The Journal of Systems and Software 243 (2027) 113107

Contents lists available at ScienceDirect

The Journal of Systems & Software

journal homepage: www.elsevier.com/locate/jss

Development of industrial Artificial-Intelligence systems in production environments: A Delphi study on benefits, issues, risks, and mitigation strategies☆

Richard May [a,*], Leonard Cassel [b], Hashir Hussain [a], Muhammad Talha Siddiqui [a], Tobias Niemand [c], Paul Scholz [d], Thomas Leich [a]

[a] Harz University of Applied Sciences, Friedrichstraße 57-59, Wernigerode, 38855, Germany
[b] Fraunhofer Institute for Production Technology IPT, Steinbachstraße 17, Aachen, 52074, Germany
[c] Siemens Mobility GmbH, Ackerstraße 22, Braunschweig, 38126, Germany
[d] Hilti AG, Werkstraße 13, Thüringen, 6712, Austria

ARTICLE INFO

Keywords:
Artificial-Intelligence systems
Development
Configuration
Challenges
Manufacturing
Production
Delphi study

ABSTRACT

The adoption of industrial Artificial-Intelligence (AI) systems in production environments has increased considerably in recent years, enabling new forms of process automation, engineering support, and data-driven decision making. However, practitioner-oriented insights regarding their practical development, configuration, and operation remain limited. To address this gap, we conducted a four-round Delphi study with 26 practitioners involved in industrial AI-system development. Using iterative expert assessments, thematic analysis, and consensus-oriented evaluation, the study synthesizes benefits, issues, risks, and mitigation strategies. The findings indicate that industrial AI systems are increasingly evolving from isolated analytical models toward integrated and highly configurable production ecosystems. While experts associated AI systems with operational and economic benefits, they simultaneously emphasized challenges related to integration complexity, configurability, dependency management, traceability, lifecycle management, and governance. In particular, dependency hell, combinatorial explosion, and configuration mismatches emerged as recurring practical concerns. Overall, the study suggests that many practical challenges of industrial AI systems no longer primarily emerge from model development itself, but from their integration and operation within heterogeneous production environments. We argue that our findings provide practice-oriented insights into the development, configuration, and operation of industrial AI systems.

1. Introduction

Industrial Artificial-Intelligence (AI) systems are increasingly integrated into production environments to support operational and engineering activities based on learning processes, such as production control, quality assurance, predictive maintenance, or process optimization (Scotti, 2019; Chryssolouris et al., 2023). In this context, their adoption is frequently associated with expectations regarding improvements in central strategic objectives of production, particularly product quality, production time, and operational costs. For example, industrial AI systems are expected to support faster and more adaptive decision making, reduce downtime, or improve process robustness (Monostori et al., 2016; Wang et al., 2018; Rakholia et al., 2024). In fact, these incentives are not limited to the production domain but also play an increasingly important role in other fields, such as construction (Sholeh et al., 2026) or healthcare (Denecke et al., 2024).

However, the development, configuration, and operation of industrial AI systems introduce substantial technical and organizational complexity (Elahi et al., 2023). Unlike isolated software applications, industrial AI systems are commonly embedded into heterogeneous production environments involving interconnected and highly configurable software components with complex dependencies on hardware systems and individual devices (Schuh and Scholz, 2019; Peres et al., 2020; Serôdio et al., 2024). This tight cyber–physical integration increases the complexity of configuring, deploying, and operating industrial AI systems, since changes in software, hardware, or operational contexts may propagate across interconnected system components and

☆ Editor: Daniel Mendez.

* Corresponding author.
E-mail addresses: rmay@hs-harz.de (R. May), leonard.cassel@ipt.fraunhofer.de (L. Cassel), u38568@hs-harz.de (H. Hussain), u38641@hs-harz.de (M.T. Siddiqui), tobias.niemand@siemens.com (T. Niemand), paul.scholz@hilti.com (P. Scholz), tleich@hs-harz.de (T. Leich).

https://doi.org/10.1016/j.jss.2026.113107
Received 14 July 2026; Received in revised form 4 September 2026; Accepted 6 September 2026
Available online 19 September 2026

Open access funding enabled and organized by Projekt DEAL.

production processes. Consequently, developing, configuring, and operating industrial AI systems becomes substantially more challenging in practice (Peres et al., 2020; Weber et al., 2023). For example, failures or inappropriate mitigation strategies may directly affect production continuity, system reliability and safety, operational efficiency, or product quality in production contexts (Steimers and Schneider, 2022; Sinha and Lee, 2024).

Prior research extensively investigates industrial AI technologies and application scenarios, including model and application development (Pimenov et al., 2023), digital twins (Bhambri et al., 2024), or Internet of Things environments (Khang et al., 2024). However, this body of work is often shaped by theory-oriented settings, prototypical implementations, or application-specific evaluations. In contrast, practitioner-based studies remain comparatively limited and frequently focus on specific domains, working areas, or use cases. We argue that broader practical experiences regarding industrial AI systems therefore remain insufficiently consolidated across industrial contexts. This gap is particularly relevant because the successful transfer of industrial AI research into production practice depends not only on technological advances, but also on a deeper understanding of practical experiences (e.g., adoption challenges) (Schuh and Scholz, 2019; Schuh et al., 2020; Scholz, 2022). At the same time, such practical perspectives become particularly valuable when they are not only collected once (e.g., through surveys), but systematically reflected upon, reconsidered, and consolidated through iterative assessments by experienced practitioners (e.g., Delphi study) (Hsu and Sandford, 2007). Hence, we argue that a structured synthesis of expert perspectives on industrial AI systems can contribute a broader empirical understanding of their development, configuration, and adoption in practice.

To address this research gap, we formulated the following two Research Questions (RQs):

- **RQ$_1$:** *What are practitioners' experiences and assessment regarding the development, configuration, and adoption of industrial AI systems in production environments?*
- **RQ$_2$:** *What are typical benefits, issues, risks, and mitigation strategies related to the development, configuration, and adoption of industrial AI systems in production environments?*

To answer these RQs, we conducted a four-round Delphi study (Hsu and Sandford, 2007) with 26 experts involved in the development, configuration, and operation of industrial AI systems in production environments. Using iterative expert assessments and thematic analysis, the study investigates practical experiences, development and configuration-related challenges, perceived opportunities, and mitigation approaches associated with industrial AI systems. Based on the collected assessments, we derive implications for research and industrial practice.

Our contributions are summarized as follows:

- We conduct a Delphi study with experts involved in the development, configuration, and adoption of industrial AI systems in production environments to systematically capture and reassess practitioner experiences.
- We synthesize expert assessments based on a thematic analysis to provide relevant benefits, issues, risks and mitigation strategies from a software-engineering perspective.
- We identify and discuss theoretical and practical implications of the derived benefits, issues, risks and mitigation strategies.
- We provide an openly accessible repository containing the study data and supplementary research artifacts.[1]

[1] www.doi.org/10.5281/zenodo.22301848

Our study's results suggests that industrial AI systems introduce opportunities and challenges that emerge throughout their integration into production environments and operational processes. Rather than being limited to isolated AI functionality, practitioner experiences are strongly connected to engineering activities, configuration practices, industrial infrastructures, and organizational conditions surrounding these systems. We expect that the insights obtained from this study contribute to a more practice-oriented understanding of industrial AI systems and support future research as well as industrial decision making in production contexts.

The remainder of this article is structured as follows. Section 2 described the background and related work. Next, Section 3 presents the design and conduct of the Delphi study. The results are reported in Section 4 and synthesized into benefits, issues, risks, and mitigation strategies in Section 5. Subsequently, Section 6 discusses implications for research and industrial practice, followed by threats to validity in Section 7. Finally, Section 8 concludes the article.

2. Background & related work

The following paragraphs provide an overview of the technological structure, engineering and configuration characteristics, as well as adoption-related aspects of industrial AI systems.

2.1. Technological structure of AI systems

From a structural perspective, industrial AI systems in production can be described through multiple interconnected technological layers (Schuh and Scholz, 2019; Schuh et al., 2019, 2020) that we organize into three dimensions:

The application dimension describes the operational context of a system, including its working areas (e.g., predictive maintenance) and use cases (e.g., tool-wear prediction).

The physical dimension comprises machines (e.g., CNC machines) and processes (e.g., milling, sawing) that generate and execute operational data.

The cyber dimension contains the technological realization of the AI system itself, including data base (e.g., vibration, force strength, sound), learning strategies (e.g., supervised learning), tasks (e.g., classification), operations (e.g., neural networks), and their implementation infrastructures (e.g., AI development frameworks, programming languages) (Schuh et al., 2020; Scholz, 2022; May et al., 2023b).

This layered technological structure highlights that industrial AI systems are not limited to isolated machine-learning models, but emerge from interactions between operational environments and processes, hardware infrastructures, and data-driven software. Consequently, changes within individual technological layers may influence interconnected system components and operational production environments.

2.2. Engineering and configuration of AI systems

Industrial AI systems are commonly engineered along processing workflows that include data collection, pre-processing, learning, evaluation, and application (Goodfellow et al., 2016; Schuh et al., 2020). Data collection refers to acquiring operational and environmental data from production environments (Wan et al., 2016), while pre-processing prepares and transforms data for subsequent analytical tasks with the goal of achieving the highest possible quality (Zhang et al., 2011, 2019). Learning describes the training of AI models based on prepared data (Goodfellow et al., 2016), while evaluation assesses the behavior and performance of the resulting model (Domingos, 2012; Schuh et al., 2020). Finally, application refers to integrating evaluated AI models into systemic production contexts (i.e., as AI systems) to generate actionable outputs that provide certain added value (Kusiak, 2018; Zhang et al., 2019).

In production environments, these steps do not occur in isolation, but depend on the configuration and integration of a variety of components, for example, industrial devices or sensors (Peres et al., 2020; Serôdio et al., 2024). Thus, engineering industrial AI systems requires coordinating AI-specific processing steps with the surrounding production infrastructure while considering dependencies between technological layers, parameter settings, data characteristics, and operational processes (Scholz, 2022; Hollender et al., 2024). This configurability introduces highly variable and interconnected system structures in which changes to individual components may affect related processes, interfaces, or system behavior (Binder et al., 2022; Mo et al., 2023; May and Adler, 2025).

2.3. Adoption of AI systems

Besides challenges in engineering and configuring industrial AI systems, their adoption is strongly influenced by the expected operational and strategic implementation incentives within production environments (Monostori, 2014; Wang et al., 2018). At the operational level, AI systems are commonly associated with objectives such as improving system robustness, detecting anomalies, optimizing resource usage, or reducing downtime (Schuh and Scholz, 2019; May et al., 2023b; Ross et al., 2024). At the strategic level, organizations further aim to improve the central production objectives of cost efficiency, production time, and product quality (Monostori, 2014; Wang et al., 2018; Plathottam et al., 2023). However, these objectives are closely interconnected and may introduce practical trade-offs during development and operation. For example, achieving high production quality together with short production times may increase operational or implementation costs, whereas strong cost reductions may negatively affect flexibility or quality requirements (Bolstorff and Rosenbaum, 2007; Orm and Jeunet, 2018). Consequently, the perceived added value of industrial AI systems represents both a major adoption incentive and a practical operational challenge in production environments (Elahi et al., 2023; Khan et al., 2023).

2.4. Related work

Research on industrial AI systems covers a broad range of technological and organizational topics within production environments. Prior studies investigate, for example, industrial AI application (Hou et al., 2017) and use case scenarios (Soori et al., 2023), key enabling technologies (Xiang et al., 2023), changes in operational processes (Mithas et al., 2022), or cost-effectiveness (Soldatos, 2024). Additional work further discusses concepts related to engineering and configuring AI systems, including research on software engineering practices in AI system development (Gopinathan, 2025), architecture frameworks for human-AI teaming (Haindl et al., 2022), configurability in sustainable manufacturing (Todescato et al., 2023), or self-configuring industrial AI systems (Rehman et al., 2021).

We acknowledge that compared to the vast amount of theoretical research there is a quite limited number of literature involving practitioner or even expert voices. Using semi-structured interviews with industrial practitioners, Kutz et al. (2022) identified practical challenges related to AI system implementation, particularly regarding data quality, operational integration, and stakeholder involvement. Alon et al. (2025) reviewed Delphi studies on AI published during the last decade, including insights on the increasing integration of AI in manufacturing. Further, Van Dyck et al. (2023) focused on a future-oriented assessment of digital twins with consideration of AI potentials. Demlehner et al. (2021) applied a Delphi study in the automotive manufacturing domain to determine promising industrial AI application scenarios and implementation potentials. Similarly, Culot et al. (2020) used a Delphi-based approach with participants from research and industry to analyze expected opportunities and barriers associated with Industry 4.0 technologies, including AI systems.

We argue that existing studies often focus on specific fields within a domain, isolated organizational aspects, or selected adoption-related topics. The closest work to our study is our previous survey on AI system development and added value perception (May et al., 2025), which is based on a subset of the first-round responses of the present Delphi study. This earlier work primarily investigated development-related challenges and perceived added value of AI systems in production environments.

In contrast to existing work, the presented Delphi study explicitly incorporates configuration-related perspectives as an increasingly relevant topic in addition to development and added value aspects. To achieve this, we use a multi-round Delphi study to systematically analyze and reassess practical experiences of industrial AI experts to synthesize them in the context of benefits, issues, risks, and mitigation strategies in the context of industrial AI systems. Through this iterative consolidation process, the study contributes consolidated practitioner-oriented insights into established and emerging challenges associated with industrial AI systems in production environments.

3. Method

To systematically collect expert perspectives, we designed a *Delphi study* oriented toward (Hsu and Sandford, 2007). The study aimed to capture experiences in developing industrial AI systems, supporting the synthesis of typical benefits, issues, risks, and mitigation strategies. Delphi studies provide a structured mechanism for collecting and refining expert knowledge in domains where consolidated empirical evidence is still scarce. Instead of relying on single-time observations (e.g., survey), the Delphi method collects insights over multiple iterations, refining them through repeated survey rounds combined with controlled feedback. This structured process allows participants to reconsider their perspectives in light of aggregated responses, which supports the gradual emergence of shared assessments and (possible) consensus across the expert panel (Hsu and Sandford, 2007; Vernon, 2009). We emphasize that the Delphi method is particularly useful for technologically complex domains, in our case industrial AI systems in production, where practical experience from practitioners represents a valuable source of knowledge. Note that across rounds, aggregated feedback from the previous iteration was provided to participants to support reflection and potential reconsideration of their responses (Hsu and Sandford, 2007).

3.1. Delphi-study design

Overall, we followed a four-phase study process, including (1) preparatory phase, (2) Delphi rounds, (3) data processing and analysis, (4) concluding and reporting (cf. Fig. 1).

Preparation. In this phase, the first round questionnaire was designed. This questionnaire and all questionnaires of the subsequent round were developed using the online-survey tool Microsoft Forms, which enabled efficient distribution of the questionnaires, structured data collection, and centralized response management. In total, the first questionnaire consisted of 18 questions. Excluding demographics, the first questionnaire comprised 49 initial items oriented toward $\mathbf{RQ_1}$. In our study, *items* denote the individual elements presented to the experts for evaluation, for example, including statements assessed via rating scales as well as options considered in ranking questions. Note that these assessments provide the empirical basis for the thematic synthesis addressing $\mathbf{RQ_2}$. Overall, the questions were distinguished into four thematic sections:

Section 1 – demographics: The first section gathered basic participant characteristics through three questions. These covered $\mathbf{Q_1}$ years of

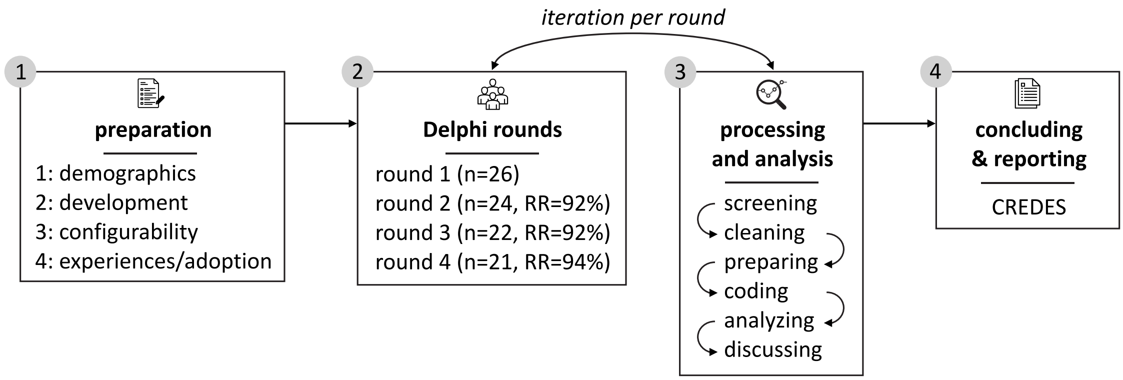

Fig. 1. Overview of the Delphi-study design and process, including preparation, four Delphi rounds (n = participants; RR = Return Rate), processing and analysis, as well as concluding and reporting.

experience in developing AI systems (single choice), $\mathbf{Q_2}$ industry background (multiple choice, including *prefer not to say* option), and $\mathbf{Q_3}$ employment role (multiple choice). Participants were also asked to create an individual pseudonym that allowed responses to be linked across the Delphi rounds while maintaining privacy. The individual pseudonyms were created using the last letter of the birth month, the second letter of the mother's first name, and the first two letters of the birthplace. In all subsequent rounds, demographics were not requested again; instead, the previously generated pseudonyms were used exclusively to associate responses with the same participant.

Section 2 – development: The second section (ten items) focused on development-related aspects of AI systems and consisted of three questions. These addressed $\mathbf{Q_4}$ the development phases perceived as most challenging (ranking), $\mathbf{Q_5}$ key development-related issues encountered in practice (free text), and $\mathbf{Q_6}$ typical strategies used to address or mitigate these issues (free text).

Section 3 – configurability: The third section (26 items) referred to practical aspects related to configuring AI systems, containing eight questions. Participants were asked to evaluate $\mathbf{Q_7}$ the awareness of configurability of AI systems (Likert scale). Subsequent questions explored $\mathbf{Q_8}$ frequently observed configuration issues (multiple choice), $\mathbf{Q_9}$ whether these issues mainly stem from the AI systems themselves or from interacting systems (single choice), $\mathbf{Q_{10}}$ how often configuration problems occur in practice (Likert scale), and $\mathbf{Q_{11}}$ the impact of configuration decisions (multiple choice). Additional questions focused on $\mathbf{Q_{12}}$ typical domains or working areas in which configuration activities take place (free text), $\mathbf{Q_{13}}$ how configuration practices differ between traditional software systems and AI systems (free text), and $\mathbf{Q_{14}}$ strategies commonly applied to mitigate configuration issues (multiple choice). Note that all multiple-choice questions offered an additional *other* option, allowing participants to complement predefined answers with further explanations in free text.

Section 4 – experiences: In the fourth section (13 items), we collected practitioners' experiences with the AI system with which they had the greatest personal involvement in their organization, assuming that this system reflects their most substantial practical experience. Here, the broader goal was to consolidate these insights after analysis and generalize them into statements for evaluation in the subsequent Delphi rounds. Participants first reported $\mathbf{Q_{15}}$ the system's primary use case (single choice) and $\mathbf{Q_{16}}$ the learning strategy applied (single choice), both including an *other* option to allow additional entries. Then, they were asked to describe $\mathbf{Q_{17}}$ the main incentives for introducing this system in their organization (free text). Finally, participants evaluated $\mathbf{Q_{18}}$ whether the implementation was considered economically worthwhile (Likert scale).

Delphi rounds. The iterative Delphi process constituted the second step of the study. All questionnaires were made accessible via Microsoft Forms, with participants receiving an invitation link by email for each round. After the first round, all open-ended responses were systematically analyzed and organized (cf. data processing and analysis). This method allowed us to synthesize the collected free-text statements into a structured set of items that served as the basis for subsequent evaluations. Then, participants were asked to rank or assess these categories in the following rounds (cf. Table 1). Questions that had originally been formulated as rating items, such as those addressing the frequency of configuration issues, remained in their original Likert-scale format. Across rounds, participants were additionally encouraged to provide comments and reflections for each thematic block, namely *development*, *configurability*, and *adoption* (created out of *experiences*). The overall questionnaire structure remained stable throughout the second, third, and fourth rounds, i.e., it consisted of the same set of 13 core questions to support comparability across these rounds. Naturally, wording and presentation of individual items and item orders were adapted based on results obtained from the previous round.

Processing and analysis. As the third step of the study, the collected responses were reviewed and analyzed by RM, HH, and MTS after each Delphi round. As an initial step, all submissions were screened to ensure completeness and sufficient response quality.

Quantitative analysis: Quantitative responses were analyzed using commonly applied procedures for Delphi studies. For questions requiring the ranking of items, we applied Kendall's coefficient of concordance (W) to assess the degree of agreement among participants regarding the relative ordering of items (Diamond et al., 2014). Kendall's W reflects how consistently participants assign similar ranks to the same set of items and ranges from 0 (no agreement) to 1 (perfect agreement). Values above 0.7 are typically interpreted as strong agreement among participants (Schmidt, 1997). To ensure statistical robustness, we considered results meaningful only when the associated significance level satisfied $p < 0.05$. For rating questions based on five-point Likert scales, responses were coded numerically from 1 (*strongly disagree*) to 5 (*strongly agree*). Then, the interquartile range (IQR) was used to measure the distribution of responses (Von Der Gracht, 2012).

Qualitative analysis: Free-text responses were reviewed after each round by RM, HH, and MTS and analyzed using thematic analysis (Clarke and Braun, 2017), supported by open-card sorting to structure emerging themes (Zimmermann, 2016). After familiarizing with the responses, HH and MTS jointly coded the data, while RM conducted an independent coding of the same responses. The coding results were subsequently compared and discussed among these authors until agreement was reached. Based on this process, items were either retained

Table 1
Overview of the Delphi study questionnaire for round 2 (n = 24), round 3 (n = 22), and round 4 (n = 21), focusing on development and configurability questions.

Section	ID	Question	Answer options
Development	Q_4	Please rank the AIS development phases according to how challenging they are.	Data preparation; design; implementation; deployment, operation; testing; change management
	Q_5	Please rank the most important challenges that arise in the development or use of AIS according to their relevance.	Data quality; system configurability; system adaptability/change management; meeting stakeholder expectations; model efficiency; system scalability; data handling/processing
	Q_6	Please rank the most important challenges that arise in the development or use of AIS according to their relevance.	Regular dataset testing; system templates; communication; improving data quality; iterative system optimization; holistic requirements engineering
	–	Further comments	Free text
Configurability	Q_7	Is there an awareness of AIS configurability and related issues in industry?	5-level Likert scale on agreement
	Q_8	Please rank the most common configuration issues encountered when deploying machine learning models in industrial settings according to their relevance.	Incorrect parameter settings; hard-ware-software mismatch; unwanted feature interactions; dependency issues; programming errors; version issues; high latency; scalability limitations; security limitations;
	Q_9	Do configuration issues arise more within AIS themselves or in combination of AIS with other systems?	In combination with other systems
	Q_{10}	How often do AIS-related configuration issues occur?	5-level Likert scale on frequency
	Q_{11}	Please rank the typical impacts of configuration issues regarding the performance and reliability of AIS according to their relevance.	Reduced accuracy; inconsistent outputs; increased downtime; computational inefficiency; system vulnerabilities; scalability issues
	Q_{12}	Please rank the working areas where configuration issues are more prevalent in AIS according to their relevance.	Feature-rich working areas; real time data working areas; security-critical working areas; monitoring systems; all working areas; less common working areas
	Q_{13}	Please rank the differences of configuration issues between AIS and traditional software systems in industrial environments.	Higher complexity; lack of legal standards; greater data dependence; usage of real time data; more challenging versioning/configuring; black-box characteristics; need for more parameters; more specific requirements
	Q_{14}	Please rank the most important solutions to overcome configuration issues in AIS according to their relevance.	Hyperparameter optimization; regular testing; data as code approach; iterative development; reducing system variability; continuous monitoring; configuration-management tools
	–	Further comments	Free text

as originally formulated, content-wise revised, or grouped into new themes. These themes were subsequently incorporated into the questionnaire of the following round or, in the case of the final round, used as the basis for the interpretation of the results.

Consensus: Consensus among participants was primarily assessed using the percentage of agreement, adopting a commonly reported threshold of 75% as an indication of agreement in Delphi studies (Diamond et al., 2014). For rating questions, consensus was additionally evaluated using the IQR. Lower IQR values indicate a stronger convergence of opinions among participants. Following common recommendations, we interpreted IQR ≤ 0.5 as evidence of consensus (Von Der Gracht, 2012; Diamond et al., 2014).

Stability: Building on recommendations for assessing stability in Delphi studies summarized by Von Der Gracht (2012), we defined thresholds to evaluate stability between Delphi rounds. For ranking questions, stability was assessed by comparing the percentage of participants assigning the same rank to an item between rounds. For categorical questions (e.g., single-choice items), stability was assessed by comparing the percentage of participants selecting each option between rounds. Responses were considered stable when the change in these percentages was $\leq 15\%$. For rating questions based on a five-point Likert scale, stability was assessed by comparing the difference between the coded mean ratings of two rounds based on the numerical codings (cf. quantitative analysis). Responses were considered stable when this difference is ≤ 0.5 (i.e., half a Likert-scale category).

Stopping criteria: The study was initially designed to comprise four Delphi rounds. Additional rounds would have been considered if sufficient agreement and stability had not been established across at least 75% of the items (Von Der Gracht, 2012; Diamond et al., 2014). Another stopping criterion concerned the participant dropout (Hoerger, 2010). For the dropout, we defined a threshold of more than 30% relative to the initial Delphi panel.

Concluding and reporting. In the final phase of the study, the findings of all Delphi rounds, however, primarily focusing on the results of round 4, were systematically concluded and reported by RM. More specifically, we reported following the Conducting and REporting of DElphi Studies (CREDES) guidelines proposed by Jünger et al. (2017). Although originally described for healthcare research, the reporting principles are not domain specific. Thus, we adopted the relevant elements and applied them to the context of industrial AI systems. To further support transparency and reproducibility, all collected data and the CREDES document are provided in an open-access repository.

3.2. Study conduct

Participants. For panel size, we aimed to recruit 10 to 20 participants, which is commonly recommended for Delphi studies to obtain robust insights (Okoli and Pawlowski, 2004; McMillan et al., 2016; Belton et al., 2021). We recruited only participants who actively work in the field of AI systems, such as programmers, data scientists, or AI-system operators. The targeted working area was the production industry. Participants were specifically recruited based on their experience with AI systems in production environments. However, we did not restrict recruitment to a specific industry in the context of these environments, such as manufacturing or logistics. We relied on four separate channels to reach potential participants:

1. a manually created mailing list (89 persons)
2. the professional network of the authors
3. the LinkedIn network of RM

Table 2
Response overview of the demographics across all Delphi rounds. Values are reported as n (%). Percentages may exceed 100% due to multiple selections.

Question	Options	Round 1	Round 2	Round 3	Round 4
Q_1: experience	0 to 2	2 (8%)	1 (4%)	1 (5%)	1 (5%)
	3 to 5	7 (27%)	6 (25%)	6 (27%)	6 (29%)
	6 to 10	7 (27%)	7 (29%)	5 (23%)	5 (24%)
	11+	10 (39%)	10 (42%)	10 (45%)	9 (43%)
Q_2: industry	Manufacturing	20 (77%)	19 (79%)	17 (77%)	16 (76%)
	Information technology	14 (54%)	14 (58%)	13 (59%)	13 (62%)
	Research	10 (39%)	10 (42%)	10 (45%)	9 (43%)
	Logistics	3 (12%)	2 (8%)	2 (9%)	2 (10%)
	Prefer not to say	1 (4%)	1 (4%)	1 (5%)	1 (5%)
Q_3: employment	Research & development	23 (87%)	22 (92%)	20 (91%)	19 (90%)
	Operations	11 (42%)	9 (38%)	8 (36%)	8 (38%)
	Quality	3 (12%)	3 (13%)	3 (14%)	3 (14%)
	Management	2 (8%)	1 (4%)	1 (5%)	1 (5%)
	Other	2 (8%)	2 (8%)	2 (9%)	2 (10%)

4. snowballing via participant networks

Delphi process. Using the defined recruitment channels, we reached at least 94 potential participants; however, the exact number of persons approached via LinkedIn and snowball sampling remains unknown. The first Delphi round was conducted from May 22 to June 22, 2024 (i.e., four weeks) and yielded 27 responses in total and 26 valid responses after data cleaning. Although this exceeded our initially targeted panel size of 10 to 20 participants, we retained all 26 eligible participants to preserve the breadth of practitioner perspectives represented in the recruited panel. Moreover, as some participant dropout is common across Delphi rounds (Hoerger, 2010), the larger initial panel was considered beneficial for maintaining a sufficient panel size throughout the subsequent rounds. All subsequent rounds were open for two weeks, with a cooling period of two weeks between consecutive rounds. Round 2 involved 24 participants (return rate 92%), followed by 22 participants in round 3 (return rate 92%). A fourth round was conducted from September 29 to October 12, 2025, after an intentionally longer interval of approximately one year. This round again allowed two weeks for participation and resulted in 21 participants (return rate 96%). The extended pause was introduced to examine the stability of expert assessments over time and to identify potential shifts in priorities after a prolonged reflection period. Thus, the fourth round was conducted to assess the stability of the previous assessments rather than in response to substantial disagreement among the experts. By the end of round 4, sufficient agreement and stability had been established across at least 75% of the items, while participant dropout remained below the predefined threshold. Therefore, no additional Delphi round was considered necessary.

The questionnaire was iteratively refined across the Delphi rounds. Round 1 included 49 items. For round 2, eleven items were removed and 42 new items were added based on expert responses. These changes resulted from systematically coding and categorizing the participants' round-1 responses; all new items were derived from this content, with no additional content introduced by the research team. Precisely, in the *development* section, one item was removed and 13 added, while 19 items were added in the *configurability* section. In the *experiences* section, all initial items were replaced by ten statements derived from the participants' free-text responses, and the section was renamed *adoption*. Thus, removed or replaced items reflect restructuring rather than the exclusion of participant-derived content. In rounds 3 and 4, no further items were added or removed. Only the ordering of ranking options was adjusted based on the previous round's results. Note that the aggregated feedback given between rounds did not result in further additions or removals of items across the subsequent rounds.

4. Results

In the following, we present the results of our Delphi study, focusing on *demographics*, *development*, *configurability*, and *adoption*, which are called *experiences* in round 1 (cf. **RQ$_1$**). While demographics are reported across all rounds, the analysis of the remaining topics focuses on the final outcomes of round 4. Data from the earlier rounds are available in our open-access repository. Except for the demographics, we report the agreement (%) for each item. Note that the IQR remained 0 for all rating questions.

4.1. Demographics

In Table 2, we show an overview of the Demographics. Generally, our panel reflected a high level of **Q$_1$** professional experience in AI-system development. Across the rounds, between 66%–67% of the experts reported more than six years of experience. Only around 6% have zero to two years of experience. Most respondents reported their **Q$_2$** industry backgrounds in the manufacturing sector (77%–79%) and information technology (54%–62%), while smaller shares indicated affiliations with research (39%–45%) and logistics (8%–12%). One participant selected the *prefer not to say* option. In terms of **Q$_3$** employment roles, research and development positions were most common among the panel (87%–92%). Further roles referred to operations (36%–42%), quality management (12%–14%), general management (4%–8%), and other (8%–10%), including, for example, risk management. Note that all participants included in our study are from Germany.

Based on the characteristics described, our expert panel can be regarded as well aligned with the focus of this study. At the same time, the number of participants remained largely stable throughout the Delphi process and, even in the round 4, slightly exceeded the panel sizes commonly reported for specialized Delphi studies (cf. Section 3.2).

We emphasize that these factors support a high level of confidence in the relevance and reliability of the collected assessments.

4.2. Development

As shown in Fig. 2, participants identified data preparation (95%) as the most **Q$_4$** challenging development phase of AI systems ($W = 0.96$). Deployment (100%), operation (76%), and change management (81%) followed, whereas testing (95%), design (91%), and implementation (91%) generally received lower ranks (all stable). Data quality (100%) was clearly perceived as the most **Q$_5$** critical development challenge ($W = 0.99$). The following challenges referred to system customization and configurability (100%), system adaptability and change management (100%), meeting stakeholder expectations (100%), and model efficiency (100%) Besides system scalability (71%) and data handling (71%), which were considered less critical overall, all other items

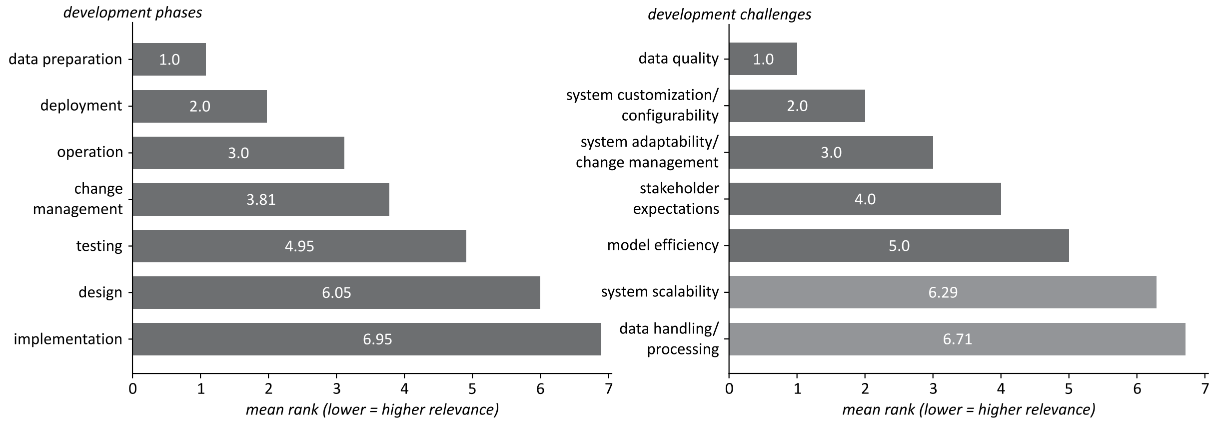

Fig. 2. Ranking of development phases (Q_4) and development-related challenges (Q_5) in round 4. Lower mean ranks indicate higher relevance. Light gray bars indicate unstable rankings.

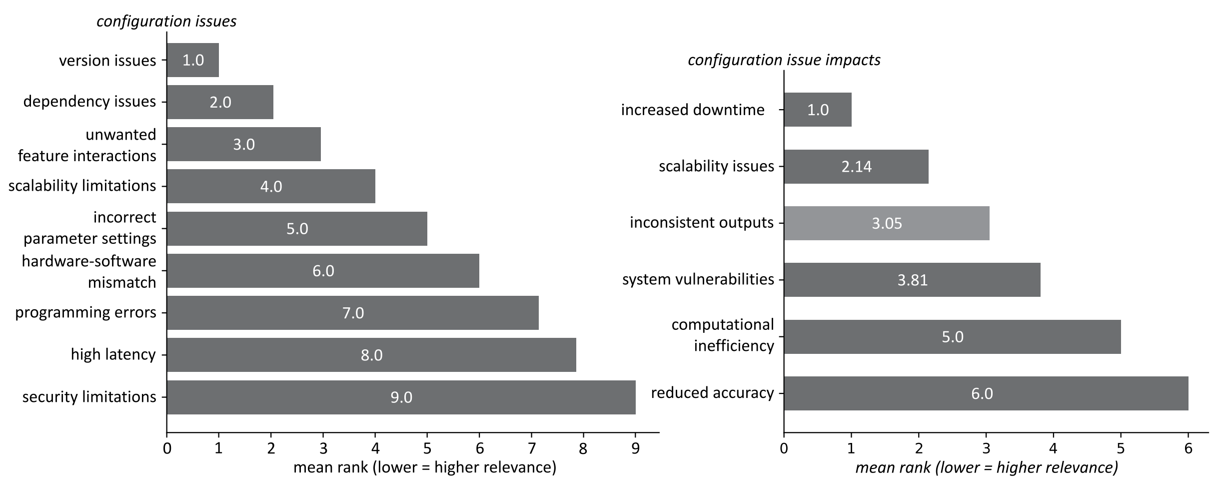

Fig. 3. Ranking of configuration issues (Q_8) and configuration issue impacts (Q_{11}) in round 4. Lower mean ranks indicate higher relevance. Light gray bars indicate unstable rankings.

remained stable. Interestingly, the perceived importance of these two fluctuated across rounds, as some participants continued to interchange their positions until the final round. Regarding $\mathbf{Q_6}$ mitigation strategies ($W = 0.96$), regular testing (100%) and the use of AI-system templates (100%) were most frequently considered appropriate strategies. Stakeholder communication (81%) and improvements in data quality (81%) followed as additional, stable strategies. Iterative system optimization (71%) and holistic requirements-engineering processes (71%) were ranked lowest and remained unstable until the final round. Ten experts provided additional comments. Specifically, they mentioned the importance of AI explainability and traceability (2), compliance with new legal regulations (2), and integration of different AI technologies (2). Other comments concerned AI-system templates (1), organizational readiness (1), preprocessing effort (1), and limited training data (1).

4.3. Configurability

Responses regarding the $\mathbf{Q_7}$ awareness of configurability in AI systems indicate stable consensus that most participants selected the neutral option, with 95% neither agreeing nor disagreeing; 5% agreed. Referring to typical $\mathbf{Q_8}$ configuration issues (cf. Fig. 3), all opinions remained stable with versioning issues most frequently selected at the first rank (100%). Dependency issues (95%) and scalability limitations (95%) followed. Less relevant issues were ranked as following (stable): unwanted feature interactions (100%), incorrect parameter configurations (100%), mismatches between hardware and software components (100%), programming errors (86%), high latency (86%), and security limitations in the target infrastructure (100%). Related to $\mathbf{Q_9}$ ($W = 0.99$), the results show that configuration issues occur primarily when AI systems interact with other systems rather than within AI systems themselves (100%). Moreover, in the context of $\mathbf{Q_{10}}$ configuration-issue occurrences, all responses indicate that these occur frequently (100%, stable). Regarding the $\mathbf{Q_{11}}$ impacts of configuration issues ($W = 0.97$), all participants agreed on increased downtime as most prominent impact (100%). As illustrated in Fig. 3, the remaining impacts were ranked as follows: scalability limitations (86%), inconsistent system outputs (67%), system vulnerabilities (81%), computational inefficiency (100%), and reduced accuracy (100%). Besides inconsistent system outputs, all items were stable. The responses on the $\mathbf{Q_{12}}$ working areas ($W = 0.98$) were all stable and showed that most participants ranked working environments characterized by high feature-richness as most dominant area in the context of configuration issues (100%). It is followed by areas relying on real-time data (81%), monitoring systems (81%), not-specified areas (100%), areas with increasing security requirements (100%), and areas in which AI systems are still uncommon (100%). The $\mathbf{Q_{13}}$ differences between AI systems and traditional cyber physical systems ($W = 0.98$) were ranked as follows (stable): higher system complexity (100%), use of real-time data (91%), greater data dependence (100%), more challenging versioning and configuring (91%), need for more parameters (91%), more specific infrastructure requirements (91%), missing legal standards (100%), and black-box characteristics (100%). The participants ranked in the context of $\mathbf{Q_{14}}$

Table 3
Assessments of adoption statements regarding industrial AI-system practice as well as operational and economic benefits, including ratings, agreement levels, and stability in Delphi round 4.

Item	Rating	n (%)	Consensus	Stability
Industrial AI-system practice				
Supervised learning is currently the most popular learning strategy for industrial environments.	Agree	21 (100%)	Yes	Yes
The prediction of machine or component failures is currently one of the most popular use cases for industrial AI systems.	Agree/disagree	17 (81%)	Yes	Yes
Operational benefits				
AI systems enhance the decision-making capabilities in industrial sectors.	Agree	18 (86%)	Yes	Yes
A main reason for the implementation of AI systems is to improve operational efficiency.	Agree	18 (86%)	Yes	Yes
A main reason for the implementation of AI systems is to reduce the downtime in industrial environments.	Agree	17 (81%)	Yes	Yes
The implementation of AI systems typically leads to an increase in production capacity and a reduction in manual labor.	Agree	21 (100%)	Yes	Yes
The implementation of AI systems improves the overall efficiency in engineering tasks (i.e., planning, production, and maintenance).	Agree	18 (86%)	Yes	Yes
A main reason for the implementation of AI systems is their positive impact on production quality.	Agree/disagree	19 (90%)	Yes	Yes
Economic benefits				
A main reason for the implementation of AI systems is to save costs.	Agree	18 (86%)	Yes	Yes
The implementation of AI systems is typically worthwhile from an economic point of view.	Agree	20 (95%)	Yes	Yes

mitigation strategies to address configuration issues ($W = 0.93$) regular testing (100%) first, followed by continuous monitoring (91%) and the use of configuration-management tools (71%, unstable). These were followed by reducing system variability (71%, unstable), data-as-code approaches (91%), hyperparameter optimization (100%), and iterative development practices (100%). Additionally, we got further comments from 17 experts. They mentioned dependency hell (5), combinatorial explosion (5), configuration reproducibility and feature traceability (4), versioning issues (3), parameter variability (2), misconfigurations (2) and operator fatigue (1).

4.4. Adoption

Out of the experiences (round 1), we derived ten statements to be assessed by the experts. In Table 3, all items and their assessments are presented. Eight items reached agreement while the participants neither agreed nor disagreed with two items. Each item reached stable consensus. Further comments concerned system integration (4), learning strategies (2), data infrastructures (2), architectures and configurability (2), human and explainability aspects (2), and system lifecycle management (1).

After describing the results, we answer $\mathbf{RQ_1}$ as follows:

> $\mathbf{RQ_1}$ **– Development, configuration, and adoption:** *Practitioners perceive industrial AI systems as challenging particularly in terms of data preparation and quality, configurability, adaptability, and system integration. Configuration issues occur frequently and predominantly emerge from interactions with other systems, with versioning, dependencies, scalability, and downtime being central concerns. At the same time, practitioners associate industrial AI adoption with operational and economic benefits, including improved efficiency, reduced downtime, increased production capacity, and cost savings.*

5. Thematic synthesis

By synthesizing the data retrieved from the questionnaire responses, we derived recurring themes related to **benefits**, **issues**, **risks**, and **mitigation strategies** that are currently part of the daily business practice of our experts (cf. $\mathbf{RQ_2}$).

5.1. Benefits

Operational and economic enablement. Industrial AI systems are perceived as operational and organizational enablers rather than isolated analytical models (Zdravković et al., 2021). Experts associated AI systems with improved operational efficiency, reduced downtime, increased production capacity, lower manual effort, and improved engineering processes. In addition, the findings imply that industrial AI systems are generally perceived as economically worthwhile and beneficial for production quality, which is already highlighted by a variety of research (Elahi et al., 2023; Plathottam et al., 2023). Altogether, these findings indicate that the practical value of AI systems is primarily associated with measurable operational improvements and economic optimization within industrial environments.

Infrastructure-level integration and scalability. There is a clear transition from isolated AI pilots toward integrated and scalable production infrastructures. Experts repeatedly highlighted the importance of modular architectures, configurable components, standardized data infrastructures, and unified data models for deploying AI systems across heterogeneous production environments. Furthermore, the integration of AI capabilities into existing operational systems such as maintenance, quality management, and manufacturing execution system platforms (Rakholia et al., 2024; Lin, 2024) was perceived as important. Hence, perceived benefits emerge not only from individual AI models, but also from their integration into larger operational ecosystems and smart factory architectures.

Human-centered decision support. While automation remains an important objective, experts perceive AI systems as supporting rather than replacing human operators. This opinion is also reflected by recent research (Rakholia et al., 2024; Tarafdar, 2025). More broadly, this human-centered perspective is increasingly emphasized in the context of Industry 5.0 (Martini et al., 2024; Yan et al., 2025). In particular, explainability and interpretability (Arrieta et al., 2020) were emphasized by our panel, especially in safety-relevant production environments where operators must understand system behavior and decision processes. These capabilities are essential for validating AI-generated recommendations, understanding unexpected system behavior, and maintaining confidence in operational decisions. Thus, successful industrial AI adoption depends not only on automation capabilities but also on the ability of operators to understand, supervise, and effectively integrate AI-supported decisions into existing production processes, reflecting the Industry 5.0 emphasis on integrating human capabilities with advanced technologies.

5.2. Issues

Integration and infrastructure complexity. A major issue of industrial AI systems concerns their integration into heterogeneous production and automation environments. Experts emphasized difficulties related to legacy infrastructures, proprietary communication protocols, and the integration of AI capabilities into existing operational systems. In particular, the combination of multiple AI modules, automation technologies, and heterogeneous runtime environments was associated with increasing system complexity and unexpected side effects. Furthermore, the results imply that many industrial AI initiatives fail not because of insufficient model capabilities, but because of challenges related to operational integration and infrastructure compatibility. Therefore, integration complexity represents a central practical barrier for industrial AI deployment in real-world production environments (Kutz et al., 2022; Sinha and Lee, 2024).

Configuration complexity. Another issue concerns the increasing configuration and variability complexity of industrial AI systems (Peres et al., 2020; Rabiser and Zoitl, 2021; Sinha and Lee, 2024). The quantitative findings already highlighted configurability, adaptability, and deployment as major challenges. These findings were strongly reinforced by the qualitative statements, where experts referred to dependency hell (Boué et al., 2023), hidden dependencies (Nyakundi, 2026), parameter variability (Eggensperger et al., 2019), combinatorial explosion (Fortz, 2023), and configuration mismatches between training and production environments (Rahman et al., 2025). In addition, experts highlighted that variability is frequently tolerated rather than systematically engineered, resulting in fragile dependencies and unstable system behavior across different production settings. In addition, the complexity of industrial AI systems extends beyond the configuration of individual models or software components. Since configuration decisions frequently affect multiple system dimensions and layers simultaneously (cf. Section 2.1), dependencies can propagate across technological boundaries and create unforeseen interactions. As a result, local configuration changes may also trigger system-wide effects that are difficult to predict, validate, and manage throughout the system lifecycle.

Data and lifecycle management challenges. The findings further show substantial challenges related to data management, traceability, and lifecycle control. Experts described high efforts for cleaning, synchronizing, and contextualizing industrial data, as well as limitations caused by missing labeled failure events. At the same time, reproducibility, versioning, and configuration traceability were described as difficult and often associated with unclear or highly manual workflows (Semmelrock et al., 2025). Several experts additionally emphasized growing requirements for long-term lifecycle management, including monitoring, retraining strategies, and configuration management. In general, the operation of industrial AI systems increasingly requires continuous lifecycle-oriented management processes that extend far beyond initial model development.

Organizational and governance limitations. The results show that organizational readiness and governance capabilities represent important practical limitations for industrial AI adoption (Kinkel et al., 2022). Experts reported that many organizations, including small and medium-sized companies, lack sufficient resources, validation procedures, and operational structures for deploying and maintaining industrial AI systems at scale. Furthermore, increasing regulatory requirements (e.g., EU AI Act, 2024) were frequently perceived as difficult to operationalize due to limited guidance and unclear implementation strategies. The findings therefore suggest that barriers to industrial AI adoption often arise from missing organizational capabilities and governance structures rather than from limitations of the AI technologies themselves.

5.3. Risks

Operational risks. The findings indicate that unresolved configuration, integration, and deployment issues can directly threaten operational stability within industrial environments. The quantitative results already identified increased downtime, scalability issues, and inconsistent outputs as major consequences of configuration-related problems, which are widely known issues in research (Xu and Zhou, 2015; Rahman et al., 2025). These findings were reinforced by qualitative expert statements describing degraded model performance, configuration mismatches between training and production environments, and even complete machine stops after deployment errors. In addition, experts repeatedly emphasized that hidden dependencies and combinatorial interactions between models, data sources, and runtime components create fragile operational environments with difficult-to-predict system behavior. Thus, our findings imply that industrial AI systems increasingly introduce risks related to operational instability and production disruptions in heterogeneous production infrastructures.

Transparency and control risks. A major risk concerns the increasing loss of transparency, traceability, and controllability of industrial AI systems (Kroll, 2021; Mora-Cantallops et al., 2021). Several times, experts stated explainability, traceability, and documentation as critical requirements, particularly in safety-relevant production environments. At the same time, the findings revealed substantial difficulties regarding versioning, reproducibility, and configuration tracking. The growing variability of configurations, parameters, models, and deployment environments further amplifies these challenges. As a result, organizations may struggle to understand which configuration states, model versions, or dependencies are responsible for specific system behaviors. Altogether, the findings indicate that industrial AI systems risk evolving into both difficult-to-trace and difficult-to-control operational ecosystems.

Human and organizational risks. There are several human-centered and organizational risks associated with industrial AI adoption, which are related to the already described issues. Experts highlighted that frequent configuration problems and false alarms may lead to operator fatigue (Rožanec et al., 2023) and reduced trust (Dorton and Harper, 2022) in AI-supported processes. Furthermore, insufficient explainability may limit operator acceptance and reduce confidence in safety-relevant decision support systems (Arrieta et al., 2020). The findings also indicate that many organizations still lack sufficient validation capabilities, governance structures, and operational readiness for deploying AI systems at scale (Uren and Edwards, 2023; Hussein et al., 2026). Consequently, we state that industrial AI risks increasingly emerge not only from technical failures, but also from organizational limitations and human operational factors.

Governance and compliance risks. Our findings also indicate growing governance and compliance risks associated with industrial AI systems. As already stated, experts referred to increasing regulatory pressure, leading to high risks regarding successful AI system implementation. In fact, risks may also arise in other regulatory areas, including security (ISO/IEC 27000, 2018). In particular, security incidents in safety-critical areas may have severe consequences, as security breaches may compromise system functionality, potentially causing environmental damage and endangering human lives (May et al., 2023a, 2024a,b). However, integrating security comes with huge challenges, such as considering increasing system complexity due to growing configuration spaces (Mellado et al., 2010; Varela-Vaca et al., 2021). Consequently, increasing configurability and operational integration can make industrial AI systems more difficult to govern, audit, and secure. This challenge is further amplified when configuration decisions, model versions, and system dependencies are insufficiently documented or traceable, limiting organizations' ability to demonstrate compliance and maintain operational control.

5.4. Mitigation strategies

Standardization and modularization. Standardization and modularization represent central mitigation strategies for managing the growing complexity of industrial AI systems. Modular architectures and components (Pfeiffer et al., 2023), reusable templates (May et al., 2023b), and standardized data infrastructures (Liang et al., 2022) are important approaches for improving scalability and operational integration. In particular, unified data models and reusable parameter configurations were perceived as important mechanisms for reducing integration effort and configuration inconsistencies across heterogeneous production environments.

Traceability and lifecycle management. Another mitigation strategy concerns the establishment of traceability and lifecycle-oriented management practices, which are ongoing challenges at the same time. Concrete approaches concerned tagging mechanisms, unique identifiers, configuration tracking, versioning approaches, and monitoring processes as important measures for improving reproducibility and operational transparency (Mora-Cantallops et al., 2021; Schlegel and Sattler, 2023). In addition, several statements emphasized the growing importance of lifecycle management strategies (Kreuzberger et al., 2023) including retraining processes, drift monitoring, and continuous configuration management. To conclude, the results suggest that effective operation of industrial AI systems depends on the ability to continuously track configurations, model versions, data changes, and system behavior throughout the system lifecycle. Without such traceability mechanisms, organizations may struggle to reproduce results, diagnose failures, or manage the evolution of deployed AI systems.

Transparency and operational oversight. Transparency and operational oversight are essential mitigation strategies for industrial AI systems. Experts stated explainability, documentation, and traceability as prerequisites for understanding system behavior, diagnosing unexpected outcomes, and validating AI-supported decisions (Arrieta et al., 2020; Tarafdar, 2025). In particular, these practices become increasingly important in safety-relevant production environments, where operators remain responsible for assessing and acting upon AI-generated recommendations.

Governance and structured reuse. Governance frameworks and organizational capability development are also relevant from the experts' perspective. While regulatory frameworks and standards can lead to issues and risks, they are also a chance for structuring industrial AI governance and operational control (Agarwal and Nene, 2024; Saleh and Abdulsalam, 2025). Investing in standardized infrastructures based on core assets (e.g., templates) and the subsequent derivation of configured variants can lead to several advantages, including systematic reuse, improved quality assurance, or reduced engineering effort (Apel et al., 2013; Rabiser and Zoitl, 2021; May and Adler, 2025). We argue that effective mitigation requires not only technical solutions, but also structured governance mechanisms, standardized engineering practices, and systematic reuse strategies.

Based on the thematic synthesis of our results, we answer **RQ$_2$** as follows:

> **RQ$_2$ – Benefits, issues, risks, and mitigation strategies:** *Relevant benefits primarily concern operational efficiency, reduced downtime, increased production capacity, and improved decision support. Issues mainly arise from integration and configuration complexity, data and lifecycle management, and organizational limitations. Risks include operational instability, limited transparency and control, human factors, and governance and compliance concerns. Mitigation strategies focus on standardization and modularization, traceability and lifecycle management, operational oversight, and structured governance and reuse.*

6. Implications

To make the insights we described more actionable, we provide additional implications for both research and practice based on them.

Research implications. For researchers, the findings suggest that future work should move beyond isolated model evaluations and place greater emphasis on the engineering and operation of industrial AI systems in production environments. Research should investigate how industrial AI systems can be engineered as configurable software ecosystems through systematic variability management, reusable core assets, and configurable platform architectures. In particular, approaches originating from software product line engineering (Apel et al., 2013), such as reusable core assets, templates, reference architectures, and configurable variants, appear promising for improving scalability, maintainability, and systematic reuse in industrial AI environments. Future studies should therefore complement performance metrics with operational qualities, such as configurability, maintainability, traceability, or reproducibility. This broader system-level perspective can help align academic evaluation criteria more closely with the practical concerns encountered by industrial practitioners.

Practical implications. For practitioners, the findings indicate that investments should primarily target operational and organizational capabilities rather than model development alone. Organizations should establish reusable AI platforms based on standardized architectures, templates, and configurable core assets that support the systematic derivation of AI system variants for different production contexts. Here, we recommend to take product line engineering approaches (Apel et al., 2013) into account. Such approaches can reduce engineering effort, improve consistency, facilitate traceability, and support the scalable deployment and maintenance of industrial AI systems across heterogeneous environments. Finally, organizations should complement these technical capabilities with structured governance mechanisms, including lifecycle management, configuration management, and clear ownership of models, data assets, and configurations, to maintain operational control and long-term system reliability.

7. Threats to validity

Next, we describe relevant threats to the internal, external, and construct validity of our study.

7.1. Internal validity

Internal validity may be influenced by researcher interpretation during the analysis and consolidation of qualitative responses. Free-text responses from the initial Delphi rounds were analyzed using thematic analysis and open-card sorting, which required us to interpret and group statements into themes and structured items. Furthermore, the final synthesis of qualitative feedback after round 4 may introduce subjective bias. To mitigate these threats, responses were independently reviewed and coded by multiple researchers and discussed until consensus was reached. The general robustness of our results is supported by the stability of the consensus process. In the final round, agreement levels were high ($W > 0.93$) with statistically significant results ($p < 0.001$), indicating a strong and reliable convergence of expert opinions that transcends individual round-to-round fluctuations. Nevertheless, another primary threat relates to the iterative nature of the Delphi process. Participants received aggregated feedback, which might trigger a bandwagon effect where experts align their assessments with emerging group tendencies rather than maintaining their initial perspectives (Bhandari and Hallowell, 2021).

7.2. External validity

The generalizability of the findings is limited by the specific study context. The Delphi panel consisted exclusively of experts involved in developing or operating industrial AI systems. While this focus ensured practical expertise, the results may not directly transfer to other AI domains, e.g., healthcare. Furthermore, the expert panel covered heterogeneous industry backgrounds, potentially limiting the transferability of findings to specific industries. However, manufacturing, information technology, and research were most strongly represented, providing a relatively consistent production-related context. Additionally, the recruitment strategy may have introduced selection bias, as practitioners already connected to these professional networks might be more inclined to participate. The panel size, however, supports the external validity within the industrial domain. Starting with 26 valid participants and maintaining high return rates (cf. Fig. 1) aligns with established recommendations for Delphi studies in specialized fields (McMillan et al., 2016; Belton et al., 2021). Despite this, a potential social-desirability bias remains, as participants might report practices that align with organizational expectations rather than actual daily-business operations.

7.3. Construct validity

Construct validity concerns whether the questionnaire items adequately represent the complex concepts of industrial AI systems. A primary threat arises from the diverse backgrounds of the practitioners; different technical roles and responsibilities can lead to varying interpretations of terms like *development challenges* or *configuration issues*. So, participants may interpret individual items differently despite identical formulations and explanations from our site. Furthermore, the operationalization of these socio-technical concepts into discrete items may not fully capture the entire depth of real-world industrial practice. To address this issue, the questionnaire was iteratively refined. Although removing or introducing items after the first round (Diamond et al., 2014) limits strict comparability between rounds, it ensures that the constructs measured are more accurately aligned with the actual experiences of the experts.

8. Conclusion

In this article, we presented a four-round Delphi study with 26 practitioners involved in the development, configuration, and adoption of industrial AI systems in production environments. Using iterative expert assessments, we consolidated practical experiences across these three areas (**RQ_1**) and, through thematic synthesis, derived recurring benefits, issues, risks, and mitigation strategies (**RQ_2**).

The findings indicate that industrial AI systems are increasingly evolving from isolated analytical models toward integrated and highly configurable operational ecosystems. While experts strongly associated AI systems with operational efficiency, production optimization, and scalable infrastructures, they simultaneously emphasized substantial challenges related to integration complexity, variability, lifecycle management, traceability, and governance. In particular, dependency-related problems, configuration mismatches, and operational integration into heterogeneous environments emerged as recurring practical concerns.

Overall, the study suggests that many practical challenges of industrial AI systems no longer primarily emerge from model development itself, but from their operational embedding into highly configurable, socio-technical production environments. Thus, future research should focus on system-centric engineering perspectives, particularly regarding variability management, lifecycle governance, explainability, and operational integration of industrial AI systems.

CRediT authorship contribution statement

Richard May: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Leonard Cassel:** Writing – review & editing, Supervision, Conceptualization. **Hashir Hussain:** Writing – review & editing, Methodology, Investigation. **Muhammad Talha Siddiqui:** Writing – review & editing, Methodology, Investigation. **Tobias Niemand:** Writing – review & editing, Conceptualization. **Paul Scholz:** Writing – review & editing, Conceptualization. **Thomas Leich:** Writing – review & editing, Project administration.

Declaration of generative AI use

During the preparation of this work the authors used ChatGPT 5.2 Pro in order to identify spelling and grammatical errors and to assist with paraphrasing. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank all experts who participated in this Delphi panel.

Data availability

The authors provide a public open-access repository on Zenodo.

References

Agarwal, A., Nene, M.J., 2024. Standardised schema and taxonomy for AI incident databases in critical digital infrastructure. In: Pune Section International Conference. PuneCon, IEEE, pp. 1–6.
Alon, I., Haidar, H., Haidar, A., Guimon, J., 2025. The future of artificial intelligence: Insights from recent Delphi studies. Futures 165, 103514.
Apel, S., Batory, D., Kästner, C., Saake, G., 2013. Feature-Oriented Software Product Lines. Springer.
Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al., 2020. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion 58, 82–115.
Belton, I., Wright, G., Sissons, A., Bolger, F., Crawford, M.M., Hamlin, I., Lūka, C.T.B., Vasilichi, A., 2021. Delphi with feedback of rationales: How large can a Delphi group be such that participants are not overloaded, de-motivated, or disengaged? Technol. Forecast. Soc. Change 170, 120897.
Bhambri, P., Rani, S., Khang, A., 2024. AI-driven digital twin and resource optimization in industry 4.0 ecosystem. Intell. Tech. Predict. Data Anal. 47–69.
Bhandari, S., Hallowell, M.R., 2021. Identifying and controlling biases in expert-opinion research: Guidelines for variations of Delphi, nominal group technique, and focus groups. J. Manag. Eng. 37 (3), 04021015.
Binder, C., Neureiter, C., Lüder, A., 2022. Towards a domain-specific information architecture enabling the investigation and optimization of flexible production systems by utilizing artificial intelligence. Int. J. Adv. Manuf. Technol. 123 (1), 49–81.
Bolstorff, P., Rosenbaum, R., 2007. Supply chain excellence: A handbook for dramatic improvement using the SCOR model. J. Supply Chain Manag. 39 (4), 38.
Boué, L., Kunireddy, P., Subotić, P., 2023. Automatically resolving data source dependency hell in large scale data science projects. In: International Conference on AI Engineering–Software Engineering for AI. CAIN, IEEE, pp. 1–6.

Chryssolouris, G., Alexopoulos, K., Arkouli, Z., 2023. Artificial intelligence in manufacturing systems. In: A Perspective on Artificial Intelligence in Manufacturing. Springer, pp. 79–135.
Clarke, V., Braun, V., 2017. Thematic analysis. J. Posit. Psychol. 12 (3), 297–298.
Culot, G., Orzes, G., Sartor, M., Nassimbeni, G., 2020. The future of manufacturing: A Delphi-based scenario analysis on industry 4.0. Technol. Forecast. Soc. Change 157, 120092.
Demlehner, Q., Schoemer, D., Laumer, S., 2021. How can artificial intelligence enhance car manufacturing? A Delphi study-based identification and assessment of general use cases. Int. J. Inf. Manage. 58, 102317.
Denecke, K., May, R., LLMHealthGroup, Rivera Romero, O., 2024. Potential of large language models in health care: Delphi study. J. Med. Internet Res. 26, e52399.
Diamond, I.R., Grant, R.C., Feldman, B.M., Pencharz, P.B., Ling, S.C., Moore, A.M., Wales, P.W., 2014. Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. J. Clin. Epidemiol. 67 (4), 401–409.
Domingos, P., 2012. A few useful things to know about machine learning. Commun. ACM 55 (10), 78–87.
Dorton, S.L., Harper, S.B., 2022. A naturalistic investigation of trust, AI, and intelligence work. J. Cogn. Eng. Decis. Mak. 16 (4), 222–236.
Eggensperger, K., Lindauer, M., Hutter, F., 2019. Pitfalls and best practices in algorithm configuration. J. Artificial Intelligence Res. 64, 861–893.
Elahi, M., Afolaranmi, S.O., Martinez Lastra, J.L., Perez Garcia, J.A., 2023. A comprehensive literature review of the applications of AI techniques through the lifecycle of industrial equipment. Discov. Artif. Intell. 3 (1), 43.
EU AI Act, 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024. Law, European Union (EU).
Fortz, S., 2023. Variability-aware behavioural learning. In: International Systems and Software Product Line Conference. SPLC, ACM, pp. 11–15.
Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016. Deep Learning. vol. 1, MIT Press, 2.
Gopinathan, V.R., 2025. Software engineering practices for AI-driven systems: From development to deployment (MLOps perspective). Int. J. Sci. Res. Technol. 8 (1), 13493–13500.
Haindl, P., Buchgeher, G., Khan, M., Moser, B., 2022. Towards a reference software architecture for human-ai teaming in smart manufacturing. In: International Conference on Software Engineering: New Ideas and Emerging Results. ICSE-NIER, ACM, pp. 96–100.
Hoerger, M., 2010. Participant dropout as a function of survey length in internet-mediated university studies: implications for study design and voluntary participation in psychological research. Cyberpsychology Behav. Soc. Netw. 13 (6), 697–700.
Hollender, M., Xu, C., Tan, R., 2024. Engineering challenges in industrial ai. In: International Conference on AI Engineering-Software Engineering for AI. CAIN, ACM, pp. 41–42.
Hou, B.-C., Yu, W.-T., Lu, X.-B., Yang, C.-w., et al., 2017. Applications of artificial intelligence in intelligent manufacturing: a review. Front. Inf. Technol. Electron. Eng. 18 (1), 86–96.
Hsu, C.-C., Sandford, B.A., 2007. The Delphi technique: Making sense of consensus. Pr. Assess. Res. Eval. 12 (1).
Hussein, R., Zink, A., Ramadan, B., Howard, F.M., Hightower, M., Shah, S., Beaulieu-Jones, B.K., 2026. Advancing healthcare AI governance through a comprehensive maturity model based on systematic review. NPJ Digit. Med. 9 (1), 236.
ISO/IEC 27000, 2018. ISO/IEC 27000:2018 – Information Technology – Security Techniques – Information Security Management Systems – Overview and Vocabulary. Standard, International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC).
Jünger, S., Payne, S.A., Brine, J., Radbruch, L., Brearley, S.G., 2017. Guidance on conducting and reporting Delphi studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. Palliat. Med. 31 (8), 684–706.
Khan, A.N., Jabeen, F., Mehmood, K., Soomro, M.A., Bresciani, S., 2023. Paving the way for technological innovation through adoption of artificial intelligence in conservative industries. J. Bus. Res. 165, 114019.
Khang, A., Abdullayev, V., Hahanov, V., Shah, V., 2024. Advanced IoT Technologies and Applications in the Industry 4.0 Digital Economy. CRC Press.
Kinkel, S., Baumgartner, M., Cherubini, E., 2022. Prerequisites for the adoption of AI technologies in manufacturing–Evidence from a worldwide sample of manufacturing companies. Technovation 110, 102375.
Kreuzberger, D., Kühl, N., Hirschl, S., 2023. Machine learning operations (mlops): Overview, definition, and architecture. IEEE Access 11, 31866–31879.
Kroll, J.A., 2021. Outlining traceability: A principle for operationalizing accountability in computing systems. In: Conference on Fairness, Accountability, and Transparency. FAccT, ACM, pp. 758–771.
Kusiak, A., 2018. Smart manufacturing. Int. J. Prod. Res. 56, 508–517.
Kutz, J., Neuhüttler, J., Spilski, J., Lachmann, T., 2022. Implementation of AI technologies in manufacturing-success factors and challenges. Hum. Side Serv. Eng. 62 (62).
Liang, W., Tadesse, G.A., Ho, D., Fei-Fei, L., Zaharia, M., Zhang, C., Zou, J., 2022. Advances, challenges and opportunities in creating data for trustworthy AI. Nat. Mach. Intell. 4 (8), 669–677.
Lin, X., 2024. Artificial intelligence in the industrial engineering. Adv. Oper. Res. Prod. Manag. 1, 1–6.
Martini, B., Bellisario, D., Coletti, P., 2024. Human-centered and sustainable artificial intelligence in industry 5.0: Challenges and perspectives. Sustainability 16 (13), 5448.
May, R., Adler, S., 2025. Experiences in developing configurable digital twin-assisted xr applications for industrial environments. In: Working Conference on Variability Modelling of Software-Intensive Systems. VaMoS, ACM, pp. 107–111.
May, R., Alex, A.J., Suresh, R., Leich, T., 2024a. Product-line engineering for smart manufacturing: A systematic mapping study on security concepts. In: International Conference on Software Technologies. ICSOFT, SciTePress, pp. 323–330.
May, R., Cassel, L., Hussain, H., Siddiqui, M.T., Niemand, T., Scholz, P., Leich, T., 2025. Adopting artificial-intelligence systems in manufacturing: A practitioner survey on challenges and added value. In: International Conference on Software Technologies. ICSOFT, SciTePress, pp. 87–98.
May, R., Gautam, J., Sharma, C., Biermann, C., Leich, T., 2023a. A systematic mapping study on security in configurable safety-critical systems based on product-line concepts. In: International Conference on Software Technologies. ICSOFT, SciTePress, pp. 217–224.
May, R., Krüger, J., Leich, T., 2024b. Sok: How artificial-intelligence incidents can jeopardize safety and security. In: International Conference on Availability, Reliability and Security. ARES, ACM, pp. 1–12.
May, R., Niemand, T., Scholz, P., Leich, T., 2023b. Design patterns for monitoring and prediction machine learning systems: Systematic literature review and cluster analysis. ICSOFT 209–216.
McMillan, S.S., King, M., Tully, M.P., 2016. How to use the nominal group and Delphi techniques. Int. J. Clin. Pharm. 38, 655–662.
Mellado, D., Fernández-Medina, E., Piattini, M., 2010. Security requirements engineering framework for software product lines. Inf. Softw. Technol. 52 (10), 1094–1117.
Mithas, S., Chen, Z.-L., Saldanha, T.J., De Oliveira Silveira, A., 2022. How will artificial intelligence and industry 4.0 emerging technologies transform operations management? Prod. Oper. Manage. 31 (12), 4475–4487.
Mo, F., Rehman, H.U., Monetti, F.M., Chaplin, J.C., Sanderson, D., Popov, A., Maffei, A., Ratchev, S., 2023. A framework for manufacturing system reconfiguration and optimisation utilising digital twins and modular artificial intelligence. Robot. Comput.-Integr. Manuf. 82, 102524.
Monostori, L., 2014. Cyber-physical production systems: Roots, expectations and R&D challenges. Procedia Cirp 17, 9–13.
Monostori, L., Kádár, B., Bauernhansl, T., Kondoh, S., Kumara, S., Reinhart, G., Sauer, O., Schuh, G., Sihn, W., Ueda, K., 2016. Cyber-physical systems in manufacturing. CIRP Ann 65 (2), 621–641.
Mora-Cantallops, M., Sánchez-Alonso, S., García-Barriocanal, E., Sicilia, M.-A., 2021. Traceability for trustworthy AI: A review of models and tools. Big Data Cogn. Comput. 5 (2), 20.
Nyakundi, N.B., 2026. Managing entanglement in AI-enabled safety-critical systems (AISCS). In: Annual Computing and Communication Workshop and Conference. CCWC, IEEE, pp. 1511–1520.
Okoli, C., Pawlowski, S.D., 2004. The Delphi method as a research tool: An example, design considerations and applications. Inf. Manag. 42 (1), 15–29.
Orm, M.B., Jeunet, J., 2018. Time cost quality trade-off problems: a survey exploring the assessment of quality. Comput. Ind. Eng. 118, 319–328.
Peres, R.S., Jia, X., Lee, J., Sun, K., Colombo, A.W., Barata, J., 2020. Industrial artificial intelligence in industry 4.0-systematic review, challenges and outlook. IEEE Access 8, 220121–220139.
Pfeiffer, J., Ruder, S., Vulić, I., Ponti, E.M., 2023. Modular deep learning. arXiv preprint arXiv:2302.11529.
Pimenov, D.Y., Bustillo, A., Wojciechowski, S., Sharma, V.S., Gupta, M.K., Kuntoğlu, M., 2023. Artificial intelligence systems for tool condition monitoring in machining: Analysis and critical review. J. Intell. Manuf. 34 (5), 2079–2121.
Plathottam, S.J., Rzonca, A., Lakhnori, R., Iloeje, C.O., 2023. A review of artificial intelligence applications in manufacturing operations. J. Adv. Manuf. Process. 5 (3), e10159.
Rabiser, R., Zoitl, A., 2021. Towards mastering variability in software-intensive cyber-physical production systems. In: International Conference on Industry 4.0 and Smart Manufacturing. ISM, vol. 180, Elsevier, pp. 50–59.
Rahman, M., Khatoonabadi, S., Abdellatif, A., Samaana, H., Shihab, E., 2025. The impact of environment configurations on the stability of AI-enabled systems. In: International Conference on Evaluation and Assessment in Software Engineering. EASE, ACM, pp. 512–523.
Rakholia, R., Suárez-Cetrulo, A.L., Singh, M., Carbajo, R.S., 2024. Advancing manufacturing through artificial intelligence: Current landscape, perspectives, best practices, challenges, and future direction. IEEE Access 12, 131621–131637.
Rehman, H.U., Chaplin, J.C., Zarzycki, L., Ratchev, S., 2021. A framework for self-configuration in manufacturing production systems. In: Doctoral Conference on Computing, Electrical and Industrial Systems. Springer, pp. 71–79.
Ross, N.S., Sheeba, P.T., Shibi, C.S., Gupta, M.K., Korkmaz, M.E., Sharma, V.S., 2024. A novel approach of tool condition monitoring in sustainable machining of Ni alloy with transfer learning models. J. Intell. Manuf. 35 (2), 757–775.

Rožanec, J.M., Križnar, K., Montini, E., Cutrona, V., Koehorst, E., Fortuna, B., Mladenić, D., Emmanouilidis, C., 2023. Predicting operators' fatigue in a human in the artificial intelligence loop for defect detection in manufacturing. IFAC-PapersOnLine 56 (2), 7609–7614.

Saleh, K., Abdulsalam, H., 2025. Operationalizing ISO/IEC 42001: Requirements and conformance evidence for AI management systems. In: International Conference on Artificial Intelligence for Sustainable Innovation. AI-SI, IEEE, pp. 1–4.

Schlegel, M., Sattler, K.-U., 2023. Management of machine learning lifecycle artifacts: A survey. ACM SIGMOD Rec. 51 (4), 18–35.

Schmidt, R.C., 1997. Managing Delphi surveys using nonparametric statistical techniques. Decis. Sci. 28 (3), 763–774.

Scholz, P., 2022. Applikation maschinell lernender Systeme in der vernetzten adaptiven Produktion (Ph.D. thesis). Apprimus Wissenschaftsverlag, RWTH Aachen.

Schuh, G., Scholz, P., 2019. Development of a framework for the systematic identification of ai application patterns in the manufacturing industry. In: Portland International Conference on Management of Engineering and Technology. PICMET, IEEE, pp. 1–8.

Schuh, G., Scholz, P., Leich, T., May, R., 2020. Identifying and analyzing data model requirements and technology potentials of machine learning systems in the manufacturing industry of the future. In: International Scientific Conference on Information Technology and Management Science. ITMS, IEEE, pp. 1–10.

Schuh, G., Scholz, P., Schorr, S., Harman, D., Möller, M., Heib, J., Bähre, D., et al., 2019. Prediction of workpiece quality: An application of machine learning in manufacturing industry. In: 6th International Conference on Computer Science, Engineering and Information Technology. CSEIT-2019, vol. 11232019, AIRCC Publishing Corporation, pp. 189–202.

Scotti, V., 2019. Artificial intelligence. IEEE Instrum. Meas. Mag. 23 (3), 27–31.

Semmelrock, H., Ross-Hellauer, T., Kopeinik, S., Theiler, D., Haberl, A., Thalmann, S., Kowald, D., 2025. Reproducibility in machine-learning-based research: Overview, barriers, and drivers. AI Mag. 46 (2), e70002.

Serôdio, C., Mestre, P., Cabral, J., Gomes, M., Branco, F., 2024. Software and architecture orchestration for process control in industry 4.0 enabled by cyber-physical systems technologies. Appl. Sci. 14 (5), 2160.

Sholeh, M.N., Hatmoko, J.U.D., Widoanindyawati, V., 2026. A decade of artificial intelligence in the construction industry: Trends, applications, and future directions. Arch. Comput. Methods Eng. 1–25.

Sinha, S., Lee, Y.M., 2024. Challenges with developing and deploying AI models and applications in industrial systems. Discov. Artif. Intell. 4 (1), 55.

Soldatos, J., 2024. Artificial Intelligence in Manufacturing: Enabling Intelligent, Flexible and Cost-Effective Production through AI. Springer.

Soori, M., Arezoo, B., Dastres, R., 2023. Machine learning and artificial intelligence in CNC machine tools, a review. Sustain. Manuf. Serv. Econ. 2, 100009.

Steimers, A., Schneider, M., 2022. Sources of risk of AI systems. Int. J. Environ. Res. Public Health 19 (6), 3641.

Tarafdar, R., 2025. Human-AI collaboration in workflow optimization: A framework for hybrid decision systems in automation-heavy industries. Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol. 11 (1), 3594–3613.

Todescato, M., Braholli, O., Chaltsev, D., Di Blasio, I., Don, D., Egger, G., Emig, J., Pasetti Monizza, G., Sacco, P., Siegele, D., et al., 2023. Sustainable manufacturing through application of reconfigurable and intelligent systems in production processes: a system perspective. Sci. Rep. 13 (1), 22374.

Uren, V., Edwards, J.S., 2023. Technology readiness and the organizational journey towards AI adoption: An empirical study. Int. J. Inf. Manage. 68, 102588.

Van Dyck, M., Lüttgens, D., Piller, F.T., Brenk, S., 2023. Interconnected digital twins and the future of digital manufacturing: Insights from a Delphi study. J. Prod. Innov. Manage. 40 (4), 475–505.

Varela-Vaca, Á.J., et al., 2021. CARMEN: A framework for the verification and diagnosis of the specification of security requirements in cyber-physical systems. Comput. Ind. 132, 103524.

Vernon, W., 2009. The Delphi technique: A review. Int. J. Ther. Rehabil. 16 (2), 69–76.

Von Der Gracht, H.A., 2012. Consensus measurement in Delphi studies: Review and implications for future quality assurance. Technol. Forecast. Soc. Change 79 (8), 1525–1536.

Wan, J., Tang, S., Shu, Z., Li, D., Wang, S., Imran, M., Vasilakos, A.V., 2016. Software-defined industrial internet of things in the context of industry 4.0. IEEE Sensors J. 16 (20), 7373–7380.

Wang, J., Ma, Y., Zhang, L., Gao, R.X., Wu, D., 2018. Deep learning for smart manufacturing: Methods and applications. J. Manuf. Syst. 48, 144–156.

Weber, M., Engert, M., Schaffer, N., Weking, J., Krcmar, H., 2023. Organizational capabilities for ai implementation—coping with inscrutability and data dependency in ai. Inf. Syst. Front. 25 (4), 1549–1569.

Xiang, W., Yu, K., Han, F., Fang, L., He, D., Han, Q.-L., 2023. Advanced manufacturing in industry 5.0: A survey of key enabling technologies and future trends. IEEE Trans. Ind. Inform. 20 (2), 1055–1068.

Xu, T., Zhou, Y., 2015. Systems approaches to tackling configuration errors: A survey. ACM Comput. Surv. 47 (4), 1–41.

Yan, J., Liu, Z., Leng, J., Zhao, J.L., Chen, C., Zhang, D., Tao, Y., Wang, Y., Liu, T., Zhang, C., et al., 2025. Human-centric artificial intelligence towards industry 5.0: Retrospect and prospect. J. Ind. Inf. Integr. 47, 100903.

Zdravković, M., Panetto, H., Weichhart, G., 2021. AI-enabled enterprise information systems for manufacturing. Enterp. Inf. Syst. 16, 668–720.

Zhang, W., Yang, D., Wang, H., 2019. Data-driven methods for predictive maintenance of industrial equipment: A survey. IEEE Syst. J. 13 (3), 2213–2227.

Zhang, G., Ye, H., Lin, Y., 2011. Modelling quality attributes in feature models in software product line engineering. In: International Conference on Software Technologies. ICSOFT, SciTePress, pp. 249–254.

Zimmermann, T., 2016. Card-sorting: From text to themes. In: Perspectives on Data Science for Software Engineering. Elsevier, pp. 137–141.