, and placement of validation checkpoints influence whether Gen-AI support becomes usable and trustworthy in routine work (Gao et al., 2024; Kim et al., 2024). For operational use, the conversational interface therefore needs to be embedded in workflow logic and implemented through adaptable development environments (Giordano and Fantoni, 2025).

Recent empirical studies illustrate this more embedded view. Recent empirical studies illustrate this more embedded view, namely by suggesting that LLM-based assistants are moving from generic conversational support toward task-specific shop-floor applications. Evidence from assembly and machine-monitoring contexts shows potential benefits for operator support, workload reduction, process execution, and technical query answering, particularly when LLMs are coupled with contextual manufacturing data or retrieval mechanisms (Colabianchi et al., 2024; Jeon et al., 2025). González-Potes et al. (2026), for example, showed that LLM-enabled agent architectures can support real-time industrial decision support when embedded within delimited supervisory logic and connected to real process data. These studies show that natural-language interfaces can provide operational value when they are linked to real tasks and data. At the same time, they mainly address specific application contexts, such as assisted assembly, conversational monitoring, and real-time decision support. They leave open the question of how conversational AI performs in broader industrial information and communication processes, where employees repeatedly shift between data retrieval, record updating, event interpretation, and recurring communication within existing organizational constraints.

In this paper, low-code deployment is treated as part of the adoption logic, not as a technical convenience. Industrial support processes often need solutions that can be configured quickly, tested with users, connected to existing tools, and adjusted without extensive custom software development. Ajimati et al. (2025) show that low-code and no-code approaches are increasingly associated with citizen development and digital transformation, while DIZEST, proposed by Shim et al. (2026), illustrates how low-code architectures can support workflow-driven AI and data analysis through modular, reusable components. This low-code perspective is especially relevant for industrial support-process contexts because the bottleneck is often not the absence of AI capability. It is the difficulty of connecting that capability to spreadsheets, communication channels, operational records, user permissions, and review routines under real organizational conditions.

2.3 Implementation Contingencies and Boundary Conditions

Conversational AI does not create operational value simply because the underlying model is capable. In industrial settings, its usefulness depends on where it is used, what information it can access, how outputs are checked, and whether people can incorporate the tool into their work routines. Recent work in digital manufacturing therefore argues that organizations should ask whether LLMs can perform a task, when their use is appropriate, and under which deployment conditions. Tan et al. (2026) frame this as a contingent decision problem in digital manufacturing supply chains, arguing that LLM use should be guided by fit with the operational context, not by general enthusiasm. This view is consistent with Holmström and Carroll (2025) and Corvello (2025), who emphasize that value depends on complementary organizational choices, including work redesign, role allocation, and governance of human-AI interaction. For industrial support workflows, this means that adoption depends on whether the selected task is bounded, whether the necessary data can be accessed, how users interact with the system, and where human control is placed.

A first requirement is access to reliable and usable operational data. Conversational AI can only support work dependably if it is connected to the systems where relevant information is stored and updated. Galdino et al. (2025), in their requirement analysis for LLM-based cognitive assistants in manufacturing quality management systems, show that deployment requirements are shaped by data structure, traceability, system interoperability, and the need for controlled outputs. Jeon et al. (2025) similarly show that conversational machine monitoring depends on retrieval mechanisms linked to real-time industrial databases. More broadly, work on grounded LLM deployment argues that retrieval-augmented generation and knowledge integration become important when organizational tasks require verifiable and context-specific information grounded in organizational records (Klesel and Wittmann, 2025; Yang et al., 2025). This is directly relevant to industrial support processes, where information is often distributed across ERP systems, spreadsheets, maintenance logs, email archives, and other semi-structured repositories. When these sources are incomplete, inconsistent, or difficult to access, the conversational layer inherits those weaknesses (Mantravadi et al., 2023; Peças et al., 2023).

A second requirement is clear governance of reliability and task boundaries. The fluency of LLM outputs does not remove known risks related to hallucination and truthfulness. Ji et al. (2023) show that hallucination remains a documented problem in natural language generation. Lin et al. (2022) demonstrate that language models can reproduce falsehoods with high confidence. For industrial operations, this makes conversational AI more appropriate for delimited, verifiable, and task-specific activities than for open-ended decision-making without review. Galdino et al. (2025) support this point by arguing that manufacturing-oriented cognitive assistants should be grounded in explicit operational requirements. Governance-oriented research reaches a similar conclusion, emphasizing safeguards, review points, and fit-for-purpose controls where outputs may influence operational decisions or controlled procedures (Kim et al., 2024; NIST, 2023; Yao et al., 2024). In practice, deployment quality depends on model performance, verification routines, human checkpoints, and the possibility of checking outputs against trusted data or standard work (Hamdani and Chihi, 2025; Ruiz et al., 2023).

A third requirement concerns users and organizational readiness. Brynjolfsson et al. (2025) found that AI assistance did not benefit all users equally, which suggests that training, digital literacy, and work experience influence the gains obtained. Ajimati et al. (2025) make a related point in the low-code literature, showing that broader adoption depends on organizational capabilities, governance, and support for non-specialist users. In industrial environments, the success of conversational AI therefore cannot be assessed only as a technical matter. It also depends on whether employees trust the system, understand its role, and can integrate it into routine work without creating new ambiguity or disruption. Research on human-LLM interaction further suggests that interface design, feedback loops, escalation logic, and visible review checkpoints influence whether users can work productively with Gen-AI systems under accountability constraints (Gao et al., 2024; Kim et al., 2024). The literature therefore supports a socio-technical view: operational value depends on the joint fit between technical feasibility, process embedding, data quality, governance, and user preparedness.

The literature identifies conversational AI as a plausible mechanism for operational support, but the empirical base remains thin for low-code interventions inside ordinary industrial workflows. Existing reviews map a broad opportunity space in digital and intelligent manufacturing, and some empirical studies report promising results in specific industrial applications. However, field-based evidence remains limited on how low-code conversational AI affects repetitive industrial support workflows under ordinary organizational constraints. This study addresses that gap by examining two recurring workflow classes, one data-centered and one communication-centered, in a maintenance-support business-process setting. The analysis focuses on efficiency, effectiveness, and the boundary conditions that shaped the gains observed.

Figure 1 links operational-support work characteristics with relevant Gen-AI/LLM mechanisms, potential workflow outcomes, and the conditions shaping value in practice.

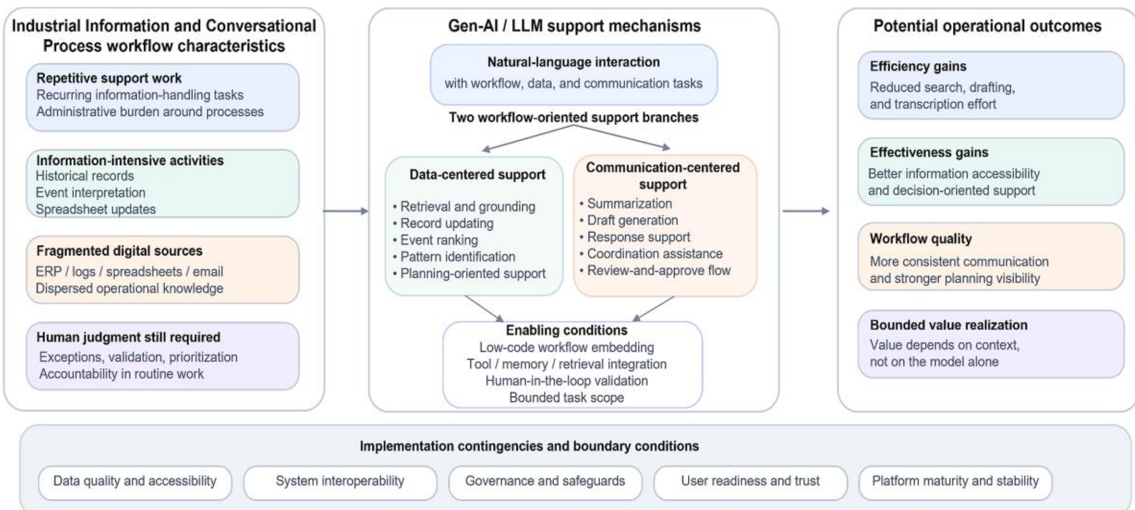

Figure 1. Literature-based framing of workflow-embedded Gen-AI support, linking workflow burdens, conversational mechanisms, operational effects, and boundary conditions.

The figure positions the empirical study by separating four elements examined in the case: task burden, conversational support mechanisms, expected workflow effects, and the conditions that may enable or limit those effects.

3. Research Method

3.1 Research Design

The paper uses an empirical single-case, intervention-oriented design to examine how a low-code Gen-AI solution affected industrial support workflows in a maintenance-support setting. The case was treated as a workflow intervention focused on task execution, data access, user validation, and observed changes in maintenance-support work. The design was chosen because the observed effects depended on local infrastructure, selected tasks, available data, and the way users worked with the agents in practice.

The research design therefore followed the intervention as it moved from problem diagnosis to tool development, testing, use, and evaluation. Baseline mapping identified recurrent information-handling burdens. Two specialized conversational agents were then developed for those bottlenecks and assessed through prototype testing, implementation-stage observation, benchmarking comparisons, and user feedback. This sequence allowed the intervention to be examined both as a technical prototype and as a change in the way maintenance-support work was performed under ordinary industrial conditions.

The single-case design was appropriate because the expected effects could not be separated from the organizational setting in which the agents were introduced. In this type of workflow-level intervention, performance depends on the AI component, data structure, platform access, task boundaries, validation routines, and user practices. The empirical work therefore combined baseline diagnosis, intervention development, controlled testing, implementation-stage observation, and performance evaluation within the same case setting. **Figure 2** summarizes the research design and empirical sequence adopted in the case.

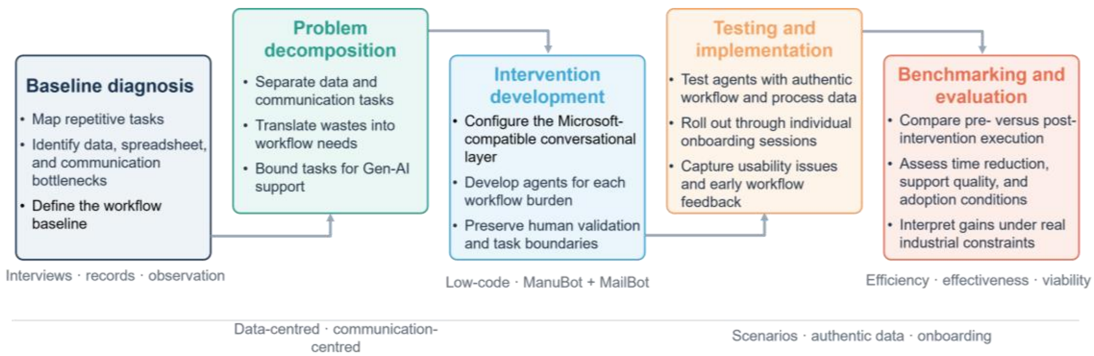

Figure 2. Research design and empirical sequence. Sequence from baseline diagnosis to intervention development, prototype testing, implementation, and evaluation.

3.2 Baseline Maintenance-Support Workflow and Problem Diagnosis

The starting point for the intervention was a maintenance-support workflow with substantial manual information work. In the case organization, the targeted work was not limited to direct technical intervention on equipment. A considerable part of the effort occurred before and after those interventions, when users had to retrieve historical records, organize operational information, update data repositories, and communicate with suppliers or internal stakeholders. These support activities influenced diagnosis, follow-up, and planning, but they were carried out through fragmented records and repeated manual handling. As a result, skilled time was absorbed by tasks that were recurrent, information-intensive, and only partly standardized.

The baseline workflow was suitable for a task-limited Gen-AI intervention because the main delays were located in information access, record handling, and communication work. The main inefficiencies were concentrated in accessing information, transforming it into usable outputs, and communicating it to others. Physical maintenance execution represented only part of the workload; much of the burden was located in the supporting information flow. This pattern was also visible in the diagnostic phase, where the existing workflow appeared as a sequential and largely manual process, with effort concentrated in extraction, transcription, validation, and repeated information handling (see **Figure 3**).

The diagnosis identified two recurrent sources of maintenance-support burden. The first was data-centered. Users needed to access, organize, update, and interpret operations-history information, but this work depended heavily on spreadsheet-based structures and manual searching. Difficulties appeared especially when users tried to retrieve historical malfunction data, track recurrent events, or understand downtime patterns. The second source was communication-centered. Maintenance personnel had to handle repetitive supplier emails, draft routine messages, and extract relevant information from message threads. These two sources of burden created avoidable delays, manual effort, and opportunities for inconsistency. They also slowed the preparation of information used in operational decisions.

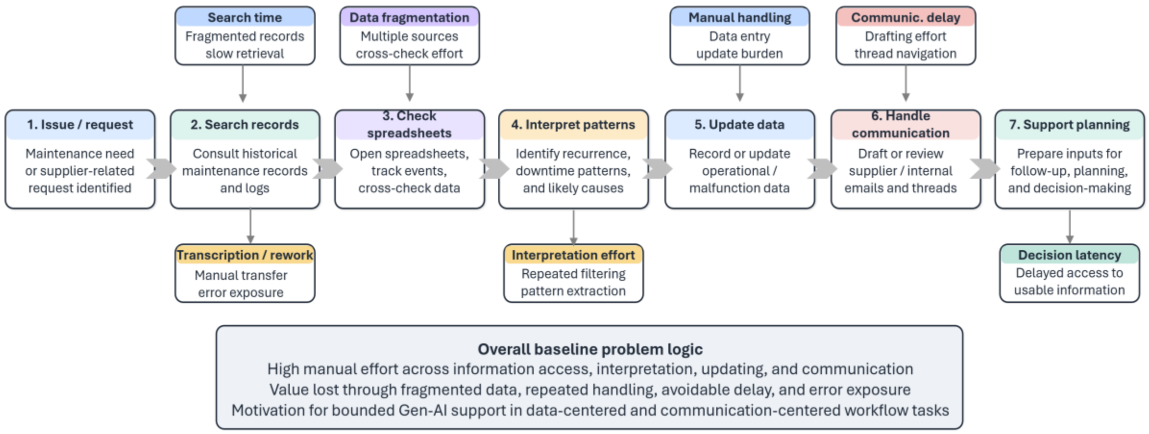

Figure 3. Baseline maintenance-support workflow and main information frictions. Pre-intervention workflow showing the main manual burden points that motivated the bounded conversational AI intervention.

The diagnosis also clarified what the intervention should not do. The aim was not to replace maintenance professionals or to turn conversational AI into an autonomous decision-maker. The intended role was narrower: to reduce effort in repetitive information and communication tasks while keeping users responsible for validation, interpretation, and final action. For that reason, each workflow intervention was preceded by the definition of task scope, operational boundaries, input and output schemas, and acceptance criteria, including completeness, correctness, and traceability. A representative test set was also assembled to validate performance offline and document recurring failure modes before broader use. In the case studied, the resulting architecture was therefore supervised: Gen-AI supported retrieval, structuring, drafting, and summarization, while maintenance personnel retained authority over verification and decisions.

The diagnosis shaped the functional structure of the solution. The solution was divided according to the two main workflow frictions found in the baseline. One part of the solution had to support interaction with operational data and basic analytical retrieval. The other had to support recurrent communication handling. This distinction led to the development of two specialized conversational agents, ManuBot and MailBot, described in the next section.

3.3 Development of the Low-Code Gen-AI Intervention: ManuBot and MailBot

The baseline diagnosis led to a low-code Gen-AI solution organized around two specialized conversational agents, ManuBot and MailBot. The solution was implemented within the organization's Microsoft-compatible digital ecosystem, so that it could work with tools already present in the maintenance-support routine. Instead of designing a single general-purpose chatbot, the architecture separated the two forms of support identified in the diagnosis: data-centered work and communication-centered work. This modular structure allowed each agent to be configured and tested with clearer task boundaries, while still sharing a conversational access layer and explicit human validation points. **Figure 4** shows this low-code architecture and its connection to existing work tools and review routines.

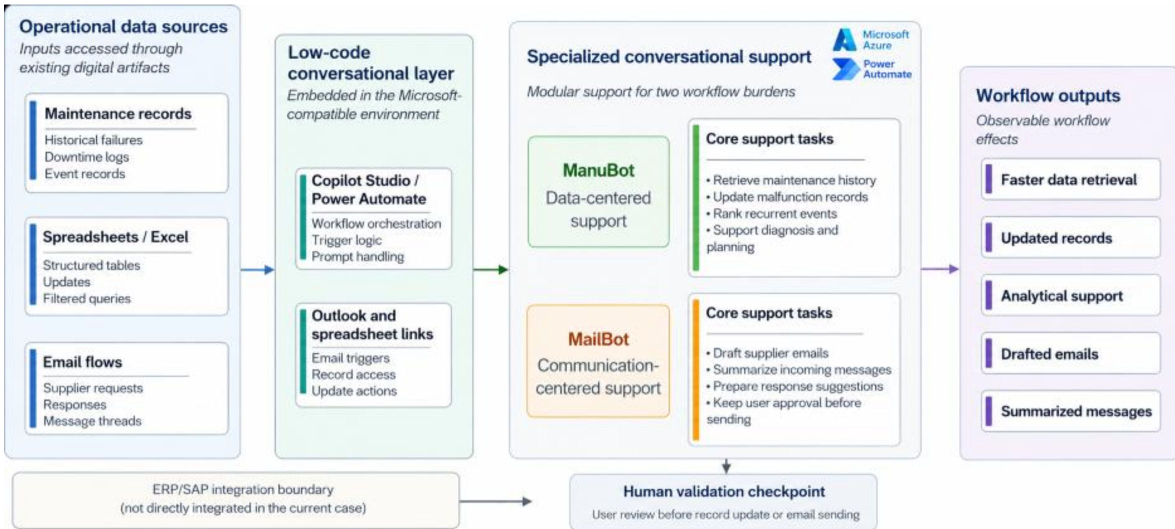

Figure 4. Low-code conversational layer supporting workflow-embedded Gen-AI intervention in industrial support processes.

ManuBot addressed the data-centered part of the maintenance-support workflow. Its role was to help users work with historical malfunction records and operations-history data without manually navigating the spreadsheet structure each time. The agent was designed around three functions identified during diagnosis: retrieving operations-history information, inserting new malfunction records into the existing spreadsheet-based structure, and producing structured outputs on recurrent events. In use, ManuBot could answer questions about fault history, recurrence, and time-related patterns. It could also support the insertion of new records and generate ranked or event-specific outputs to assist diagnosis and preventive-planning activities. Before testing, the source dataset had to be prepared so that dates and event sequences could be interpreted more consistently, since parts of the company file were incompletely structured. ManuBot therefore combined conversational querying with limited spreadsheet interaction. Users checked retrieved information and validated inserted records, so ManuBot operated as a supervised support tool within the maintenance-data workflow. ManuBot was tested with authentic maintenance data before routine use. **Figure A1** provides a complementary implementation view, showing the workspace configuration, the limited update-data action, and the resulting insertion of a new record.

MailBot addressed the communication-centered part of the workflow. The objective was to reduce the time spent on repetitive supplier-email handling by turning drafting and thread navigation into a review-and-approval task. MailBot was configured to detect specific incoming emails, summarize relevant content, generate draft responses, and support sending after user validation. It could also use contextual information already available in the workflow, including supplier-related references, to make the correspondence more structured and relevant. The prototype relied on predefined trigger logic, and the final decision to send remained with the user. This design kept MailBot within a supervised communication-support role. The implementation material documents the trigger-based activation, draft-generation, and user-validation flow, including the use of Outlook "Send an email (v2)" and Copilot Studio trigger logic. **Figure A2** summarizes how incoming supplier emails were converted into a review-and-send process within the low-code environment.

Both agents were built in the same Microsoft-compatible low-code environment because the practical aim was to connect conversational support to the tools already used by the maintenance team, especially spreadsheets, Outlook-based communication, and other Microsoft-linked components. This choice avoided heavy custom software development and allowed adjustments during prototyping. Development followed a bounded sequence: task definition, representative testing, and explicit review points. This kept the solution aligned with the available data structures, the organization’s permissions, and the need for human oversight. The resulting scope was limited to the tasks defined during diagnosis. ManuBot reduced the burden of accessing, updating, and interpreting maintenance-related data, while MailBot reduced repetitive communication handling. Neither agent was designed to replace user judgment or to operate without human review. **Figure 5** summarizes the functional scope of the two agents by distinguishing their roles, inputs, outputs, and control logic.

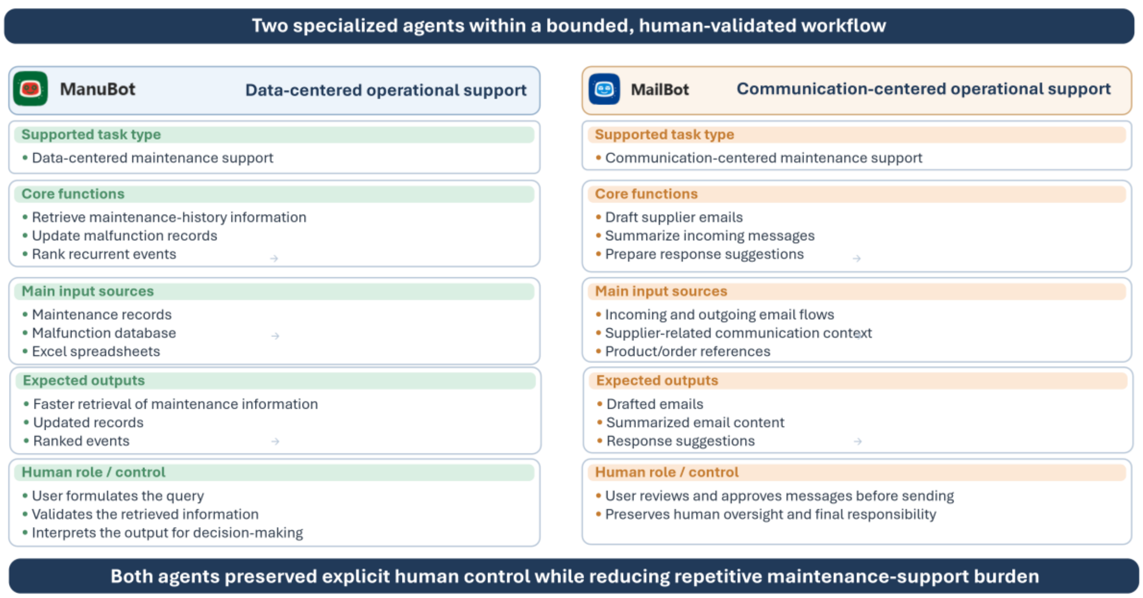

Figure 5. Functional scope of the conversational AI intervention. Visual summary of the task type, core functions, input sources, outputs, and human-control logic of ManuBot and MailBot.

3.4 Empirical Procedure and Evaluation Measures

The empirical work was organized around the same sequence followed by the intervention in the company: diagnosing the baseline workflow, developing and testing the two agents, introducing them into routine work, and evaluating the changes observed. The pre-intervention workflow was examined through observation, interviews, and operational documents, with attention to the tasks that repeatedly created manual burden and information friction. ManuBot and MailBot were then tested separately with authentic data and representative work scenarios, before being introduced in monitored implementation sessions with members of the maintenance team. This procedure captured two aspects of the agents: their technical behavior during testing and their effect on how selected support tasks were carried out in practice. The approach is consistent with established case-research logic in operations management and qualitative theory building (Eisenhardt and Graebner, 2007; Stuart et al., 2002; Voss et al., 2002).

Evidence was collected from baseline workflow observations, semi-structured interviews, prototype-testing records, implementation-stage feedback, and benchmarking comparisons. The empirical material included baseline workflow observations, semi-structured interviews, prototype-testing records, implementation-stage feedback, and benchmarking comparisons between selected pre-intervention and post-intervention tasks. These sources documented how the agents behaved during testing, how users worked with them, and how selected task times changed after implementation. The analysis therefore drew on process evidence, user experience, and task-level comparison. This triangulated evidence base was used to strengthen the link between the diagnosed workflow problems, the design of the intervention, and the outcomes observed after implementation, in line with case-study guidance on construct validity (Stuart et al., 2002; Voss et al., 2002).

The evaluation focused on three dimensions. Efficiency was assessed through task-completion time and the reduction of manual effort in activities such as email handling, record retrieval, and data handling. Effectiveness was assessed through the accessibility of maintenance information, the usability of agent outputs, support for diagnosis and planning, and the perceived usefulness of communication support. Implementation viability was assessed through user feedback, consistency of adoption, and the practical conditions affecting sustained use, including system integration, permissions, and digital readiness. These dimensions were selected to evaluate specific workflow-support tasks in the maintenance setting. The evaluation therefore asked how the agents performed in real maintenance-support work, using direct task evidence and implementation feedback, which is consistent with recent work arguing that the usefulness of LLM-based industrial support depends on workflow fit, operational requirements, and deployment context (Galdino et al., 2025; Tan et al., 2026).

3.5 Scope, Boundaries, and Validity Considerations

The empirical scope of the paper is deliberately bounded. The analysis does not address AI adoption across the whole company, nor does it evaluate conversational AI as a general-purpose industrial technology. It focuses on a low-code Gen-AI intervention used in industrial support workflows, within a maintenance-related support case. More specifically, the empirical work examines two recurring classes of support activity: a data-centered class, represented by ManuBot, and a communication-centered class, represented by MailBot. This focus keeps the analysis close to the level at which the intervention was actually observed: specific tasks, information flows, communication routines, and user validation points.

The technological scope is also limited. ManuBot and MailBot were designed to assist selected tasks within the workflow, not to make autonomous operational decisions. Human oversight remained part of the operating logic, particularly when users validated inserted data, interpreted outputs, and approved outgoing emails. The intervention was also implemented within the constraints of the organization's existing digital environment, including available data structures, platform permissions, and the absence of full ERP integration. These conditions are not secondary details; they define the boundary within which the results should be interpreted.

Several measures were used to strengthen the validity of the case evidence. The empirical procedure combined workflow diagnosis, prototype testing, implementation-stage observation, benchmarking, and user feedback, following established guidance for case research in operations management and organizational studies (Stuart et al., 2002; Voss et al., 2002). The use of multiple evidence sources helped connect the initial workflow problems, the design of the two agents, and the outcomes observed after implementation. This triangulation supports analytical interpretation of the case, while avoiding claims of statistical generalization (Eisenhardt, 1989; Eisenhardt and Graebner, 2007).

The single-case design supports analytical interpretation of the observed workflow intervention. The purpose is not to claim that the same gains will occur in every industrial support process. Rather, the case shows how a bounded conversational AI intervention was designed, introduced, and evaluated under a specific set of technical and organizational conditions. The results and discussion therefore interpret the observed gains together with the constraints identified in the case, including data structure, platform permissions, user validation, and incomplete ERP integration.

4. Results

This section reports the task-level results obtained after introducing ManuBot and MailBot in the maintenance-support setting. The analysis is organized around three aspects: changes in time and manual effort, changes in the quality and accessibility of support work, and the constraints that affected implementation. The aim is not to present the agents as generic AI solutions, but to document what changed in the specific workflow where they were tested. The results are therefore reported at the level of the tasks targeted by the intervention, in line with recent calls for more empirical evidence on LLM use in industrial operations (Ouerghemmi and Ertz, 2025; Tan et al., 2026).

4.1 Efficiency Gains in Industrial Support Workflows

Before the intervention, several maintenance-support tasks required users to move manually between spreadsheets, operations-history records, and email threads. Time was spent searching records, checking information across files, interpreting recurrent events, updating data, and drafting routine messages for suppliers or internal stakeholders. This configuration placed much of the effort in information access, transcription, and follow-up, reducing the time available for direct value-adding maintenance work.

The clearest time reduction appeared in the communication-centered stream. MailBot reduced the preparation and sending of recurrent supplier emails from approximately 12-15 minutes to 2-3 minutes per message. This corresponds to a reduction of up to 80% in the time required for that task. On days with a higher volume of supplier communication, the estimated saving reached up to one hour of work. In practical terms, the task changed from manual drafting and thread navigation into a shorter review-and-approval step. This freed time that could be redirected to maintenance activities with greater operational value.

The data-centered stream also showed a measurable reduction in execution time. ManuBot reduced the time needed to retrieve and query operations-history information by approximately 50%. The gain was relevant because the previous workflow required users to find the appropriate records manually, interpret partially structured data, and, in some cases, transfer or re-enter information across files. By allowing users to interact with malfunction records and historical event data through natural language, ManuBot reduced the amount of manual searching and cross-checking required for recurrent support tasks. The agent did not automate the complete information workflow, but it shortened a frequent and time-consuming part of it.

The efficiency gains were therefore concentrated in bounded and repetitive tasks, not in maintenance work as a whole. MailBot performed better where the communication pattern was recurrent and user validation could be inserted without slowing the workflow. ManuBot was useful where maintenance records could be accessed and interpreted through conversational queries. The common feature in both streams was not general AI capability, but the match between the agent, the data or communication source, and a clearly defined support task.

Figure 6 summarizes the case results by distinguishing efficiency gains, effectiveness gains, and the boundary conditions that shaped the observed outcomes. It also shows that the value of the intervention was not limited to time reduction, since support quality, data quality, system integration, and user adoption conditions also influenced the results.

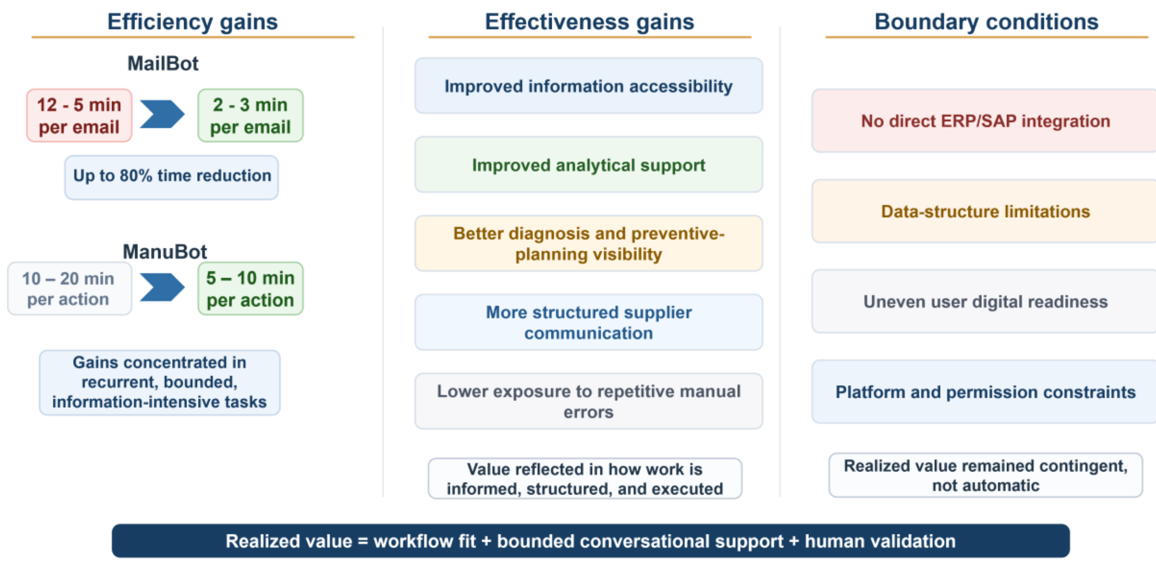

Figure 6. Efficiency and effectiveness outcomes, boundary conditions, and implementation implications across ManuBot and MailBot.

4.2 Effectiveness Gains in Information Accessibility, Analytical Support, and Communication Handling

The intervention also changed how maintenance-support information was accessed, interpreted, and used. In the data-centered stream, ManuBot made operations-history information easier to consult because users no longer had to rely only on manual navigation of spreadsheet records. The agent returned structured outputs such as recurrent event counts, average downtime associated with specific occurrences, and rankings of malfunction patterns. These outputs helped convert dispersed spreadsheet content into information that could be used more directly for diagnosis and planning. This type of result is consistent with recent arguments that LLM-based assistants should be evaluated through decision-relevant outputs linked to specific work tasks (Galdino et al., 2025; Ma et al., 2025). **Figure 7** illustrates two representative ManuBot outputs: one ranking recurrent malfunction events and another combining occurrence counts with average time consumed for a specific event.

The practical effect of ManuBot was especially visible in maintenance-related decision support. Feedback collected during implementation suggested that users could identify recurrent stoppages more easily and use historical data with less effort when preparing preventive-maintenance actions. The agent also improved visibility of periods that could require more staffing or closer attention, which supported earlier preparation of operational resources. ManuBot shortened access to data and made the retrieved information easier to interpret for diagnosis, planning, and anticipation.

MailBot produced effectiveness gains through a different route. Its main contribution was not analytical retrieval, but more structured handling of supplier communication. In addition to drafting outgoing emails, the agent summarized incoming messages and helped users identify relevant information in message threads. The workflow remained explicitly human-in-the-loop: users reviewed and approved messages before sending. Even with this control step, the recurrent effort involved in drafting, checking, and navigating supplier emails was reduced. The communication process therefore became faster, more consistent, and easier to manage without removing user responsibility for the final message.

The intervention also reduced exposure to some manual information-handling errors. In the baseline workflow, users repeatedly transcribed information, moved between spreadsheets, and transferred content across sources. These actions created opportunities for inconsistency, omission, or re-entry mistakes. After implementation, users reported simpler access to historical information and less dependence on repeated manual transfer, especially in ManuBot-supported tasks. The case does not show that errors were eliminated, and that claim should not be made. It does show, however, that conversational support reduced some of the repetitive search, transfer, and re-entry work that had previously increased the risk of manual inconsistency.

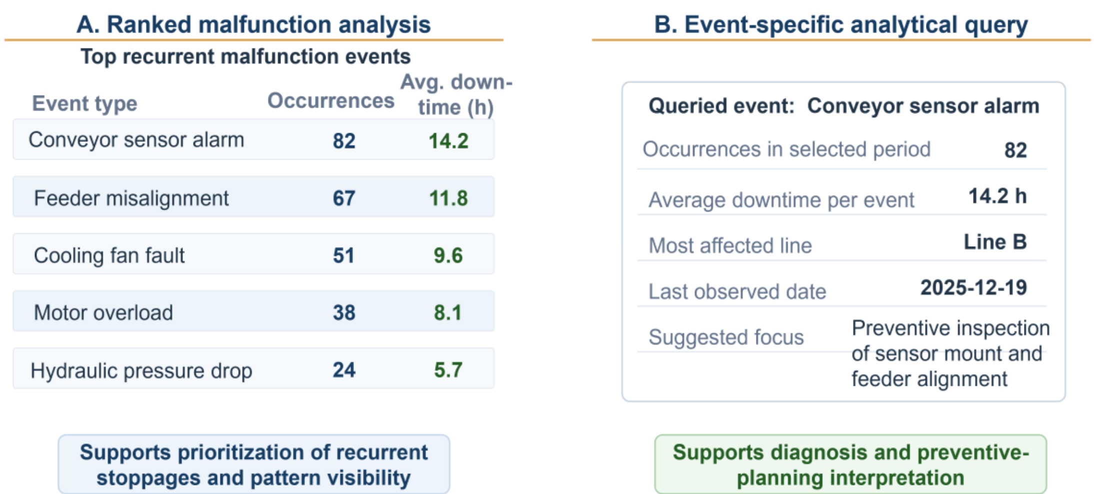

Figure 7. Representative analytical outputs generated by ManuBot. Examples of ranking-oriented and event-specific maintenance-data analysis supporting diagnosis and preventive planning.

The effectiveness evidence adds a second layer to the time results. ManuBot improved access to maintenance-history information and supported diagnosis and preventive planning. MailBot reduced the burden of recurrent supplier communication and made email handling more structured. Across the two agents, users spent less effort reaching information, obtained outputs that were easier to use, and handled recurrent communication with fewer manual steps. This interpretation aligns with recent work arguing that LLM-based industrial assistants should be assessed through operational usefulness, support quality, and fit with real work conditions (response speed or automation potential are important but not a first priority) (Galdino et al., 2025; Tan et al., 2026).

4.3 Boundary Conditions and Implementation Constraints

The measured reductions and workflow changes occurred under non-ideal implementation conditions. The agents were tested and used in a real maintenance-support environment, where the existing systems, data quality, permissions, and user capabilities shaped what could be achieved. For this reason, the observed improvements should be read together with the constraints that limited their scale and consistency. This case evidence matches literature that treats task fit, integration, and governance as determinants of industrial LLM deployment (Galdino et al., 2025; Tan et al., 2026).

The most visible limitation was the absence of direct SAP integration. Users indicated that the conversational agents would become more useful, especially in communication and information workflows, if they were connected more deeply to the company's ERP environment. Without that connection, some work still had to be bridged manually between systems, which limited end-to-end automation. In practical terms, the agents reduced specific search, drafting, and retrieval tasks, but they did not remove the need for users to move across enterprise systems when information was not available through the low-code workflow.

Data structure also affected performance. During early ManuBot testing, the maintenance dataset required adjustment because incompletely structured date fields affected the agent's ability to interpret the records correctly. This showed that conversational performance depended on the language model and on the consistency, format, and accessibility of the underlying maintenance data. The limitation is aligned with recent work on grounded LLM deployment and retrieval-based support in data-dependent organizational tasks (Klesel and Wittmann, 2025; Yang et al., 2025).

User readiness was another source of variation. The implementation evidence showed that the gains were not experienced equally by all maintenance professionals. Some users adopted the tools quickly and reported clear benefits, whereas others needed more time to become familiar with the new interaction mode and used the agents less consistently. The value obtained therefore depended partly on user confidence, digital familiarity, and ability to incorporate the agents into routine work. This result is consistent with prior findings on adoption heterogeneity and low-code implementation capability (Ajimati et al., 2025; Brynjolfsson et al., 2025).

Platform and permission constraints also affected the final configuration. The broader orchestration concept, which would have unified the two support streams under a single entry point, could not be implemented because the required permissions were not available. In addition, evolving platform features occasionally introduced instability or required adjustments to functions that had previously worked correctly. These issues reduced the possibility of a smoother and more integrated conversational layer.

The case produced measurable task-level changes, but only within specific technical and organizational boundaries. Gains were stronger when Gen-AI was applied to clearly delimited support activities, when the relevant data could be accessed, and when user validation could be kept inside a controlled workflow. Gains weakened when ERP disconnection, data-format problems, permission limitations, platform instability, or uneven user readiness interrupted the integration into routine work. The case therefore reinforces a condition-dependent interpretation of low-code Gen-AI adoption in industrial support processes (Ajimati et al., 2025; Galdino et al., 2025; Tan et al., 2026).

5. Discussion

5.1 Main Interpretation of the Findings

The case suggests that low-code Gen-AI produced observable task-level changes only when it was placed close to a well-defined support task. In this study, the strongest improvements did not appear in broad maintenance activities or in open-ended technical decision-making. They appeared in repetitive support work, where users spent time retrieving information, checking records, drafting supplier messages, and moving between existing digital tools. This is an important distinction. The observed effect came from connecting the conversational model to a specific workflow problem, not from using the model as a standalone technology.

In the case organization, ManuBot and MailBot became useful because they were inserted into existing work routines: spreadsheet-based maintenance records, Outlook-mediated communication, and a Microsoft-compatible low-code environment. The agents did not replace the maintenance team's judgement; they reduced the effort needed to reach, structure, and communicate information before users made or approved decisions. This reading matches recent work showing that Gen-AI produces more credible productivity gains when it is embedded in organizational routines (Brynjolfsson et al., 2025; Calvino et al., 2025; Corvello, 2025).

5.2 Workflow-Level Sources of Value

The first mechanism was the reduction of information and communication work surrounding maintenance decisions. ManuBot and MailBot did not automate physical maintenance interventions. They shortened or structured the information and communication work used to diagnose, follow up, and plan maintenance interventions. ManuBot reduced the effort involved in retrieval and structuring of maintenance-history information. MailBot reduced the effort involved in drafting and handling recurrent supplier communication. In both cases, human validation, interpretation, and final action remained part of the workflow. This distinction matters because it indicates that conversational AI is most useful in industrial contexts where information friction is high, but accountability must remain with the user. The case documents field evidence in an area still strongly shaped by reviews, conceptual discussions, and architecture-level proposals (Ma et al., 2025; Ouerghemmi and Ertz, 2025).

The second mechanism was the connection between measured time reduction and task usability. The time savings were substantial in the tasks directly addressed by the agents, especially supplier-email preparation and maintenance-data retrieval. However, speed alone does not fully explain the contribution of the intervention. In the data-centered workflow, ManuBot made operational information easier to access and interpret, supporting diagnosis, visibility of recurrent events, and preventive-planning decisions. In the communication-centered workflow, MailBot shortened message preparation and reduced the effort needed to follow repeated supplier exchanges. The agents reduced task execution time and produced outputs that users could apply in routine maintenance decisions.

5.3 Contingencies Shaping Realized Value

The case also shows that the gains were conditional rather than automatic. MailBot and ManuBot produced measured time reductions and reported workflow changes, but these effects depended on the technical and organizational setting in which the agents were used. The absence of direct SAP integration limited end-to-end automation, incomplete data structures affected some retrieval functions, and differences in digital readiness influenced how consistently users adopted the tools. These limitations were part of the result, not merely implementation details. They indicate that low-code Gen-AI deployment in industrial support work depended on ERP connectivity, usable maintenance records, user validation routines, and user capability, which is consistent with recent work on industrial Gen-AI deployment and grounded cognitive assistants (Ajimati et al., 2025; Galdino et al., 2025; Klesel and Wittmann, 2025). **Figure 8** separates the elements observed in the case: the targeted workflow burden, the bounded functions assigned to the agents, the implementation conditions, and the resulting efficiency and effectiveness effects. The figure shows where the observed effects were obtained: in clearly defined support tasks linked to usable data sources, existing work routines, and explicit validation points. Without these conditions, the same technology may remain useful as a prototype but weaker as a sustained operational tool.

Three points follow from the case evidence. First, the case documents an embedded conversational AI intervention in a maintenance-support workflow. Second, it separates two mechanisms of task-level change: data-centered retrieval and communication-centered drafting. Third, it shows how the observed effects were limited by SAP disconnection, data-structure problems, platform permissions, and differences in user readiness. For managers, the practical lesson is to start with a narrow workflow problem avoiding a broad model-driven transformation agenda. The practical lesson is to begin with recurrent, information-intensive support tasks where the data sources, workflow boundaries, user validation points, and expected gains can be made explicit (Holmström and Carroll, 2025; Ouerghemmi and Ertz, 2025).

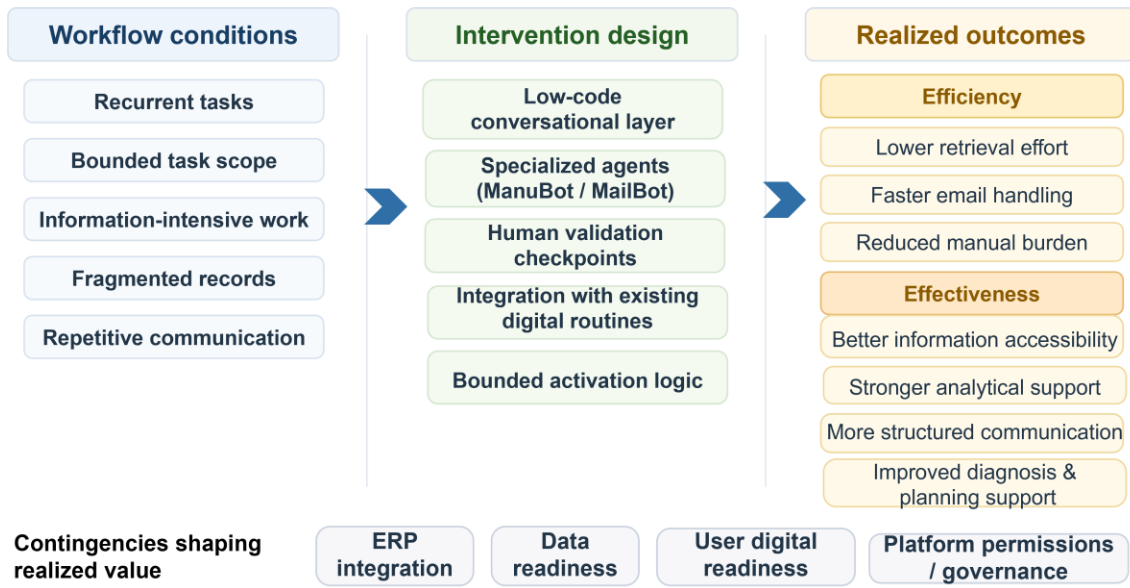

Figure 8. Mechanism of realized value in workflow-embedded conversational AI. Conceptual synthesis linking workflow characteristics, bounded conversational support, and socio-technical contingencies to realized efficiency and effectiveness gains.

5.4 Transferable Design Principles for Workflow-Embedded Conversational AI in Industrial Support Processes

Although the empirical setting was maintenance support, the case points to four design principles for other industrial support processes. The first is to start with bounded support tasks whose boundaries, data sources, expected outputs, and validation points can be defined before deployment. The tasks selected for implementation should be repetitive, information-intensive, and sufficiently delimited for their boundaries, data sources, expected outputs, and validation points to be defined before deployment. Recent industrial cases report similar patterns in delimited contexts, such as assembly assistance, task decomposition, and manufacturing question answering (Colabianchi et al., 2024; Giordano and Fantoni, 2025; Ruiz et al., 2023). A second principle is to design the conversational solution around the main source of workflow burden. The present case distinguished between data-centered and communication-centered work, and this distinction led to two specialized agents: one for data-centered retrieval and one for communication-centered drafting. This modular approach is relevant because different types of workflow friction require different support logics. A data-centered task may need retrieval, structuring, and analytical outputs, while a communication-centered task may need thread interpretation, drafting, and review support.

A third principle is to keep human validation visible at critical points. This is especially important when the output of the conversational system may affect diagnosis, communication with suppliers, record integrity, or follow-up decisions. In the case studied, the agents supported retrieval, drafting, structuring, and summarization, but users retained responsibility for checking and approving outputs. This principle is consistent with recent work on adaptive human-computer interaction in industrial environments, where usability and control remain central to effective adoption (Hamdani and Chihi, 2025).

A fourth principle concerns integration. Model capability is only one part of the result. In this case, the observed effects also depended on connections with enterprise information structures, digital routines, permissions, and data-quality conditions. This point is consistent with studies emphasizing connected information infrastructures and smoother digital integration in industrial settings (Amaral and Peças, 2021a; Mantravadi et al., 2023). Industrial adoption should therefore be evaluated through the selected task, the data sources available to the agent, the enterprise systems it can access, and the points where users check or approve outputs.

6. Conclusions

This paper analyzed a maintenance-support case in which a low-code Gen-AI layer was connected to two recurring support tasks: querying and updating maintenance-history data and handling supplier-email exchanges. The intervention had a defined operational scope, providing conversational access, drafting, and review support within existing work routines, while maintenance personnel retained responsibility for operational decisions and physical actions. In relation to the research question, the case shows measured time reductions and reported workflow changes when the agents were connected to specific information and communication tasks already present in the workflow. The gains were most visible in the data-centered and communication-centered streams addressed by ManuBot and MailBot, although their magnitude depended on the technical and organizational conditions of the case.

The case evidence points to two main effects. First, MailBot and ManuBot reduced manual effort and execution time in supplier-email preparation and maintenance-data retrieval. Second, users reported easier access to historical malfunction information, clearer visibility of recurrent events, and more structured email routines. These gains were not automatic, nor should they be read as evidence of broad autonomous AI capability. They depended on clearly delimited task design, human oversight, data quality, system integration, platform permissions, and user preparedness.

The paper documents a workflow-embedded conversational AI intervention in an industrial support process, including the tasks selected, the agents configured, the time reductions measured, and the constraints encountered during implementation. The case treats Gen-AI adoption as a targeted workflow intervention: the agents were assigned to bounded support activities, and users retained validation and approval points. This perspective helps explain why the intervention generated gains in the observed case: the agents were not used as open-ended chatbots, but as tools connected to records, communication routines, and review steps. For practitioners, the implication is to start with narrow support tasks where the records, communication routines, user checks, and expected time savings can be specified before deployment.

The case also responds to a literature still dominated by reviews, conceptual proposals, and architecture-level claims by showing how a low-code conversational AI intervention behaved during implementation. The evidence is especially relevant because the intervention operated with incomplete ERP integration, uneven user readiness, platform permissions, and data-structure limitations. These constraints are part of the contribution because they clarify the conditions under which the observed gains were obtained.

The study has limitations. It is based on a single industrial case, the intervention was still at an early implementation stage, and the results were shaped by the specific technical and organizational setting of the company. Future research should compare similar interventions in other industrial support domains, test alternative deployment architectures, and examine how governance routines, data structures, ERP connectivity, and human-AI interaction modes affect long-term use. This direction is aligned with recent literature on low-code adoption, LLM grounding, and deployment contingencies (Ajimati et al., 2025; Klesel and Wittmann, 2025; Tan et al., 2026).

Appendix

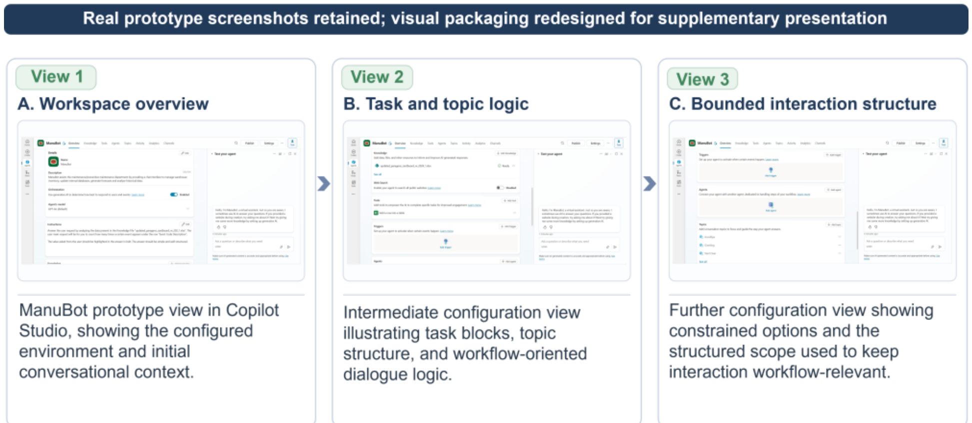

Figure A1. Complementary ManuBot interface and update-data example. Illustrative sequence showing the workspace, bounded update action, and resulting table entry.

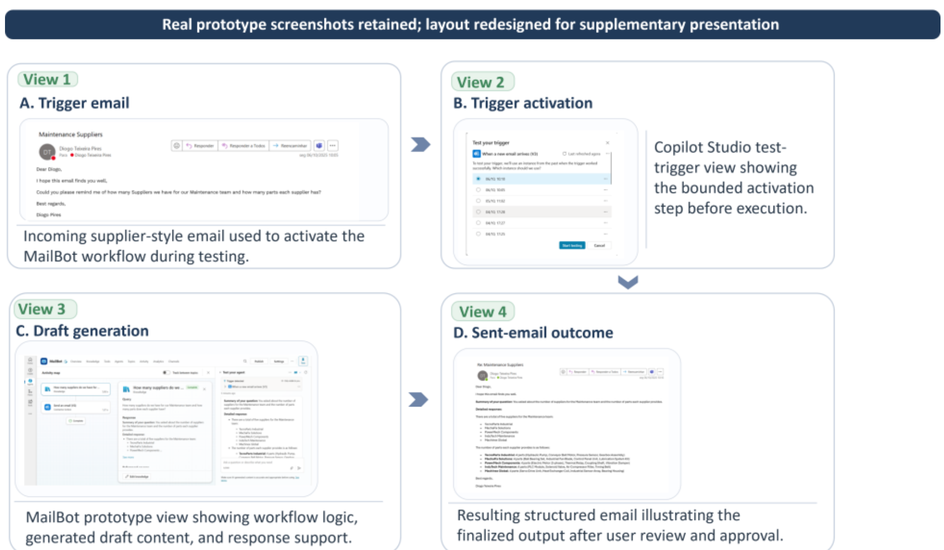

Figure A2. Complementary MailBot workflow for trigger-based email handling. Illustrative sequence showing trigger activation, draft generation, user validation, and sent-email outcome.

Conflicts of Interest

The authors do not have conflicts of interest.

Acknowledgments

The authors acknowledge Fundação para a Ciência e a Tecnologia (FCT) for its financial support via LAETA (project https://doi.org/10.54499/UID/50022/2025).

This work has been supported by the European Union under the Next Generation EU, through a grant of the Portuguese Republic's Recovery and Resilience Plan (PRR) Partnership Agreement, within the scope of the project PRODUTECH R3 - "Agenda Mobilizadora da Fileira das Tecnologias de Produção para a Reindustrialização", Total project investment: 166.988.013,71 Euros; Total Grant: 97.111.730,27 Euros.

AI Disclosure

During the preparation of this manuscript, the authors used ChatGPT to identify typographical errors, inconsistencies in terminology, and content repetition. The authors reviewed and edited the manuscript after using the tool and accept full responsibility for the final content.

References

Ali, M.A., Dornaika, F., & Charafeddine, J. (2025). Agentic AI: a comprehensive survey of architectures, applications, and future directions. *Artificial Intelligence Review*, *59*(1), 11. https://doi.org/10.1007/s10462-025-11422-4

Ajimati, M.O., Carroll, N., & Maher, M. (2025). Adoption of low-code and no-code development: a systematic literature review and future research agenda. *Journal of Systems and Software*, *222*, 112300. https://doi.org/10.1016/j.jss.2024.112300

Amaral, A., & Peças, P. (2021a). SMEs and industry 4.0: two case studies of digitalization for a smoother integration. *Computers in Industry*, *125*, 103333. https://doi.org/10.1016/j.compind.2020.103333

Amaral, A., & Peças, P. (2021b). A framework for assessing manufacturing SMEs industry 4.0 maturity. *Applied Sciences*, *11*(13), 6127. https://doi.org/10.3390/app11136127

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. *The Quarterly Journal of Economics*, *140*(2), 889-942. https://doi.org/10.1093/qje/qjae044

Calvino, F., Reijerink, J., & Samek, L. (2025). The effects of generative AI on productivity, innovation and entrepreneurship. *OECD Artificial Intelligence Papers,* No. 39, OECD Publishing, Paris. https://doi.org/10.1787/b21df222-en

Colabianchi, S., Costantino, F., & Sabetta, N. (2024). Assessment of a large language model based digital intelligent assistant in assembly manufacturing. *Computers in Industry*, *162*, 104129. https://doi.org/10.1016/j.compind.2024.104129

Corvello, V. (2025). Generative AI and the future of innovation management: a human centered perspective and an agenda for future research. *Journal of Open Innovation: Technology, Market, and Complexity*, *11*(1), 100456. https://doi.org/10.1016/j.joitmc.2024.100456

Donauer, M., Peças, P., & Azevedo, A. (2015). Identifying nonconformity root causes using applied knowledge discovery. *Robotics and Computer-Integrated Manufacturing*, *36*, 84-92. https://doi.org/10.1016/j.rcim.2014.12.012

Eisenhardt, K.M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532-550. https://doi.org/10.5465/amr.1989.4308385

Eisenhardt, K.M., & Graebner, M.E. (2007). Theory building from cases: opportunities and challenges. *Academy of Management Journal*, *50*(1), 25-32. https://doi.org/10.5465/amj.2007.24160888

Fettke, P., & Di Francescomarino, C. (2025). Business process management and artificial intelligence: literature survey and future research. *KI-Künstliche Intelligenz*, *39*(2), 67-79. https://doi.org/10.1007/s13218-025-00891-y

Galdino, M., Hamann, T., Abdelrazeq, A., & Isenhardt, I. (2025). Large language model-based cognitive assistants for quality management systems in manufacturing: a requirement analysis. *Engineering Reports*, *7*(10), e70437. https://doi.org/10.1002/eng2.70437

Gao, J., Gebreegziabher, S.A., Choo, K.T.W., Li, T.J.J., Perrault, S.T., & Malone, T.W. (2024). A taxonomy for human-LLM interaction modes: an initial exploration. In *2024 Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (pp. 1-11). ACM SIGCHI. Hawaii, USA. https://doi.org/10.1145/3613905.3650786

Giordano, V., & Fantoni, G. (2025). Decomposing maintenance actions into sub-tasks using natural language processing: a case study in an Italian automotive company. *Computers in Industry*, *164*, 104186. https://doi.org/10.1016/j.compind.2024.104186

González-Potes, A., Martínez-Castro, D., Paredes, C.M., Ochoa-Brust, A., Mena, L.J., Martínez-Peláez, R., Félix, V.G., & Félix-Cuadras, R.A. (2026). Hybrid AI and LLM-Enabled agent-based real-time decision support architecture for industrial batch processes: a clean-in-place case study. *AI*, *7*(2), 51. https://doi.org/10.3390/ai7020051

Hamdani, R., & Chihi, I. (2025). Adaptive human-computer interaction for industry 5.0: a novel concept, with comprehensive review and empirical validation. *Computers in Industry*, *168*, 104268. https://doi.org/10.1016/j.compind.2025.104268

Holmström, J., & Carroll, N. (2025). How organizations can innovate with generative AI. *Business Horizons*, *68*(5), 559-573. https://doi.org/10.1016/j.bushor.2024.02.010

Jeon, J., Sim, Y., Lee, H., Han, C., Yun, D., Kim, E., Nagendra, S.L., Jun, M.B.G., Kim, Y., Lee, S.W., & Lee, J. (2025). ChatCNC: conversational machine monitoring via large language model and real-time data retrieval augmented generation. *Journal of Manufacturing Systems*, *79*, 504-514. https://doi.org/10.1016/j.jmsy.2025.01.018

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y.J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, *55*(12), 1-38. https://doi.org/10.1145/3571730

Jin, W., Wang, N., Zhang, L., Tian, X., Shi, B., & Zhao, B. (2025). A review of AI-driven automation technologies: latest taxonomies, existing challenges, and future prospects. *Computers, Materials & Continua*, *84*(3), 3961-4018. https://doi.org/10.32604/cmc.2025.067857

John, L., Sampayo, M., & Peças, P. (2021). Lean & green on industry 4.0 context - contribution to understand L&G drivers and design principles. *International Journal of Mathematical, Engineering and Management Sciences*, *6*(5), 1214-1229. https://doi.org/10.33889/ijmems.2021.6.5.073

Kim, T.S., Ignacio, M.J., Yu, S., Jin, H., & Kim, Y.G. (2024). UI/UX for generative AI: taxonomy, trend, and challenge. *IEEE Access*, *12*, 179891-179911. https://doi.org/10.1109/access.2024.3502628

Klesel, M., & Wittmann, H.F. (2025). Retrieval-augmented generation (RAG). *Business & Information Systems Engineering*, *67*(4), 551-561. https://doi.org/10.1007/s12599-025-00945-3

Lin, S., Hilton, J., & Evans, O. (2022). TruthfulQA: measuring how models mimic human falsehoods. In *2022 Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics* (pp. 3214-3252). ACL. Dublin, Ireland. https://doi.org/10.18653/v1/2022.acl-long.229

Ma, Y., Zheng, S., Yang, Z., Zheng, P., Leng, J., & Hong, J. (2025). Leveraging large language models in next generation intelligent manufacturing: retrospect and prospect. *Journal of Manufacturing Systems*, *82*, 809-840. https://doi.org/10.1016/j.jmsy.2025.07.019

Mantravadi, S., Srai, J.S., & Møller, C. (2023). Application of MES/MOM for industry 4.0 supply chains: a cross-case analysis. *Computers in Industry*, *148*, 103907. https://doi.org/10.1016/j.compind.2023.103907

Martinho, R., Lopes, J., Jorge, D., de Oliveira, L.C., Henriques, C., & Peças, P. (2022). IoT based automatic diagnosis for continuous improvement. *Sustainability*, *14*(15), 9687. https://doi.org/10.3390/su14159687

National Institute of Standards and Technology (NIST). (2023). Artificial intelligence risk management framework (AI RMF 1.0). *U.S. Department of Commerce*. https://doi.org/10.6028/nist.ai.100-1

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187-192. https://doi.org/10.1126/science.adh2586

Ouerghemmi, C., & Ertz, M. (2025). Integrating large language models into digital manufacturing: a systematic review and research agenda. *Computers*, *14*(8), 318. https://doi.org/10.3390/computers14080318

Peças, P., Encarnação, J., Gambôa, M., Sampayo, M., & Jorge, D. (2021). PDCA 4.0: a new conceptual approach for continuous improvement in the industry 4.0 paradigm. *Applied Sciences*, *11*(16), 7671. https://doi.org/10.3390/app11167671

Peças, P., Faustino, M., Lopes, J., & Amaral, A. (2022). Lean methods digitization towards lean 4.0: a case study of e-VMB and e-SMED. *International Journal on Interactive Design and Manufacturing, 16*(4), 1397-1415. https://doi.org/10.1007/s12008-022-00975-1

Peças, P., John, L., Ribeiro, I., Baptista, A.J., Pinto, S.M., Dias, R., Henriques, J., Estrela, M., Pilastri, A., & Cunha, F. (2023). Holistic framework to data-driven sustainability assessment. *Sustainability*, *15*(4), 3562. https://doi.org/10.3390/su15043562

Ruiz, E., Torres, M.I., & del Pozo, A. (2023). Question answering models for human-machine interaction in the manufacturing industry. *Computers in Industry*, *151*, 103988. https://doi.org/10.1016/j.compind.2023.103988

Sampayo, M., & Peças, P. (2022). CPSD2: a new approach for cyber-physical systems design and development. *Journal of Industrial Information Integration*, *28*, 100348. https://doi.org/10.1016/j.jii.2022.100348

Shao, M., Basit, A., Karri, R., & Shafique, M. (2024). Survey of different large language model architectures: trends, benchmarks, and challenges. *IEEE Access*, *12*, 188664-188706. https://doi.org/10.1109/access.2024.3482107

Shim, C., Gim, J., Kim, Y., & Chae, Y. (2026). DIZEST: a low-code platform for workflow-driven artificial intelligence and data analysis. *SoftwareX*, *33*, 102519. https://doi.org/10.1016/j.softx.2026.102519

Stuart, I., McCutcheon, D., Handfield, R., McLachlin, R., & Samson, D. (2002). Effective case research in operations management: a process perspective. *Journal of Operations Management*, *20*(5), 419-433. https://doi.org/10.1016/s0272-6963(02)00022-0

Tan, B.Q., Kang, K., & Zhong, R.Y. (2026). When to use large language models for digital manufacturing in supply chains? *International Journal of Production Economics*, *295*, 109932. https://doi.org/10.1016/j.ijpe.2026.109932

Varriale, V., Cammarano, A., Michelino, F., & Caputo, M. (2025). Critical analysis of the impact of artificial intelligence integration with cutting-edge technologies for production systems. *Journal of Intelligent Manufacturing*, *36*(1), 61-93. https://doi.org/10.1007/s10845-023-02244-8

Voss, C., Tsikriktsis, N., & Frohlich, M. (2002). Case research in operations management. *International Journal of Operations & Production Management*, *22*(2), 195-219. https://doi.org/10.1108/01443570210414329

Yang, W., Some, L., Bain, M., & Kang, B. (2025). A comprehensive survey on integrating large language models with knowledge-based methods. *Knowledge-Based Systems*, *318*, 113503. https://doi.org/10.1016/j.knosys.2025.113503

Yao, Y., Duan, J., Xu, K., Cai, Y., Sun, Z., & Zhang, Y. (2024). A survey on large language model (LLM) security and privacy: the good, the bad, and the ugly. *High-Confidence Computing*, *4*(2), 100211. https://doi.org/10.1016/j.hcc.2024.100211

Zhang, C., Xu, Q., Yu, Y., Zhou, G., Zeng, K., Chang, F., & Ding, K. (2025). A survey on potentials, pathways and challenges of large language models in new-generation intelligent manufacturing. *Robotics and Computer-Integrated Manufacturing*, *92*, 102883. https://doi.org/10.1016/j.rcim.2024.102883