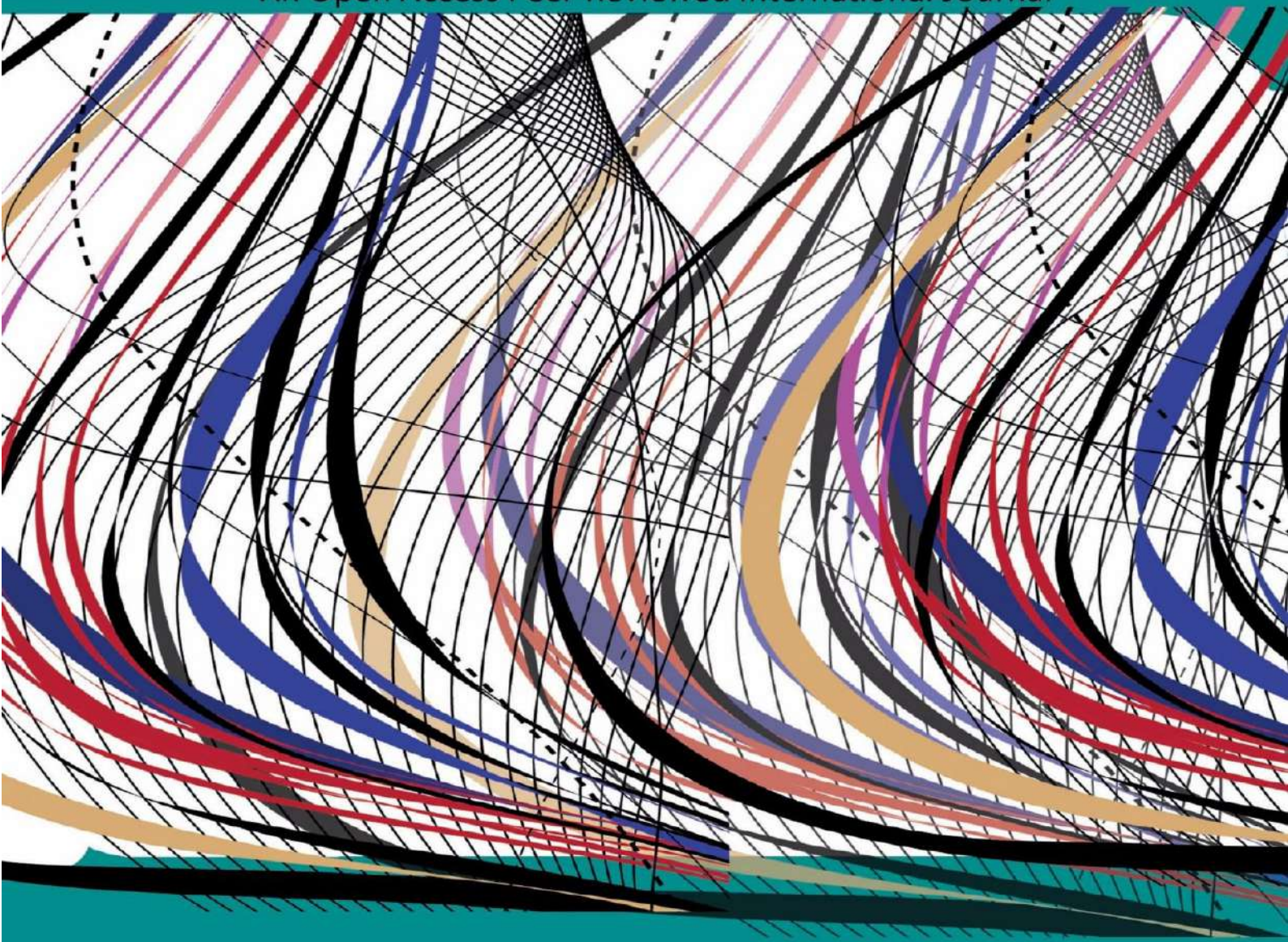
ISSN : 2454-1311

International Journal of Advanced Engineering, Management and Science

Journal CrossRef DOI: 10.22161/ijaems

(IJAEMS)

An Open Access Peer-Reviewed International Journal



Vol-12, Issue- 2 | Mar-Apr 2026

Issue DOI: 10.22161/ijaems.122

INFOGAIN PUBLICATION

https://www.ijaems.com/ | ijaems.editor@gmail.com

International Journal of Advanced Engineering, Management and Science (IJAEMS)

(ISSN: 2454-1311)

DOI: 10.22161/ijaems

Vol-12, Issue-2

March - April, 2026

Publisher

Infogain Publication

Email: ijaems.editor@gmail.com ; editor@ijaems.com

Web: www.ijaems.com

Editorial Board/ Reviewer Board

Vol-12, Issue-2, March-April, 2026

(10.22161/ijaems.122)

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.1

Theoretical Foundations of Decision-Making for Implementing Artificial Intelligence Technologies in Software Products

Elena Levi

Director of Product at Payoneer, Giv'atayim, Israel

Received: 22 Jan 2026; Received in revised form: 21 Feb 2026; Accepted: 27 Feb 2026; Available online: 02 Mar 2026

Abstract *— The study examines theoretical foundations for managerial decision-making on the implementation of artificial intelligence technologies in software products under conditions of accelerated prototyping, AI-supported software engineering workflows, and data-driven product development. The objective is to construct a conceptual decision model that links decision theory, organizational readiness, AI governance, and modern software product management practices. Within the research, existing approaches to AI-based decision-making, organizational AI implementation, and AI-driven software engineering are systematized. The conditions for the reliable deployment of AI functionality into production-grade software are analyzed. Special attention is paid to data quality, digital maturity, and responsible AI governance as determinants of adoption decisions. The methodological base combines a targeted review of recent scientific literature, comparative analysis of conceptual frameworks, and synthesis of a multi-level decision model for product leaders. The conclusions outline the stages and criteria for decision-making on AI implementation in software products, and provide practical guidelines for product leaders and software engineering managers seeking to evaluate AI opportunities, structure experimentation, and align AI-enabled prototyping with long-term product strategy and trust.*

Keywords *— **artificial intelligence implementation, software products, decision-making, product management, data-driven decisions, digital maturity, responsible AI governance, AI prototyping, software engineering, organizational adoption.***

I. INTRODUCTION

Artificial intelligence has moved from experimental pilots to routine components of commercial software products, including SaaS platforms, marketing technology systems, and data-intensive services. AI-supported prototyping and AI-enabled code generation shorten the path from idea to interface, but, by themselves, do not answer whether a given AI feature deserves investment, integration into the architecture, and long-term support from engineering teams. Product leaders face not only the question of technical feasibility but also the challenge of structuring decisions about when, where, and how AI should be embedded into software products without eroding trust, maintainability, or customer value.

Existing research on AI and decision-making concentrates on multi-criteria decision-making, data-driven optimization, and human–AI collaboration. Studies of AI implementation in organizations describe antecedents, barriers, and consequences at the levels of processes, infrastructure, and people, yet rarely delve into the concrete decision logic of product teams responsible for software features. At the same time, work on AI-driven software engineering highlights fast-growing possibilities for AI-assisted coding, testing, and maintenance, but usually focuses

on engineering practices rather than strategic product choices.

The purpose of the article is to develop a theoretical model of decision-making for implementing AI technologies in software products that integrates decision theory, organizational research on AI implementation, responsible AI governance, and data-centric product management. Three research tasks are formulated:

1) To systematize theoretical approaches to AI-related decision-making at individual, team, and organizational levels relevant for software product development.

2) To identify determinants of organizational readiness for AI implementation in software products, including digital maturity, data quality, and governance structures, through synthesis of recent empirical and conceptual studies.

3) To construct an integrated decision model for product leaders that links AI prototyping, experimentation, risk assessment, and responsible deployment across the product lifecycle.

Scientific novelty lies in focusing theoretical analysis specifically on the decision process of product leaders responsible for AI features, integrating organizational AI implementation frameworks with AI-driven software engineering and responsible AI governance, and reinterpreting them as a coherent decision-support structure for managing AI transformations in software products.

Contribution

This study contributes to research on artificial intelligence and information systems in three ways. First, it systematizes theoretical perspectives on AI-related decision-making and explicitly adapts them to the context of software product implementation, an area underdeveloped in the existing literature. Second, it integrates organizational AI adoption frameworks, AI-driven software engineering research, and responsible AI governance into a unified, multi-level decision model for product leaders. Third, it reframes AI implementation decisions as an ongoing product lifecycle process rather than a one-time technological choice, highlighting the roles of data readiness, digital maturity, and governance in sustaining AI-enabled software products.

II. MATERIALS AND METHODS

The theoretical basis of the study comprises ten scientific sources published between 2021 and 2025, selected through targeted searches in major academic databases (Scopus, Web of Science, ScienceDirect, MDPI, and others), using combinations of keywords related to artificial intelligence, decision-making, software engineering, organizational adoption, and governance.

M. Alenezi and M. Akour [1] analyze AI-driven innovations across the software engineering lifecycle and describe how AI tools support code generation, debugging, testing, and maintenance, emphasizing their influence on productivity, code quality, and development speed in professional environments. A. Al-Surmi, M. Bashiri, and I. Koliousis [2] develop an AI-based decision-making framework that combines marketing and IT strategies, using artificial neural networks to explore optimal strategic configurations and their impact on operational performance. G. Amoako and co-authors [3] propose a conceptual framework where AI systems affect entrepreneurial decision quality through customer preferences, industry benchmarks, and employee involvement, treating AI as a mediator of better strategic choices in emerging markets.

K. Al-Bukhaiti and co-authors [4] present a comprehensive review of decision-making in the age of AI, covering theoretical decision frameworks, multi-criteria decision-making methods, and computational tools, and analysing human–AI collaboration in complex decisions. K. Kovič and co-authors [5] examine the adoption of AI software in European manufacturing companies and show that the level of Industry 4.0 readiness and digital infrastructure strongly influences AI implementation. At the same time, basic firm characteristics play a minor role. M.C.M. Lee and colleagues [6] conduct a systematic literature review of AI implementation in organizations, identify themes across organizational, technological, information systems, and people-related dimensions, and propose a conceptual framework of AI implementation antecedents, challenges, and outcomes.

O. Neumann and co-authors [7] investigate AI adoption in public organizations using comparative

case studies, highlighting governance, capabilities, and public value orientation as determinants of implementation decisions. An OECD report [8] characterizes AI adoption in firms across OECD and partner countries, documenting cross-country differences in the share of enterprises using AI and linking adoption to firm size, sector, and digital capabilities. E. Papagiannidis, P. Mikale, and K. Conboy [9] conduct a scoping review of responsible AI and propose a framework for responsible AI governance structured around structural, relational, and procedural practices across the AI lifecycle. B. Sattari and co-authors [10] design a theoretical framework for data-driven AI decision-making to enhance asset integrity management in the oil and gas sector, combining expert knowledge and AI/ML techniques to identify critical determinants in safety-critical infrastructures.

These materials collectively provide theoretical and empirical insights into AI-supported decision-making, AI adoption in organizations, software engineering practices, governance requirements, and sector-specific risk management, which are reinterpreted in the article from the viewpoint of product leaders making implementation decisions for AI features in software products.

The research methodology relies on a narrative and comparative review of the selected sources, structured thematic coding of decision-related concepts, and conceptual modelling. A comparative analysis is applied to identify convergences and differences among organizational AI implementation frameworks, AI-driven software engineering practices, and responsible AI governance models. The findings are synthesized into a multi-level decision model using methods of conceptual abstraction, typologization of decision criteria, and analytic generalization. The analysis focuses on extracting decision constructs relevant to product management in software companies, including evaluating AI opportunities, assessing readiness conditions, considering risk and governance considerations, and balancing accelerated prototyping with long-term product sustainability.

III. RESULTS

The theoretical analysis indicates that decision-making regarding AI implementation in software products follows a multi-level structure, combining individual, team, organizational, and ecosystem dimensions. The general decision-making frameworks summarized by K. Al-Bukhaiti and co-authors [4] indicate that contemporary decision theory is shifting from purely rational–calculative models toward integrative approaches, where multi-criteria optimization, bounded rationality, and human–AI collaboration coexist. In the context of product management, this means that choices about AI features cannot be reduced to a single metric. Product leaders simultaneously weigh customer value, data availability, technical feasibility, organizational capabilities, regulatory exposure, and long-term maintainability, while interacting with AI tools that both inform and constrain their judgments.

G. Amoako and co-authors [3] emphasize that AI-enhanced decision-making in entrepreneurial settings depends on how AI systems interact with customer preference signals, industry benchmarks, and employee involvement. Translated into software product practice, this suggests that AI functionalities in products should not be evaluated solely by internal efficiency or novelty, but by their capacity to surface reliable customer insights, position the product against industry standards and competitors, and mobilize cross-functional teams around shared evidence. When product leaders interpret model outputs alongside behavioral data and user research, AI becomes not a replacement for product sense but an amplifier of structured learning about markets and customers.

The organizational literature on AI implementation offers a structured framework for assessing readiness and constraints. The systematic review by M.C.M. Lee and colleagues [6] identifies organizational, technological, information systems, and people-related dimensions that shape AI implementation, including strategic alignment, governance structures, IT architecture, data infrastructure, skills, and change management mechanisms. In parallel, K. Kovič and co-authors [5] empirically show that the decisive factor for AI software adoption in manufacturing is not company size or position in the supply chain, but Industry 4.0

readiness and the presence of integrated digital infrastructures. For product leaders in software companies, these findings signal that decisions on AI features have to be embedded in a broader assessment of organizational digital maturity: whether data pipelines are trustworthy, whether deployment and monitoring processes can accommodate AI models, and whether the team has the competencies to operate and refine AI-enabled functionality over time.

Evidence from OECD enterprise surveys, synthesized in the 2025 report [8], provides an empirical backdrop for these theoretical considerations. The data show that, in 2023, the share of enterprises using at least one AI technology varies significantly across European countries, with Nordic economies achieving notably higher adoption rates than the EU average, while Southern European countries lag. This dispersion reflects differences in digital infrastructure, skills, and strategic focus, underscoring that AI implementation decisions depend on the surrounding ecosystem, including the availability of cloud services, regulatory expectations, and customer readiness. For software vendors operating in multiple markets, such heterogeneity directly influences which AI features are worth building, which must remain optional, and where the commercial return on AI investment is realistic.

Figure 1 illustrates that AI adoption across enterprises remains uneven despite overall growth. This heterogeneity reinforces the need for product-level decision frameworks that account for customer readiness and market context. AI features that deliver value in digitally mature environments may yield limited benefits or operational friction in less prepared segments, making selective, staged implementation strategies necessary.

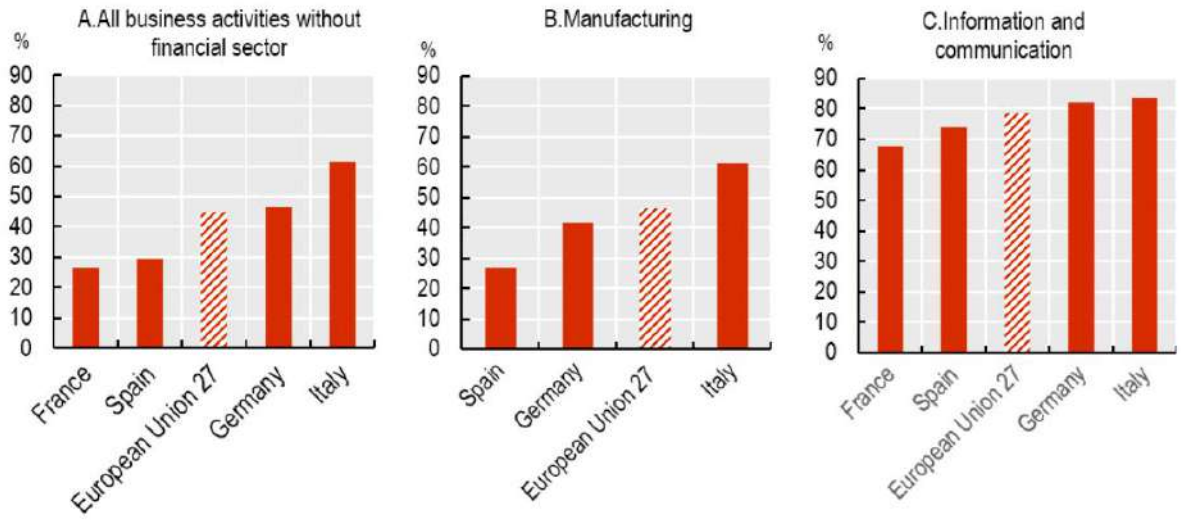


Fig.1. Share of enterprises that use at least one AI technology in selected European countries (2023), adapted from OECD [8].

AI adoption across enterprises is growing but remains uneven, underscoring the need for explicit decision frameworks at the product level. Where client organizations have low digital readiness, AI-heavy features may create more friction than value; in digitally mature client segments, the absence of AI capabilities can undermine competitiveness. Product leaders therefore face a portfolio of decisions: which AI opportunities to pursue for advanced customers, what minimum AI functionality is required to stay credible, and when to postpone AI implementation in favor of strengthening underlying data and workflow capabilities.

Studies focusing on AI-driven software engineering add another dimension to the field. M. Alenezi and M. Akour [1] demonstrate that AI tools already affect many phases of the software development lifecycle: code generation by large language models, intelligent debugging, automated test generation, refactoring recommendations, and predictive estimates of effort. These capabilities enable accelerated prototyping and AI-enabled code

generation, where high-fidelity prototypes emerge from natural-language prompts in days rather than weeks. For decision-making, this changes the cost structure of experimentation: product managers can validate user flows, messaging, and initial interactions with minimal engineering effort. At the same time, the review highlights unresolved issues of trustworthiness, explainability, and integration across tools. From a theoretical standpoint, AI-assisted prototyping moves some decisions from the realm of abstract planning to empirically observable behavior much earlier. Still, it does not eliminate the need for rigorous evaluation before features are advanced into robust production services.

Operational research on AI-based decision-making provides concrete mechanisms for structuring complex choices. A. Al-Surmi, M. Bashiri, and I. Koliousis [2] propose a three-phase framework where AI models, built on artificial neural networks, explore performance implications of different combinations of marketing and IT strategies and support operations managers in selecting optimal strategies. Their work demonstrates how AI can be integrated into the decision-making process itself, not just as a feature in products. Applied to software product strategy, such models can support simulation of alternative AI feature roadmaps, pricing schemes, or levels of automation, helping product leaders understand performance trade-offs before committing engineering resources.

Sector-specific research in high-risk environments illustrates how theoretical AI decision frameworks operate under stringent reliability and safety constraints. B. Sattari and co-authors [10] design a data-driven framework for AI-supported decisions in asset integrity management, combining expert knowledge, machine learning, and keyword analysis to construct reaction networks and identify the most influential determinants of integrity in oil and gas systems. The central conclusion is the necessity of systematically linking AI outputs to domain knowledge, explicit risk tolerances, and inspection procedures. Translated to software products, particularly in regulated domains such as finance, healthcare, or marketing with strict compliance requirements, AI features must be justified not only by predictive accuracy but by their integration into controlled workflows, monitoring, and human oversight.

Responsible AI governance research adds a normative layer to decision-making. E. Papagiannidis, P. Mikalef, and K. Conboy [9] distinguish structural, relational, and procedural governance practices that operationalize responsible AI principles across the AI lifecycle. Structural practices concern roles, committees, and policies; relational practices govern stakeholder engagement and transparency; procedural practices regulate how models are designed, validated, deployed, and monitored. For product leaders, this framework implies that AI implementation decisions require explicit consideration of the following responsibility dimensions: fairness, transparency, robustness, accountability, and societal impact. AI features that accelerate user workflows but undermine transparency or create opaque dependencies on external models introduce hidden liabilities that product roadmaps must recognize.

Public-sector evidence from O. Neumann and co-authors [7] indicates that governance, capabilities, and public value orientation strongly influence AI adoption in public organizations. Their comparative case study reveals that even when technical AI solutions are available, decision-makers postpone or scale back implementation when interpretability is lacking, data quality is insufficient, or organizational mandates stress accountability and citizen trust. For software suppliers whose products serve both private and public clients, these findings underline that AI implementation decisions must be filtered through the lens of client governance expectations, not just technical feasibility or market hype.

Integrating the reviewed strands, a conceptual decision model for AI implementation in software products can be outlined. At the strategic level, product leaders formulate AI-related objectives, including differentiation, cost efficiency, personalization, and risk management. At the discovery level, AI-driven prototyping and AI-enabled code generation, as described by M. Alenezi and M. Akour [1], generate candidate solutions and interaction patterns with minimal investment. At the evaluation level, decision-makers apply multi-criteria frameworks inspired by K. Al-Bukhaiti and co-authors [4] and A. Al-Surmi et al. [2], balancing user

value, technical feasibility, data readiness, compliance, and governance requirements. At the organizational level, the conditions identified by M.C.M. Lee et al. [6], K. Kovič et al. [5], and the OECD [8] define whether the firm and its customers are ready to sustain AI features over time. At the governance level, structural, relational, and procedural practices, as described by E. Papagiannidis and colleagues [9], determine the acceptable configurations of AI functionality in relation to trust and responsibility.

To make the synthesized decision logic operational and readable for product teams, the proposed model is presented schematically (Figure 2). The diagram separates decision layers across the product lifecycle and makes explicit the information flows linking strategic intent, AI-enabled discovery, evaluative gates, organizational readiness constraints, governance controls, and production feedback loops.

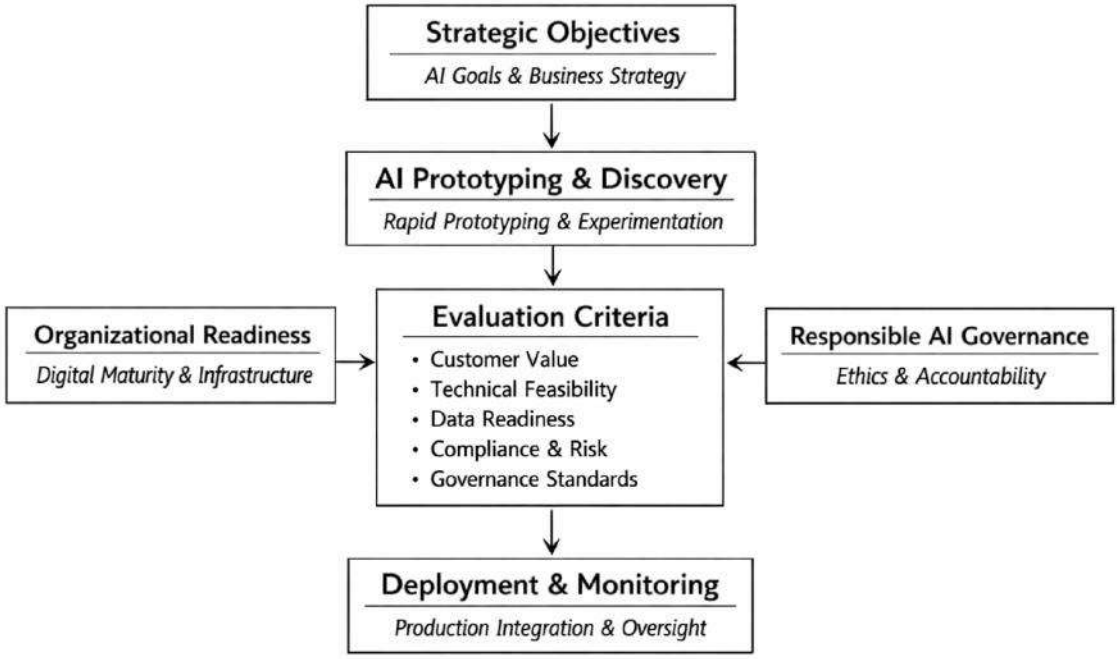


Fig.2. Multi-level decision model for implementing AI technologies in software products (conceptual synthesis based on [1; 2; 4–6; 8–10]).

Within this model, data quality and data-centric product management occupy a central position. Without reliable, well-understood data, AI features in software products degenerate into unstable prototypes that behave inconsistently across customers and environments. The frameworks of B. Sattari et al. [10] demonstrate that in safety-critical contexts, decisions rest on explicitly modeled relationships between variables and controlled data flows; an analogous discipline is required in business products whose decisions affect customers' finances, marketing outcomes, or reputations. For product leaders, investing in monitoring pipelines, data contracts, and feedback loops becomes a precondition for sustainable AI implementation, even if such work does not immediately yield visible interface changes.

In summary, the literature's theoretical foundations converge on a view of AI implementation decisions as structured, multi-level, and deeply intertwined with organizational maturity, governance, and data-centric practices. AI-generated prototypes accelerate discovery, but decisions on which prototypes evolve into trustworthy, scalable features depend on a layered evaluation of value, risk, and feasibility that extends beyond the surface allure of prototyping AI development.

IV. DISCUSSION

The synthesized decision model reveals several tensions that product leaders must address when considering the implementation of AI in

software products. The first tension lies between the speed of prototyping and the depth of evaluation. AI-driven tools, as described by M. Alenezi and M. Akour [1], reduce the marginal cost of new ideas. Prompt-based prototypes produced via AI-enabled code generation support early demonstrations and stakeholder engagement. At the same time, organizational research indicates that successful AI implementation relies on digital maturity, governance, and data quality, which evolve slowly and necessitate deliberate investment. This asymmetry leads to an overproduction of prototypes relative to the organization's ability to industrialize them. A structured decision framework helps distinguish experiments intended to inform learning from those intended for production, preventing the product portfolio from becoming cluttered with fragile AI components.

A second tension concerns the locus of intelligence in decision-making. Multi-criteria and AI-based decision methods discussed by K. Al-Bukhaiti et al. [4] and A. Al-Surmi et al. [2] show that AI can support configuration of strategies and evaluation of options, yet do not eliminate the need for human interpretation, especially when criteria conflict or when data are incomplete. In product management practice, this means that algorithmic scoring of AI opportunities — based on estimated impact, confidence, effort, and risk — must remain subordinate to structured conversation within cross-functional teams. Human judgment connects AI-generated insights with qualitative knowledge of customers, markets, and organizational constraints.

To make the synthesis explicit, Table 1 brings together the main theoretical lenses used in the sources and their implications for product-level AI implementation decisions.

Table 1. Theoretical lenses relevant for AI implementation decisions in software products (based on [1–4; 6; 9])

Theoretical lens	Core focus for AI decisions	Implications for product leaders
Multi-criteria decision-making and optimization	Structuring trade-offs across multiple quantitative and qualitative criteria	Use scoring models and scenario analysis to compare AI feature options under uncertainty and competing goals.
Organizational implementation frameworks	Antecedents, challenges, and outcomes of AI deployment in organizations	Evaluate AI features against the firm's and its customers' digital maturity, skills, processes, and change readiness.
AI-driven software engineering	Automation of coding, testing, and maintenance through AI tools	Treat AI tools as accelerators of discovery and refactoring, not substitutes for architectural decisions.
Entrepreneurial decision-making	Influence of AI systems on opportunity evaluation and strategic choices	Align AI features with customer preferences, industry benchmarks, and internal stakeholder engagement.
Responsible AI governance	Principles and practices for ethical and accountable AI	Embed governance gates into product decision processes for AI features across the lifecycle.

Organizational implementation research offers insight into readiness and obstacles, but is largely agnostic to the specifics of software product lifecycles. AI-driven software engineering research excels at describing tooling and automation, but says little about product strategy or governance. Responsible AI frameworks define "how to behave" with AI, rather than "what to build" in specific

products. Only by combining these lenses can product leaders produce decisions that are both theoretically grounded and practically actionable.

Empirical findings on AI adoption further refine the decision space. K. Kovič and co-authors [5] show low but growing AI software usage in manufacturing, with adoption strongly associated with Industry 4.0 readiness, while OECD data [8] reveal heterogeneous adoption rates across countries and sectors. For software vendors building AI-enabled products, these results suggest that a single, globally uniform AI roadmap is rarely optimal. Instead, decision-making must account for clusters of customers with different readiness profiles: some prioritize advanced AI capabilities, while others prefer stability and transparent workflows.

To clarify the connection between theoretical categories and concrete decision criteria in software product management, Table 2 summarizes key groups of criteria for AI implementation decisions and links them to evidence from the considered sources.

Table 2. Decision criteria for implementing AI functionalities in software products across the product lifecycle (based on [1–10])

Criterion group	Typical evaluative questions for product decisions	Representative evidence
Customer value and behavior	Does the AI feature solve a recurring customer problem, improve outcomes, or meaningfully reduce effort relative to alternatives?	Entrepreneurial decision models; AI-enabled operational performance.
Data readiness and quality	Are data sources complete, reliable, and stable over time? Are labeling, drift, and feedback loops manageable with existing resources?	Asset integrity frameworks; organizational implementation themes.
Digital and organizational maturity	Do engineering, DevOps, and MLOps practices support secure deployment, monitoring, and rollback of AI components?	Industry 4.0 readiness and AI adoption; implementation review; public sector adoption.
Governance and risk	Are fairness, transparency, privacy, and compliance requirements satisfied? Are accountability and escalation paths defined?	Responsible AI governance; public sector AI cases.
Experimentation and learning	Can the AI feature be evaluated through experiments, A/B tests, or quasi-experiments with clear success metrics and guardrails?	Decision-making frameworks; AI-based decision models; AI in software engineering.
Ecosystem and market conditions	Do target markets and client organizations have sufficient AI adoption, skills, and regulatory clarity to use and trust the feature?	OECD enterprise adoption data; sectoral readiness analyses.

Many AI ideas that appear attractive in demos fail when evaluated against data readiness, governance, or ecosystem conditions. Conversely, certain less spectacular AI functions — for example, invisible optimization of internal workflows or data quality — may score highly on sustainability and risk criteria, making them strong candidates for prioritization even if they are less visible in marketing materials.

From a product leadership perspective, the findings support several practical interpretations. First, AI-generated prototyping should be treated as an extension of product discovery rather than as evidence of implementation readiness. Prototypes

help refine understanding of the problem, language, and user flows, but theoretical and empirical work on AI implementation cautions against equating prototype success with operational viability. Second, data-centric product management emerges as a structural prerequisite for AI strategies. Studies on asset integrity [10] and responsible governance [9] show that robust decisions depend on traceable, explainable data flows; in product terms, investment into telemetry, event models, and monitoring must precede large-scale AI rollouts. Third, decision-making on AI features benefits from explicit governance checkpoints, such as review boards or decision templates, that require product teams to articulate problem statements, expected impact, risk mitigation plans, and monitoring schemes before committing scarce engineering capacity.

Ultimately, the reviewed literature suggests that responsible AI and competitive advantage are not mutually exclusive. E. Papagiannidis and co-authors [9] argue that structural, relational, and procedural governance practices can be designed to support innovation rather than suppress it. When product leaders incorporate these practices into their decision frameworks, they establish a repeatable process for evaluating AI opportunities, including initial market discovery with AI prototypes, data readiness assessment, structured multi-criteria evaluation, governance review, and staged rollout with ongoing monitoring. For organizations experiencing transformation fatigue, where employees are subjected to continuous waves of "AI initiatives," a transparent, theory-informed decision-making process helps filter noise, reduce random experimentation, and focus attention on AI features that support coherent product strategies and trustworthy customer relationships.

Limitations and Directions for Future Research. This study is conceptual and does not empirically test the proposed decision model. While grounded in recent empirical and review-based research, the framework has not been validated through case studies, surveys, or longitudinal analyses of software product organizations. In addition, the focus on software and SaaS products may limit transferability to hardware-intensive or highly regulated embedded systems. Future research could empirically examine decision-making processes surrounding AI features, assess the relative weight of different decision criteria, and explore how governance and data maturity influence AI success across product portfolios.

V. CONCLUSION

This research reconstructs and integrates theoretical foundations for decision-making on the implementation of artificial intelligence technologies in software products.

Based on ten recent scientific sources, a multi-level decision model is synthesized that links individual and organizational decision frameworks, AI-driven software engineering tools, data-centric thinking, and responsible AI governance. The model views AI-generated prototyping as a powerful tool for discovery, but assigns significant weight to data readiness, digital maturity, governance practices, and ecosystem conditions when selecting AI features for industrialization.

The first research task is addressed through the systematization of decision-making frameworks, AI implementation studies, AI-based decision models, and responsible governance concepts, which collectively demonstrate that AI decisions cannot be reduced to isolated technical choices but instead depend on a structured evaluation of multiple criteria. The second task is to identify the determinants of organizational readiness, including Industry 4.0 maturity, architectural and data infrastructure, skills, and public value orientation, which condition the feasibility and sustainability of AI in software products. The third task leads to an integrated decision model for product leaders that filters AI ideas through customer value, data quality, governance, and ecosystem readiness, supported by explicit experimentation and learning mechanisms.

For product leaders and software engineering managers, the results suggest a practical trajectory: treat AI prototypes and AI-enabled code generation outputs as inputs into structured decision processes; build data and governance foundations ahead of large-scale AI integration; calibrate AI roadmaps to client readiness; and institutionalize multi-criteria and responsible AI perspectives in product decision forums. For researchers, the proposed model opens a path to empirical studies of decision processes

surrounding AI features in real product organizations, including the measurement of how governance, data quality, and digital maturity influence the success of AI implementations within software portfolios.

ACKNOWLEDGMENT

No external funding was received for this study.

REFERENCES

[1] Alenezi, M., & Akour, M. (2025). AI-driven innovations in software engineering: A review of current practices and future directions. Applied Sciences, 15(3), 1344. https://doi.org/10.3390/app15031344

[2] Al-Surmi, A., Bashiri, M., & Koliousis, I. (2021). AI based decision making: Combining strategies to improve operational performance. International Journal of Production Research, 60, 1–23. https://doi.org/10.1080/00207543.2021.1966540

[3] Amoako, G., Omari, P., Kumi, D. K., Agbemabiase, G. C., & Asamoah, G. (2021). Conceptual framework—Artificial intelligence and better entrepreneurial decision-making: The influence of customer preference, industry benchmark, and employee involvement in an emerging market. Journal of Risk and Financial Management, 14(12), 604. https://doi.org/10.3390/jrfm14120604

[4] Wei, J., Qi, S., Wang, W., Jiang, L., Gao, H., Zhao, F., Al-Bukhaiti, K., & Wan, A. (2025). Decision-making in the age of AI: A review of theoretical frameworks, computational tools, and human-machine collaboration. Contemp. Math., 6, 2089–2112. https://doi.org/10.37256/cm.6220256459

[5] Kovič, K., Tominc, P., Prester, J., & Palčič, I. (2024). Artificial intelligence software adoption in manufacturing companies. Applied Sciences, 14(16), 6959. https://doi.org/10.3390/app14166959

[6] Lee, M. C. M., Scheepers, H., Lui, A. K. H., & Ngai, E. W. T. (2023). The implementation of artificial intelligence in organizations: A systematic literature review. Information and Management, 60(5), 103816. https://doi.org/10.1016/j.im.2023.103816

[7] Neumann, O., Guirguis, K., & Steiner, R. (2024). Exploring artificial intelligence adoption in public organizations: A comparative case study. Public Management Review, 26(1), 114–141. https://doi.org/10.1080/14719037.2022.2048685

[8] OECD/BCG/INSEAD. (2025). The adoption of artificial intelligence in firms: New evidence for policymaking. OECD Publishing. https://doi.org/10.1787/f9ef33c3-en

[9] Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. The Journal of Strategic Information Systems, 34(2), 101885. https://doi.org/10.1016/j.jsis.2024.101885

[10] Sattari, B., et al. (2022). A theoretical framework for data-driven artificial intelligence decision making for enhancing the asset integrity management system in the oil and gas sector. Journal of Petroleum Science and Engineering, 208, 109295. https://doi.org/10.1016/j.jlp.2021.104648

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.2

Fuzzy PID Control Performance Analysis for Robotic Arm System Research

Zhi-Jian Tan, Ho-Sheng Chen*

School of Sciences, Guangdong University of Petrochem Technology (GDUPT), Maoming 525000, China
*Corresponding author: hschen98.tw@gmail.com

Received: 25 Jan 2026; Received in revised form: 22 Feb 2026; Accepted:26 Feb 2026; Available online: 03 Mar 2026

Abstract *— The need for robotic arm control in the realm of industrial automation is growing as technology progresses. In view of the challenges of speed step, external disturbance and joint friction leading to the loss in control accuracy in robotic arm trajectory tracking, the classic PID control system is difficult to meet the requirements of high-precision operation. In order to maximize the dynamic responsiveness, this research suggests a fuzzy PID control scheme [1]. In the MATLAB/simulink environment, the KUKA KR6R700 six-axis robotic arm serves as the simulation object. The results demonstrate that compared with the traditional PID control system, the fuzzy PID control system shortens the joint adjustment time by ≥43.72% (reduced by 2.37s), considerably improves the fit of the robotic arm joint trajectory, and has less trajectory deviation and smoother transition. Its capacity to retain great robustness and trajectory tracking even in challenging work situations is particularly impressive. According to research, the suggested approach can significantly increase robotic arms' motion control accuracy in automated production workshops.*

Keywords *— **Robotic arm, PID control, fuzzy control***

I. INTRODUCTION

Robotic arms, a key component of industrial automation, are indispensable in automated production lines because of their small size, manageable cost, and adaptable deployment. Robotic arm trajectory tracking has long been plagued by issues such speed step changes, external disturbances, and reduced control precision due to joint friction. However, their control performance directly influences their working accuracy. Conventional PID control methods are inadequate to deal with nonlinear elements such joint friction, unknown external disturbances, and model parameter changes. This can easily result in excessive overshoot, delayed response, trouble removing steady-state errors, and even oscillations. This study suggests a fuzzy PID control system as a workable and efficient option for high-precision, high-efficiency, and high-reliability motion control of robotic arms in automated production workplaces.

II. ESTABLISHMENT OF KINEMATIC MODEL AND MOTION TRAJECTORY OF ROBOTIC ARM

The kinematic model of the robotic arm is the basis for describing its position and posture changes. The KUKA KR6R700 six-axis robotic arm, which is commonly used in automated production workshops, is taken as the research object. [2] This robot has a serial robotic arm with six degrees of freedom. The structure includes a base joint, waist joint, upper arm joint, forearm joint, wrist rotation joint and wrist swing joint. The KR6R700 robotic arm has a load capacity of 6kg and a repeatability of ±0.03 mm, which is suitable for precision assembly tasks in automated production workshops. The structural parameters of the robotic arm are shown in Table 1.

Table 1. Structural parameters of robotic arm

Parameter	Wrist swing joint	Wrist rotation joint	Forearm joint	Upper arm joint	Lumbar joint	Base joint
Maximum rotation Angle/(◦)	±360	±360	±130	±150	±185	±360
Member length/mm	85	90	210	480	110	0

The robotic arm consists of multiple links and joints. The movement of each joint can be described by the joint angle, as shown in Figure 1.

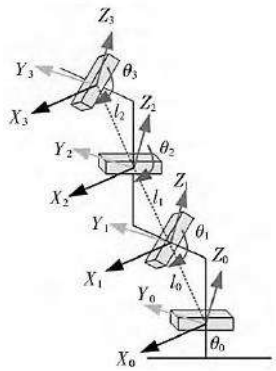


Fig.1. Kinematic model of the robotic arm

Let the link length of the robotic arm is $l = \{l_0, l_1, l_2, \dots, l_n\}$, the Joint angle is $\theta = \{\theta_0, \theta_1, \theta_2, \dots, \theta_n\}$, The position of the end effector that controls the movement of the robotic arm $\{x, y, z\}$ can be calculated using kinematic equations, as shown in equation (1).

$$\begin{cases} x = l_0 \cos\theta_0 + l_1 \cos(\theta_0 + \theta_1) + \cdots + \\ \qquad l_n \cos(\theta_0 + \theta_1 + \ldots + \theta_n \\ y = l_0 \sin\theta_0 + l_1 \sin(\theta_0 + \theta_1) + \cdots + \\ \qquad l_n \sin(\theta_0 + \theta_1 + \ldots + \theta_n) \\ \qquad z = \lambda_0 + \lambda_1 + \ldots + \lambda_n \end{cases} \quad (1)$$

where $(\lambda_0, \lambda_1, \ldots \lambda_n)$ represents the offset (cm) of the robotic arm link in the Z-axis direction.

To achieve precise motion control of the robotic arm, its motion trajectory needs to be planned. The motion trajectory can be divided into three stages: the initial acceleration stage, the intermediate constant speed stage, and the final deceleration stage. Assuming the total motion time of the robotic arm is T, and the acceleration time of the initial and final stages is t_1, then in the initial stage, the motion trajectory of the serial robotic arm is a parabola with constant angular acceleration. The change law of the joint angle $\theta(t)$ can be expressed as equation (2).

$$\theta(t) = \theta_0 + \frac{at^2}{2}, (0 \le t \le t1) \quad (2)$$

where θ_0 is the initial joint angle and a is the angular acceleration.

The change law of the joint angle is given by equation (3).

$$\theta(t) = \theta(t_1) + \varepsilon(t - t_1), (t_1 < t < T - t_1) \quad (3)$$

where ε is the angular velocity during the uniform velocity phase.

The robotic arm descends to the target point along a parabolic trajectory, and the change in joint angles follows the formula (4).

$$\theta(t) = \theta(T - t_1) + \varepsilon(T - t_1) - \frac{\alpha(T - t_1)^2}{2}, \quad (T - t_1 < t < T) \quad (4)$$

By planning the motion trajectory as described above, the speed jump during the transition from acceleration to constant speed in the robotic arm can be eliminated. This allows for smoother movement of the robotic arm and reduces trajectory tracking errors.

Based on the above analysis of the motion trajectory and modeling of the robotic arm joint drive system, the input angle $\theta(s)$ and output angle $U(s)$ of the robotic arm joint motion are calculated, and the transfer function formula (5) is established.

$$G(s) = \frac{U(s)}{\theta(s)} = \frac{s + 5}{s^2 + 4s + 5} \quad (5)$$

where $U(s)$ is the Laplace transform of the joint output angle and $\theta(s)$ is the Laplace transform of the joint angle input.

We can obtain the undamped natural frequency of the system is $\omega_n = \sqrt{5} \approx \frac{2.236rad}{s}$ and the damping ratio is $\xi = \frac{2}{\sqrt{5}} \approx 0.894$. Its control system is a minimum-phase second-order underdamped system.

Dynamic performance analysis shows that the system has excellent transient response characteristics: extremely small overshoot. After zero-point correction, it is about 0.5%-1%), the rise time is t_r≈2.6 s, the settling time is t_s (±5% error band) about 1.5 s, the convergence speed is fast and the operation is stable; however, since the system is a zero-type system (without an integral element), there is a steady-state error of e_{ss}=0.5 for a unit step input. The frequency domain analysis results are shown in Figures 2 and 3.

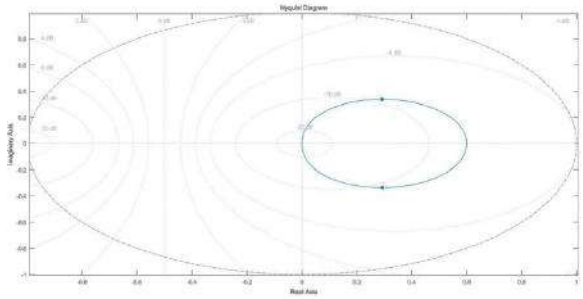


Fig.2. Nyquist Diagram

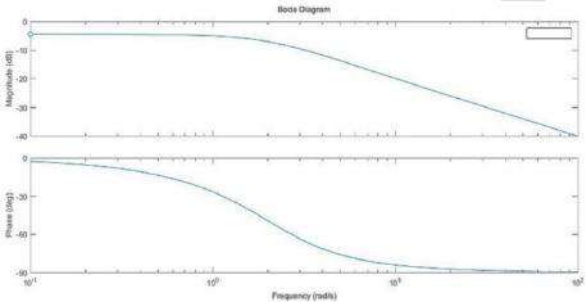


Fig.3. Bode Diagram

The system amplitude expression is as $|G(j\omega)| = \frac{\sqrt{\omega^2+25}}{\sqrt{\omega^4+6\omega^2+25}}$ and phase angle is $\angle G(j\omega) = \tan^{-1}\left(\frac{\omega}{5}\right) - \tan^{-1}\left(\frac{4\omega}{5-\omega^2}\right)$. Its amplitude-frequency response exhibits monotonically decaying across the entire frequency range with an amplitude consistently ≤0dB, no resonant peak (due to $\xi > 0.707$), a phase margin of 180°, and an amplitude margin of GM = ∞. Furthermore, the closed-loop characteristic curve is entirely located in the left half-plane. This indicates that the system possesses absolute stability and high resistance to high-frequency interference. Overall, this system demonstrates outstanding dynamic stability and is suitable for applications requiring moderate tracking accuracy from the robotic arm while prioritizing operational stability and interference resistance.

III. TRADITIONAL PID CONTROL SYSTEM PRINCIPLE

PID control is the most common classical control method in control systems. [3] It uses the set value and the actual output value to form the control deviation, and combines the deviation with proportional (P), integral (I) and derivative (D) to form the control quantity, so as to control the controlled object and make its system output accurately and stably track the target value. Due to its simple structure, good robustness and no need for an accurate system mathematical model, PID control is widely used in industrial automation and other fields. The working principle of the PID control system is shown in Figure 4.

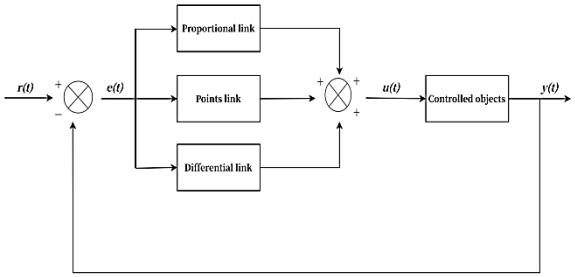


Fig.4. Working principle of PID control system

The expression for the traditional PID control as.

$$u(t) = K_p e(t) + k_i \int e(t)d(t) + k_d \frac{\mathrm{d}e(t)}{\mathrm{d}t} \qquad (6)$$

where K_P is the proportional coefficient, K_I is the integral coefficient, K_D is the derivative coefficient, $e(t)$ is the error between the set value and the actual output value, and $u(t)$ is the controller output.

IV. TRADITIONAL PID CONTROL SYSTEM PRINCIPLE

Fuzzy PID control algorithm is a new type of control method that combines traditional PID control with fuzzy logic theory [4]. It combines the reliability of traditional PID control with the advantages of fuzzy logic theory in dealing with uncertainty and nonlinear problems. Its control system can dynamically adjust the PID parameters according to

the real-time running state of the controlled object, thereby effectively overcoming the limitations of traditional PID control in dealing with complex systems such as nonlinearity, time-varying and large delay. The core of the algorithm is to adjust the proportional (K_P), integral (K_I) and derivative (K_D) coefficients online adaptively according to the real-time error (e) and error change rate (ec) of the system using a set of pre-designed fuzzy control rules, so that the system can obtain good dynamic performance with small overshoot, fast response and strong robustness. The working principle of the fuzzy PID control system is shown in Figure 5.

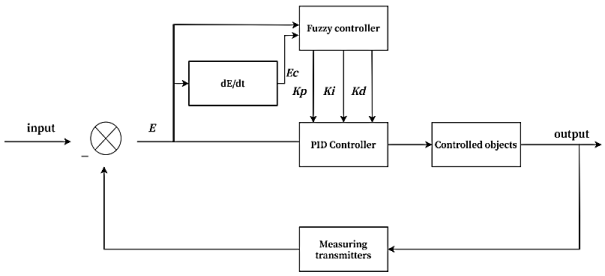


Fig.5. Working principle of fuzzy PID control system

The fuzzy controller mainly consists of three modules: fuzzification, fuzzy inference, and defuzzification as Figure 6.

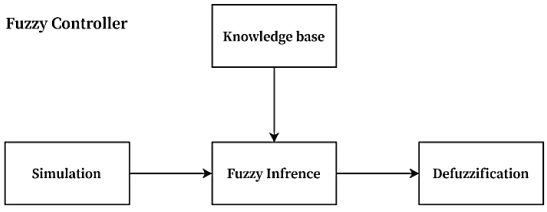


Fig.6. Fuzzy rule base structure framework diagram

As shown in Figures 5 and 6, the system introduces a fuzzy control rule based on the traditional PID control loop. The rule takes the system deviation e and its rate of change ec as inputs, performs fuzzy inference calculations to output the PID parameter corrections (ΔK_p, ΔK_i, and ΔK_d) as shown in Equation (7), thereby achieving online automatic optimization of PID parameters.

$$\begin{cases} \Delta K_p = K_p + K_{pp} \\ \Delta K_i = K_i + K_{ii} \\ \Delta K_d = K_d + K_{dd} \end{cases} \quad (7)$$

In the fuzzy rule setup, the membership functions of input variables all adopt smoothly curved Gaussian functions, with the fuzzy linguistic values divided into five levels: {Negative Big (NB), Negative Medium (N), Zero (ZE), Positive Small (P), Positive Big (PB)}. The output variables are the adjustment amounts $\triangle K_p$, $\triangle K_d$, and $\triangle K_i$ for the three PID parameters. To ensure computational simplicity and output clarity, the membership functions for output variables employ simple triangular functions, with fuzzy linguistic values categorized into four levels: {Extremely Small (MS), Small (S), Medium (B), Extremely Big (MB)}. Fuzzy rules are established based on the fuzzy rule table, as shown in Table (1). The system dynamically adjusts the PID parameters in real-time according to pre-defined fuzzy control rules to adapt to different operational states of the robotic arm.

V. CONTROL SYSTEM SIMULATION EXPERIMENT

To investigate the impact of PID controllers on the control system of robotic arms, a simulation model of joint angle changes in PID control of robotic arms was first constructed based on MATLAB/Simulink environment (as shown in Figure 8) for simulation experiments [5]. In this model, the KUKA KR6R700 industrial robot is taken as the research object, and the input angle of the robotic arm joint is set to 120°, while the output results reflect the angle changes of the robotic arm joint. Through the model and dynamic simulation process, the working principle of the PID controller and the impact of parameter adjustment on the system's dynamic performance can be intuitively understood.

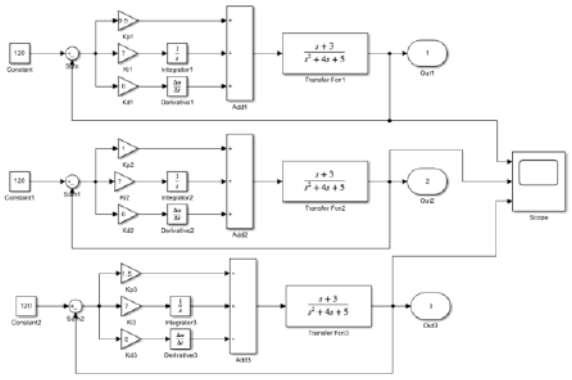


Fig.7. MatLab/Simulink

In the study of the influence of the comparative coefficient K_p on the dynamic characteristics of the robotic arm control system, control variables were set in SIMULINK simulation: integral coefficient K_i and differential coefficient K_d were set as initial values (K_i=6, K_d=0.4), the input angle of the robotic arm joint was set to 120 °, and the proportional coefficient K_p was set to three different values of 0.8, 1, and 1.2, respectively. The dynamic characteristic curves of the control system were analyzed under different proportional coefficients K_p. As shown in Figure 8. Within the selected range of K_p values, the output angle response of the robotic arm can gradually converge to the target angle of 120° after experiencing brief fluctuations, indicating overall stability of the system. Further analysis of the dynamic characteristics reveals that as the value of K_p increases, the maximum overshoot of the system shows a decreasing trend, and the response time required to reach steady state is correspondingly shortened, resulting in an improvement in response speed. This rule indicates that an appropriate increase in the proportional coefficient K_p can optimize the control system's regulation capability. In PID control, the integral and derivative stages are the core components for achieving "zero static error regulation" and "leading suppression of overshoot". The influence of the integral coefficient K_i and differential coefficient K_d of the PID controller on the dynamic characteristics of the robotic arm control system was studied. Simulation experiments were conducted using the control variable method: the input angle of the robotic arm was set to 120°, and the proportional parameter K_p=1. The integral parameter K_i and differential parameter K_d were adjusted separately to observe the changes in the dynamic characteristics of the control system. The corresponding simulation results are shown in Figures 8, 9 and 10

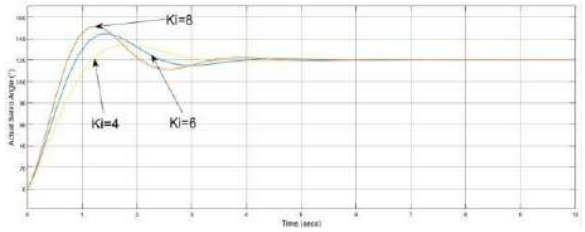


Fig.8. The Influence of Scale Coefficient K_p on the Dynamic Characteristics of Control System

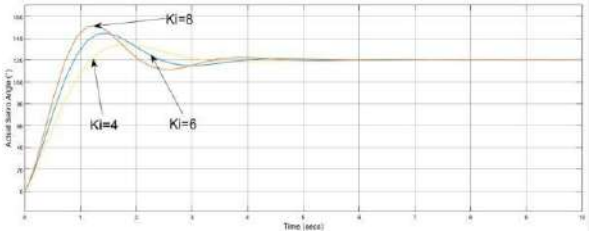


Fig.9 The Influence of Integral Coefficient K_i on the Dynamic Characteristics of the Control System

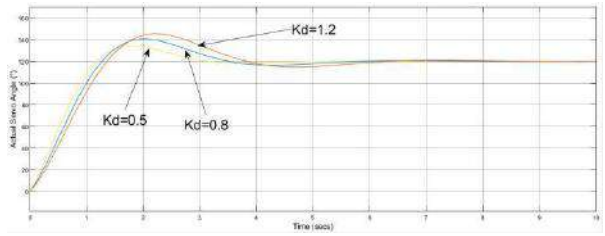


Fig.10. The Influence of Differential Coefficient K_d on the Dynamic Characteristics of Control System

As shown in Figure 9, when K_i is set to 4, 6, and 8, the control system can ultimately converge to the target angle of 120°. However, the dynamic performance shows a significant increase in maximum overshoot and adjustment time as Ki increases. The frequency of system oscillations increases, and the stability gradually weakens. From the perspective of control principles, the integration process eliminates static errors by accumulating deviations. However, if the value of K_i is too high, it will accelerate the accumulation rate of errors, increase the phase lag of the system, break the balance between "deviation adjustment response feedback", and cause oscillation instability. Therefore, K_i needs to seek a balance

between static error elimination ability and dynamic stability.

From Figure 10, it can be observed that as the differential coefficient K_d increases (with K_d values of 0.4, 0.8, and 1.2), the maximum overshoot of the system increases, the response speed slows down, and the adjustment time also increases accordingly. In the analysis of PID control principle, the differential coefficient K_d can mainly predict the future error, provide the response to the error change rate, play the role of advance regulation and enhance the stability of the system. However, if the differential coefficient K_d is too large, the system will become unstable and prone to introducing high-frequency noise. Therefore, the differential coefficient K_d should be appropriately selected.

Study the influence of PID controller and fuzzy PID controller on the dynamic characteristics of robotic arm control system. Based on MATLAB/Simulink environment, taking KUKA KR6R700 industrial robot as the research object, a traditional PID control simulation model and a fuzzy PID control simulation model of the robotic arm control system are established. The joint input angle of the robotic arm is set to 120°, and the controller parameters are set as proportional coefficient (K_p=0.8), integral coefficient (K_i=8), and differential coefficient (K_d=1.2). A dynamic performance comparison analysis of the two control strategies is carried out.

In traditional PID control models, the research object is directly controlled through the proportional integral derivative process. The fuzzy PID control model introduces a fuzzy logic controller with "e" and "ec" as inputs. After fuzzification, fuzzy inference, and deblurring, the parameters K_p, K_i, and K_d are adaptively adjusted online to achieve dynamic optimization control of the joint angle changes of the robotic arm. Comparison simulation model between PID control and fuzzy PID control, as shown in Figure 11; Comparison of system dynamic performance between fuzzy PID control and PID control, as shown in Figure 12

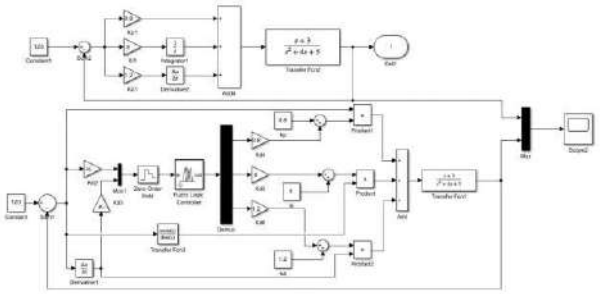


Fig.11. Comparison simulation model between PID and fuzzy PID control

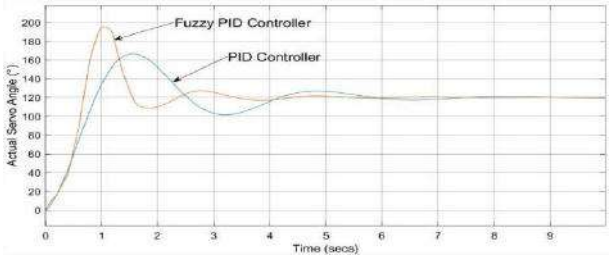


Fig.12. The Influence of PID Control and Fuzzy PID Control on the Dynamic Characteristics of the System

From the perspective of control principles, PID control systems have fixed parameters, which limits their adaptability to controlled objects with nonlinearity and time-varying characteristics such as robotic arms. They are prone to problems such as overshoot, large amplitude, or slow adjustment time; The fuzzy PID control system achieves real-time online adjustment of PID parameters through fuzzy rules, significantly improving the robustness and dynamic performance of the system. Experiments have shown that fuzzy PID control, with its parameter adaptive mechanism, effectively compensates for the shortcomings of PID control systems in nonlinear time-varying systems. Enable the robotic arm to have higher positioning accuracy, fast response speed, and stable dynamic performance in complex working environments. This provides a more advantageous control solution for precise trajectory tracking and operation of robotic arms.

Study the influence of fuzzy rule levels on the dynamic performance of robotic arm control systems, based on MATLAB/Simulink environment, using KUKA KR6R700 industrial robot as the research object. In the design of a fuzzy PID control system, error e and error change rate ec are used as inputs for the fuzzy PID, and ΔK_p, ΔK_i, and ΔK_d are used as

outputs. Fuzzy Logic Toolbox is used to construct 3, 5, and 7 level fuzzy inference systems (FIS) with 9, 25, and 49 rules, respectively. To eliminate the interference of membership function types on the dynamic performance of the control system, Gaussian membership functions are used for input variables and triangular membership functions are used for output variables, with only the density of language variable partitioning adjusted; The simulation model of the fuzzy PID control system for the robotic arm under different levels of fuzzy rules is shown in Figure 13; The influence of fuzzy rule levels on the dynamic performance of robotic arm control systems is shown in Figure 14

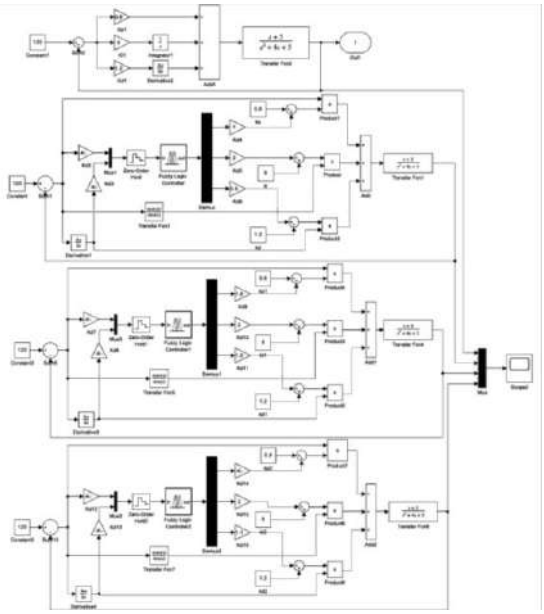


Fig.13. Simulation models of different levels of fuzzy PID

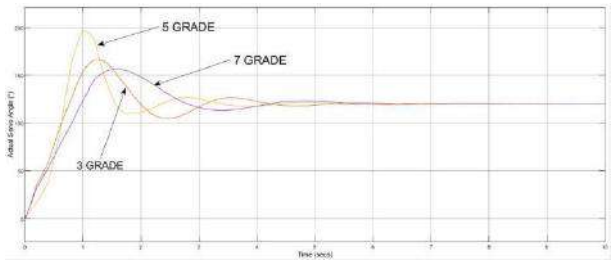


Fig.14. The Influence of Fuzzy Rule Levels on the Dynamic Performance of Robot Arm Control Systems

According to the simulation results, the rise time of the 3-level fuzzy PID control system is 0.78 seconds, but its fuzzy rule division is relatively simple, and the accuracy of the angle error and error change rate of the robotic arm is insufficient, resulting in overshoot reaching 0.383%, adjustment time of 3.71 seconds, and poor dynamic adjustment performance; Although the 7-level fuzzy PID control system has improved accuracy by increasing the number of rules and reducing overshoot to 0.30%, excessive rules can easily lead to rule redundancy and logical contradictions, resulting in delayed system response. The rise time is extended to 0.98s, the adjustment time is also increased to 3.81s, and the overall dynamic response speed is significantly reduced; The 5-level fuzzy PID control system achieves a good balance between rule refinement and system response characteristics. It not only shortens the rise time to 0.68s and the adjustment time to 3.05s, but also controls the overshoot within a reasonable range of 0.625%, making it a good fuzzy rule level scheme suitable for this robotic arm control system. In summary, it can be concluded that in the rule design of fuzzy PID control systems, the number of fuzzy rule levels is not necessarily better. Excessive increase in levels can lead to rule redundancy and contradictions, which in turn can reduce system control performance; The optimal control effect can only be achieved when the partition density of fuzzy rules matches the dynamic characteristics of the controlled system and a balance point between performance and rule rationality is found.

In addition, to investigate the trajectory tracking capability and stability of the PID control system and fuzzy PID control system mentioned above. Based on MATLAB/Simulink environment, establish traditional PID control simulation model and fuzzy PID control simulation model for robotic arm control system, and set a step disturbance input signal at time=10s. The dynamic performance comparison chart of the system is shown in Figure 15.

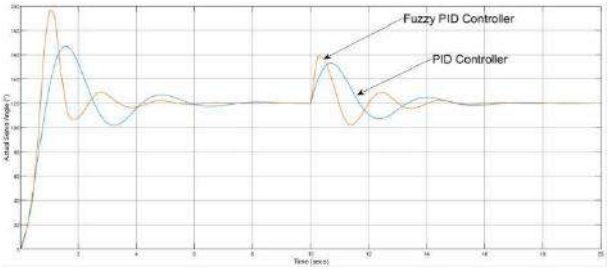


Fig.15. Dynamic characteristic curve of control system under interference signal

Based on the above dynamic performance indicators,

the trajectory tracking performance of the PID control system and the fuzzy PID control system is compared. Compared with the traditional PID control system, the fuzzy PID control system improves the response speed by 10.59% (reducing 0.32s) under the influence of interference signals, significantly improves the joint trajectory fit of the robotic arm, and reduces trajectory deviation. Especially in the face of complex work environments, it maintains strong robustness and trajectory tracking ability.

VI. CONCLUSION

This study focuses on the complex working environment of automated production workshops, where the trajectory tracking of robotic arms suffers from long-term problems such as speed steps, external disturbances, and reduced control accuracy caused by joint friction. The traditional PID control system lacks the ability to handle nonlinear factors and cannot meet the requirements of high-precision operations. Therefore, a fuzzy PID adaptive control method is proposed. The MATLAB/SIMULINK simulation experiment on the KUKA KR6R700 six axis robotic arm shows that the proposed method significantly improves control performance. Compared with traditional PID control systems, the joint adjustment time of the robotic arm is shortened by ≥ 43.72% (reduced by 2.37 seconds), and the joint trajectory fit of the robotic arm is significantly improved. The trajectory deviation is small and the change is smoother. Especially in the face of complex work environments, it maintains strong robustness and trajectory tracking ability. Research has shown that the proposed solution can effectively improve the motion control accuracy of robotic arms in automated production workshops.

ACKNOWLEDGEMENTS

This research was supported by "Guangdong Scienceand Technology Program", "Research on Investigation ofthe Multimodal Intelligent Flying Electric Motorcycle's Autonomous Motion. (NO. 2024A0505050019)", theirhelp in teaching the research.

REFERENCES

[1] Wang Shuyan, Shi Yu, Feng Zhongxu. (2011). Research on control method based on fuzzy PID controller. *Mechanical Science and Technology for Aerospace Engineering*, 30(01), 166–172. DOI:10.13433/j.cnki.1003-8728.2011.01.035.

[2] Niu Yaqing. (2025). Research on optimization of robotic arm equipment control in automated production workshops. *Mechanical Management and Development*, 40(10),181–183. DOI:10.16525/j.cnki.cn141134/th.2025.10.067.

[3] Fan Guoping. (2005). *Design and research of intelligent PID control system* [Master's thesis, Zhejiang University of Technology].

[4] Tian Ye. (2021). Design and implementation of a fuzzy control system for robotic arms. *Journal of Xichang University (Natural Science Edition)*, 35(03), 48–50+74. DOI:10.16104/j.issn.1673-1891.2021.03.010.

[5] Su Ming, Chen Lunjun, Lin Hao. (2004). Fuzzy PID control and its MATLAB simulation. *Modern Machinery*, (04), 51–55. DOI:CNKI:SUN:XDJX.0.2004-04-024.

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.3

Comparative analysis of static and dynamic approaches to constructing factor investment strategies

Abdelmadjid Laouedj

Quantitative Researcher, Jersey City, USA

Received: 22 Jan 2026; Received in revised form: 24 Feb 2026; Accepted:27 Feb 2026; Available online: 03 Mar 2026

Abstract *— The article compares static and dynamic approaches to constructing factor investment strategies, focusing on performance, risks, and practical implementation. The aim of the study is to formalize criteria for selecting a factor portfolio architecture that accounts for the cyclicality of premia, liquidity constraints, and transaction costs, and to clarify under which conditions a more complex dynamic specification is justified relative to fixed rules. The relevance of the work is driven by the widespread adoption of factor investing and, simultaneously, by mounting doubts regarding the persistence of premia and the out-of-sample transportability of factor-timing results. The scientific contribution lies in a comprehensive typology of static and dynamic schemes (exposure timing, volatility scaling, regime filters), an explicit treatment of premium degradation channels via costs, capacity, and model risk, and the rationale for a hybrid design with a robust static core and a minimally parametric dynamic risk-management layer. The main findings indicate that static strategies are superior in terms of interpretability and robustness to overfitting, whereas dynamic constructions can increase return density per unit of risk at the cost of greater sensitivity to estimation errors and to the structure of trading frictions. The article is intended for portfolio managers, risk managers, and researchers engaged in developing and testing factor strategies.*

Keywords *— **factor investing, static factor strategies, dynamic factor strategies, factor timing, volatility scaling***

I. INTRODUCTION

Factor investing is a portfolio construction method in which the selection and weights of securities are determined by formalized rules designed to harvest persistent risk premia associated with measurable asset characteristics. In the applied literature, this approach is often implemented through long–short portfolios constructed on the basis of characteristics that are empirically related to expected returns (Kagkadis et al., 2023). In the broader context of capital management, factor premia are viewed as an autonomous layer of exposures that can be overlaid on a baseline equity and bond portfolio in order to increase the likelihood of achieving long-term objectives and to smooth the trajectory of wealth. This idea is formalized in studies that analyze factor exposures as a strategic complement to traditional asset allocation, including assessments of the impact of extreme adverse episodes (Cavaglia et al., 2022).

The value of factor investing stems from its discipline and reproducibility. The investor obtains a transparent language of risk, since exposures are described by a limited set of systematic drivers rather than by a list of individual issuers and narratives. At the same time, vulnerability arises from the cyclicality of premia and shifts in the market environment. Survey studies on factor premia in adjacent segments, including government bonds, emphasize that the evidence for factors depends on samples, methods, and observation windows, and therefore the question of premium persistence is methodologically central (Bektić et al., 2020). This provides a practical motivation for comparing strategy-construction architectures, since the same factors may exhibit

different returns and risks depending on how the procedure for forming and maintaining exposure is specified.

The problem of choosing an approach to constructing factor strategies centers on the dilemma between rule invariance and market-state adaptation. Static constructions fix the method of measuring signals and the rebalancing frequency, making their properties easier to interpret and to control. Dynamic constructions allow factor weights or trading intensity to vary in response to changing conditions, relying on the notion of time-varying expected premia. Recent research on factor timing shows that the predictability of factor-portfolio returns can be economically meaningful, and models that exploit information contained in portfolio characteristics, in one empirical design, explain about 67% of the variation in factor-portfolio returns and deliver an annual Sharpe ratio of approximately 0.73 for the corresponding trading rules (Kagkadis et al., 2023). These results simultaneously strengthen interest in dynamic methods and sharpen the question of their reliability and out-of-sample robustness, which sets the stage for the subsequent comparative analysis.

II. MATERIALS AND METHODOLOGY

The research material is based on a corpus of academic studies on factor premia and their time-series stability, as well as on practices for constructing long–short and long-only portfolios. To formulate the problem and operationalize factors, the study draws on evidence regarding reproducible premia over long horizons and across different samples, including a global examination of a broad set of premia (Baltussen et al., 2021) and a survey of factor approaches in related asset classes, which underscores the dependence of conclusions on test design and observation period (Bektić et al., 2020). To establish the context of portfolio applications of factors as an exposure layer in multi-asset portfolios, the analysis incorporates results on the role of factors in strategic asset allocation and on sensitivity to extreme adverse episodes (Cavaglia et al., 2022).

The empirical motivation for comparing static and dynamic architectures relies on evidence of the predictability of factor-portfolio returns through portfolio characteristics and on the potential economic significance of factor timing (Kagkadis et al., 2023), as well as on the literature concerning the time-variation of premia and the fragility of gains once execution frictions are taken into account (Haddad et al., 2020).

The methodology implements a comparative analysis of two classes of factor-strategy construction and focuses on differences in mechanisms for maintaining exposures, sources of risk, and channels through which results may deteriorate. The static approach is interpreted as a fixed algorithm for measuring signals and calendar-based rebalancing, contrasted with dynamic rules that permit time-varying factor weights, risk scaling, and regime filters (Lioui & Tarelli, 2020), including the impact of sector neutrality as a structural parameter of implementation (Ehsani et al., 2023).

For dynamic constructions, three applied directions are distinguished, which serve as objects of comparison at the level of decision logic and feasible trade-offs: factor-exposure timing (Kagkadis et al., 2023), volatility-based scaling (Nucera & Uhl, 2022), and regime schemes tied to inflation and macroeconomic states (Baltussen et al., 2023), with separate consideration of the role of transaction costs and rebalancing-boundary optimization as sources of differences between architectures (El Bernoussi & Rockinger, 2022; Bai et al., 2025).

III. RESULTS AND DISCUSSION

In modern asset-pricing theory, factor premia are understood as the average excess returns of portfolios formed on the basis of ex-ante characteristics that are persistently related to subsequent returns and risk. In practical constructions, such characteristics are grouped into several families (Baltussen et al., 2021). These include value characteristics that reflect the relative cheapness of a firm based on accounting and market metrics; momentum characteristics that capture the continuation of recent return trends; size characteristics that separate small-capitalization companies from large ones; quality characteristics that proxy for profitability, robustness of outcomes, and the issuer's investment discipline; and low-volatility characteristics that select securities with more subdued price dynamics.

In empirical surveys based on extensive historical data, factors are treated as reproducible sorting rules applicable across asset classes, and dozens of premia are shown to retain economic meaning under uniform out-of-sample verification. Across a global universe of markets, the analysis of 24 premia over a 217-year horizon shows that most survive out-of-sample testing without pronounced degradation in average effect (Baltussen et al., 2021).

Explanations for the existence of factor premia typically revolve around two families of mechanisms. Risk-based arguments relate the premium to compensation for systematic risks that intensify in adverse states of the economy and financial markets, so that the expected premium co-moves with the price of risk. Behavioral arguments attribute the premium to persistent expectation errors and to limits to arbitrage, which cause over- and underpricing to be corrected slowly and generate predictability. For both lines of reasoning, the temporal stability of the premium is critical, since it distinguishes a structural effect from a statistical accident.

Studies of the dynamics of factor returns reveal considerable time-variation in expected premia and simultaneously demonstrate that the benefits of attempts at systematic timing often prove fragile due to statistical uncertainty and execution frictions (Haddad et al., 2020). The problem of cyclicality manifests in the ability of individual factors to enter prolonged periods of weak premium realization and to experience rare but severe loss episodes (Dierkes & Krupski, 2022). For momentum strategies, a characteristic pattern of sharp performance deterioration has been documented in reversal phases after market downturns.

Static approaches in factor strategies rely on a prespecified algorithm for forming exposures, in which the set of characteristics, the rules for ranking securities, and the rebalancing calendar are determined before live implementation and change very little thereafter. Within such a framework, the researcher or manager chooses a metric for each characteristic and specifies a procedure for constructing portfolios from sorted groups of securities, which is then repeated at regular intervals. A canonical example is a construction in which stocks are sorted by size and a second characteristic, and long–short portfolios are then formed from several groups with constant shares in sub-portfolios. This logic is standard for empirical factor portfolios and serves as a baseline for discussing what is forfeited when fixed weights ignore time variation in premia and market exposures (Lioui & Tarelli, 2020).

A typical implementation of static design begins by partitioning the universe into quantiles or deciles along the chosen characteristic, after which portfolios are formed with long positions in the upper groups and, where permitted, short positions in the lower groups. A long-only variant is usually constructed as a tilt toward the top group, subject to concentration and risk constraints. The long–short variant is often supplemented by requirements for neutrality to market risk and to sector tilts to approximate a pure factor exposure and reduce dependence on sectoral conditions. At the same time, sector neutrality is a structural parameter that affects the efficacy of long–short and long-only implementations differently. In one empirical study, it is shown that preserving the sector component more often improves the properties of long-only strategies, whereas sector-neutralization more often enhances the performance of long–short portfolios (Ehsani et al., 2023). More flexible static constructions replace coarse partitioning at fixed thresholds with the construction of characteristic portfolios, in which the aggregation of signals and asset weights is estimated more parsimoniously from the data and allows for explicit limits on individual positions (McGee & Olmo, 2022).

The main strength of static approaches lies in their transparency and reproducibility. They are easier to verify, to explain, and to monitor, since strategy behavior follows from a small set of rules and does not require continual recalibration. An additional practical advantage is the predictability of turnover and rebalancing costs, which is important under limited liquidity and explicit cost control. Theoretical–empirical analysis of rebalancing with transaction costs shows that fixed-weight restoration can preserve attractive risk–return trade-offs at realistic cost levels, although dominance depends on the structure of autocorrelations and cross-correlations among assets (El Bernoussi & Rockinger, 2022).

A key limitation of static design manifests as extended adverse cycles of factor premia, which create the risk of prolonged drawdowns and the erosion of

investor confidence. For the value premium, a multi-year period of weak performance has been documented, where, as of mid-2020, a drawdown on the order of 55% relative to growth stocks is discussed, illustrating the phenomenon of factor winters and raising the question of how long it is acceptable to wait for premium recovery under invariant rules (Arnott et al., 2021). Static Factor Strategy Implementation is shown in Figure 1.

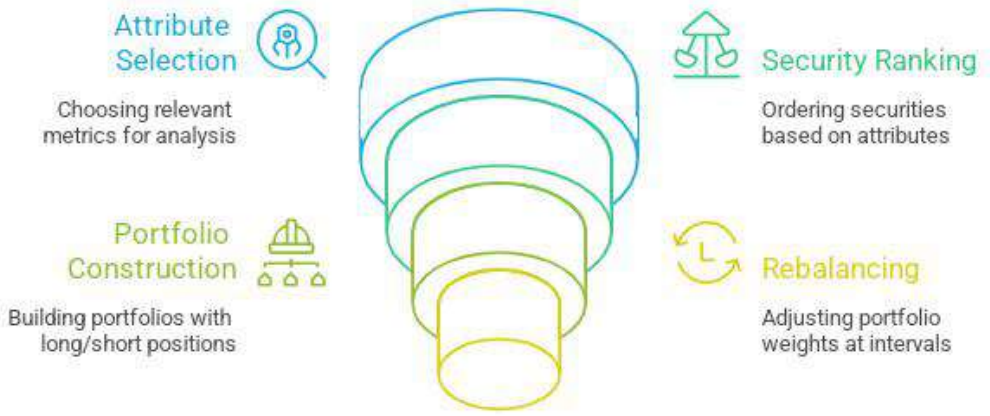


Fig. 1. Static Factor Strategy Implementation

A factor strategy becomes dynamic when it incorporates rules that adjust exposures in response to the market's observed state or to internal indicators of the strategy. In a static construction, a signal generates a portfolio according to a fixed scheme and is then maintained on a calendar basis. In a dynamic construction, a management layer is superimposed on the signal, allowing for reconfiguration of factor weights, risk scaling, switching of constraint sets, and changes in trading speed. Such an architecture presumes that expected factor premia and their risks follow their own dynamics and that these dynamics can be measured through variables observable at the decision time. Empirical results on the timing of factor portfolios show that characteristics aggregated at the factor-portfolio level contain information about future returns and permit the construction of strategies that systematically differ from fixed-weight schemes in their realized performance trajectories (Kagkadis et al., 2023).

The most widely used class of dynamics concerns factor-exposure timing, in which factor weights change in response to predictors reflecting valuation, momentum, sentiment, and other features of the market environment. A second class is associated with volatility targeting and risk scaling, whereby exposures are increased in tranquil periods and reduced in turbulent periods. Studies devoted to volatility scaling for factor portfolios document improved predictability of factor returns and substantial time-variation in performance characteristics, while translating this predictability into robust factor rotation remains limited and sensitive to portfolio-construction methodology (Nucera & Uhl, 2022). A more applied branch of this approach develops volatility-targeting mechanisms that incorporate transaction costs via rebalancing bands, where optimization aims to reduce costs without losing risk control, and shows that even the simple introduction of a no-trade zone alters the performance profile of the dynamic mechanism (Bai et al., 2025).

A third class of dynamics is based on regime filters that use indicators of inflation, macroeconomic phase, liquidity, and market volatility to select factor tilts that are more appropriate for current conditions. These filters are motivated by the observation that factors behave differently across deflationary, high-inflation, and stagflation regimes, as well as across regimes with different trend persistence and different costs of capital. Empirical analysis of premia on assets and factors across inflation regimes shows that regime comparisons provide a meaningful foundation for conditional exposure management and help formalize the risk of adverse cycles that, in a static strategy, are perceived as protracted factor winters (Baltussen et al., 2023).

A fourth class of dynamics is associated with adaptive selection thresholds and rebalancing

frequency, in which the strategy varies the intensity of intervention to maintain a compromise between deviations from target weights and trading costs. The risks of dynamics are concentrated in regime-classification errors, in the accumulation of estimation error, and in the latent dependence of outcomes on execution costs. As a result, every source of added return in a dynamic construction simultaneously constitutes a channel of model risk (Nuriyev et al., 2024). Figure 2 illustrates Dynamic Factor Strategies.

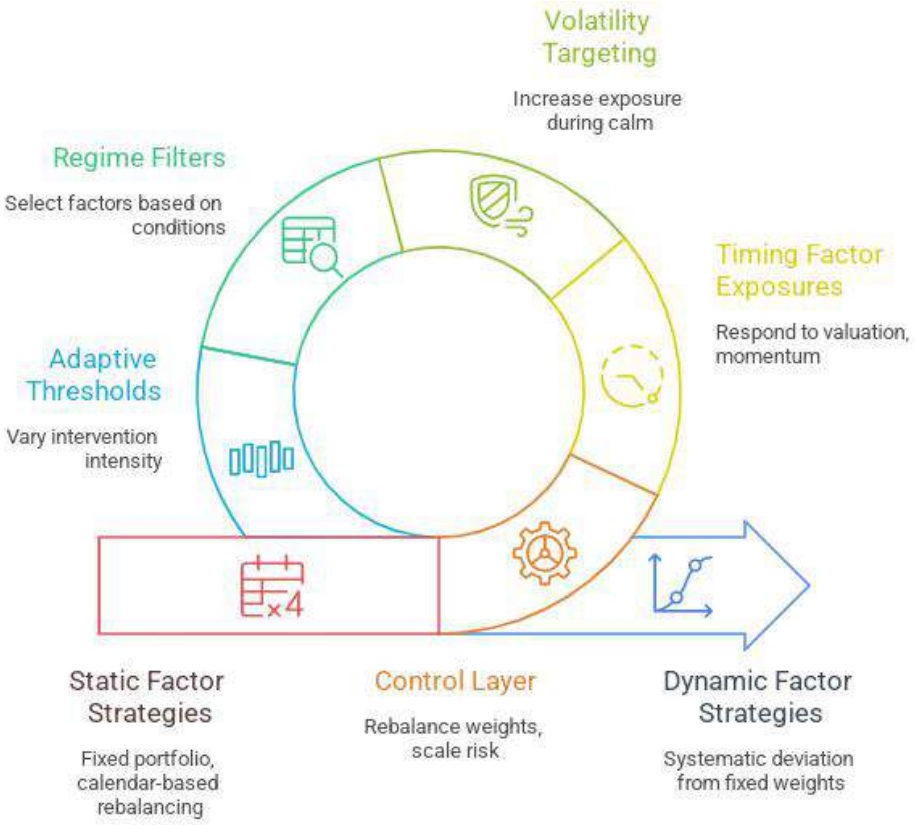


Fig. 2. Dynamic Factor Strategies

The comparison of static and dynamic approaches to expected return begins with the observation that, in both cases, the source of performance is the same: compensation for systematic exposures. The difference lies in how the strategy attempts to transform the conditional variability of premia into realized returns. A static construction treats the premium as a long-horizon quantity and seeks to accumulate it through continuous exposure. This mechanism is advantageous when the premium materializes slowly and is relatively evenly distributed over time.

A dynamic construction allows the premium to depend on market conditions, valuations, and the uncertainty regime. It aims to amplify exposure during favorable phases and attenuate it during unfavorable ones. The potential benefit is associated with a higher density of return per unit of risk. The potential loss is associated with misclassification of phases and with the possibility that the adaptation process itself destroys the effect owing to lags and trading frictions. Premium robustness in a static scheme rests on the stability of factor definition and on the persistence of its economic rationale. Premium robustness in a dynamic scheme also requires the stability of conditional relationships linking market states to future premia.

Risk criteria highlight the contrast more sharply, since the risk of factor strategies is concentrated in rare episodes when usual relationships break down. Static strategies are characterized by a predictable pattern of volatility and drawdowns, yet they remain vulnerable to extended adverse periods and to abrupt crashes that occur when regimes shift, liquidity spikes, or positions are forcibly unwound by market participants. Dynamic

strategies introduce mechanisms to mitigate such episodes by reducing exposure or redistributing risk. When tuned successfully, this reduces drawdown depth and the probability of extreme losses. When tuning is unsuccessful, an additional layer of risk emerges: procyclicality, in which exposure reduction occurs after losses have already accumulated, thereby locking in damage. There is also the risk of frequent switching, which transforms rare events into a sequence of moderate losses. Tail risk in a dynamic architecture arises from a combination of market turbulence and control errors. Tail risk in a static architecture is closer to a pure form of factor risk, with the potential for collapse.

Cross-factor correlations and diversification properties are particularly important in stress periods, when many return sources begin to move in tandem. A static multifactor portfolio typically assumes that different factors will offset one another. However, in stressed phases, correlations may increase, and the expected diversification benefit may weaken. A dynamic approach seeks to account for such shifts and reallocate risk to reduce dependence on generalized market stress. At the same time, precisely in stress periods, statistical estimates of correlations and risks become less reliable.

This leads to the criterion of parameter and estimation-error sensitivity. Static strategies usually feature fewer degrees of freedom and a lower probability of hidden overfitting. Their errors are more often linked to poor choices of characteristics, data biases, and persistent structural changes in the market. Dynamic strategies involve more control parameters. They are more sensitive to the quality of estimates, to indicator stability, and to the choice of observation windows. Consequently, even for the same factor idea, dynamic implementation requires stricter discipline in robustness testing. Otherwise, adaptivity becomes a source of noise-driven decisions that appear plausible in backtests but disintegrate when the environment changes. Fig. 3 compares static and dynamic strategies.

Fig. 3. Comparison of Static and Dynamic Strategies

The implementation of factor strategies begins with portfolio turnover and immediately confronts transaction costs. Even with identical signals, static and dynamic architectures generate different trading profiles, since dynamics amplify the frequency and magnitude of weight changes. Costs arise through bid–ask spreads, market impact, and the limited availability of securities for borrowing on the short side. In the short book, costs become endogenous because borrowing fees rise precisely when demand for short positions clusters. In real-world settings, strategy performance is determined by how closely

the expected incremental return matches the total cost of maintaining exposure. In dynamic constructions, an additional cost layer arises associated with regime switches and frequent risk adjustments, making trading speed, rather than just target weights, a central object of control.

Liquidity and capacity constraints delineate the feasible domain of factor premia. A strategy may be effective at a small scale but deteriorate as capital grows, because market impact increases more than linearly and affects precisely those securities where the signal is strongest. This is particularly visible in small-cap segments and in constructions that require short positions in low-signal securities. Capacity depends on the breadth of the universe, permissible deviations from neutrality constraints, and the holding horizon. Dynamic strategies often claim an advantage through more active risk management, yet their capacity may be lower due to the need to move capital more swiftly across factors and segments. As a result, the choice of architecture must be aligned with market depth, instrument availability, and acceptable levels of hidden concentration.

Model risk and overfitting issues arise from the temptation to turn a dynamic strategy into a set of knobs tuned to past data. The more parameters there are, the easier it is to obtain an impressive historical track record, and the lower the probability that it will repeat once the regime changes. The danger is amplified when regime indicators are employed that are themselves unstable and subject to revision. Macroeconomic time series are revised; inflation and output estimates are updated; liquidity indicators depend on methodology; even market microstructure metrics may change meaning due to regulatory reforms and alterations in trading microstructure.

Distinguishing structure from noise requires a rigorous robustness-testing discipline. A genuine regularity manifests itself across samples, markets, and reasonable parameter variations. Noise disappears under minor changes in the observation window, threshold, or normalization scheme. An additional signal is provided by economic coherence, whereby the mechanism has a clear channel through which it affects risks and participants' expectations.

Proper testing and comparison require a design that mirrors real-time decision-making. In-sample performance is suitable for diagnosing the idea and for an initial filter of hypotheses. Out-of-sample testing assesses viability. Step-ahead evaluation over time reduces the risk of information leakage and forces the model to make decisions under conditions of uncertainty closer to those in practice. The portability of results should be assessed across markets and asset classes, since local regularities often reflect institutional features of a specific venue.

The modeling of costs and constraints must be embedded in the research from the outset, including the dependence of costs on volume and on market conditions. Stress tests and scenario analysis examine how the strategy behaves in regimes where correlations and liquidity change abruptly. Performance attribution decomposes the contribution of timing decisions from that of risk scaling and separates the impact of market direction from the factor premium per se.

On this basis, hybrid solutions are constructed. A stable core sets long-term factor tilts, while a dynamic layer manages risk and constrains drawdown depth. A model committee allocates responsibility across several simple rules, thereby reducing dependence on a single regime classifier. The principle of minimal dynamics calls for a small number of parameters and clear application conditions, so that adaptation remains interpretable and controllable.

Practical recommendations converge on aligning the architecture with the investor's objectives and capabilities. A static approach suits those for whom explainability, compliance, and a long horizon are paramount, as well as those constrained in instruments and execution budget. A dynamic approach suits those who possess appropriate infrastructure, can employ derivative instruments, and prioritize risk management in adverse regimes. The choice can be structured through a set of questions concerning acceptable drawdowns, tolerance for extended periods of weak performance, availability of shorting, turnover limits, and the share of performance that can be entrusted to a model requiring continuous validation and oversight.

IV. CONCLUSION

The comparative analysis of static and dynamic approaches to constructing factor investment strategies shows that in both cases, the core of performance remains the idea of systematic exposures to measurable asset characteristics and the discipline of formalized rules. In static design, the advantage lies in interpretability and reproducibility, as well as in turnover predictability and cost control, which make such an architecture convenient for verification and operation under liquidity and compliance constraints. At the same time, empirical time variation of factor premia and the phenomenon of extended adverse cycles reinforce the importance of robustness, since even economically coherent premia can pass through periods of deep drawdowns, creating the risk of confidence loss and forced strategy unwinding in an unfavorable phase.

Dynamic constructions expand the decision space through exposure timing, volatility targeting, and regime filters, attempting to exploit the conditional variability of premia and convert it into higher returns per unit of risk. Empirical findings point to predictability in factor returns at the portfolio-characteristics level and thus support the motivation for adaptive methods, yet also reveal fragile out-of-sample transportability, sensitivity to methodology, and increased dependence on trading frictions and estimation error. Ultimately, the choice of architecture becomes a problem of aligning investment horizon, infrastructure, and acceptable model risk, while a proper comparison requires a testing design that reflects real-time decision-making and embeds costs, constraints, and stress behavior as coequal elements of evaluation.

REFERENCES

[1] Arnott, R. D., Harvey, C. R., Kalesnik, V., & Linnainmaa, J. T. (2021). Reports of Value's Death May Be Greatly Exaggerated. *Financial Analysts Journal*, *77*(1), 44–67. https://doi.org/10.1080/0015198x.2020.1842704

[2] Bai, Z., Pachamanova, D., Steblovskaya, V., & Wallbaum, K. (2025). Target volatility strategies: optimal rebalancing boundary for transaction cost minimization. *Financial Markets and Portfolio Management*. https://doi.org/10.1007/s11408-025-00486-5

[3] Baltussen, G., Swinkels, L., van Vliet, B., & van Vliet, P. (2023). Investing in Deflation, Inflation, and Stagflation Regimes. *Financial Analysts Journal*, *79*(3), 5–32. https://doi.org/10.1080/0015198x.2023.2185066

[4] Baltussen, G., Swinkels, L., & Van Vliet, P. (2021). Global factor premiums. *Journal of Financial Economics*, *142*(3), 1128–1154. https://doi.org/10.1016/j.jfineco.2021.06.030

[5] Bektić, D., Hachenberg, B., & Schiereck, D. (2020). Factor-based investing in government bond markets: a survey of the current state of research. *Journal of Asset Management*, *21*(2), 94–105. https://doi.org/10.1057/s41260-020-00156-3

[6] Cavaglia, S., Scott, L., Blay, K., & Gupta, T. (2022). Equity factors for multi-asset class portfolios: a strategic asset allocation perspective. *Journal of Asset Management*, *23*, 100–113. https://doi.org/10.1057/s41260-022-00262-4

[7] Dierkes, M., & Krupski, J. (2022). Isolating momentum crashes. *Journal of Empirical Finance*, *66*, 1–22. https://doi.org/10.1016/j.jempfin.2021.12.001

[8] Ehsani, S., Harvey, C. R., & Li, F. (2023). Is Sector Neutrality in Factor Investing a Mistake? *Financial Analysts Journal*, *79*(3), 95–117. https://doi.org/10.1080/0015198x.2023.2196931

[9] El Bernoussi, R., & Rockinger, M. (2022). Rebalancing with transaction costs: theory, simulations, and actual data. *Financial Markets and Portfolio Management*, *37*, 121–160. https://doi.org/10.1007/s11408-022-00419-6

[10] Haddad, V., Kozak, S., & Santosh, S. (2020). Factor Timing. *The Review of Financial Studies*, *33*(5), 1980–2018. https://doi.org/10.1093/rfs/hhaa017

[11] Kagkadis, A., Nolte, I., Nolte, S., & Vasilas, N. (2023). Factor Timing with Portfolio Characteristics. *The Review of Asset Pricing Studies*, *14*(1), 84–118. https://doi.org/10.1093/rapstu/raad010

[12] Lioui, A., & Tarelli, A. (2020). Factor Investing for the Long Run. *Journal of Economic Dynamics and Control*, *117*, 103960. https://doi.org/10.1016/j.jedc.2020.103960

[13] McGee, R. J., & Olmo, J. (2022). Optimal characteristic portfolios. *Quantitative Finance*, *22*(10), 1853–1870. https://doi.org/10.1080/14697688.2022.2094282

[14] Nucera, F., & Uhl, B. (2022). The impact of volatility scaling on factor portfolio performance and factor timing. *Journal of Asset Management*, *23*(6), 522–533. https://doi.org/10.1057/s41260-022-00279-9

[15] Nuriyev, D., Duan, S., & Yi, L. (2024). Augmenting Equity Factor Investing with Global Macro Regimes. *Proceedings of the 5th ACM International Conference on AI in Finance*, 445–452. https://doi.org/10.1145/3677052.3698620

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.4

A Novel Design Approach for Extended Modular Intelligent Robots

He-Rong Lee[1], Wei Lee[2], Ming-Hui Huang[3], Chun-Chi Chang[4], Ho-Sheng Chen[5*]

[1]Guangzhou Institute of Science and Technology, China
Email: chelung168@gmail.com
[4]National Taipei University of Business
Email: czoe2772@gmail.com
[2,3,5]School of Sciences, Guangdong University of Petrochem Technology (GDUPT), Maoming 525000, China
Corresponding author: Ho-Sheng Chen
Email:hschen98.tw@gmail.com

Received: 27 Jan 2026; Received in revised form: 26 Feb 2026; Accepted:28 Feb 2026; Available online: 05 Mar 2026

***Abstract*—** *This paper presents the design of a reconfigurable multi-form robot integrating UAV, hexapod biomimetic, and wheeled structures based on modular principles. The system utilizes an Arduino platform with Mega2560 microcontroller for joint motor control and a dedicated flight control module for brushless motor regulation. Physical prototyping is achieved through 3D printing technology, with detachable joint and arm structures enabling flexible module replacement and expansion. Ultrasonic sensors are incorporated to provide obstacle avoidance functionality. The proposed design demonstrates a practical approach to developing adaptable robotic systems for diverse application scenarios.*

Keywords*—** ***Modular robot; Reconfigurable mechanism; Arduino; 3D printing; Obstacle avoidance

I. INTRODUCTION

Since the 21st century, with the continuous development of science and technology, scholars at home and abroad have conducted increasingly in-depth research on robot technology. Robot technology has become a focus of attention for technology workers and is considered one of the high-tech fields that are of great significance for the development of emerging industries in the future. Robots have broad application prospects in fields such as agricultural planting, rescue and detection, military security, home services, and intelligent transportation [1,2,3,6].

In these fields, robots need to complete different types of homework tasks in dynamic, unknown, and unstructured complex environments, which puts high demands on their environmental adaptability, environmental perception, autonomous control, and human-computer interaction [4,5].

Based on the above analysis and description, this article takes a multi form robot as the research object, with six wing, six legged, and four wheeled models as the main body. Mechanical structure design, control system design, and kinematic analysis are carried out to complete corresponding simulation experiments and robot experiments, and experimental results are obtained [7-10].

II. STRUCTURAL DESIGN OF HEXACOPTER MODULE

The six-wing UAV combines the structure of a hexapod robot, featuring six propellers and six mechanical legs arranged in a hexagonal configuration [11-13]. Adjacent rotors rotate in opposite directions to counteract torque, allowing attitude control through speed adjustment. An aluminum alloy frame ensures lightweight durability and stable flight. The mechanical legs serve as

landing gear for shock absorption and enable loading, transport, and grasping functions. This design enhances performance and adaptability in hazardous environments (Figure 1) [14,15].

Fig.1. Structure of Unmanned Aerial Vehicle

The components n=3, 1, 2, and 3 of the mechanism are thigh, calf, and foot end components, respectively. I, II, and III are the rotating pairs of the mechanism, all of which are low pairs and do not have high pairs. According to the formula for calculating institutional degrees of freedom.

$$F = 3n - 2p_l - p_h \quad (1)$$

where n is the number of active components, p_l is the low parity of the institution; p_h is the high parity number of the institution.

Therefore, the degree of freedom F of a hexapod robot's single leg is 3, which can achieve the internal and external expansion, flexion and extension, and pitch movements of the single leg. The thigh drives the servo to rotate around the z-axis at the root joint to achieve internal and external expansion and contraction movements. Drive the servo at the calf and thigh to move in the z-axis direction to achieve pitch motion. The servo is driven to move in the y-axis direction at the foot and calf ends to achieve flexion and extension movements. At the same time, the mechanical leg uses magnets and snap fasteners at the joint between the thigh and calf joints to achieve rapid disassembly and assembly of the leg structure. The design drawings of the mechanical leg are shown in Figures 2 and 3.

Fig.2. Leg Structure

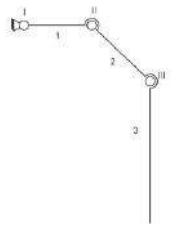


Fig.3. Mechanical principle

The front and rear wheels of the four-wheel drive car have power, and the engine output torque can be distributed on all the wheels in different proportions according to the driving road conditions to improve the driving ability of the four-wheel drive car. The four-wheel drive car controls its motion through a 24 channel development board with STM32 core as the autonomous control. The PID algorithm is used to correct the stability of the car's motion. The proportion, integral, and derivative of the deviation are linearly combined to form the control quantity of the four-wheel drive car. The control quantity is used to control the servo of the controlled object, which is the four-wheel drive car. The PID process of servo control is shown in Figure 4.

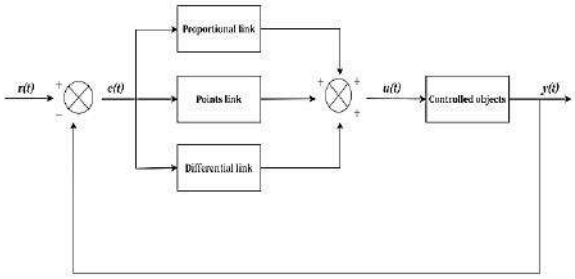


Fig.4. Servo control system of PID process

Control the aircraft's ascent, descent, rotation, and hovering movements through the thrust of six brushless motors. During the flight, the flight control system of the hexacopter drone sends commands to each electric motor to control their speed and output thrust. The combination of these instructions can enable a hexacopter drone to fly in different directions, such as forward, backward, left turn, right turn, etc., and hover in the air. For ascent and descent, the hexacopter drone achieves this by controlling the total thrust of all electric motors. When the thrust of the electric motor increases, the

aircraft will ascend; On the contrary, when the thrust decreases, the aircraft will descend. By adjusting the thrust of each electric motor, the aircraft can be deflected or rotated. During the flight, the flight control system of the hexacopter drone continuously detects the aircraft's attitude and automatically adjusts the speed of the electric motor as needed to maintain the balance and stability of the aircraft. In summary, the hexacopter drone achieves accurate control of the aircraft by controlling the speed and thrust of each electric motor, enabling operations such as flight, hovering, and turning in the air.

The rigid body model includes kinematics and dynamics. Dynamics calculates the relationship between sum force and acceleration, kinematics calculates the relationship between velocity and position, and dynamics calculates acceleration for use in kinematics [9]. Input the throttle of six motors and output the lift and torque of the body. From this, the following formula can be derived:

$$\begin{aligned} T_m\dot{\varpi}_1(t) + \varpi_1(t) &= C_m\sigma_1(t) + \varpi_m \\ T_m\dot{\varpi}_2(t) + \varpi_2(t) &= C_m\sigma_2(t) + \varpi_m \\ T_m\dot{\varpi}_3(t) + \varpi_3(t) &= C_m\sigma_3(t) + \varpi_m \\ T_m\dot{\varpi}_4(t) + \varpi_4(t) &= C_m\sigma_4(t) + \varpi_m \\ T_m\dot{\varpi}_5(t) + \varpi_5(t) &= C_m\sigma_5(t) + \varpi_m \\ T_m\dot{\varpi}_6(t) + \varpi_6(t) &= C_m\sigma_6(t) + \varpi_m \end{aligned} \quad (2)$$

From Eq. (2), can be concluded that

$$T = c_T(\varpi_1^2 + \varpi_2^2 + \varpi_3^2 + \varpi_4^2 + \varpi_5^2 + \varpi_6^2) \quad (3)$$

The resultant moment is:

$$\begin{aligned} \tau_x &= \sqrt{3}/2dc_T(-\varpi_1^2 + \varpi_2^2 + \varpi_3^2 - \varpi_4^2 - \varpi_5^2 + \varpi_6^2) \\ \tau_y &= \sqrt{3}/2dc_T(\varpi_1^2 + \varpi_2^2 + \varpi_3^2 - \varpi_4^2 - \varpi_5^2 - \varpi_6^2) \\ \tau_z &= c_M(\varpi_1^2 - \varpi_2^2 + \varpi_3^2 - \varpi_4^2 + \varpi_5^2 - \varpi_6^2) \end{aligned} \quad (4)$$

The symbols of the dynamic model and the values of the dynamic model constants are the meanings of the parameters in equations (2), (3), and (4), as shown in Table 1.

Table 1. Symbol Explanation of Dynamic Model

variable	symbol	unit
the throttle of each motor	$\boldsymbol{\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6}$	Normalization [0,1]
motor speed	$\boldsymbol{\varpi_1, \varpi_2, \varpi_3, \varpi_4, \varpi_5, \varpi_6}$	rad/s
lift generated	$\mathbf{F_1, F_2, F_3, F_4, F_5, F_6}$	N
induced drag generated	$\mathbf{M_1, M_2, M_3, M_4, M_5, M_6}$	N
the torque generated by the lift	$\boldsymbol{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6}$	N·m
the reverse torque generated	$\boldsymbol{\tau_1, \tau_2, \tau_3}$	N·m
resultant moment on the axis	**89.96**	N·m
Combined lift force	**T**	N

Given a throttle, the motor will gradually reach the target speed, approximating the motor as a first-order system. Therefore, Eq. (3) shows that the stable speed is linearly related to the throttle:

$$T_m d\varpi(t)/dt + \varpi(t) = C_m\sigma(t) + \varpi_m \quad (5)$$

is equivalent to

$$\dot{\varpi}(t) = d\varpi(t)/dt = \frac{1}{T_m}\left(C_m\sigma(t) + \varpi_m - \varpi(t)\right) (6)$$

By using the above formula and Simulation model of Extended Modular Intelligent Robots based on MATLAB/Simulink to test the motor model of a six wing unmanned aerial vehicle, the conclusion can be drawn that the larger the motor time constant, the smaller the speed change, as shown in Figure 5 and 6.

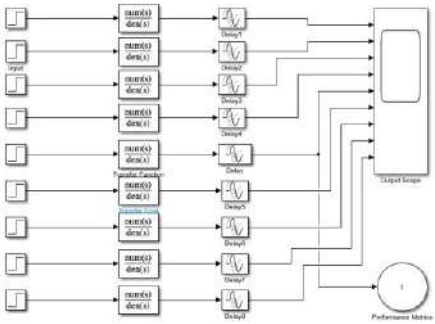


Fig.5. Simulation model of Extended Modular Intelligent Robots based on MATLAB/Simulink

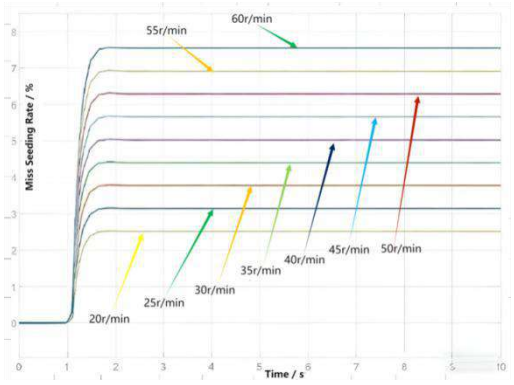


Fig.6. Test results of motor model for six wing unmanned aerial vehicle based on MATLAB

If the lift generated by a single propeller is proportional to the square of the rotational speed:

$$F = c_T \varpi^2 \tag{7}$$

The combined lift generated by the six propellers is:

$$T = c_T(\varpi_1^2 + \varpi_2^2 + \varpi_3^2 + \varpi_4^2 + \varpi_5^2 + \varpi_6^2) \tag{8}$$

The physical quantity that produces a rotational effect on an object can be divided into the moment of force on the axis and the moment of force on the point, that is, M=d × F. Among them, d is the position vector from the axis of rotation to the point of force, and F is the vector force; Torque is also a vector. According to the torque formula:

$$\begin{aligned} M_1 &= d \times F_1 \\ M_{1x} &= -\frac{\sqrt{3}}{2} dF_1 = -\frac{\sqrt{3}}{2} dc_T \varpi_1^2 \\ M_{1y} &= \frac{\sqrt{3}}{2} dF_1 = \frac{\sqrt{3}}{2} dc_T \varpi_1^2 \end{aligned} \tag{9}$$

XY axis torque:

$$\begin{aligned} \tau_x &= \frac{\sqrt{3}}{2} dc_T(-\varpi_1^2 + \varpi_2^2 + \varpi_3^2 - \varpi_4^2 - \varpi_5^2 + \varpi_6^2) \\ \tau_y &= \frac{\sqrt{3}}{2} dc_T(\varpi_1^2 + \varpi_2^2 + \varpi_3^2 - \varpi_4^2 - \varpi_5^2 - \varpi_6^2) \\ \tau_z &= c_M(\varpi_1^2 - \varpi_2^2 + \varpi_3^2 - \varpi_4^2 + \varpi_5^2 - \varpi_6^2) \end{aligned} \tag{10}$$

Drones should be studied for dynamics based on dynamic coordinates. According to the Euler equation of a rigid body, the absolute derivative in dynamic coordinates can be expressed as: $\sum M^b = I\dot{W}^b + W^b \times (IW^b)$, where (p, q, r are the roll, pitch, and yaw velocities in the body coordinate system, respectively):

$$I = \begin{bmatrix} I_x & 0 & 0 \\ 0 & I_y & 0 \\ 0 & 0 & I_z \end{bmatrix} \tag{11}$$

According to the above equation:

$$\begin{aligned} M_x &= I_x \dot{p} + (I_z - I_y)qr \\ M_y &= I_y \dot{q} + (I_x - I_z)pr \\ M_z &= I_z \dot{r} + (I_y - I_x)pq \end{aligned} \tag{12}$$

By simplifying, we can obtain:

$$\begin{bmatrix} M_{Tx} \\ M_{Ty} \\ M_{Fz} \end{bmatrix} = \begin{bmatrix} l(F_6 - F_4 - F_2) \\ l(F_5 - F_3 - F_1) \\ -M_{D1} + M_{D2} - M_{D3} + M_{D4} - M_{D5} + M_{D6} \end{bmatrix} = \begin{bmatrix} lb(\Omega_6^2 - \Omega_4^2 - \Omega_2^2) \\ lb(\Omega_5^2 - \Omega_3^2 - \Omega_1^2) \\ d(-\Omega_1^2 + \Omega_2^2 - \Omega_3^2 + \Omega_4^2 - \Omega_5^2 + \Omega_6^2) \end{bmatrix} \tag{13}$$

where, b is the rotor lift coefficient, d is the rotor drag coefficient, and represents the rotor speed.

III. ROBOT PROGRAMMING AND SIMULATION

Zide is a graphical programming tool for multi-channel servos dominated by PWM signals. The upper computer can directly use a 24 channel servo control board to write data for the 18 motion servos of the robot, and can group the written data into action groups. The Zide upper computer page is shown in Figure 7, which saves a lot of workload and time for subsequent programming work.

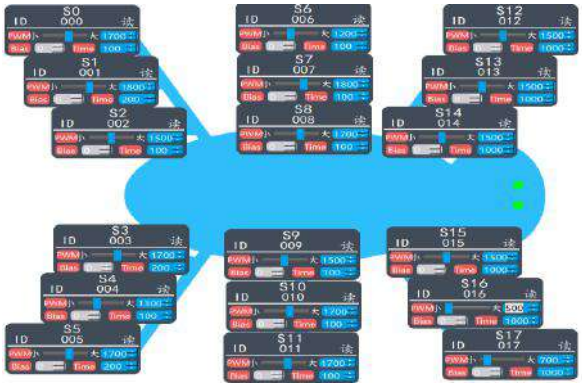


Fig.7. Zide upper computer interface diagram

We will label the 18 servos of the robot according to S0-S17, and classify these labels according to the six legs of the robot. Use SO-S2 as leg 1, S3-S5 as leg 2, S6-S8 as leg 3, S9-S11 as leg 4, S12-S14 as leg 5, and S15-S17 as leg 6. The driving motor of the wheels uses a 360 ° servo motor, which is controlled by

PWM signals. The servo motors of the four wheels are calibrated as S1-S4 as shown in Figure 8. S0 is the left front wheel, S2 is the right front wheel, S1 is the left rear wheel, and S3 is the right rear wheel. The actions of the car are divided into four parts: forward, rear wheel, left turn, and right turn.

Fig.8. Arrangement diagram of small car servo motor

Taking forward and left turn actions as an example, the data of the car's actions are shown in Table 2

Table 2. Data of small car action servo

	S_0	S_1	S_2	S_3
forward	2500	2500	2500	2500
turn left	50	50	2500	2500

First, install the servo into the installation slots of each leg, and then connect the three joints into the main body using the servo disc and nuts. Cut the carbon brazing tube into six sections and connect them to each other through carbon tube connection blocks and motor mounts. Before installing the ultrasonic sensor into the main body, as shown in Figure 9

Fig.9. The Extended Modular Intelligent Robots

IV. CONCLUSION

This article proposes a design and manufacturing method for an extended class multimodal robot, which integrates three modules: hexapod robot, hexacopter drone, and four-wheel cart. By designing a detachable structure, different forms can be switched, and a six legged leg structure and a wheeled structure for the car can be designed using Solid Works. And the corresponding driving electrical components were selected. The designed parts were produced using 3D printing technology. Then, FDM 3D printer technology was used to print and assemble the parts. In terms of control, use Arduino and Linkboy platforms to write relevant posture and motion control programs, use Arduino 2560 and 24 channel servo control board, and combine with ultrasonic obstacle avoidance module, relay, lighting, etc. The robot action group was programmed and debugged using Zide servo programming software, and the flight power parameters of the brushless motor were debugged using flight control parameter tuning software. Finally, the robot's form was assembled and tested.

REFERENCES

[1] Wang Yu. (2021). Gait planning and control system design of hexapod robot (Master's thesis). Hubei University of Technology. DOI:10.27131/d.cnki.ghugc.2021.000155.

[2] Liu Yan, Zheng Luying, Sun Haoyang, Ma Bingran, Zhang Shenghan. (2018). Motion analysis of hexapod robot under triangular gait. Journal of Qingdao University, 33(03), 38-42+46. DOI:10.13306/j.1006-9798.2018.03.007.

[3] Sun Xun. (2022). Research on gait design of hexapod robot in complex slope environment (Master's thesis). Harbin University of Science and Technology. DOI:10.27063/d.cnki.ghlgu.2022.001126.

[4] Ding Kai. (2016). Research on bionic locomotion gait planning and control system of hexapod robot (Master's thesis). Southeast University.

[5] Chen Rui, Chen Nanxing, Huang Wenjun, Wang Zhanhao, Ma Changyun. (2020). Hexapod robot applied to post-disaster rescue and patrol measurement. Electronics World, (09), 132-133. DOI:10.19353/j.cnki.dzsj.2020.09.070.

[6] Fan Dingjia. (2019). Design of flight control system for six-rotor UAV (Master's thesis). North University of China.

[7] Yang Jiankang. (2020). Design and implementation of grasping system for rotor UAV (Master's thesis). Nanjing University of Information Science and Technology. DOI:10.27248/d.cnki.gnjqc.2020.000378.

[8] Qiu Pengrui. (2022). Research on vision-based pose estimation and flight control of quadrotor UAV (Master's thesis). Kunming University of Science and Technology. DOI:10.27200/d.cnki.gkmlu.2022.000029.
[9] Shao Linwen, Liao Fang, Ding Liming, Shu Wei, He Zhiqian. (2021). PID simulation design of quadrotor UAV flight control based on MATLAB. Shanxi Electronic Technology, (05), 43-46.
[10] Zhang Xuejin, Qin Yi. (2015). Research and application of four-wheeled car based on Arduino. Today's Electronics, (06), 57-59.
[11] Li Zhiqiang, Kang Qinqing, Xiao Yuliang, Li Bengao, Huang Ming, Wang Dongtao. (2022). Design and implementation of intelligent car based on Arduino. Wireless Internet Technology, 19(16), 43-46.
[12] Fu Xiaoyun. (2021). Design and development of intelligent obstacle avoidance car based on Arduino typical sensors. Precision Manufacturing and Automation, (02), 25-29. DOI:10.16371/j.cnki.issn1009-962x.2021.02.006.
[13] Jin Ma, Ding Liang, Gao Haibo, Su Yang, Zhang Pinjia. (2023). Dynamics modeling and simulation of a hexapod robot with a focus on trajectory prediction. Journal of Intelligent & Robotic Systems, 108(1).
[14] Yao Shun, Yao Yanan, Liu Ran, Liu Chao. (2022). A portable off-road crawling hexapod robot. Journal of Field Robotics, 39(6).
[15] Wangyue Di, Xin Tong, Zhigang Li. (2022). Tracking control of the six-rotor UAV with input dead-zone and saturation. IAENG International Journal of Applied Mathematics, 52(4).

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.5

Models and Concepts of AI Agents in Financial Operations for Autonomous Payroll Processing

Shanmuka Siva Varma Chekuri

Data Engineer, American Software Group (ASG), United States, New Jersey

Received: 29 Jan 2026; Received in revised form: 25 Feb 2026; Accepted:02 Mar 2026; Available online: 06 Mar 2026

Abstract *— The study examines models and architectural concepts of AI agents embedded in financial operations to support autonomous payroll processing in multi-region, compliance-sensitive environments. The research focus lies on agentic patterns that orchestrate payroll data flows end-to-end over ACID-compliant lakehouse platforms, feature stores, and metadata-driven orchestration layers. The work synthesizes current literature on AI agents in finance, autonomous decision systems, data lakehouse architectures, and feature-store-centric machine learning pipelines, and combines it with design patterns emerging in production-grade payroll and HR systems. Particular attention is given to deterministic computation graphs, idempotent pipelines, concurrency-safe merge patterns, and real-time observability for reconciliation and audit. The article aims to formulate a conceptual model of an autonomous payroll stack, outline classes of domain-specific agents, and identify their limitations and governance needs. The material is intended for researchers and practitioners in AI, data engineering, and financial systems design.*

Keywords *— **autonomous payroll, AI agents, financial data engineering, lakehouse architecture, feature store, metadata-driven orchestration, idempotent pipelines, compliance automation, multi-agent systems, real-time reconciliation.***

I. INTRODUCTION

Rapid diffusion of AI agents into financial services reshapes how organizations design payment, settlement, and payroll infrastructures. Large institutions gradually replace rule-based batch pipelines with systems where autonomous software agents monitor data streams, enforce policies, and trigger transactions with limited human supervision. Recent studies underline that agentic AI in finance no longer affects only analytics; it structurally changes intermediation, risk transfer, and operational workflows in payments and asset management [1].

Global payroll represents a particularly demanding field for such transformation. Multi-country organizations operate diverse labor codes, tax regimes, benefits schemes, and collective agreements. Errors propagate into banking interfaces, general ledgers, and regulatory reporting, creating exposure that cannot be mitigated purely through manual control. Industry and academic work on generative and autonomous agents in finance points toward architectures where specialized agents continuously validate calculations, monitor anomalies, and maintain regulatory alignment across jurisdictions.

The purpose of the article is to systematize models and concepts of AI agents suited for autonomous payroll processing and to connect them with modern financial data-engineering patterns. The research pursues three tasks:

1) to classify functional archetypes of AI agents that act across the payroll lifecycle, from eligibility determination to payment reconciliation;
2) to relate these archetypes to a reference lakehouse- and feature-store-centric architecture that supports ACID transactions,

schema evolution, and multi-region data harmonization;

3) to assess the limitations, risk factors, and governance requirements that arise when such agents participate in production payroll flows under regulatory scrutiny.

The novelty of the work lies in unifying three lines of research that are usually treated separately: conceptual analyses of AI agents in finance and decentralized markets, surveys of data lakehouse architectures and sustainable financial data platforms, and technical work on feature stores and real-time ML pipelines.The article reinterprets these results through the lens of global payroll, where deterministic behavior, traceability, and explainability are not optional but foundational design requirements.

II. MATERIALS AND METHODS

The study draws on a corpus of recent work dealing with AI agents in finance, autonomous decision systems, and financial data platforms. I. Aldasoro [1] analyses how artificial intelligence, including emerging AI agents, alters key financial system functions and stability concerns, providing a macro-level framing for agentic finance. K. Allam [2] examines cloud-based data hubs and SQL pipelines for real-time financial analytics, illustrating practical paths from monolithic warehouses to more flexible streaming and hub-and-spoke architectures in financial institutions. L. Ante [3] investigates autonomous AI agents in decentralized finance, identifying agent typologies, coordination patterns, and governance tensions that inform the design of autonomous financial workflows. J. de la Rúa Martínez [4] introduces the Hopsworks feature store and describes an FTI (feature–training–inference) architecture that relies on centralized feature governance and mixed offline/online serving, which is directly relevant for payroll-related models. A. A. Harby [5] and co-author present a survey of data lakehouse architectures in Information Systems, synthesizing how lakehouses merge warehouse reliability with data-lake scalability and how ACID transactions, unified metadata, and open table formats contribute to AI-ready platforms. I. Hettiarachchi [6] focuses on generative AI agents in finance and analyzes operational disruption scenarios, including multi-agent systems built on large language models. S. A. Ionescu [7] investigates sustainable data architectures for modern financial institutions, highlighting lakehouse-style designs, hybrid cloud deployment, and governance for resilient financial data infrastructures. W. Liang [8] and co-authors discuss scalable and secure feature stores in real-time ML pipelines, emphasizing feature governance, low-latency serving, and security constraints in production ML. K. U. S. Somarathna [9] proposes an agent-based simulation approach for human resource management as a complex system, providing conceptual foundations for multi-agent reasoning over workforce and payroll-related variables. Z. Yordanova [10] and Y. Hristizov explore the evolution of financial analysis toward AI agents, mapping how autonomous systems assume decision authority in financial workflows.

For the article, comparative analysis is applied across these sources to extract converging patterns of agent design, data-platform architecture, and governance in financial systems. Conceptual modeling is used to define a taxonomy of AI agent types for payroll operations and to map their interactions onto an ACID-compliant lakehouse with an enterprise feature store. Structural–functional analysis supports decomposition of the autonomous payroll lifecycle into discrete, agent-addressable tasks. Synthesis and abstraction form the basis for the reference architecture and for the conceptual figure that integrates findings from the surveyed works. Throughout, the reasoning is aligned with real-world constraints of regulated payroll environments, including idempotent processing, concurrency-safe merge patterns, multi-region data residency, and traceable computation graphs.

III. RESULTS

Research on AI agents in finance indicates that agentic systems progress from decision-support tools to semi-autonomous and fully autonomous operators embedded in financial infrastructure. Aldasoro et al. highlight that generative models and AI agents intersect with core payment and settlement functions, with particular relevance for information acquisition, signal extraction, and execution of rule-driven actions [1]. Yordanova and Hristizov describe a similar

trajectory for financial analysis, where systems evolve from static analytics to goal-directed agents that interpret data, propose actions, and implement them within predefined bounds [10]. When these insights are transferred to payroll, they point toward a layered architecture where multiple specialized agents cooperate to maintain payroll correctness, compliance, and timeliness across regions.

First, the literature supports clear differentiation between strategic agents, operational agents, and infrastructure-level agents. Generative AI agents described by Hettiarachchi operate at strategic and tactical levels, synthesizing information, detecting patterns, and proposing interventions in areas such as trading and risk [6]. By analogy, an autonomous payroll environment benefits from strategic agents that monitor long-term cost dynamics, overtime trends, and regional exposure, while operational agents focus on rule execution, calculation validation, and workflow automation. Ante's typology in decentralized finance, which distinguishes AI agents by autonomy degree and governance arrangement, provides a template for defining payroll-specific agent roles across eligibility checks, accrual derivation, tax calculation, payment scheduling, and anomaly triage [3].

Second, data-engineering work on lakehouses and sustainable financial architectures indicates that autonomous agents reach full potential only when grounded in a unified, high-quality data plane. Harby and Zulkernine show that data lakehouses combine warehouse-style ACID guarantees with lake-style scalability and schema flexibility [5]. Ionescu demonstrates similar trends in financial institutions, where hybrid lakehouse designs simplify integration of transactional stores, risk systems, and historical archives while maintaining strong governance [7]. For payroll, such a lakehouse hosts canonical employee, contract, and transaction records across jurisdictions, with slowly changing dimensions (SCD-2) capturing historical versions of employment terms. Concurrency-safe merge patterns and ACID transactions protect against partial updates when agents simultaneously apply corrections, reclassifications, or back-dated adjustments.

Feature-store research adds a further layer to this picture. De la Rúa Martínez and co-authors describe an FTI architecture in which a feature store supplies consistent features to training and inference pipelines, eliminating offline/online skew [4]. Liang et al. emphasize that secure, real-time feature stores standardize feature computation, control access, and support low-latency inference in production ML pipelines [8]. In payroll automation, such a feature store aggregates contract attributes, time-sheet aggregates, historical payments, tax brackets, and compliance signals into reusable features. Validation agents then consume these features deterministically, applying rules and ML-based estimators within a computation graph whose nodes and edges are fully versioned and observable. This architecture supports idempotent pipelines: if an upstream data source sends duplicate or corrected events, agents recompute downstream states without double-paying or corrupting balances.

Figure 1 integrates these strands into a conceptual architecture for autonomous payroll on AI-ready financial infrastructure. The diagram follows the structure of lakehouse and feature-store architectures described by Harby & Zulkernine and de la Rúa Martínez et al., adapted to the payroll domain [4–5].

At the foundation, an ACID-compliant lakehouse consolidates raw HR, time-tracking, benefits, banking, and tax data into harmonized tables partitioned by region and period. Above this plane, the enterprise feature store exposes curated payroll features: gross-to-net factors, regulatory thresholds, eligibility indicators, and audit metrics. A metadata-driven orchestration layer defines deterministic computation graphs that describe how agents traverse states: ingestion, normalization, feature computation, rule evaluation, payment execution, and reconciliation. On top of this layer, several classes of AI agents operate:

- Ingestion and normalization agents align heterogeneous feeds, enforce schema evolution policies, and record lineage for each transformation step. Their behavior reflects practices described in sustainable financial data architectures and real-time financial analytics pipelines [2,7]
- Policy and compliance agents encode labor laws, tax rules, and internal policies as executable rule graphs. Building on ideas from agentic

finance, they interpret regulatory signals, update rule sets, and flag conflicts between jurisdictions [1,6].

- Calculation and anomaly-detection agents compute payroll outputs, compare them against historical baselines, and use ML models trained over feature-store data to identify anomalies such as outlier net pay, misclassified contracts, or missing contributions [4,8,10].
- Reconciliation and audit agents align payroll results with general ledger postings, bank movements, and third-party system data. Insights from work on autonomous agents and decentralized finance—where agents coordinate across multiple ledgers—inform the design of these reconciliation flows [3].

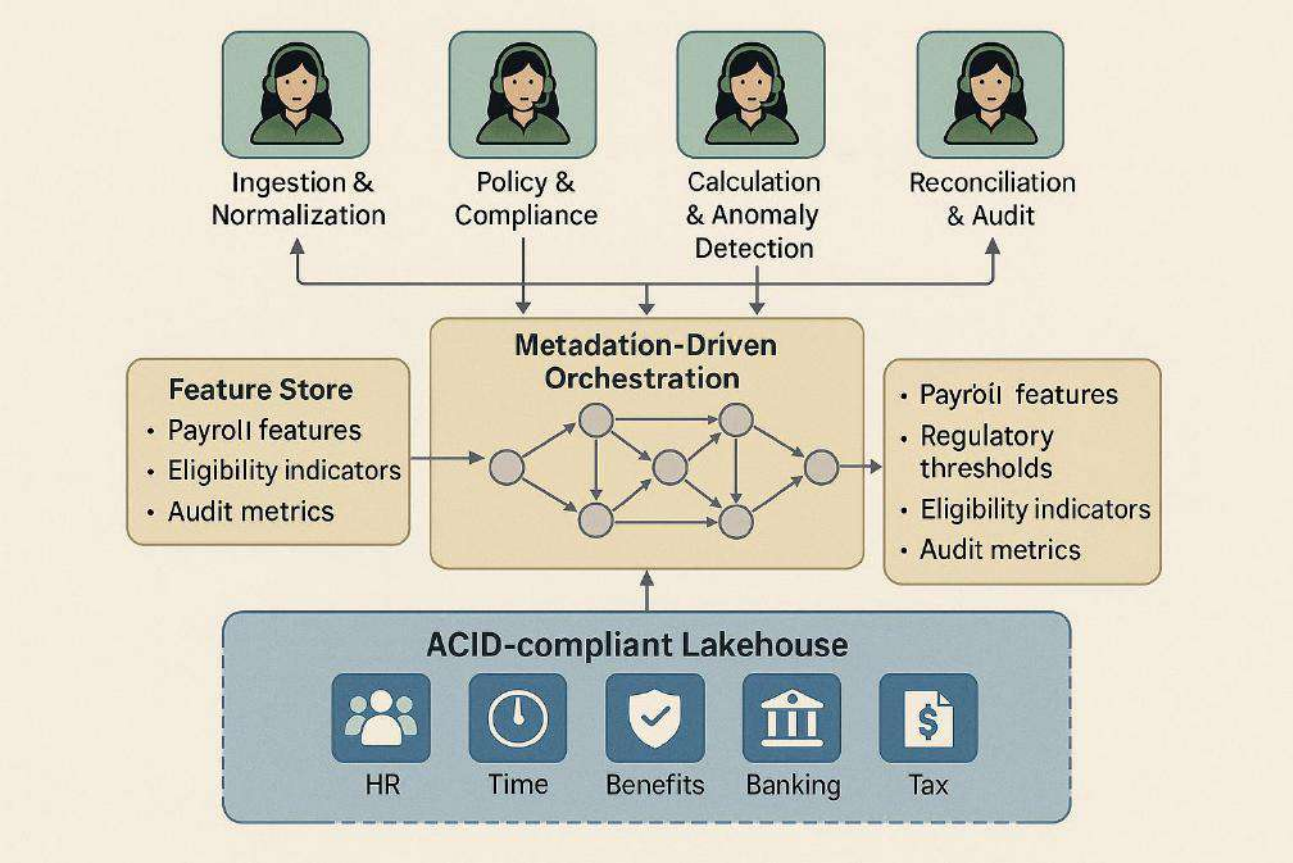


Fig.1. Conceptual multi-agent lakehouse architecture for autonomous payroll processing (compiled by author based on [4–5,7–8]).

Agent-based modeling work in HR, such as the simulation of human resource management strategies by Somarathna, shows how agent interactions and local decision rules generate emergent organizational dynamics [9]. Translating such ideas into payroll, one can model agents representing employees, managers, regulators, and systems, and simulate how modifications to policies, data-quality regimes, or scheduling logic influence error rates, rework, and compliance risk. This perspective motivates design of autonomous payroll agents not as isolated scripts but as entities embedded in a socio-technical system with feedback from human operators and regulators.

The surveyed literature further clarifies that such agentic payroll environments depend on real-time observability and lineage. Harby and Ionescu both emphasize unified metadata layers and catalogues that track table schemas, versions, and access patterns [5,7]. In a payroll setting, every decision made by an agent—such as assigning a tax code or resolving a conflict between overlapping allowances—must be traceable back to data versions, rule definitions, and model parameters in force at the decision time. Feature-store architectures add fine-

grained lineage at the level of individual features and their transformation pipelines [4,8]. Together, these elements make post-hoc audit, rollback, and re-simulation feasible, which is central for highly regulated payroll operations.

Finally, several works highlight sustainability, security, and governance as cross-cutting concerns. Ionescu points out that financial institutions seek architectures that balance performance with energy consumption and operational resilience [7]. Liang et al. stress secure access control, encryption, and isolation in feature stores [8]. Aldasoro et al. and Hettiarachchi address systemic risk, regulatory adaptation, and governance for AI agents in finance [1,6]. For autonomous payroll, these insights translate into requirements for least-privilege access for agents, verifiable guardrails around payment initiation, and explicit alignment with responsible-AI frameworks. In combination, the materials enable a coherent conceptual result: an autonomous payroll system constructed as a set of domain-specific agents operating over an AI-ready, lakehouse-based data platform with feature-store-centric ML, deterministic orchestration, and full observability.

IV. DISCUSSION

The conceptual architecture outlined above presents both opportunities and constraints when reconciled with the surveyed literature. Works on AI agents in finance consistently signal a shift from passive analytics toward systems that act as economic participants. Aldasoro et al. describe AI agents as entities that influence price formation, risk distribution, and stability channels [1]. Yordanova and Hristizov extend this picture by arguing that AI agents increasingly undertake analytical, forecasting, and decision-making functions previously delegated to financial analysts [10]. In payroll, a similar migration from advisory dashboards to execution-capable agents is technically feasible but institutionally constrained by labor law, collective bargaining, and organizational risk appetite.

Table 1 summarizes how insights from the reviewed financial-agent literature map onto concrete responsibilities in the payroll lifecycle. The mapping synthesizes themes from generative AI agents in finance, decentralized-finance agents, and financial-analysis agents [1,3,6,10].

Table 1. Functional allocation of AI agents across the payroll lifecycle (based on [1,3,6,10])

Payroll lifecycle stage	Dominant agent capabilities	Illustrative literature link
Data ingestion and normalization	Schema alignment, anomaly detection in source feeds, policy-aware filtering	Generative AI agents for data preparation and monitoring [6,10]
Eligibility and accruals	Rule evaluation, scenario exploration, explanation generation	AI agents as decision-support systems in finance [1,10]
Gross-to-net computation	Deterministic rule execution, conflict resolution between overlapping rules	Autonomous agents coordinating financial operations [3,6]
Compliance and controls	Constraint checking, alert triage, justification logging	Agentic AI for regulatory functions and governance [1,6,10]
Payment execution and routing	Workflow orchestration, exception management, retry logic	Operational agents coordinating financial transactions [3]
Reconciliation and audit	Multi-ledger matching, variance analysis, trace reconstruction	Multi-agent coordination in financial ecosystems [1,3]

The table illustrates that many agent capabilities required for payroll already appear in more general financial domains, though they must be adapted to specific payroll artifacts such as pay periods, earnings codes, and jurisdiction-specific tax constructs. The literature on DeFi agents emphasizes coordination across multiple ledgers and governance frameworks, which parallels the need to reconcile payroll with ERPs, banking systems, and statutory portals [3]. Generative AI work, in turn, suggests that

explanation-oriented agents can help bridge the gap between autonomous operation and human oversight by producing narrative justifications for complex decisions [6,10].

The second cluster of findings concerns the underlying data platform. Lakehouse surveys show that successful deployments rely on a convergence of ACID table formats, unified catalogs, and fine-grained governance, which together support both analytical and operational workloads [5,7]. Feature-store work adds another layer by demonstrating how shared, governed feature repositories enable consistent training and inference across batch and streaming contexts [4,8]. Table 2 compares three architectural patterns that emerge from the reviewed materials in relation to autonomous payroll: a traditional warehouse-centric stack, a lakehouse without a feature store, and a lakehouse with an enterprise feature store and metadata-driven orchestration.

Table 2. Architectural patterns for payroll automation and their suitability for AI agents (based on [2,4–5,7–8])

Architecture pattern	Data characteristics	Suitability for autonomous payroll agents
Legacy warehouse-centric ETL	Rigid schema-on-write, nightly batches, limited history of schema changes	Constrains agent autonomy; limited real-time feedback; brittle under schema evolution
Lakehouse without feature store	ACID tables over semi-structured data, scalable storage, partial streaming support	Supports foundational observability, but ML features are duplicated across teams; higher risk of offline/online skew
Lakehouse + enterprise feature store + orchestration	Unified tables with SCD-2 dimensions, governed feature entities, deterministic orchestration graphs	Aligns with agentic operation: reproducible decisions, idempotent pipelines, low-latency feature access, strong lineage

The comparison reflects arguments from Allam, Harby, Ionescu, de la Rúa Martínez, and Liang regarding the shift from rigid ETL pipelines toward flexible, metadata-centric architectures in finance [2,4–5,7–8]. In the first pattern, agents are constrained to operate on static snapshots; recomputation requires full batch reruns, and partial failures often force manual intervention. In the second pattern, agents gain real-time data access, but absence of a feature store fragments feature logic across codebases, undermining determinism. The third pattern brings together lakehouse guarantees, feature governance, and metadata-driven orchestration, creating an environment where agents can recompute fallible steps, replay transactions, and rely on a stable semantic layer.

Despite these advantages, the surveyed works on AI agents and financial architectures repeatedly caution about systemic risks and governance gaps. Aldasoro et al. draw attention to feedback loops between AI-driven decision systems and financial stability [1]. Hettiarachchi underscores that generative AI agents introduce new operational-risk dimensions around explainability, drift, and adversarial behavior [6]. Liang and colleagues stress attack surfaces around feature-store access, data poisoning, and leakage of sensitive features [8]. Translated to payroll, these concerns manifest as mis-classified workers, systematically biased rule encoding, or silent failures in anomaly detection during periods of regulatory change. Mitigation requires defense-in-depth: layered validation agents, independent monitoring services, and human approval gates for high-impact operations such as payment release and retroactive corrections.

A further dimension emerging from Ionescu's work concerns sustainability and resource efficiency [7]. Agentic systems in finance can demand intensive compute resources, particularly when generative models or simulation-heavy decision policies are involved. For a global payroll platform running on a lakehouse, design choices around feature computation frequency, model refresh cadence, and aggregation granularity influence both cost and energy consumption. Combining sustainable-architecture principles with feature-store governance encourages approaches such as incremental feature

updates, adaptive sampling for anomaly detection, and load-aware orchestration—choices that preserve responsiveness for critical controls while limiting unnecessary resource use.

Finally, literature on agent-based modeling in HR suggests that socio-technical alignment matters as much as technical soundness [9,10]. Payroll processes involve HR, finance, compliance, and workforce representatives whose trust in autonomous agents depends on transparency, traceability, and controllable failure modes. The conceptual model derived in the results section acknowledges this by assigning explanation-oriented responsibilities to certain agents and by grounding all decisions in auditable data and rule artifacts. The surveyed materials thus jointly support a view of autonomous payroll not as a fully self-driving system, but as a layered, agent-enabled platform where automation and governance are deliberately co-designed.

V. CONCLUSION

The synthesis of recent research on AI agents in finance, decentralized markets, data lakehouse architectures, sustainable financial data platforms, and feature-store-centric ML pipelines suggests that autonomous payroll processing becomes practicable when three conditions are met. First, functional responsibilities across the payroll lifecycle are assigned to specialized agents whose behaviors are bounded by explicit rules, structured features, and deterministic computation graphs rather than opaque heuristics. Second, these agents operate over an AI-ready financial data platform that combines an ACID-compliant lakehouse, an enterprise feature store, and metadata-driven orchestration, enabling idempotent pipelines, concurrency-safe merge patterns, and comprehensive lineage. Third, governance, security, and sustainability measures are integrated into the design from the outset, addressing concerns highlighted in the literature regarding explainability, systemic risk, and resource consumption.

The tasks set in the introduction are thereby resolved: a taxonomy of agent types tailored to payroll operations is formulated, their interaction with modern financial data architectures is described, and their limitations and governance implications are analyzed using evidence from the surveyed works. For practitioners in data engineering and AI who design payroll and financial platforms, the article provides a conceptual blueprint that aligns agentic automation with regulatory requirements and enterprise-grade reliability, while leaving room for future research on empirical evaluation, formal verification, and large-scale deployment outcomes.

REFERENCES

[1] Aldasoro, I., Gambacorta, L., Korinek, A., Shreeti, V., & Stein, M. (2024). Intelligent financial system: How AI is transforming finance (BIS Working Paper No. 1194). Bank for International Settlements.

[2] Allam, K. (2024). Cloud-based data hubs and SQL pipelines for real-time financial analytics. International Journal of Emerging Trends in Computer Science and Information Technology.

[3] Ante, L. (2024). Autonomous AI agents in decentralized finance: Market dynamics, application areas, and theoretical implications. SSRN preprint. https://doi.org/10.2139/ssrn.5055677

[4] de la Rúa Martínez, J., et al. (2024, June). The Hopsworks feature store for machine learning. In Proceedings of the 2024 ACM SIGMOD International Conference on Management of Data (SIGMOD '24).

[5] Harby, A. A., & Zulkernine, F. H. (2025). Data lakehouse: A survey and experimental study. Information Systems, 127, 102460.

[6] Hettiarachchi, I. (2025). The rise of generative AI agents in finance: Operational disruption and strategic evolution. International Journal of Engineering Technology Research & Management, 9(4).

[7] Ionescu, S. A. (2025). Engineering sustainable data architectures for modern financial institutions. Electronics, 14(8), 1650.

[8] Liang, W., Taofeek, A., Rajuroy, A., Blessing, M., & Hamzah, F. (2025). Developing scalable and secure feature stores in real-time machine learning pipelines. Manuscript in preparation / preprint.

[9] Somarathna, K. U. S. (2020). An agent-based approach for modeling and simulation of human resource management as a complex system: Management strategy evaluation. Simulation Modelling Practice and Theory, 104, 102118.

[10] Yordanova, Z., & Hristizov, Y. (2025). The evolution of financial analysis: From manual methods to AI and AI agents. Economics – Innovative and Economics Research Journal, 13(3), 219–239.

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.6

Visual Communication in Academic Marketing: The Impact of Digital Assets on Stakeholder Perception

Simran Ratnani

Assistant Director, Marketing University of North Texas, Dallas, USA

Received: 04 Feb 2026; Received in revised form: 03 Mar 2026; Accepted:08 Mar 2026; Available online: 12 Mar 2026

Abstract *— The article examines how visual communication practices in academic marketing shape stakeholder perception when universities deploy digital assets across websites and short-form video platforms. Relevance is driven by intensified competition for attention, heightened sensitivity to credibility signals in digitally mediated interactions, and expanding use of AI-supported creative production. Novelty lies in integrating evidence on visual design richness and processing outcomes, institutional visual identity effects, and the trust consequences of AI-related disclosures into a single analytic account of stakeholder judgment formation. The study aims to explain how specific categories of digital assets — identity systems, still imagery, motion graphics, short videos, and AI-generated creative — alter perceived credibility, institutional reputation, and behavioral intentions among prospective students, current students, alums, donors, and partners. The work applies analytical synthesis of recent peer-reviewed research and open research outputs, using comparative analysis and structured source review. The conclusion formulates design and governance implications for academic marketing teams seeking higher persuasion efficiency under trust constraints.*

Keywords *— **visual communication, academic marketing, digital assets, stakeholder perception, university branding, short-form video, trust, AI disclosure, design richness, misinformation.***

I. INTRODUCTION

Academic marketing increasingly depends on visual communication because stakeholder judgments about universities are frequently formed in environments where direct experience is limited and informational asymmetries remain high. In such settings, digital assets—visual identity elements, imagery, interface micro-visuals, and platform-native video—operate as credibility cues that compress complex institutional attributes into fast, interpretable signals. The growing use of short-form video in institutional outreach adds pressure for rapid meaning-making. At the same time, attention competition increases incentives for visually "richer" executions that can either strengthen interpretation or overload it. A parallel development concerns AI-assisted content production and the governance of AI-marked communication, where disclosure and labeling practices influence trust and ad evaluations in non-trivial ways.

Aim – to analytically explain how the composition and governance of digital assets in academic marketing affects stakeholder perception. Objectives:

1. To systematize digital asset categories used in academic marketing and link them to perception pathways grounded in recent empirical marketing and communication research.
2. To clarify how trust, credibility, and authenticity judgments shift when AI disclosures or AI-content labels are present in promotional communication.

3. To derive design and governance implications for academic marketing teams that balance attention capture with credibility preservation under platform conditions dominated by rapid consumption and misinformation concerns.

Novelty – an integrated analytic model that connects visual design richness and processing outcomes, institutional visual identity and reputation formation, and AI-related transparency mechanisms within a stakeholder-oriented academic marketing frame.

II. MATERIALS AND METHODS

The analysis relies on a curated set of recent scholarly sources addressing digital marketing assets and capability deployment, visual design richness and processing consequences, university visual identity and reputation management, short-form video marketing and trust, and AI-generated content transparency and disinformation pressures. D. Hagen, A. Risselada, B. Spierings, J. W. J. Weltevreden, and O. Atzema [1] examined how digital marketing activities and resources influence adoption and update behavior for core digital channels, supporting the asset-governance logic in distributed organizations. Y. Bashirzadeh, R. Mai, and C. Faure [2] investigated how combining visual design elements can produce enrichment or clutter, informing the efficiency boundary for "rich" academic creatives. P. Dwitasari and colleagues [3] analyzed how visual identity elements shape perceptions of institutional identity and reputation, directly anchoring academic branding claims. J. L. Grigsby and colleagues [4] tested how generative-AI disclosures affect trust and ad attitudes, offering evidence for transparency trade-offs in promotional messaging. F. Li and colleagues [5] measured how AI-generated content labels influence perceived accuracy, credibility, and sharing intention in misinformation settings, supplying a labeling mechanism relevant to governance. A. Ishino, Y. Nakao, K. Kokubu, and H. Okada [6] studied how images in corporate reporting affect impressions, contributing methodological precedent for visual content analysis transferable to academic marketing artifacts. D. Stoica and colleagues [7] addressed reputation management drivers in higher education, connecting stakeholder perception to communication strategy. N. Rawangngam and colleagues [8] modeled consumer response in TikTok marketing with explicit attention to trust under rapid consumption. L. Theodorakopoulos and colleagues [9] reviewed interactive viral marketing through big-data analytics while discussing misinformation and deepfake risks, framing the integrity threat space. A. López-Borrull and colleagues [10] mapped generative AI's role in disinformation, strengthening the rationale for trust-preserving design and disclosure.

Methods: structured analytical synthesis, comparative method across communication settings, and critical analysis of empirical findings and conceptual frameworks from the selected sources.

III. RESULTS

Across the reviewed literature, stakeholder perception in academic marketing emerges as a function of how digital assets manage two competing pressures: meaning intensification and credibility preservation. Universities deploy digital assets to compress institutional quality cues (academic rigor, student outcomes, research culture, inclusion, and employability) into recognizable visual signals that stakeholders can process quickly. Visual identity systems—logos, color regimes, typographic consistency, photography style, and template discipline—support coherence; empirical evidence in higher education branding shows that visual identity management influences perceived institutional identity and reputation, and alignment between internal and external perception supports stronger institutional positioning [3]. A reputational pathway becomes visible: visual identity coherence reduces interpretive ambiguity, which supports reputational stability when stakeholders encounter fragmented information streams across multiple channels [7].

Digital assets do not act only as identifiers; they modify cognitive processing conditions. Research on visual design richness indicates that single visual design elements can contribute positive "enrichment," while combinations can generate "clutter," interrupt processing flow, and reduce effectiveness [2]. For academic marketing, this yields an actionable boundary: adding motion overlays, pictographs, stickers, emoji-like elements, and dense iconography to already information-heavy university

messages can shift the audience from interpretive gain to overload. The exact mechanism applies to multi-claim program ads: when scholarship information, rankings, campus lifestyle, and employability outcomes are visually layered in one creative, the probability of clutter-mediated skepticism increases [2]. From a stakeholder perspective, overload can be misread as persuasion pressure, which harms credibility—particularly for audiences with high involvement, such as prospective graduate students, faculty recruits, research sponsors, and donors.

Short-form video assets add a second processing constraint: speed. TikTok-oriented marketing research emphasizes that trust becomes central in short-form video settings because content is rapidly consumed and interactions are largely intangible, strengthening reliance on heuristics and source cues [8]. In academic marketing, this means stakeholder perceptions may be disproportionately influenced by creator identity signals (official university channel versus student ambassador versus third-party media), production choices that communicate authenticity, and narrative structures that convey transparency rather than sales intensity. Under these conditions, visual communication succeeds when it creates verifiable micro-evidence: glimpses of real facilities, unedited classroom moments, faculty presence, and consistent branding markers that connect the clip to institutional ownership, reducing misattribution risk in fast-scrolling feeds [8].

A separate but converging pathway concerns AI-supported creative production and transparency practices. Evidence from advertising experiments indicates that AI disclosures can reduce trust and worsen ad attitudes, and that the design focus (intangible versus tangible cues) interacts with disclosure effects [4]. Translating this to academic marketing, an AI disclosure placed on a scholarship campaign featuring AI-generated student portraits introduces a credibility tax: stakeholders may infer artificiality, question representativeness, and reduce perceived institutional sincerity—mainly when the message relies on intangible promises (belonging, transformation, "future-ready" identity) rather than tangible evidence (labs, faculty expertise, curriculum, accreditation artifacts) [4]. Complementary evidence from randomized experiments on AI-generated content labels in misinformation contexts shows that labels can have modest average effects, interact with content type, and be perceived as driven by profit motives. It may shape how audiences distinguish AI-generated from human-generated content without straightforwardly improving credibility [5]. In academic marketing, where persuasion intersects with public trust, a naive "label-and-forget" approach does not address reputational risk; labeling must align with asset choice and claim type.

Integrity threats further constrain the asset strategy. Analyses of interactive viral marketing highlight how deepfakes, false endorsements, and manipulative tactics complicate platform trust, motivating stronger verification and governance in marketing practice [9]. Mapping studies of generative AI's disinformation impact reinforce that synthetic media can erode trust and accelerate misattribution at scale [10]. Academic brands face an exposure profile distinct from consumer brands: reputational harm extends to admissions conversion, alums giving, partnership formation, and faculty recruitment, so the cost of a trust shock is multiplied across stakeholder groups. Under this condition, digital assets that prioritize credibility signals outperform those optimized only for reach.

The findings cohere into a perception mechanism: digital assets influence stakeholder perception through (i) identity coherence, (ii) processing fluency versus clutter, (iii) trust heuristics in short-form video, and (iv) transparency governance for AI-supported assets. A practical implication follows: the optimal asset mix is not "maximum richness," but "evidence-weighted richness," in which visual complexity is calibrated to achieve verifiability and stakeholder risk sensitivity.

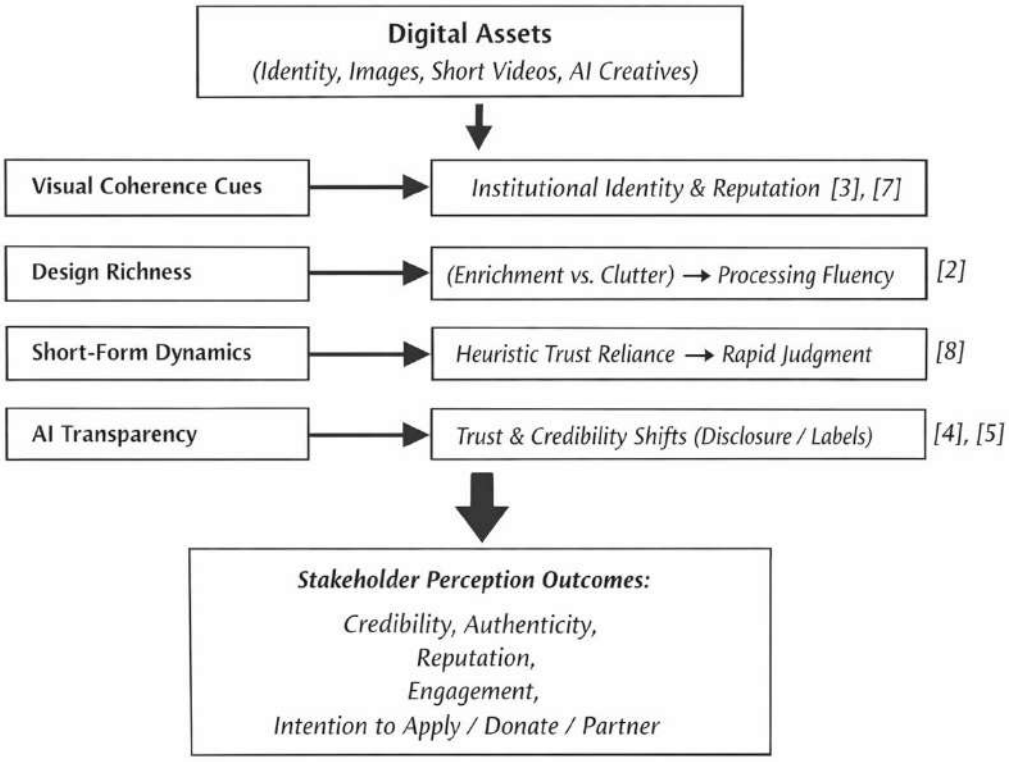


Fig.1. Evidence-weighted visual communication mechanism for academic marketing

Taken together, the reviewed evidence indicates that stakeholder perception is most resilient when academic marketing assets couple recognizability with verifiable cues and avoid the enrichment-to-clutter tipping point identified in digital visual design research. The results also show that AI transparency practices shape credibility, with message intangibility and content type influencing this effect, implying that disclosure and labeling require channel-sensitive governance rather than uniform application. Finally, the short-form environment intensifies heuristic evaluation, so consistency of institutional signifiers and proof-like micro-evidence becomes a functional substitute for slower, deliberative verification. This effect is especially consequential under pressure from disinformation.

IV. DISCUSSION

The synthesized evidence supports a stakeholder-centered interpretation: academic marketing assets function as governance instruments rather than merely creative outputs. Visual identity coherence serves as a stabilizer of institutional meaning across multiple channels, consistent with findings that visual identity affects perceived institutional identity and reputation [3] and that reputation management in higher education is linked to communication strategy and stakeholder perceptions [7]. From this lens, a university's digital asset library should be treated as reputational infrastructure: templates, photography standards, motion rules, and disclosure conventions define how institutional claims are encoded into visuals and how those claims survive platform redistribution.

A central tension concerns the pursuit of attention through visual richness. Empirical work on visual design elements shows that combining elements can shift perception toward clutter, weakening processing flow [2]. Academic marketers often face pressure to condense multiple proof points into a single creative. The evidence suggests that this compression can backfire when stakeholders are evaluating high-stakes decisions such as enrollment or partnership. A disciplined approach would allocate richness to assets where enrichment is likely (single-claim visuals, clear narratives) and restrict richness where clutter risk rises (multi-claim posters, dense carousel frames).

A second tension concerns AI-assisted content production. Experimental evidence indicates that AI disclosure can depress trust and ad attitudes [4], whereas labeling in misinformation settings shows nuanced, interaction-dependent effects rather than a universal trust boost [5]. For universities, the rational policy is not blanket avoidance of AI, but risk-tiered governance: AI-generated or AI-edited assets should be preferentially used for low-identity, low-representational-risk visuals (abstract illustrations, conceptual diagrams, non-human imagery) and avoided for assets that represent lived student experience unless high-fidelity provenance and transparency standards are in place [4], [5].

Short-form video expands both opportunity and vulnerability. Trust plays a decisive part in TikTok-style environments because rapid consumption amplifies reliance on credibility heuristics [8]. The literature on viral marketing and disinformation pressures emphasizes that synthetic media and manipulative tactics can degrade platform trust [9], [10], implying that academic marketing must embed authenticity signals and verification cues into the video grammar itself. This favors formats such as "day-in-the-life" clips anchored in recognizable campus landmarks, faculty micro-explanations tied to identifiable facilities, and consistent institutional signifiers that reduce the risk of impersonation.

Table 1. Digital-asset risk profile in academic marketing and literature-grounded mitigations [2–5; 7–10]

Asset pattern in academic marketing	**Primary perception risk**	**Mechanism reported in the literature**	**Practical mitigation grounded in sources**
High-density multi-claim visuals (many icons, overlays, micro-text)	Skepticism via overload	Richness combinations can produce clutter and disrupt processing	Reduce simultaneous design elements; separate claims into distinct creatives
AI-generated human imagery (students, faculty likeness proxies)	Authenticity loss; trust penalty	AI disclosures reduce trust; labels do not guarantee credibility gains	Prefer tangible evidence visuals; restrict AI humans; apply clear transparency and provenance notes
Short-form video optimized for virality without verification cues	Misattribution; credibility erosion	Trust reliance is high under rapid consumption; disinformation pressures increase risk	Embed institution identifiers; emphasize verifiable scenes; use consistent visual identity grammar
Fragmented visual identity across channels	Reputation inconsistency	Visual identity influences perceived institutional identity/reputation; reputation depends on communication strategy.	Centralize asset standards; enforce template discipline; align internal and external identity cues.

The highest reputational exposure concentrates in assets that combine representational sensitivity (depicting people) with AI assistance, and in assets that prioritize attention capture over verifiability. Both cases connect to experimentally documented trust penalties under AI disclosure [4] and to the heuristic processing patterns of short-form environments [8].

Table 2. Evidence-linked evaluation indicators for digital assets in academic marketing[2–5; 7–10]

Evaluation indicator	Operational meaning in academic marketing	Literature anchor
Perceived credibility message credibility	The stakeholder believes that the communication is reliable and not manipulative	AI labels and disclosure influence credibility judgments in experimental settings
Trust in source/brand trust	Confidence in the institution's integrity and competence is signaled through assets	Trust functions as a determinant in short-form platform environments
Processing fluency versus clutter	Ease of interpreting the message without overload	Enrichment–clutter trade-off under visual design richness
Institutional identity and reputation perception	Stakeholder impression of "who the institution is" and its standing	Visual identity effects on identity/reputation; reputation management dimensions include stakeholder perception and communication strategy
Integrity risk exposure	Probability that assets amplify misinformation dynamics or impersonation risk	Viral marketing ethics and deepfake/disinformation mapping highlight systemic threats.

Table 2 supports a measurement logic for non-experimental analytical work: even without primary data collection, academic marketing research can evaluate digital assets by mapping them to validated constructs used in recent studies—credibility, trust, processing outcomes, and reputation-linked perception—while situating governance concerns within the documented disinformation environment.

The discussion therefore supports a governance interpretation: academic marketing teams improve outcomes not by maximizing creative novelty, but by controlling how assets encode identity, manage processing demands, and maintain credibility under AI-mediated uncertainty. Evidence on visual identity and higher-education reputation suggests that coherence across touchpoints stabilizes interpretation. At the same time, research on richness and short-form trust explains why overdesigned creatives and unverifiable claims can trigger skepticism more quickly than in slower media. When combined with disclosure and labeling findings, this implies a structured decision logic: representationally sensitive assets warrant stricter provenance and transparency rules, whereas low-identity visuals can accommodate AI assistance with lower reputational exposure.

V. CONCLUSION

The analysis addressed the stated objectives by structuring academic marketing digital assets into perception-relevant categories and linking them to processing outcomes, clarifying that AI transparency mechanisms introduce trust trade-offs that depend on claim type and representational sensitivity, and deriving design and governance implications under short-form and misinformation pressures. The results support an evidence-weighted asset strategy: visual richness should be calibrated to the level of verifiability. In contrast, AI-assisted assets require risk-tiered governance, with priority given to tangible evidence cues and consistent institutional identity markers to protect stakeholder trust.

In applied terms, the article yields a stakeholder-oriented blueprint for academic marketing practice: curate asset libraries as

reputational infrastructure, calibrate visual richness to claim verifiability, and treat AI transparency as a design-and-governance parameter rather than a compliance afterthought. The proposed evidence-weighted approach follows directly from convergent findings on enrichment versus clutter, institutional visual identity effects, disclosure and labeling impacts on credibility, and trust heuristics in rapid-consumption platforms. Such an approach supports more stable stakeholder perceptions—credibility, reputation, and intention—across web, social, and short-form ecosystems, where misinformation risks and synthetic media have become persistent externalities.

REFERENCES

[1] Hagen, D., Risselada, A., Spierings, B., Weltevreden, J. W. J., & Atzema, O. (2022). Digital marketing activities by Dutch place management partnerships: A resource-based view. *Cities, 123*, 103548. https://doi.org/10.1016/j.cities.2021.103548

[2] Bashirzadeh, Y., Mai, R., & Faure, C. (2022). How rich is too rich? Visual design elements in digital marketing communications. *International Journal of Research in Marketing, 39*(1), 58–76. https://doi.org/10.1016/j.ijresmar.2021.06.008

[3] Dwitasari, P., Zulaikha, E., Hanoum, S., Alamin, R. Y., & Lee, L. (2025). Internal perspectives on visual identities in higher education: A case study of top-ranked universities in Indonesia. *F1000Research, 13*, 1535. https://doi.org/10.12688/f1000research.159232.2

[4] Grigsby, J. L., Michelsen, M., & Zamudio, C. (2025). Service ads in the era of generative AI: Disclosures, trust, and intangibility. *Journal of Retailing and Consumer Services, 84*, 104231. https://doi.org/10.1016/j.jretconser.2025.104231

[5] Li, F., & Yang, Y. (2024). Impact of artificial intelligence-generated content labels on perceived accuracy, message credibility, and sharing intentions for misinformation: Web-based, randomized, controlled experiment. *JMIR Formative Research, 8*, e60024. https://doi.org/10.2196/60024

[6] Nakao, Y., Ishino, A., Kokubu, K., & Okada, H. (2024). Exploring visual communication in corporate sustainability reporting: Using image recognition with deep learning. *Corporate Social Responsibility and Environmental Management, 31*(4), 3210–3234. https://doi.org/10.1002/csr.2735

[7] Stoica, D., Patriche, C.-C., David, S., Beldiman, C. M., & Dragomir Bălănică, C. M. (2025). Knowledge-driven reputation management in higher education institutions. *Journal of Innovation & Knowledge, 10*(6), 100811. https://doi.org/10.1016/j.jik.2025.100811

[8] Rawangngam, N., Pongsakornrungsilp, S., Pongsakornrungsilp, P., & Moghadas, S. (2025). TikTok marketing strategies and consumer response: A structural equation modeling study on purchase intention in Thailand. *Journal of Theoretical and Applied Electronic Commerce Research, 20*(4), 319. https://doi.org/10.3390/jtaer20040319

[9] Theodorakopoulos, L., Theodoropoulou, A., & Klavdianos, C. (2025). Interactive viral marketing through big data analytics, influencer networks, AI integration, and ethical dimensions. *Journal of Theoretical and Applied Electronic Commerce Research, 20*(2), 115. https://doi.org/10.3390/jtaer20020115

[10] López-Borrull, A., & Lopezosa, C. (2025). Mapping the impact of generative AI on disinformation: Insights from a scoping review. *Publications, 13*(3), 33. https://doi.org/10.3390/publications13030033

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.7

Application of Industrial Internet of Things in Fastener Forming Process Quality

Chih-Wei Hsu

Guangzhou Institute of Science and Technology, Guangzhou, China
e120247013@gmail.com

Received: 03 Feb 2026; Received in revised form: 05 Mar 2026; Accepted:10 Mar 2026; Available online: 13 Mar 2026

Abstract *— In the manufacturing industry, the current common practice for quality traceability and anomaly detection is to purchase a large number of inspection devices and implement full inspection of semi-finished and finished products to fulfill quality assurance commitments to customers. However, this approach consumes significant production time, labor, and related resources. Therefore, to save costs, many factories currently adopt "manual sampling inspection" for quality monitoring, but this method fails to achieve comprehensive quality management. To prevent the escalation of losses caused by defective products and to improve yield rates, full inspection of all finished products at a high cost becomes necessary.The Industrial Internet of Things (IIoT)-embedded fastener forming process system, through sensor deployment and machine networking to collect big data, can avoid decision-making models that previously relied solely on experienced operators and the substantial losses caused by delayed responses. This system eliminates the need to maintain inventory and tie up capital to handle large quantities of scrapped products. The approach described in this paper can halt machine operation before anomalies occur, preventing the escalation of losses. More importantly, through the recording of historical data, it enables traceability of factors related to personnel, machines, materials, methods, measurement, and environments – namely, 5M1E. Big data analysis of processing conditions helps identify the root causes of anomalies and implement effective countermeasures to achieve optimal condition settings, while also extending the manufacturing lifecycle.*

Keywords *— **Manual sampling inspection, Industrial Internet of Things (IIoT), Fastener forming process, Sensor deployment, Machine networking, Big data.***

I. INTRODUCTION

Screws and nuts are collectively referred to as fasteners, which are commodities widely used in applications ranging from basic civilian needs to high-tech fields and industries. Although the fastener industry is classified as a traditional industry, the consumption volume of fasteners is regarded as an indicator of a country's industrial development. According to data from Zion Market Research [1], the global industrial fastener market size was USD 92.56 billion. It is projected to reach USD 141.88 billion by 2032, exhibiting a Compound Annual Growth Rate (CAGR) of 4.86% during the forecast period from 2024 to 2032. The report provides a comprehensive analysis of the market, highlighting factors influencing market growth, potential challenges, and possible opportunities in the industrial fastener market over the next decade.

The prospects for the fastener industry are promising. However, a characteristic of traditional industries is the pursuit of mass production scale; greater production capacity naturally leads to greater production benefits, but when anomalies occur, the resulting losses are also substantial. Traditionally, the fastener forming process could only rely on numerous patrolling inspectors and the experiential

judgment of senior on-site technicians, as the machinery did not transmit key information, making it impossible to know the machine's status during production. Consequently, besides being labor-intensive, this approach also led to issues such as the mixing of defective products, and the inability to immediately resolve machine or other anomalies when they occurred. This resulted in an inability to improve fastener processing quality, excessive scrap, and reduced equipment uptime due to damage, compounded by delayed customer deliveries, leading to significant losses.

This study employs an Industrial Internet of Things (IIoT) approach, involving sensor deployment and machine networking to collect big data, embedded within the fastener forming process for real-time monitoring of fastener production. This effectively reduces the current heavy reliance on manual inspection and addresses machine production problems. The implementation process of using IIoT in the fastener forming process enables timely detection of quality anomalies and ensures maximum machine uptime. Simultaneously, it allows for advance scheduling of production downtime, online quality control, preventing defective products from mixing with good ones, and reducing customer complaint compensation.

II. LITERATURE REVIEW

The research and writings of previous scholars on topics such as "sensor data collection in metal forming processes" and "industrial internet of things applications" were collected, using the research findings of many predecessors as the theoretical foundation for this study.

2.1 Sensor Data Collection in Metal Forming Processes

The Industry 4.0 revolution stems from advancements in digitalization, enabling the acquisition of information on workpiece surface roughness and forming tool wear to improve productivity, reduce costs, or maintain constant tool life, making factories smarter and more efficient. Giampiccolo et al. [2] suggested that Small and Mid-size Enterprises (SMEs) implement data-driven methods, manage data processes, and review the impact of data quality to address key data challenges within companies (such as data volume and data accuracy). Their research developed a general model for the evolution of result accuracy from sheet metal forming data functions, subsequently generating data content and missing data, and proposed the concept of economic data volume for data infrastructure scaling.

Casbas-Gimenez et al. [3] utilized the versatility of piezoelectric sensors in measurement technology and enabled the sequential development of wireless and networked solutions by embedding sensors directly into machines, fixtures, and tools. In their study, piezoelectric sensors were used during ongoing operations to monitor the mechanical strength of metal structures in real-time. Through the control of mechanical components, the strength was studied using piezoelectric sensors, measuring the structural behavior and geometric changes under bending loads, and quantifying loads and micro-strains in specific time-frequency domains under operating conditions. Furthermore, modeling metal forming processes is a fundamental task in production engineering. Due to recent technological developments, a wide variety of models are available and continuously expanding, making it important to select the appropriate model.

Additional articles outline the classification and characterization of metal forming modeling and common models, and introduce the model selection process. Based on this model classification, the upcoming challenges in modeling constraints for various related metal forming processes are discussed [4,5]. Ralph et al. [6] utilized Digital Twins within a common system, using industrial software for data logging and analyzers for automated further processing, monitoring and controlling forming units with different technological maturity levels, and explored the possibility of connecting metal forming machines, simulation programs, and automation software using finite element analysis simulation procedures.

The above literature reviews relevant theories of metal forming processes and outlines how to design integration and response for fastener research.

2.2 Industrial Internet of Things Applications

Generally speaking, the overall architecture of the Industrial Internet of Things can be divided into

three layers: the perception layer, the network layer, and the application layer. First, equipment collects data through the perception layer, which is then transmitted from the network layer to the application layer for the most effective use [7]. Consequently, in studying production systems focused on data driving, equipment utilizes sensors for data collecting. The collected data primarily pertains to information relevant to the product lifecycle. Subsequently, through data transmission, individual pieces of equipment communicate with each other, aggregating the various aspects of information collected. This mainly involves using machine networking communication technologies to transmit product-related data in real-time to the network layer, enabling interconnection between machines or communication between humans and machines. The network layer is primarily constructed on wireless communication networks. Its main function is to virtualize the information collected by the perception layer, allowing the collected information to be freely accessed via the cloud, and to determine the transmission of information gathered by the perception layer to the cloud, finally passing it to the application layer for processing. The network structure can be divided into intranet and internet. The intranet is the so-called local area network, where all sensors establish an internal network to exchange information and perform data storage. Next, the cloud carries out data processing, as well as advanced techniques like data mining, to analyze the data collected by multiple sensors, further understanding relevant information throughout the product lifecycle. The external network is the internet, which connects the local area network to larger networks via routers, and then transmits the collected information outward through the internet.

With the advancement of sensor technology, during complex manufacturing production processes, production and testing equipment rapidly generate vast amounts of data. An increasing amount of data within manufacturing plants can now be comprehensively collected, including production history data, equipment parameter data, factory utility data, and measurement quality parameter data. However, the manufacturing shop floor faces both visible and invisible problems. The former includes equipment malfunctions, product defects, excessive cycle times, poor Overall Equipment Effectiveness (OEE), etc. The latter includes equipment performance degradation, component wear, insufficient raw materials, etc. [8]. Typically, visible problems in the manufacturing process can be resolved through continuous improvement, whereas invisible problems like equipment performance degradation can only be detected after a malfunction occurs.

In the perception layer, physical equipment needs to support real-time information acquisition, communication devices should provide high-speed transmission of heterogeneous information, and the shop floor must ensure rapid reconfiguration and adaptability. Huisingh et al. [9] noted that due to the Internet of Things and advanced information technologies, such as Radio Frequency Identification (RFID) tags and smart sensors, these are widely used in daily production and management within manufacturing enterprises, generating a huge volume of data impacting Product Lifecycle Management (PLM) processes. Du [10] stated that with the advent of Industry 4.0, current industrial processes are transforming into intelligent processes. Particularly, many modern industrial processes are equipped with meticulously designed sensors to collect process-related data in order to discover existing or emerging faults within the process.

III. IMPLEMENTATION PROCESS OF THE IIOT-EMBEDDED FASTENER FORMING PROCESS

The screw manufacturing process primarily proceeds from wire rod to wire drawing, heading, threading, heat treatment, and plating. The subject of this study is the heading machine. A smart fastener factory first receives data at the perception layer. This data is then transmitted through the network layer and connected to the application layer. Within the network layer, big data analysis is conducted using artificial intelligence methods to obtain analytical results. Subsequently, the application layer integrates these results and delivers them to the visualization system of the control center.

3.1 Installation of Sensing Components

After an initial requirements investigation, on-site inspection, and assessment of the machine's

automation capabilities, it was found that the heading machine lacked any sensor data. Therefore, the evaluation, setup, and installation of sensors for the foundational IoT layer were confirmed. The components included the use of a Dynamic Force Sensor (DFS). This is a patented design by the Industrial Technology Research Institute (ITRI), Taiwan. It is a force sensing device featuring a double-end suspended buckle-ring structure. This design enhances the uniformity of force applied to the piezoelectric element, reducing damage caused by uneven force, and improves the convenience of disassembly and assembly. Its positioning structure design also enhances assembly precision. Furthermore, through the reading circuit, sensor module, and interface testing environment, it enables real-time monitoring and display of the impacted output force's time curve, sensitivity, and variability. The piezoelectric dynamic force sensor is shown in Figure 1 (Source: ITRI).

Fig 1: Piezoelectric Dynamic Force Sensor

The dynamic force sensor uses highly sensitive piezoelectric ceramic materials. Its structural design incorporates a mix slot to enhance the uniformity of force distribution under high impact loads and improve component reliability. The circuit design employs a quasi-static charge retention technique to read the sensing signals, reducing signal distortion and increasing the usable signal bandwidth. Simultaneously, the module system integrates the sensor readout circuit, PLC communication circuit, and sensing signal acquisition and computation unit. Online analysis software has been developed to judge the good or bad status based on the pressure curve during the process, achieving the effect of full inspection.

Many current studies integrate sensors directly into machinery, fixtures, and tools. However, before the deployment and installation of sensors in the fastener forming process can become operational, a comprehensive understanding of the heading machine's operation and maintenance procedures is required. The component molds are fixed, and processing one product requires two strokes of reciprocating mechanical action. During the continuous operation and control of mechanical parts, sensors are used to monitor the mechanical strength of the metal structure in real-time [3]. Therefore, the appropriate placement of sensors on the heading machine is critically important.

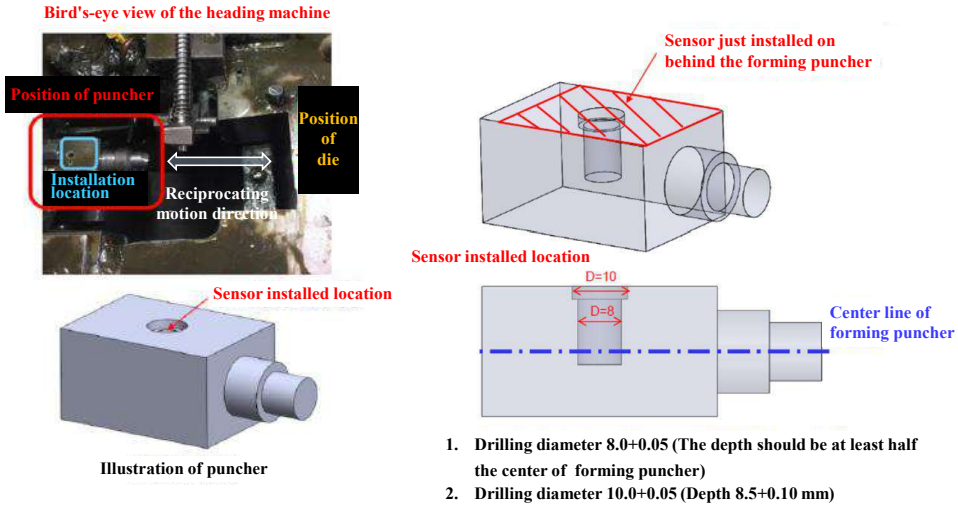


Fig 2 ： Installation Position at the Punch End

3.2 Evaluation of Sensor Installation Options Correctly selecting the appropriate location for sensors to obtain representative data requires detailed evaluation, strictly considering the "actual on-site processing operations" and "machine structural characteristics." The following three options, A, B, and C, were evaluated:

Option A: Sensor installed at the punch end (as shown in Figure 2)

(1) Actual On-Site Processing Operations: Although a sensor could be installed on the punch, the punch frequently needs to be replaced due to production line changeovers for different screw products or wear of the tooling. This would involve repeatedly detaching and attaching the sensor, causing cumbersome recalibration and affecting accuracy. Therefore, this location is unsuitable for sensor installation.

(2) Machine Structural Characteristics: No impact.

(3) Summary of Feasibility Assessment Results: This option is not adopted.

Option B: Sensor installed directly behind the forming hole in the mold end (as shown in Figure 3)

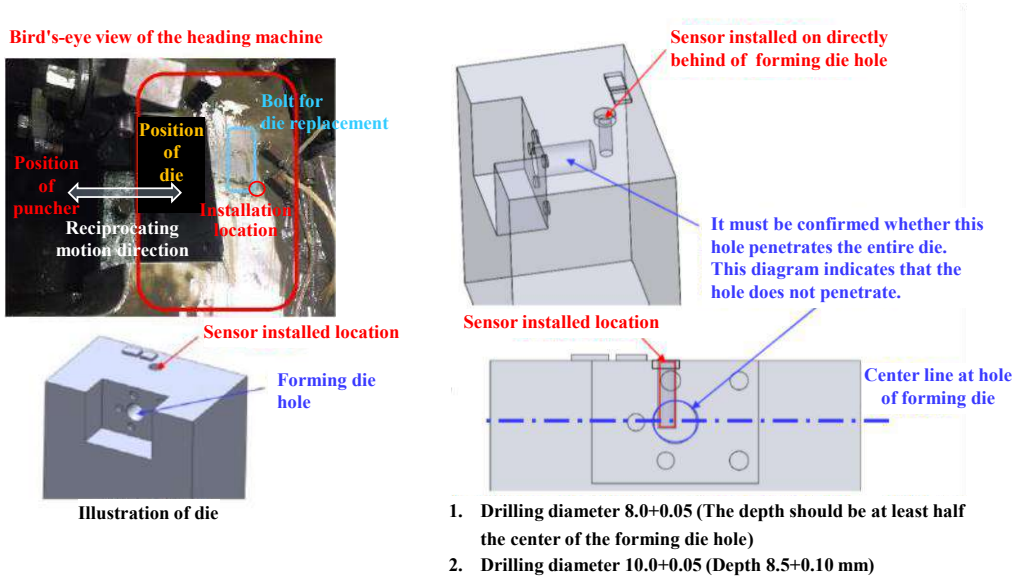


Fig 3: Installation Position Directly Behind the Forming Hole at the Mold End

(1) Actual On-Site Processing Operations: During mold changes or production line changeovers, bolts on the rear end of the machine need to be replaced. Installing a sensor at this location would interfere with the bolt replacement work.

(2) Machine Structural Characteristics: The impact point of the punch is shallow. The sensor must be installed flush against the recess. However, being too close to the recess poses a potential risk of structural failure.

(3) Summary of Feasibility Assessment Results: This option is not adopted.

Option C: Sensor installed on both sides of the forming hole at the mold end

(1) Actual On-Site Processing Operations: No impact.

(2) Machine Structural Characteristics: No impact.

(3) Feasibility Assessment Result: This option is adopted.

Therefore, the selected dynamic force sensor was installed on the top of the mold, aligned with the reciprocating movement direction of the punching tool. Figure 4 shows the position of the installed sensor assembly within the heading machine.

3.3 Method of Sensor Installation

Hot melt adhesive is used inside the hole at the installation location, and then protected on the surface with metal AB adhesive. The external part of the wiring is protected, and a protective sleeve is added. Regarding the wiring layout, the cables are routed along the same paths as existing wiring to

ensure they are not interfered with. The trigger switch primarily relies on the action of the punch rod to initiate the start and end points of vibration signal collection, as well as the counting function. A protective sleeve is added to the external part of the wiring at the installation location, leading to integration with the hardware (as shown in Figure 5). All of these tasks require close communication with experienced on-site technicians. The details are as follows:

During high-speed machine operation, lubricating oil is added to protect and cushion mechanical friction parts. The reciprocating action generates high temperatures, leading to oil mist and fumes. To prevent the lubricating oil mist from penetrating the piezoelectric sensor and interfering with detection, hot melt adhesive is used at the drilled installation site, and a layer of metal-specific AB adhesive is applied on the surface for enhanced isolation. Simultaneously, a protective sleeve is added around the outer edge of the wiring to shield the exposed cables. Furthermore, the routing of the sensor cables is positioned identically to the wiring of other components in the original mechanism (located at the side to ensure no interference during machine operation), as shown in Figure 6.

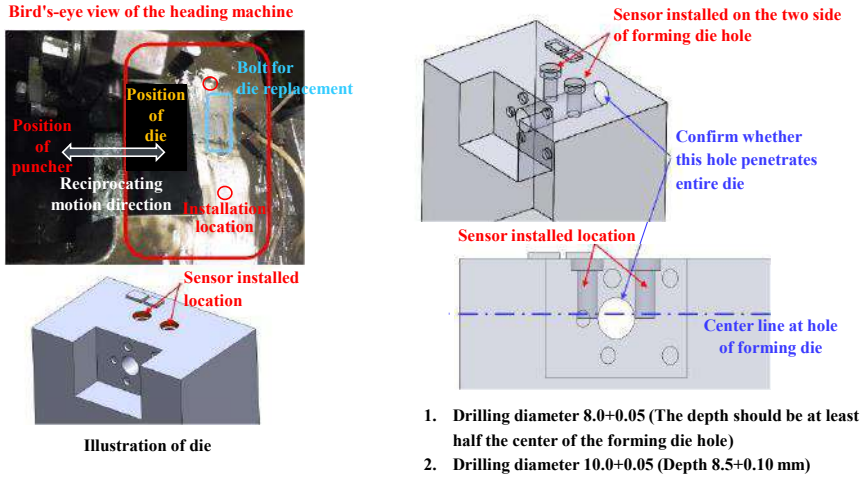


Fig 4 : Installation Position on Both Sides of the Forming Hole at the Mold End

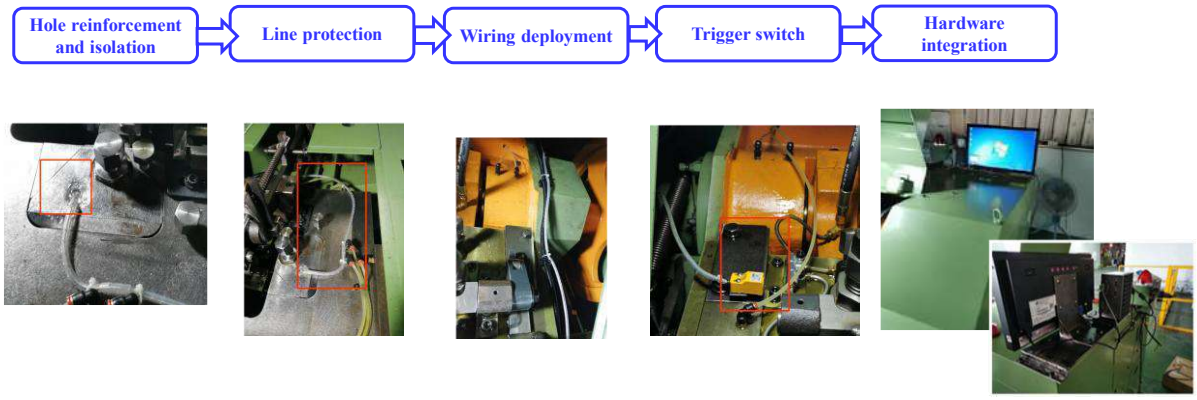


Fig 5: Method of Sensor Installation

Fig 6: Reinforcement and Isolation at the Drilled Installation Site, Outer Wiring Protective Sleeve, and Cable Routing

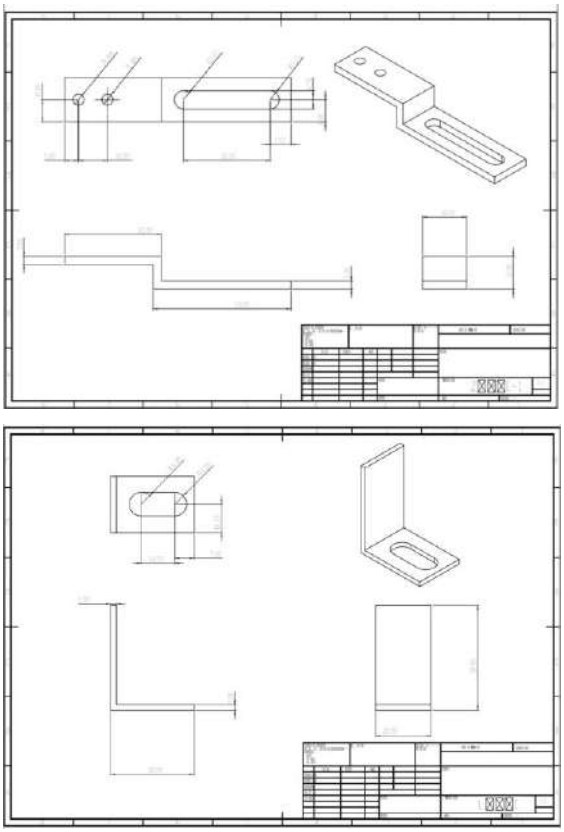


Fig 7: Engineering Design of Z-shaped Bracket & L-shaped Bracket

The trigger switch calculates the heading count and captures the signal start and end time periods to facilitate the formation of representative waveforms. The installation design principle avoids drilling holes in the machine to prevent damaging its original functional structure. Instead, it utilizes the machine's existing screw holes for fixation. A Z-shaped bracket and an L-shaped bracket fixture are fabricated and added onto the original positions. Screw holes are drilled into these fixtures, and the trigger switch is mounted and secured onto them. The engineering design drawing for the trigger switch is shown in Figure 7.

After completing the setup of the trigger switch device (as shown in Figure 8), a protective sleeve is added to the external part of the wiring at the trigger switch installation location to protect the exposed cables. The cable routing is kept the same as the routing position of other existing wiring to ensure no interference. The same principle applies to the protective sleeve and routing for the sensor wiring.

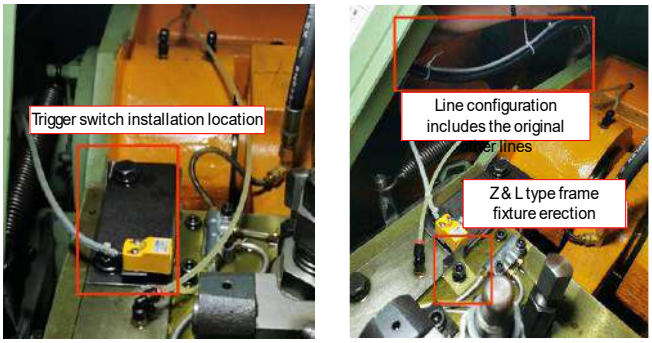


Fig 8: Trigger Switch Device Setup

Regarding hardware integration, engineering designs for components such as the computer base, monitor stand, and host unit chassis are illustrated in the related engineering drawings shown in Figure 9, facilitating the completion of sensor data collection.

The computer host and monitor are installed at the rear end of the machine operation panel, in a clearly visible location. The monitor position (front), the fixing points for the computer host and monitor (back), and the cable fixing points on the back are shown. A socket for the computer will be added later, followed by cable management, as illustrated in Figure 10.

When installing the computer and precision computing hardware, attention must be paid to on-site environmental factors such as machine vibration, high temperatures, and oil stains. During the implementation of this study, machine vibration once interfered with the sensor information received by the computer (resulting in garbled data). This issue was resolved by relocating the setup to a different position.

3.4 Data Visualization

First, an analog-to-digital converter is used to convert the detected analog signals into data waveform curves. Figure 11 shows the data browser, which serves as a basis for verifying the correctness of the waveform data. Chen, et al. [11] indicate that special sensors are used for data collection in production areas, and the data collection interface should be adjustable and scalable, as well as compatible with the visualization system for the data collection process. Furthermore, the monitoring system can effectively collect data from devices installed on the equipment and rapidly transmit the data to cloud servers in the network layer for further processing and visualization [12].

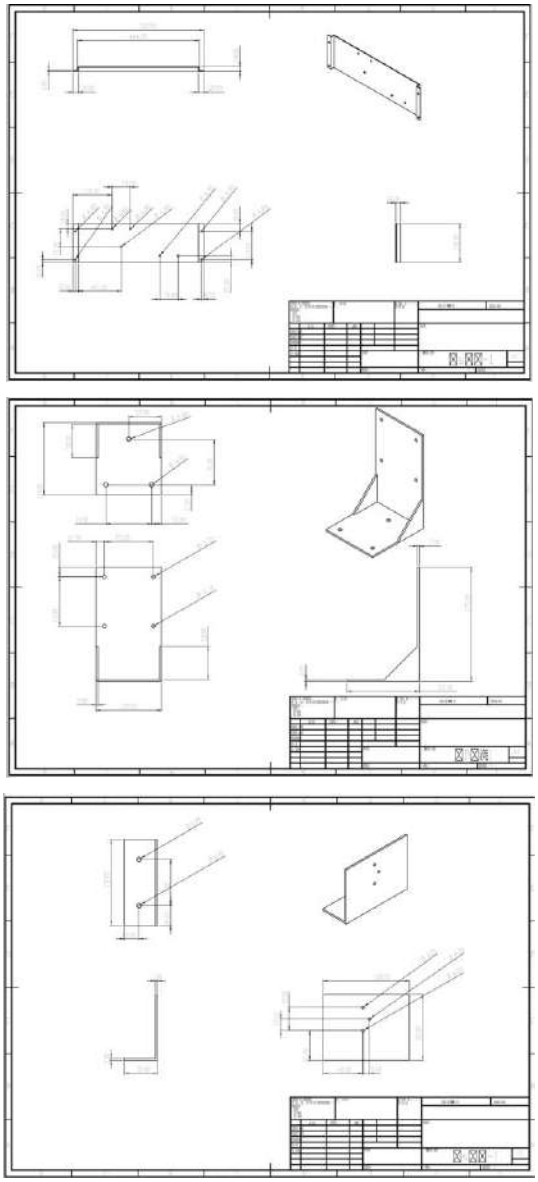


Fig 9: Engineering Design of Base & Monitor Stand & Host Unit Chassis

Fig10: Monitor Position (Front) & Computer Host and Monitor Mounting Bracket (Back) & Rear Cabling

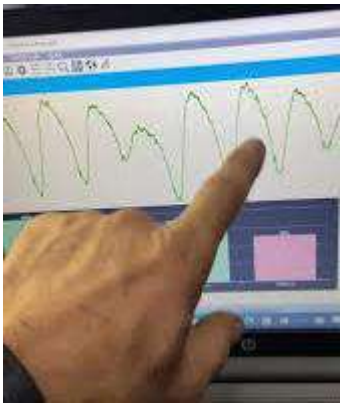


Fig11: Data Browser

During the installation process, the dynamic force changes generated during the cycle time of the heading operation are collected as relevant data through the dynamic force sensor. The collected data are then converted, and the most suitable start and end points for individual data waveform curves are extracted. In this way, each consecutive waveform pattern can be compared, analyzed, and verified between normal and abnormal forms, completing the normal waveform data and correctness verification.

3.5 Cyber-Physical Agent (CPA)

Piezoelectric sensors are installed to collect production data, detecting pressure signals and production information (such as time). The CPA universal data collection module is then used to gather data and perform simple data preprocessing (data segmentation, data cleaning). The CPA's data collection module, data preprocessing module, and even the algorithm application module are designed in a universal plug-in type format. The data collection module needs to integrate the communication method of the piezoelectric sensor, the data preprocessing module must preprocess the analog pressure signals, and the algorithm module establishes the failure mode of the heading punch.

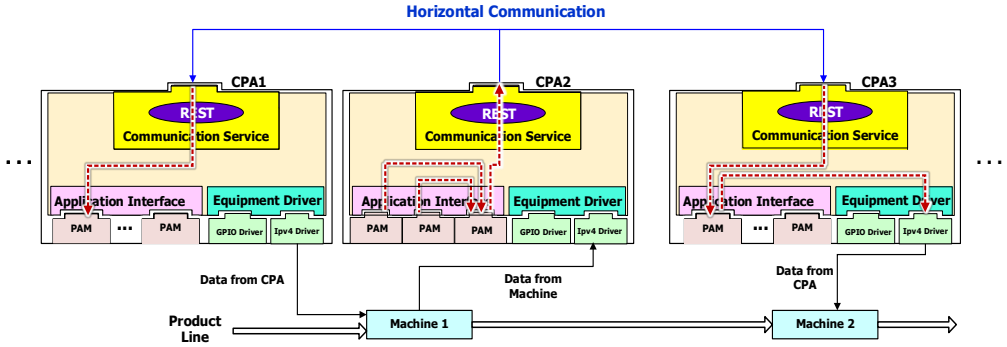


Fig.12 Design and Integration Architecture of the Cyber-Physical Agent

Regarding data collection, data preprocessing, and even the subsequent application of artificial intelligence algorithms, this study utilizes a Cyber-Physical Agent on the heading machine. The design and integration architecture of the Cyber-Physical Agent is shown in Figure 12 (Source: FS-TECH CO., LTD.). The Cyber-Physical Agent is similar to the research by Giampiccolo et al. [2] on sheet metal forming machines, which investigated the impact of data management processes and data quality on preventive maintenance and was equipped with an integrated acquisition system to record failure histories.

A universal IIoT architecture is constructed, incorporating horizontal and vertical communication designs required for future scalability. The key technological element is the CPA, which enables horizontal and vertical communication integration, providing a crucial foundation for machine intelligence and servitization, as explained below:

(1) Enabling Machine Intelligence: Enables machines to communicate and exchange data with each other. In the future, all production equipment and workpieces will be equipped with various sensors, requiring close information exchange during production or assembly to achieve precision manufacturing, assembly, and traceability. It allows for the collection of all product quality control items or yield rates for causal analysis.

(2) Enabling Machine Servitization: A critical aspect of the IoT and intelligent mechanization. It makes the CPA more accessible for use by other devices or

platforms by providing a web service to transmit machine production status data. Any authorized platform can use this web service to request machine production status data and implement necessary applications.

Furthermore, the data acquisition module of the CPA in this study needed to integrate the communication method of the dynamic force sensor. Gangoiti et al. [13] pointed out that using distributed agents with a versatile, customizable, and scalable multi-agent architecture can ensure the Quality of Service (QoS) provided by automated systems. Therefore, a universal IIoT architecture was built, and the data communication interface was connected to the dynamic force sensor data of the heading machine to verify data correctness. For further expansion, including the design for horizontal communication among multiple Cyber-Physical Agents (CPAs) and vertical communication for numerous machines, the core technology revolves around integrating machine intelligence and services, providing an important foundation.

IV. CONCLUSION

For a long time, to ensure process quality, production line personnel have conducted sampling inspections through manual patrols. Additionally, market trends are gradually shifting towards high-value fasteners, primarily applied in industries such as aerospace, wind and solar green energy, and medical sectors. These are characterized by high-mix, low-volume production and demand higher quality precision levels. This necessitates achieving full in-process online inspection and embedding intelligent sensing modules with machine health prediction capabilities into the IIoT to help ensure process quality, reduce equipment damage, and lower labor costs.

Traditional full inspection methods, using the cold heading machine in the fastener industry as an example, can be categorized into Off-Line Metrology (OLM), where the precision of processed parts is inspected in a laboratory, and In-Line Metrology (ILM), where the machine's own measuring tools are used to directly measure part precision. However, besides the high inspection costs, OLM suffers from measurement time delay. This delay means that when processing quality deteriorates, it cannot be known and corrected immediately, leading to mass scrap. ILM, on the other hand, incurs machine time delay, reducing the machine's available processing time and thus lowering its overall equipment effectiveness (OEE).

By connecting with the large-scale information systems that most companies already possess, vast amounts of data (e.g., sensor data) can be transmitted in real-time to the required systems and applications. Meanwhile, MES and ERP focus on production topics, output, targets, and scheduling. With each improvement monitored and predicted, the goals increasingly align with the systems, data models, and automated processes and equipment. IoT-enabled data acquisition can enhance almost any equipment or its operation. With enhanced data acquisition and the IoT, the aggregation of ERP-MES systems also takes a major step forward. These developments are suggested as designs for a powerful core to make factories intelligent and responsive.

These application examples demonstrate the advantages of real-time capability and precision, effectively assisting factories in upgrading quality from sampling inspection to a full inspection level. This not only improves production quality and increases production efficiency but also reduces overall operational costs. Furthermore, by understanding which types of machine failures cause the majority of problems, companies can formulate correct action plans.

REFERENCES

[1] Zion Market Research, 2023. Global pharmaceutical continuous manufacturing market report. New York: Market Research Store.

[2] Giampiccolo, S., Mairot, N., Masry, Z.A., Omri, N. and Zerhouni, N., 2020. Industrial data management strategy towards an SME-oriented PHM. Journal of Manufacturing Systems, 56, 23–36.

[3] Casbas-Gimenez, J., Gil-Hernandez, D., Muñoz-Navascues, O., Murillo, N., Perez-Alfaro, I. and Sanchez-Catalan J.C., 2020. Low-cost piezoelectric sensors for time domain load monitoring of metallic structures during operational and maintenance processes. Sensors, 20(5), 1471.

[4] Brosius, A., Ghiotti, A., Groche, P., Kinsey, B. L., Liewald, M., Madej, L., Min, J., Volk, W. and Yanagimoto, J., 2019. Models and modelling for

process limits in metal forming. CIRP Annals, 68(2), 775–798.

[5] Jazdi, N., Jung, T., Lindemann, B., Sahlab, N., Schloegl, W., Talkhestani, B.A. and Weyrich, M., 2019. An architecture of an intelligent digital twin in a Cyber-Physical production system. Journal of Automatisierungstechnik, 67(9), 762–782.

[6] Ralph, B.J., Schwarz, A. and Stockinger, M., 2020. An implementation approach for an academic learning factory for the metal forming industry with special focus on digital twins and finite element analysis. Procedia Manufacturing, 45, 253–258.

[7] Imran, M., Li, D., Shu, Z., Tang, S., Vasilakos, A.V., Wan, J. and Wang, S., 2016. Software-defined Industrial Internet of Things in the context of Industry 4.0. IEEE Sensors Journal, 16(20), 7373–7380.

[8] Beese, A.M., De, A., DebRoy, T., Elmer, J.W., Milewski, J.O., Mukherjee, T., Wei, H.L., Wilson-Heid, A., Zhang, W. and Zuback, J.S., 2018. Additive manufacturing of metallic components–Process, structure and properties. Progress in Materials Science, 92, 112–224.

[9] Huisingh, D., Liu, Y., Ren, S., Sakao, T. and Zhang, Y., 2017. A framework for Big Data driven product lifecycle management. Journal of Cleaner Production, 159, 229–240.

[10] Du, X., 2019. Fault detection using bispectral features and one-class classifiers. Journal of Process Control, 83, 1–10.

[11] Chen, B., Wan, J., Shu, L., Li, P., Mukherjee, M. and Yin, B., 2018. Smart factory of industry 4.0: Key technologies, application case, and challenges. IEEE Access, 6, 6505–6519.

[12] Milas, N., Mourtzis, D. and Vlachou, A., 2018. An Internet of Things-based monitoring system for shop-floor control. Journal of Computing and Information Science in Engineering, 18(2), 021005.

[13] Gangoiti, U., Iriondo, N. and Priego, R., 2017. Agent-based middleware architecture for reconfigurable manufacturing systems. The International Journal of Advanced Manufacturing Technology, 92, 1579–1590.

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.8

Policy Convergence and Explainable Stability in Orchestrated Multi-Agent LLM Trading Frameworks

Shourya Dewansh

Student, Wentworth Institute of Technology, Boston, MA

Received: 05 Feb 2026; Received in revised form: 06 Mar 2026; Accepted:09 Mar 2026; Available online: 14 Mar 2026

Abstract *— This article examines orchestrated multi-agent large language model trading frameworks through the paired lenses of policy convergence and explainable stability. The topic has gained urgency because recent financial agent systems increasingly rely on distributed deliberation, memory, critique, and risk filtering, while evaluation practice still privileges return metrics over decision coherence and rationale persistence. The aim is to formulate an analytical framework for assessing whether specialized agents move toward a stable executable policy and whether the accompanying explanation remains consistent under workflow pressure, market variability, and iterative revision. The study draws on recent work in explainable AI for finance, explainable reinforcement learning, multi-agent reinforcement learning, quantitative trading, and LLM-based financial agents. Comparative source analysis, conceptual structuring, analytical synthesis, and cross-framework interpretation were applied. The analytical section identifies four determinants of stability: evidence discipline, deliberative topology, risk-gating logic, and adaptive memory control. The resulting framework supports research design, model auditing, and architecture selection for high-stakes AI trading systems.*

Keywords *— **multi-agent LLMs, algorithmic trading, policy convergence, explainable stability, financial AI, agent orchestration, explainable reinforcement learning, trading system evaluation, risk-aware automation, decision traceability***

I. INTRODUCTION

The current stage of financial AI research is marked by a visible shift from isolated predictive models toward coordinated agent assemblies that divide analytical labor across market interpretation, risk screening, memory handling, and decision synthesis. That shift has intensified a methodological problem: performance figures alone do not reveal whether a trading policy emerges from disciplined inter-agent agreement or from unstable narrative negotiation. In finance, where execution errors are magnified by volatility, leverage, and latency, a system's internal coherence matters alongside profitability. Recent surveys on financial explainability and reinforcement learning point to the same unresolved tension: high-capacity models often produce strong outputs while leaving decision pathways insufficiently transparent for audit, intervention, or post-trade diagnosis.

The purpose of the article is to develop an analytical model for evaluating policy convergence and explainable stability in orchestrated multi-agent LLM trading frameworks. Three tasks structure the study:

1) to differentiate policy convergence from raw prediction agreement and links it to executable coordination under market constraints.
2) to identify the architectural conditions under which explanation traces remain stable across

debate, revision, memory updates, and risk control.

3) to formulate an evaluation scheme suitable for comparing recent trading-oriented multi-agent systems without reducing assessment to return statistics.

The novelty lies in joining two lines of inquiry that are often treated separately: multi-agent coordination on one side and explainable financial AI on the other. The article argues that stable execution in LLM trading systems depends less on the presence of many specialized agents than on the disciplined relation among evidence intake, deliberation order, gating rules, and adaptive revision.

II. MATERIALS AND METHODS

The source base combines ten recent publications that jointly cover explainability, reinforcement learning, financial decision systems, and orchestrated LLM agents. P.-D. Arsenault, S. Wang, and J.-M. Patenaude [1] systematize explainable AI approaches for financial time-series forecasting and distinguish interpretability from explanation design. H.S. Jung and H. Lee [2] present a zero-shot multi-agent trading architecture with explicit rationale generation and backtested interpretability checks. X. Li, Y. Zeng, X. Xing, J. Xu, and X. Xu [3] analyze a simulated-trading multi-agent system in which forward-looking evaluation and meeting-based coordination structure decision formation. Z. Ning and L. Xie [4] provide a broad survey of multi-agent reinforcement learning, focusing on coordination problems, benchmark environments, and implementation burdens. L. Saulières [5] reviews explainable reinforcement learning and proposes a taxonomy centered on explanation targets and delivery modes. S. Sun, R. Wang, and B. An [6] synthesizes reinforcement learning research for quantitative trading and clarifies the relation among objectives, environments, and action spaces. C.-A. Wang, S.-H. Huang, C.-T. Chen, and Y.-T. Fang [7] studied reinforcement learning in finance under concept drift and showed why stability cannot be detached from environmental regime change. Y. Xiao, E. Sun, D. Luo, and W. Wang [8] introduce a trading-firm-like multi-agent architecture with analyst debate, risk management, and manager approval. W.J. Yeo, W. Van Der Heever, R. Mao, E. Cambria, R. Satapathy, and G. Mengaldo [9] review explainable AI for finance from the perspectives of transparency requirements, method families, and sector-specific constraints. Y. Yu, Z. Yao, H. Li, Z. Deng, Y. Cao, Z. Chen, J.W. Suchow, R. Liu, Z. Cui, Z. Xu, D. Zhang, K. Subbalakshmi, G. Xiong, Y. He, J. Huang, D. Li, and Q. Xie [10] formulate an LLM multi-agent financial framework with conceptual verbal reinforcement and hierarchical communication.

The study applies comparative analysis, source analysis, conceptual structuring, analytical synthesis, and cross-framework interpretation. A comparative analysis was used to identify recurring coordination patterns across trading-oriented agent systems. Source analysis was used to track how recent literature defines explainability, reinforcement learning stability, and multi-agent decision structure. Conceptual structuring was used to derive the paired constructs of policy convergence and explainable stability. Analytical synthesis connected these constructs to financial execution logic, memory design, and risk-screening procedures. Cross-framework interpretation aligned architecture choices with likely stability outcomes in trading environments.

III. RESULTS

A review of recent multi-agent financial systems shows that policy convergence should not be reduced to simple agreement among agents. In trading architectures, convergence has a narrower and more operational meaning: divergent local interpretations must be transformed into a single executable policy that remains aligned with risk constraints, evidence quality, and market timing. A multi-agent system may display verbal consensus while still producing unstable trade selection if its agreement rests on weak evidence routing, excessive prompt sensitivity, or unfiltered memory carryover. For that reason, convergence is better understood as a property of the orchestration layer rather than of individual model outputs. The trading literature already points in this direction. Financial reinforcement learning surveys treat environment design, reward shaping, and action specification as structural determinants of agent behavior [6]. In

contrast, MARL surveys show that coordination quality depends on communication form, information sharing, and task decomposition [4]. In LLM-based trading systems, these determinants reappear in natural-language form through meetings, debates, supervisory agents, and memory-mediated revision [3; 8; 10].

Within this analytical frame, explainable stability refers to the persistence of a decision rationale across iterative refinement. A rationale is stable when the system preserves the same causal story about evidence, risk, and action after critique, summarization, and managerial synthesis. It becomes unstable when later stages alter the justification while preserving the final action, or preserve the justification while changing the action without a traceable reason. Financial explainability surveys show why this distinction matters. In high-risk domains, explanation does not function as decorative transparency; it serves auditability, trust calibration, and intervention support [1; 9]. Explainable reinforcement learning reaches a similar conclusion from another direction by distinguishing what is being explained from how that explanation is delivered [5]. Applied to trading, this means that an intelligible rationale must retain semantic continuity from analyst signals to final order logic. Recent work on explainable multi-agent trading [2] makes this issue explicit by evaluating rationale quality alongside market outcomes.

Recent agent frameworks permit a more precise decomposition of convergence mechanics. TradingAgents organizes analysts, adversarial researchers, traders, a risk management team, and a manager into a staged pipeline where evidence is first diversified, then contested, and only afterward converted into execution [8]. FinCon uses a manager–analyst hierarchy with conceptual, verbal reinforcement, meaning that prior reasoning episodes are transformed into reusable belief updates [10]. QuantAgents adds simulated trading and repeated meetings, shifting the system away from purely retrospective reflection toward anticipatory policy formation [3]. Jung and Lee's explainable zero-shot framework places a meta-agent above heterogeneous specialist agents, using rationale synthesis as the bridge between modalities and action [2]. Taken together, these systems suggest that convergence improves when three conditions are met: evidence streams remain functionally differentiated, contradiction is processed before action selection, and a dedicated gate compresses deliberation into a bounded policy. Where one of these conditions is absent, systems drift toward either parallel monologues or premature consensus.

The literature further indicates that convergence depends on the sequence in which disagreement is resolved. If debate begins before evidence normalization, agents argue from incomparable informational bases. If risk control enters too late, the system may optimize narrative plausibility rather than tradability. If memory updates occur before contradiction resolution, unstable beliefs harden into future prompts. In this respect, policy convergence in financial LLM systems resembles a constrained reduction process: heterogeneous signals are progressively compressed through specialist interpretation, adversarial testing, risk adjudication, and managerial synthesis. The value of orchestration lies in how it orders these reductions. QuantAgents formalizes this through meetings dedicated to market analysis, strategy development, and risk alerts [3]. TradingAgents formalizes it through specialized analyst and researcher teams, followed by approval from traders and managers [8]. FinCon formalizes this through hierarchical communication and verbal reinforcement that selectively propagate revised beliefs [10]. Stable policy formation emerges when this reduction path remains explicit and reversible.

A second determinant concerns the relation between explanation and memory. LLM trading systems rely heavily on summaries, reflections, and episodic records. Those mechanisms preserve strategic continuity, yet they introduce a latent instability: a compressed memory trace may survive longer than the evidence that initially justified it. Once that occurs, future decisions inherit an explanation artifact rather than a verified analytical premise. FinCon addresses this issue by selectively propagating conceptual belief updates [10]. QuantAgents addresses it through several memory types connected to meetings and tools [3]. From the standpoint of explainable reinforcement learning, such designs raise a direct question: Does memory store observations, evaluations, or decisions already interpreted by prior agents [5]? The answer matters

because explainable stability deteriorates when the system cannot separate factual retention from argumentative residue. A stable architecture, therefore, requires memory typing, update thresholds, and clear boundaries between raw evidence and post-debate abstraction.

A third determinant lies in risk-gating logic. Financial decision systems differ from generic multi-agent settings because convergence toward a policy is not sufficient; the policy must remain admissible under exposure limits, portfolio discipline, and regime sensitivity. TradingAgents places a risk management team before managerial approval [8]. FinCon embeds risk control into episodic self-critique and belief revision [10]. QuantAgents gives risk control its own analyst and meeting space [3]. This recurrent design choice suggests that stable convergence in trading is inseparable from veto structure. A framework without a dedicated gate may still converge, yet it converges toward an action language rather than an execution policy. In financial terms, the difference is substantial: one system ends with a persuasive narrative, while the other ends with a controllable order hypothesis. Explainable stability rises when the rationale for accepting or blocking a trade is recorded at the same level of granularity as the trade thesis itself.

Environmental variability adds another layer of instability. Trading systems do not operate in stationary settings; evidence distributions shift, sentiment regimes break, and action values decay. Reinforcement learning literature on quantitative trading has long treated non-stationarity as a structural problem [6]. Recent work on concept drift in financial RL sharpens this point by distinguishing between gradual and sudden drift and linking adaptation to sentiment-aware restructuring, curriculum learning, and knowledge distillation [7]. For orchestrated LLM trading frameworks, the implication is clear: a policy may appear convergent within one regime while remaining fragile across adjacent regimes. Explainable stability, therefore, requires more than coherent debate at a single decision point. It requires the explanation schema to be resilient to changing market distributions. If the system revises actions after regime change, the rationale for revision must remain legible; if it retains the prior action, the rationale for persistence must remain equally readable. Stability without adaptive transparency degenerates into rigid consistency.

These observations support an analytical distinction between local convergence and systemic convergence. Local convergence refers to agreement within one deliberation cycle. Systemic convergence refers to the architecture's repeated ability to transform heterogeneous evidence into bounded decisions without explanation collapse. The latter is the more demanding property and the one that matters for deployable trading systems. Surveys of explainable AI in finance warn that transparency mechanisms often remain external to the model's logic [1; 9]. Multi-agent financial systems partially remedy that weakness by exposing intermediate reasoning stages through agent interactions [2; 3; 8; 10]. Yet exposure alone does not ensure stability. A transcript of unstable reasoning remains unstable. What matters is whether the architecture constrains how evidence enters, how disagreement is handled, how risk vetoes are applied, and how memory is rewritten.

On this basis, policy convergence in orchestrated multi-agent LLM trading frameworks may be defined as the architecture's capacity to reduce heterogeneous market interpretations into a single risk-bounded, executable decision through explicit, reviewable coordination steps. Explainable stability may be defined as the persistence of a coherent, auditable rationale across those coordination steps and across subsequent adaptation episodes. The conjunction of these two properties forms the most defensible analytical criterion for contemporary LLM trading systems. A framework lacking convergence remains operationally noisy; a framework lacking explainable stability remains epistemically fragile.

To consolidate the analytical observations discussed above, Figure 1 presents the proposed model of policy convergence and explainable stability in orchestrated multi-agent LLM trading frameworks. The scheme captures the ordered transition from heterogeneous market inputs to a bounded policy proposal and then to post-decision monitoring. Its logic emphasizes that stability is produced through a structured sequence rather than through isolated agent competence. In this sequence, specialized interpretation, contradiction processing, risk filtering, and explanation tracing form an integrated chain in

which each stage constrains the next one and preserves the auditability of the final decision.

Fig. 1. Analytical model of policy convergence and explainable stability in orchestrated multi-agent LLM trading frameworks (adapted from [3; 8; 10])

The model shows that convergence becomes analytically meaningful only when the final policy remains connected to the reasoning path that produced it and when subsequent revisions under regime change remain traceable. Viewed in this way, the architecture of the system determines whether coordination yields a governable trading policy or merely a temporary narrative agreement. This interpretation prepares the transition to the discussion section, where the comparative implications of recent frameworks are examined with respect to coordination design, explanation integrity, and sources of structural instability.

IV. DISCUSSION

The analytical reconstruction above suggests that recent multi-agent financial systems are converging toward a shared architectural grammar, though they still differ in how they manage disagreement, memory, and adaptation. The comparison in Table 1 condenses those differences into the dimensions most relevant for the present study: coordination form, convergence mechanism, explanation strategy, and the primary source of instability. Table 1 is derived from the ten-source corpus and focuses on architecture-level implications rather than reported benchmark superiority.

Table 1. Comparative architecture patterns related to convergence and explainable stability (synthesized from [1–10])

Primary coordination form	Main convergence mechanism	Explanation mechanism	Likely instability source
Meta-agent synthesis across modality specialists	Central integration of heterogeneous signals	Natural-language rationale generated by the meta-agent	Explanation drift between modality-level evidence and final synthesis
Meeting-based coordination with analyst specialization	Repeated deliberation plus simulated trading feedback	Explicit meeting structure and multi-memory workflow	Memory contamination across meetings; unstable transfer from simulation to execution
Trading-firm hierarchy with analyst/researcher/trader/manager chain	Adversarial research followed by risk filtering and approval	Intermediate team outputs expose deliberation stages	Premature consensus if adversarial evidence is weakly grounded
Manager–analyst hierarchy with	Selective propagation of revised beliefs	Verbal reinforcement preserves belief	Overcompression of belief updates into future prompts

conceptual verbal reinforcement		updates in textual form	
RL-oriented financial decision systems	Reward discipline and adaptation under market change	Post hoc interpretation of policy behavior	Regime drift and reward-policy mismatch
Explainability and XRL taxonomies	No direct execution mechanism; conceptual support for evaluation	Formal separation of targets, methods, and reliability of explanations	External explanations detached from internal decision logic

Table 1 shows that convergence and explainable stability are related, but they do not collapse into a single property. Architectures with stronger supervisory compression tend to produce clearer final policies, though they risk hiding unresolved disagreement inside summary layers. Architectures with richer deliberation expose more of the reasoning path, though they may preserve conflict without resolving it into an admissible order. Survey-based sources [1; 5; 9] help explain why both tendencies remain incomplete: explanation quality depends on alignment between what is described and where that information enters the decision process. When a trading framework generates narratives after policy selection, transparency remains cosmetic. When explanation artifacts participate in memory and veto logic, transparency becomes operational.

A second issue concerns evaluation design. Existing trading studies still privilege financial output metrics, whereas the present analysis argues for a broader matrix centered on decision integrity. Table 2 translates this argument into an applied evaluation scheme suitable for architecture review, internal audit, or pre-deployment stress testing. Each criterion targets a failure point that emerges repeatedly across recent studies: unstable evidence routing, summary distortion, unmanaged regime change, and untraceable veto logic.

Table 2. Proposed evaluation matrix for analytical assessment of orchestrated LLM trading systems (analytically derived from [1–10])

Evaluation dimension	What should be inspected	Failure signature	Practical implication
Evidence discipline	Traceability from source signal to agent claim	Claims survive after source inconsistency is detected	Weak auditability and inflated confidence
Deliberative convergence	Reduction of disagreement across coordination rounds	Final action appears without a resolved contradiction	Action selection rests on narrative compression rather than agreement
Risk-gating integrity	Visibility of vetoes, overrides, and exposure checks	Trade approval lacks explicit risk justification	Execution remains difficult to govern in regulated settings
Memory hygiene	Separation of raw evidence, summaries, and learned beliefs	Old summaries override current evidence	Historical residue distorts present decisions
Regime resilience	Stability of rationale under drift and market shifts	The same explanation persists despite a clear regime break	False consistency hides adaptation failure
Explanation continuity	Semantic alignment between analyst-level reasoning and final order logic	Final rationale rewrites earlier evidence without justification	Post-trade analysis loses causal coherence

The matrix in Table 2 clarifies the main limitation of a purely analytical article: without controlled experiments, it cannot rank architectures by measurable superiority. Yet such a format remains productive because recent literature already reveals repeated structural tensions that merit conceptual consolidation. The absence of a unified vocabulary has slowed rigorous comparison between LLM agent frameworks and reinforcement-learning-based trading systems. The paired constructs proposed here address that gap by linking execution discipline to explanation persistence. In the professional field of AI-driven trading, the practical implication is straightforward. An architecture review should examine where agreement is produced, how it is bounded, what kind of memory is retained, and whether explanations survive adaptation pressure. Systems that satisfy these conditions are better positioned for internal governance, model risk review, and human override design.

V. CONCLUSION

The analysis showed that convergence in multi-agent LLM trading frameworks refers to reducing heterogeneous interpretations into a single risk-bounded, executable decision through explicit coordination steps. The article found four such conditions: disciplined evidence routing, ordered contradiction processing, explicit risk gating, and controlled memory revision. In response, the article proposed an analytical matrix that integrates convergence, explanation continuity, regime resilience, and auditability within a single assessment framework. The main inference is that deployable trading intelligence depends on the joint presence of convergent policy formation and stable explanation traces. If one property is absent, the system either remains operationally noisy or becomes difficult to govern after deployment.

REFERENCES

[1] Arsenault, P.-D., Wang, S., & Patenaude, J.-M. (2025). A survey of explainable artificial intelligence (XAI) in financial time series forecasting. *ACM Computing Surveys, 57*(10), Article 265, 1–37. https://doi.org/10.1145/3729531

[2] Jung, H. S., & Lee, H. (2026). Explainable zero-shot trading using multi-agent LLM architecture: A backtested approach for Bitcoin price. *Information Processing & Management, 63*(2, Part B), 104466. https://doi.org/10.1016/j.ipm.2025.104466

[3] Li, X., Zeng, Y., Xing, X., Xu, J., & Xu, X. (2025). QuantAgents: Towards multi-agent financial system via simulated trading. In *Findings of the Association for Computational Linguistics: EMNLP 2025* (pp. 17438–17464). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.findings-emnlp.945

[4] Ning, Z., & Xie, L. (2024). A survey on multi-agent reinforcement learning and its application. *Journal of Automation and Intelligence, 3*(2), 73–91. https://doi.org/10.1016/j.jai.2024.02.003

[5] Saulières, L. (2025). *A survey of explainable reinforcement learning: Targets, methods and needs*. arXiv. https://doi.org/10.48550/arXiv.2507.12599

[6] Sun, S., Wang, R., & An, B. (2023). Reinforcement learning for quantitative trading. *ACM Transactions on Intelligent Systems and Technology, 14*(3), Article 44, 1–29. https://doi.org/10.1145/3582560

[7] Wang, C.-A., Huang, S.-H., Chen, C.-T., & Fang, Y.-T. (2026). Financial reinforcement learning under concept drift based on knowledge distillation and curriculum learning. *Decision Support Systems, 203*, 114624. https://doi.org/10.1016/j.dss.2026.114624

[8] Xiao, Y., Sun, E., Luo, D., & Wang, W. (2024). *TradingAgents: Multi-agents LLM financial trading framework*. arXiv. https://doi.org/10.48550/arXiv.2412.20138

[9] Yeo, W. J., Van Der Heever, W., Mao, R., et al. (2025). A comprehensive review on financial explainable AI. *Artificial Intelligence Review, 58*, Article 189. https://doi.org/10.1007/s10462-024-11077-7

[10] Yu, Y., Yao, Z., Li, H., Deng, Z., Cao, Y., Chen, Z., Suchow, J. W., Liu, R., Cui, Z., Xu, Z., Zhang, D., Subbalakshmi, K., Xiong, G., He, Y., Huang, J., Li, D., & Xie, Q. (2024). FinCon: A synthesized LLM multi-agent system with conceptual verbal reinforcement for enhanced financial decision making. *Advances in Neural Information Processing Systems, 37*, 137010–137045. https://doi.org/10.48550/arXiv.2407.06567

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.9

Fuzzy PID Intelligent Control of System Model

Hung Yu Chen[1], Zi Xian Huang[2], Ming Hui Huang[3], Zou Zheng Hao Dong[4], Ho Sheng Chen[5*]

[1]Guangzhou Institute of Science and Technology, China
Email: ronwisely@outlook.com
[2,3,4,5]School of Sciences, Guangdong University of Petrochem Technology (GDUPT), Maoming 525000, China
Corresponding author: Ho-Sheng Chen
Email: hschen98.tw@gmail.com

Received: 13Feb 2026; Received in revised form: 12 Mar 2026; Accepted:17 Mar 2026; Available online: 21 Mar 2026

Abstract— *To improve the control performance of conventional PID controllers in nonlinear and time-varying uncertain systems, a fuzzy neural network PID control algorithm is proposed. This approach integrates the nonlinear control advantages of fuzzy logic with the self-learning and adaptive characteristics of neural networks to achieve real-time online tuning of PID parameters. Based on the mathematical model of an electro-hydraulic servo control system, a fuzzy neural network PID control system model is developed. Taking the control system of a transplanting manipulator as a case study, simulations are conducted using Matlab/Simulink, with the transfer function of the electro-hydraulic servo control system serving as the controlled object. The tracking performance of different control systems is evaluated by comparing traditional PID control, Type-1 fuzzy PID control, interval Type-2 fuzzy PID control, and fuzzy neural network PID control methods employing various membership functions. Simulation and experimental results demonstrate that the fuzzy neural network PID controller exhibits superior control performance compared to traditional PID control, Type-1 fuzzy PID control, and interval Type-2 fuzzy PID control in the target system.*

Keywords— ***fuzzy neural network; PID control; electro-hydraulic servo system; transplanting manipulator; Matlab/Simulink simulation***

I. INTRODUCTION

the introduction of the paper should explain the nature of the problem, previous work, purpose, and the contribution of the paper. The contents of each section may be provided to understand easily about the paper.

At present, electro-hydraulic servo system is widely used in industrial production and aerospace [1]. In fields such as mechanical manufacturing, this technology employs a combined electric-hydraulic approach to achieve precise motion control. As a primary method for high-performance closed-loop control in hydraulic systems, its core mechanism involves motor-driven oil pumps that draw hydraulic fluid into pipelines. Electronic signals from the controller regulate hydraulic valves, which then direct the fluid to actuators like cylinders or motors. The fluid entering the cylinder pushes pistons to drive load movement [2]. It has the characteristics of high power-to-weight ratio, fast response, high stiffness and low maintenance cost. The characteristics of fast response of electrical system and high output power of hydraulic system can be reflected in the electro-hydraulic servo system [3]. In the control of electro-hydraulic servo systems, widely used PID control is relatively easy to implement and can achieve good results. However, conventional PID controllers lack adaptive parameter adjustment

capabilities. When dealing with nonlinear and time-varying uncertain electro-hydraulic servo systems, there remains room for optimization and improvement. This necessitates enhancing control performance beyond the basic PID controller framework [4,5].

To investigate which control scheme is superior when compared with conventional PID controllers, Type 1 fuzzy PID controllers, interval-type Type 2 fuzzy PID controllers, and fuzzy neural network PID controllers in handling uncertain nonlinear time-varying systems, we will select an uncertain nonlinear time-varying system as the control object. Given the widespread application of electro-hydraulic servo systems, they serve as a representative example of such nonlinear time-varying systems. Jin X., Chen K., Zhao Y., et al. have previously developed simulation models for transplanting robotic arms in their research. The transplanting robotic arm employs a hydraulic control system featuring an electro-hydraulic servo valve. A mathematical model was developed to simulate fuzzy PID control, validating the feasibility of applying this controller to robotic arm operations. Building upon the transfer function of the controlled system, four controller designs were developed using fuzzy control theory and automatic control principles: conventional PID controller, Type 1 fuzzy PID controller, interval-type Type 2 fuzzy PID controller, and a fuzzy neural network PID controller [7-10].

II. SYSTEM DESCRIPTION

The electro-hydraulic servo system operates on Pascal's law. It is primarily categorized into valve-controlled and pump-controlled types, both sharing similar control logic for regulating output position, force, or speed. This study focuses on the valve-controlled system, as illustrated in Fig. 1, specifically designed to control the position output of the hydraulic cylinder piston.

The power source for electro-hydraulic servo systems typically originates from the hydraulic pump at the actuators front end, which generates pressure differentials in the fluid through high-speed operation. Displacement sensors and pressure sensors respectively transmit position signals from the end effector and hydraulic pressure signals back to the controller. The controller then outputs servo electrical signals to the electro-hydraulic servo valve via a servo amplifier, thereby controlling the end effector as illustrated in Fig. 2.

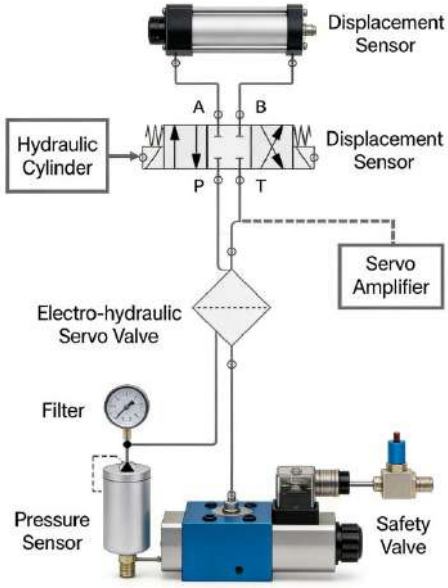


Fig.1. Valve controlled electro-hydraulic servo system diagram

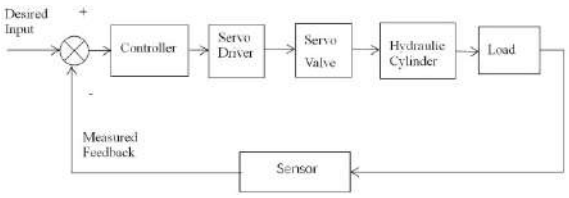


Fig.2. Closed-loop control block diagram of electro-hydraulic servo system

The power source of an electro-hydraulic servo system is usually the hydraulic pump at the front end, which creates pressure increase in the fluid through the high-speed operation of the pump. The position signal and liquid pressure signal of the end effector are fed back to the controller using displacement sensors and pressure sensors, respectively. Finally, the controller outputs the servo electrical signal to the electro-hydraulic servo valve through a servo amplifier to control the end effector. In addition, the control of the electro-hydraulic servo control system is also affected by the flow of liquid between the valve port, valve, and hydraulic cylinder, as well as the load on the hydraulic cylinder

itself during work. Therefore, the three fundamental time-domain equations of the electro-hydraulic servo system are given, the valve port flow equation:

$$q_L = K_q x_v - K_c p_L \quad (1)$$

The continuity equation of hydraulic cylinder flow:

$$q_{\mathrm{L}} = A_{\mathrm{p}} \frac{\mathrm{d}x_{\mathrm{p}}}{\mathrm{d}t} + C_{\mathrm{tp}} p_{\mathrm{L}} + \frac{V_{\mathrm{t}}}{4\beta_{\mathrm{e}}} \frac{\mathrm{d}p_{\mathrm{L}}}{\mathrm{d}t} \quad (2)$$

The force balance equation of hydraulic cylinder:

$$A_{\mathrm{p}} p_{\mathrm{L}} = m_{\mathrm{t}} \frac{\mathrm{d}^2 x_{\mathrm{p}}}{\mathrm{d}t^2} + B_{\mathrm{p}} \frac{\mathrm{d}x_{\mathrm{p}}}{\mathrm{d}t} + K x_{\mathrm{p}} + F_{\mathrm{L}} \quad (3)$$

After applying the Laplace transform to the three major time-domain equations mentioned above, we obtain the three fundamental frequency-domain equations of the electro-hydraulic servo system:

Valve port flow equation:

$$Q_L = K_q X_V - K_c P_L \quad (4)$$

where Q_L represents the flow rate through the load, K_q is the servo valve flow gain, X_V is the effective spool displacement, K_c is the servo valve flow pressure gain, and P_L is the load pressure

The continuity equation of hydraulic cylinder flow:

$$Q_L = A_p s X_p + C_{tp} P_L + \frac{V_t}{4\beta_e} s P_L \quad (5)$$

Where A_p is the effective area of the hydraulic piston during operation, X_p is the piston displacement during operation, C_{tp} is the total leakage coefficient of the hydraulic cylinder, β_e is the effective bulk modulus of the oil, and V_t is the total volume of the two chambers of the hydraulic cylinder.

The force balance equation of hydraulic cylinder:

$$A_p P_L = m_t s^2 X_p + B_p s X_p + K X_p + F_L \quad (6)$$

Considering the terminal load as a variable load composed of mass, spring, and damping, which includes three parts: inertia, elasticity, and damping. Without considering the influence of external load forces, the force balance equation of the hydraulic cylinder can be obtained based on Newton's second law and where m_t is the equivalent load mass, B_p is the load viscous friction coefficient, K is the load elastic coefficient, and F_L is the force acting on the piston rod of the hydraulic cylinder. Based on the effective spool displacement X_V in the above three fundamental frequency domain equations and the piston displacement X_p during operation, the open-loop transfer function of the system is obtained as follows:

$$\mathrm{G(s)} = \frac{\mathrm{X_p}}{\mathrm{X_v}} = \frac{\frac{K_{\mathrm{q}}}{A_1}}{s\left(\frac{s^2}{\omega_{\mathrm{h}}^2} + \frac{2\zeta_{\mathrm{h}}}{\omega_{\mathrm{h}}} s + 1\right)} \quad (7)$$

where A_1 represents the equivalent working area of the hydraulic cylinder piston rod during displacement, ζ_{h} denotes the damping ratio of the hydraulic cylinder, and ω_{h} signifies the natural frequency of the hydraulic cylinder. The natural frequency ω_{h} can be calculated as follows:

$$\omega_h = \sqrt{\frac{4\beta_e A_P^2}{V_t m_t}} \quad (8)$$

The damping ratio ζ_{h} can be calculated as follows:

$$\zeta_h = K_{ce} \sqrt{\frac{m_t \beta_e}{V_t A_p^2}} + \frac{B_P}{4} \sqrt{\frac{V_t}{m_t \beta_e A_P^2}} \quad (9)$$

The closed-loop transfer function can be obtained

$$\phi(S) = \frac{G(s)}{1 + G(s)} = \frac{K_v}{\left(\frac{s^3}{\omega_h^2} + \frac{2\zeta_h}{\omega_h} s^2 + s + K_v\right)} \quad (10)$$

The structure of the transplanting robotic arm is shown in the Fig. 3, where 1. Conductive liquid injection hole, 2. Hydraulic pump, 3. Power output shaft, 4. Rotating power output bracket, 5. Clamp jaw fixing device, 6. Oil pipeline outlet, 7. Clamp jaw, 8. Spring, 9. Clamping device, and 10. Rotating sleeve, respectively.

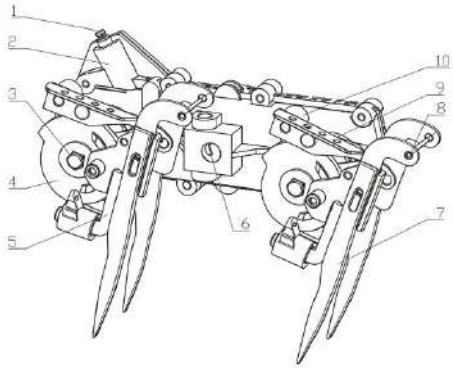


Fig.3. The ransplanting robot structure

Jin X, Chen K, Zhao Y, et al. constructed a simulation model for a transplanting robotic arm in their research [11]. The robotic arm employed hydraulic control incorporating an electro-hydraulic servo valve, and the obtained mathematical model was

utilized to complete the simulation of fuzzy PID control. Therefore, this paper refers to the simulation configuration parameters introduced by Jin X, Chen K, Zhao Y, et al. in their article for data calculation. As shown in Table 1.

Table. 1. Simulation parameters of electro-hydraulic servo system [11]

variable	value	variable	value
Q_L	209	m	0.24
K_q	5	F_L	$11e^4$
C_{tp}	$1.5e^{-10}$	β_e	$7e^8$
V_t	$5.8e^{-5}$	A_p	$4.4e^{-3}$
B_P	$2e^5$	K_h	$2.5e^7$
ω_m	185	K_f	$2.5e^{-5}$

Ultimately, the closed-loop transfer function of the control object can be obtained

$$G(s) = \frac{208.75}{s^3 + 14.31s^2 + 447.53s + 208.75} \quad (11)$$

III. CONTROLLER DESIGN

PID controllers are widely used in modern control fields due to their simple structure and ease of implementation. Before designing a PID controller, we must first clarify its components: a conventional PID controller mainly consists of three parts: proportional, integral, and derivative. Based on this, we obtain the closed-loop system transfer function in Eq. 11, and the time-continuous equation of the PID controller can be obtained

$$G(s) = \frac{U(s)}{E(s)} = K_p\left(1 + \frac{1}{K_I s} + K_D s\right) \quad (12)$$

where K_p, K_I, and K_D represent the proportional coefficient, integral time (also known as integral coefficient), and derivative time (also known as derivative coefficient) of the controller, respectively. These components adjust the output signal by monitoring the error between the feedback signal and the set value, ensuring that the system response reaches the desired value. The design of a Type-1 fuzzy PID controller (T1FPID) is used to enable the controlled system to track the target signal. It can be seen that the overall structure of the controller consists of a Type-1 fuzzy controller, a PID controller, and a controlled object electro-hydraulic servo system, in addition to a feedback loop.

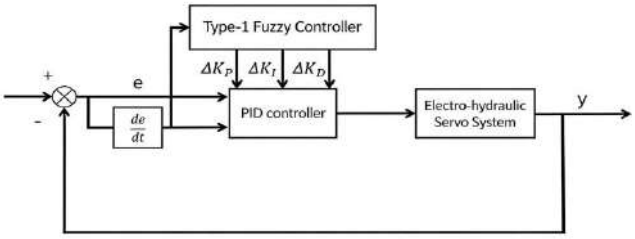


Fig. 4. Structure diagram of type 1-fuzzy PID controller

The structure of the Type 1 fuzzy controller is illustrated in Fig. 5 below, primarily consisting of three main components: fuzzification, fuzzy inference, and defuzzification. From the perspective of input and output signal forms, the fuzzy controller of this system is of the multi-input multi-output type. Its complete design process can be mainly divided into steps such as determining fuzzy variables, establishing the range of the universe of discourse, discretization, fuzzification, establishing fuzzy rules, and defuzzification output. Each step will be introduced and analyzed separately below.

From Fig. 4, we can see that the input of the Type 1 fuzzy controller consists of two parts: the system error e and the rate of change of error de/dt (hereinafter defined as ec). The output consists of three parts: ΔK_P, ΔK_I, and ΔK_D. Therefore, we have determined the variables involved in the fuzzification process in this controller.

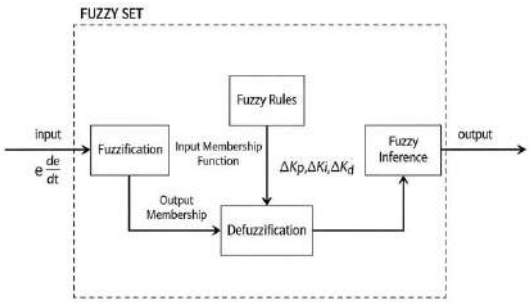


Fig.5: Structure of Type-1 Fuzzy Controller

Compared to the aforementioned Type-1 fuzzy PID controller, the structure of the Interval Type-2 fuzzy PID controller is identical, as illustrated in Fig. 6. It can be observed that the overall structure of the controller consists of an Interval Type-2 fuzzy

controller, a PID controller, and the controlled object, which is an electro-hydraulic servo system.

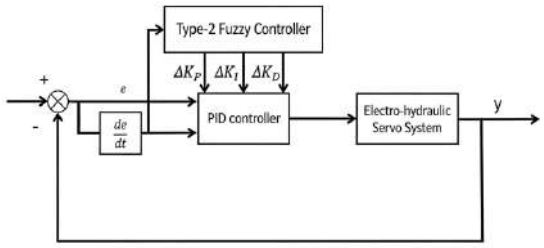


Fig. 6. Structure diagram of type 2-fuzzy PID controller

In the representation of fuzzy sets, the Type-1 fuzzy controller employs a fixed-shape membership function to represent fuzzy sets. It handles system uncertainties through fuzzy logic and fuzzy inference, but may perform inadequately when faced with significant uncertainties. In the Interval Type-2 fuzzy controller, fuzzy sets are represented using intervals, meaning the fuzzy set for each input variable is represented as an interval rather than a specific membership function. This approach enables better handling of system uncertainties, as intervals can describe uncertainties within a certain range. In the design presented in this section, the interval representation of fuzzy sets in the Interval Type-2 fuzzy controller is formed by shifting the membership function of the Type-1 fuzzy controller, representing an extended form of the Type-1 fuzzy controller's membership function. Its functional characteristics are based on the Gaussian membership function. To understand the fuzzy sets in the interval type-2 fuzzy controller, we can assume that we have two Gaussian functions, denoted as $G_1(x)$ and $G_2(x)$, respectively. Their mathematical expressions are as follows:

$$G_1(x) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \quad (13)$$

$$G_2(x) = \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} \quad (14)$$

where σ_i, μ_i, i=1, 2, represents the standard deviation of the corresponding Gaussian function and the mean of the corresponding Gaussian function, respectively. The shape of the Gaussian function will change with the change of the mean and standard deviation. As shown in Fig. 7, the interval type-2 Gaussian membership function, the standard deviation and mean of the two Gaussian functions are as follows:

$$\sigma_1 = 0.7078,\ \sigma_2 = 0.56624,\ \mu_1 = 2,\ \mu_2 = 2.$$

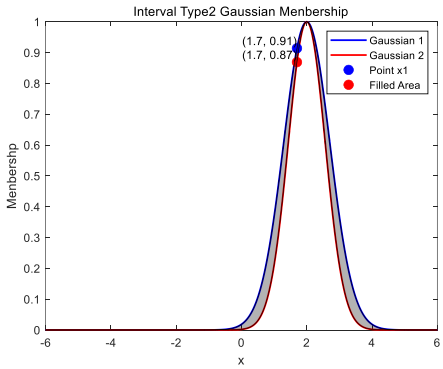


Fig. 7. Calculation of interval Type-2 Gaussian membership function

After summing $G_1(x)$ and $G_2(x)$ and taking their average, the resulting arithmetic mean is the actual membership function. Substituting x_0=1.7 into Eq. (13) and (14), we obtain $G_1(x)$ =0.91 ; $G_2(x)$ =0.87. Therefore, the actual membership degree at the point x_0=1.7 is:

$$G_f(x) = \frac{G_1(x) + G_2(x)}{2} = \frac{0.91 + 0.87}{2} = 0.89 \quad (15)$$

Similarly, after determining the mathematical expression of the membership function, it is necessary to establish the fuzzy domains for the controller's input and output. The interval type-2 fuzzy controller has two inputs and three outputs, making it a multi-input multi-output system. The inputs to the interval type-2 fuzzy controller include the system error e and the rate of change of error ec. The outputs of the interval type-2 fuzzy controller include ΔK_P, ΔK_I, and ΔK_D. Next, the domain ranges for the system error e and the rate of change of error ec are set, both of which are [-6,6]. Subsequently, the domains for ΔK_P, ΔK_I, and ΔK_D are set to [1,3], [1,3], and [0.5,2], respectively, as Fig. 8.

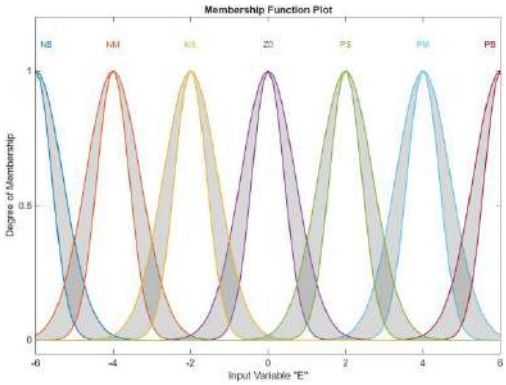


Fig. 8. Membership function of Interval Type2-Fuzzy PID

IV. SIMULATION

To simulate the system model, we constructed systems such as conventional PID control, Type-1 fuzzy PID control, and interval Type-2 fuzzy PID control in the Simulink simulation environment of Matlab according to the controller design content. In addition, to explore the signal tracking capabilities of conventional PID control, Type-1 fuzzy PID control, and interval Type-2 fuzzy PID control, we designed input step signals with variations. Firstly, to facilitate the analysis and comparison of the control capabilities of the three control systems, we integrated them into the same simulation environment as shown in Fig. 9, Simulation of Multiple Controllers in Simulink.

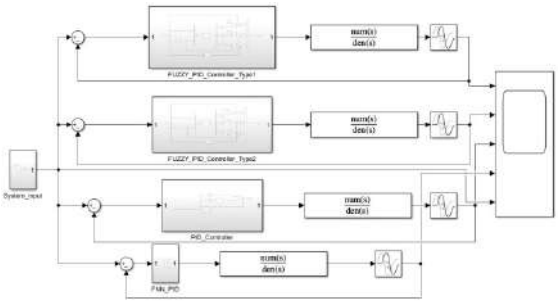


Fig. 9. Simulation of multiple controllers in Simulink

The initial value of the input signal is Position=1.8mm. When the simulation progresses to time=7s, the signal undergoes a step pull-up, raising the target signal value to Position=3mm. After a duration of 2s, at simulation time time=9s, the signal is pulled back to the initial position of 1.8. The signal tracking performance of each controller is observed, and the simulation effect of the system is shown in Fig. 10, which illustrates the simulation effects of various controllers.

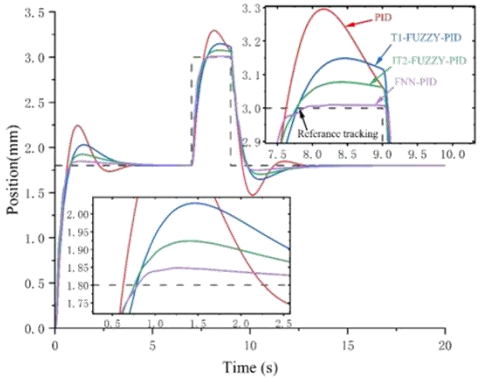


Fig. 10. Output response curve diagram of multiple controllers

V. CONCLUSION

This paper proposes a control strategy based on a fuzzy neural network PID algorithm for real-time dynamic tuning of PID parameters in system models characterized by time-varying uncertainties and nonlinearities. The proposed method integrates the advantages of both fuzzy control and neural networks. The fuzzy controller leverages existing expert experience to accelerate the initialization and learning process of the neural network, while the neural network enhances the self-learning and adaptive capabilities of the fuzzy controller, enabling real-time online tuning of PID parameters. To address the issue of low output precision in transplanting manipulators caused by external disturbances, a mathematical model of a fuzzy neural network PID control system is developed based on a conventional PID controller. Simulation analysis is conducted using Matlab and Simulink. By comparing the control performance of conventional PID control, type-1 fuzzy PID control, interval type-2 fuzzy PID control, and the proposed fuzzy neural network PID control, the results demonstrate the relative superiority of the fuzzy neural network intelligent control strategy for the target system model.

ACKNOWLEDGEMENTS

This research was supported by "Guangdong Scienceand Technology Program", "Research on Investigation ofthe Multimodal Intelligent Flying Electric Motorcycle's Autonomous Motion. (NO.

2024A0505050019)", theirhelp in teaching the research.

REFERENCES

[1] Alimhillaj, P., Lamani, L., & Vedova, D. L. D. M. (2024). Simulation of an aerospace electrohydraulic servomechanism with different Coulomb friction models. *Journal of Physics: Conference Series*, 2716(1).

[2] Mintsa, A. H., Eny, E. G., Senouveau, N., et al. (2023). Optimal tuning PID controller gains from Ziegler-Nichols approach for an electrohydraulic servo system. *Journal of Engineering Research and Reports*, 25(11), 158-166.

[3] Borase, R. P., Maghade, D. K., Sondkar, S. Y., et al. (2021). A review of PID control, tuning methods and applications. *International Journal of Dynamics and Control*, 9, 818-827.

[4] Carvajal, J., Chen, G., & Ogmen, H. (2000). Fuzzy PID controller: Design, performance evaluation, and stability analysis. *Information Sciences*, 123(3-4), 249-270.

[5] Nguyen, A. T., Taniguchi, T., Eciolaza, L., et al. (2019). Fuzzy control systems: Past, present and future. *IEEE Computational Intelligence Magazine*, 14(1), 56-68.

[6] Xiaolin, D., Pengfei, Z., Wanlu, Z., et al. (2023). Fuzzy adaptive PID speed controller design for modern elevator traction machine. *Energy Reports*, 9(S8), 175-183.

[7] Jin, X., Chen, K., Zhao, Y., et al. (2020). Simulation of hydraulic transplanting robot control system based on fuzzy PID controller. *Measurement*, 164.

[8] Castillo, O., & Melin, P. (2012). A review on the design and optimization of interval type-2 fuzzy controllers. *Applied Soft Computing Journal*, 12(4), 1267-1278.

[9] Wu, D., & Mendel, J. M. (2011). On the continuity of type-1 and interval type-2 fuzzy logic systems. *IEEE Transactions on Fuzzy Systems*, 19(1), 179-192.

[10] Li, M., Li, J., Li, G., et al. (2022). Analysis of active suspension control based on improved fuzzy neural network PID. *World Electric Vehicle Journal*, 13(12), 226.

[11] Jin X ,Chen K ,Zhao Y , et al. Simulation of hydraulic transplanting robot control system based on fuzzy PID controller [J]. Measurement, 2020, 164 (prepublish)

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.10

Top Challenges for Manufacturing Enterprises through 2030: A Structured Review and Mitigation Operating Model

Dmitrii Voistrocheko

Lean Six Sigma Black Belt, LLC Robotek, Moscow, Russia
Email: dmvoist@gmail.com

Received: 12 Feb 2026; Received in revised form: 15 Mar 2026; Accepted: 18 Mar 2026; Available online: 21 Mar 2026

Abstract *— Manufacturing enterprises through 2030 face compounded pressures: persistent supply chain volatility, continued cost pressure, rapid skills disruption, growing operational technology (OT) cybersecurity exposure, and intensifying sustainability and carbon-related compliance requirements. Although these challenges are widely documented, many organizations struggle to convert them into coherent, scalable transformation programs. This paper presents a structured narrative review and introduces a Mitigation Operating Model (MOM) that links external pressures to internal managerial failure modes and governance mechanisms. Evidence is synthesized from authoritative industry outlooks, global benchmark programs on industrial transformation, policy and standards bodies, and peer-reviewed research on resilience, digital transformation, and performance measurement behavior. The review identifies six recurring failure modes that inhibit scaling: unclear decision rights, weak KPI architecture and incentive alignment, low data integrity, fragmented transformation portfolios, insufficient capability building, and under-managed interfaces between operations and risk/compliance. The MOM consolidates actionable levers – data stewardship, cadence-based performance management, portfolio stage-gates, OT risk controls, and compliance-ready measurement – into a blueprint for prioritizing initiatives under constraints of capital and talent. The review concludes that the highest-leverage interventions institutionalize ownership and measurement integrity before scaling digital/AI or compliance programs, enabling simultaneous gains in cost, resilience, and readiness through 2030.*

Keywords *— **cybersecurity, governance, manufacturing challenges, resilience, skills disruption***

I. INTRODUCTION

Manufacturing enterprises through 2030 are operating in an environment where multiple external pressures interact rather than occur in isolation. Industry outlooks anticipate continued supply chain risks, disruptions, and elevated costs as persistent constraints for manufacturers' operational performance [1]. In parallel, workforce disruption remains high: forward-looking labor market analyses for the 2025–2030 horizon report substantial shifts in the skill composition required by employers, reinforcing the need for systematic upskilling and capability building [2]. Operational technology (OT) cybersecurity has also become an availability and safety issue, as OT-focused security guidance explicitly addresses unique OT performance, reliability, and safety requirements in connected industrial environments [3]. Finally, for firms exposed to regulated export markets, sustainability is increasingly converted from aspiration into auditability and compliance. In the European Union (EU), the Carbon Border Adjustment Mechanism (CBAM) increases pressure for auditable embedded-emissions measurement and reporting as it transitions into the definitive regime from 2026 onward [4], [5].

At the system level, industry remains a material contributor to global energy-related emissions, supporting the strategic relevance of decarbonization and measurement readiness [6].

Existing reports and studies frequently list manufacturing challenges, but they often do not translate them into a repeatable management system for mitigation and scaling. In practice, stalled transformations are frequently driven by organizational mechanisms—ownership ambiguity, KPI/incentive distortion, weak data integrity, and fragmented portfolios—rather than by the absence of technical solutions.

This paper aims to: (a) synthesize key cross-industry challenges affecting manufacturing enterprises through 2030; (b) identify managerial failure modes that explain why mitigation efforts fail to scale; and (c) propose a Mitigation Operating Model (MOM) mapping pressures → failure modes → governance mechanisms → expected outcomes.

II. REVIEW METHODOLOGY

This study follows a structured narrative review ("PRISMA-lite") suitable for cross-domain synthesis of operations, digital transformation, cybersecurity, and compliance evidence.

Evidence was collected from four categories: (1) industry outlooks and executive/sector analyses providing current-state constraints and near-term baseline conditions (e.g., cost and supply chain disruption) [1]; (2) benchmark programs and applied transformation compendiums documenting scaling patterns and operating disciplines in high-performing plants [7]; (3) policy and standards bodies defining applicable requirements and recommended practices (OT cybersecurity; manufacturing cybersecurity profiles; CBAM rules) [3]–[5], [8]; and (4) peer-reviewed research validating key mechanisms, especially resilience strategies, digital transformation enablers, and performance measurement dysfunction [9]–[12]. The horizon "through 2030" is treated as a forward-looking planning window grounded in sources that explicitly cover 2025–2030 (e.g., skills outlooks), alongside current-state manufacturing outlooks and standards/regulation shaping operating constraints between now and 2030.

Search terms combined "manufacturing" with: resilience, supply chain disruption, digital transformation barriers, data-driven transformation, performance measurement dysfunctional behavior, OT security, skills disruption, carbon compliance, and CBAM. Academic databases and publisher portals were used to identify peer-reviewed literature; policy/standards documents were selected from official issuers; and industry outlooks and benchmark publications were selected from established publishers.

Sources were included if they (a) addressed manufacturing operations or industrial transformation; (b) connected pressures to outcomes (cost, service, resilience, cyber risk, compliance readiness); (c) provided actionable mechanisms (frameworks, maturity models, practices); and (d) were either recent (primarily 2020–2026 for contextual evidence) or were active standards/regulatory guidance applicable during the period to 2030. Sources were excluded if they were promotional without traceable methodology, duplicates of the same underlying content, or not directly applicable to manufacturing execution and governance.

The synthesis used three coding passes: (1) pressure coding (external challenges); (2) failure-mode coding (internal managerial/organizational mechanisms); and (3) lever coding (governance and operating mechanisms), consolidated into the MOM.

III. CHALLENGE TAXONOMY (THROUGH 2030)

To provide a structured overview of the pressures discussed in this section, Table 1 summarizes the main manufacturing challenges through 2030, their typical operational symptoms, managerial implications, and indicative monitoring metrics.

Table 1: Challenge taxonomy and managerial implications through 2030. These challenges are discussed in greater detail below.

External pressure (through 2030)	**Typical operational symptoms**	**Managerial implication (why it's hard)**	**Example monitoring metrics**
Supply chain volatility & resilience	Schedule instability; expediting; inventory buffers; supplier variability	Requires cross-functional decision rights and rapid trade-off routines, not only forecasting	OTIF; lead-time variability; supplier PPM; inventory turns
Cost pressure & margin compression	Standard vs actual cost gaps; rework/scrap; overtime; unplanned downtime	Cost programs fail when KPIs/incentives reward output over accuracy and capability	COPQ; OEE; conversion cost/unit; scrap rate
Skills disruption & talent scarcity	Long ramp-up; shortages in maintenance/automation/planning; dependence on "heroes"	Needs institutional learning system (skill matrix, certification, standard work governance)	Time-to-competence; training hours/FTE; FPY; maintenance backlog
Digital scaling bottlenecks ("pilot trap")	Pilots succeed but don't scale; low adoption; inconsistent definitions; fragmented data	Scaling is a governance and portfolio problem (ownership, KPI hygiene, replication)	Use-case scale rate; adoption %; data quality score; benefit realization rate
OT cybersecurity as continuity risk	Expanded remote/vendor access; patching constraints; incomplete asset inventory; recovery delays	Must be governed jointly by operations–engineering–IT with uptime/safety constraints	Asset coverage %; segmentation coverage; incident MTTR; remote access exceptions
Sustainability & carbon compliance readiness	Need auditable product carbon data; supplier data requests; reporting overhead	Primarily a measurement and data governance challenge; requires audit-ready workflows	Energy intensity; emissions data completeness; supplier data coverage; audit findings

Supply chain volatility remains a structural constraint on operational performance, driving renewed emphasis on resilience, flexibility, and rapid reconfiguration. Benchmark evidence from industrial transformation programs highlights holistic transformation at scale, including resilience outcomes rather than isolated efficiency improvements [7]. Peer-reviewed resilience literature similarly identifies flexible strategies (e.g., sourcing and capacity flexibility) as meaningful contributors to resilience performance [9]. Operational symptoms include schedule instability, expediting, inventory buffers, and supplier variability. The managerial implication is that resilience requires cross-functional governance and decision rights that enable rapid trade-offs, not only improved forecasting.

Cost pressure remains persistent due to disrupted logistics, volatile inputs, and the structural cost of variability (rework, scrap, changeovers, downtime). Industry outlooks emphasize elevated costs and disruptions as continuing constraints for manufacturers [1]. Operational symptoms include gaps between standard and actual costs, "hidden factory" losses, and constrained improvement

budgets. The managerial implication is that cost programs underperform when measurement systems and incentives prioritize short-term output over accuracy, learning, and capability.

The workforce challenge is not only headcount scarcity but also skill composition change. Forward-looking analyses for the 2025–2030 period emphasize substantial shifts in required skills, increasing the urgency of systematic upskilling [2]. As shown in Deloitte's 2025 Manufacturing Industry Outlook - Fig. 1, labor market tightness in manufacturing eased during 2024, and July marked the first month since May 2021 when the number of unemployed individuals in manufacturing exceeded the number of job openings. However, this shift should not be interpreted as a resolution of the sector's talent problem. Instead, it suggests that the challenge is evolving from pure labor scarcity toward a broader issue of capability availability, role fit, and workforce adaptability.

Fig. 1: Easing labor market tightness in manufacturing does not eliminate the long-term skills challenge. Source: Adapted from Deloitte, 2025 Manufacturing Industry Outlook

Operational symptoms include shortages in maintenance, automation, and planning roles; long ramp-up times; and reliance on key individuals. The managerial implication is that training must be institutionalized (skills matrices, certification pathways, standard work governance) rather than treated as episodic.

Many manufacturers achieve pilot success but fail to scale across plants or value streams. Benchmark programs emphasize scaling mechanisms and operating discipline rather than isolated showcases [7]. Peer-reviewed research indicates that organizational and cultural factors play a critical role in enabling data-driven transformation and performance impact [10]. Operational symptoms include fragmented architectures, low adoption, inconsistent definitions, and benefits that are not sustained. The managerial implication is that without data ownership, KPI hygiene, and portfolio governance, digital investments become disconnected tools.

OT cybersecurity is increasingly intertwined with uptime and safety in connected industrial environments. OT-specific security guidance

addresses OT constraints and recommends practices aligned to industrial conditions [3]. Manufacturing-specific cybersecurity profiles also emphasize sector-aligned outcomes and prioritized practices [8]. Operational symptoms include expanded remote/vendor access, inconsistent patching, and incomplete asset inventories. The managerial implication is that OT risk must be governed as an operations and safety issue, not solely an IT function.

Sustainability increasingly requires auditable measurement across products and supply chains. Industry emissions remain material, keeping decarbonization and measurement on the executive agenda [6]. For EU-exposed firms, CBAM increases the need for embedded-emissions data and reporting processes, with the definitive regime beginning from 2026 onward [4], [5]. Operational symptoms include demand for product carbon traceability, supplier emissions information requests, and competing capex priorities. The managerial implication is that sustainability becomes a measurement and governance problem as much as an engineering problem.

IV. MANAGERIAL FAILURE MODES (WHY MITIGATION STALLS)

Across the reviewed evidence, initiatives frequently fail not because solutions are unknown, but because execution systems are insufficient. Six recurring failure modes were identified. Table 2 consolidates the identified failure modes, their typical manifestations, and the corresponding high-leverage governance controls.

Table 2: Managerial failure modes and recommended governance controls.

Failure mode	How it manifests	High-leverage controls (MOM levers)
Unclear decision rights	Slow decisions; inconsistent standards; escalation overload	RACI for data/process owners; decision cadences; escalation rules
KPI/incentive misalignment	Metric gaming; distorted reporting; local optimization	KPI hierarchy; leading/lagging split; KPI audit; incentive redesign
Low data integrity	Planning noise; poor costing; low trust in analytics	Data stewardship; definitions; measurement audits; single source of truth
Fragmented portfolio	Too many pilots; low scale rate; change fatigue	Stage gates; benefit verification; replication playbooks; portfolio board
Capability gap	Dependence on experts; regression after turnover	Skill matrix; certification; standard work governance; learning routines
Risk/compliance interface gaps	Cyber controls bypassed for uptime; reporting scramble	Integrated OT governance; asset inventory/segmentation; audit-ready emissions workflow

Unclear decision rights and accountability: ambiguity over who owns master data (BOM/routings), downtime taxonomy, supplier remediation, OT segmentation, and compliance reporting leads to delayed decisions, inconsistent standards, and "everyone and no one" responsibility.

Weak KPI architecture and incentive misalignment: performance measurement systems can produce dysfunctional consequences when measures become targets, encouraging gaming, misreporting, and effort substitution [11]. Broader performance measurement literature reinforces that metric design and governance influence behavior and outcomes [12].

Low data integrity: inconsistent definitions and poor data quality undermine planning, costing, analytics credibility, and sustainability reporting readiness. Data integrity gaps also weaken trust and reduce adoption of digital tools.

Fragmented transformation portfolios: too many initiatives without stage gates, benefit verification, and replication playbooks yields pilot saturation but limited enterprise scale. This also increases change

fatigue and reduces credibility of transformation programs.

Insufficient capability building: training is not integrated into daily operations; standard work is not governed; knowledge remains localized. Improvements depend on "heroes" and degrade with turnover.

Under-managed operations–risk/compliance interfaces: OT security is treated as IT-only, sustainability as reporting-only, and supply chain risk as procurement-only, creating gaps at interfaces and weak controls.

V. MITIGATION OPERATING MODEL (MOM)

The Mitigation Operating Model translates external pressures into governance mechanisms that address failure modes and enable scalable outcomes. The MOM is structured into six operating layers and is intended as a management blueprint rather than a list of tools. The conceptual structure of the proposed Mitigation Operating Model is illustrated in Figure 2.

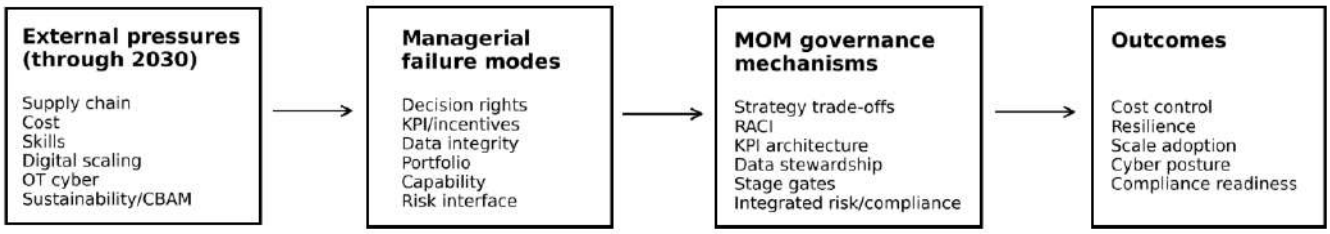


Fig. 2: Conceptual structure of the Mitigation Operating Model (MOM).

Strategic intent and trade-offs: define an explicit optimization logic (e.g., cost vs service vs resilience vs CO_2) and manage trade-offs to prevent KPI overload and conflicting local targets. Translate intent into a limited set of enterprise priorities with clear ownership.

Decision rights and accountability (RACI discipline): establish ownership for master data (BOM/routings), downtime and quality taxonomies, supplier risk actions, OT cybersecurity controls (asset inventory, segmentation, remote access), and compliance reporting workflows (e.g., embedded emissions). This layer reduces ambiguity, accelerates decisions, and enables stable standards.

KPI architecture and incentive alignment: design KPI hierarchies with leading and lagging indicators; define rules preventing local optimization; introduce periodic KPI audits for gaming risk; ensure incentives do not reward data distortion. Evidence on dysfunctional measurement consequences supports governance beyond metric selection [11], [12].

Data integrity and measurement governance: create data stewardship roles, a single source of truth for critical objects, and measurement audits. This layer is foundational for planning, digital scaling, and compliance. Typical mechanisms include data dictionaries, governance councils, stewardship KPIs, and systematic correction workflows.

Transformation portfolio and scaling mechanism: implement portfolio stage gates (idea → pilot → scale), benefit verification, and replication playbooks. Benchmark evidence emphasizes scaling patterns and replication across sites [7]. Portfolio governance should include intake and prioritization criteria, stage-gate reviews, benefit standards, and replication packages (standard work, training, technical steps).

Risk and compliance integration into operations: operationalize OT cybersecurity using OT-specific guidance and manufacturing profiles [3], [8]. Integrate sustainability measurement and reporting into operational routines to meet compliance requirements such as CBAM [4], [5].

VI. DISCUSSION

The synthesis indicates that the most damaging pattern is treating each challenge as independent and "solvable by projects," while the real constraint is the absence of a scalable operating model. Digital transformation depends strongly on organizational

culture and governance enabling data-driven behavior, consistent with peer-reviewed findings on cultural enablers of data-driven transformation [10]. Resilience strategies require cross-functional routines and governance, supported by systematic review evidence that flexible strategies contribute to resilience performance [9]. OT cybersecurity and sustainability compliance further shift from technical topics into governance: OT security must respect reliability and safety constraints and therefore requires integrated operations–engineering–IT governance [3]. CBAM compliance increases the need for auditable measurement and reporting processes and supplier data workflows [4], [5]. The interdependence of these pressures is illustrated in Figure 3.

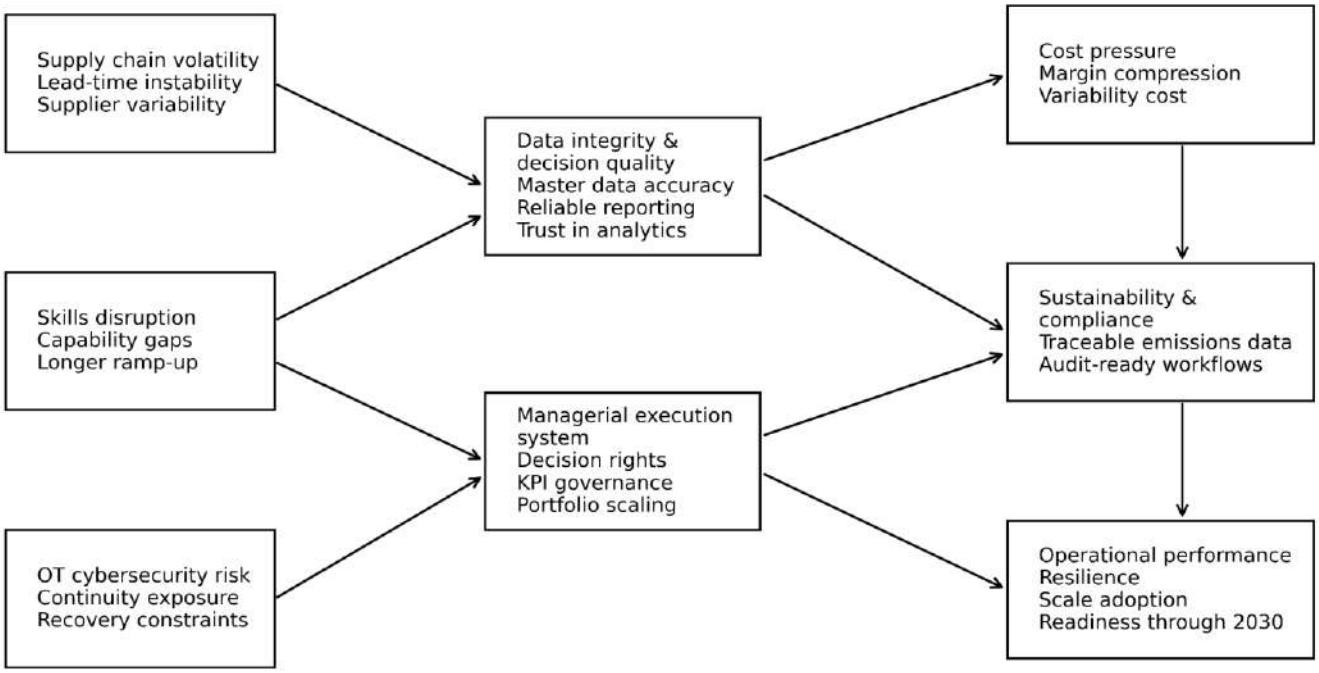


Fig. 3: Interdependence of manufacturing challenges through 2030.

A pragmatic sequencing under capital and talent constraints is: (1) clarify decision rights and KPI governance to reduce metric distortion and accelerate decisions; (2) fix measurement integrity and data ownership to restore trust in numbers and enable planning and analytics; (3) scale transformation via portfolio governance using stage gates and replication playbooks to avoid pilot saturation; and (4) integrate OT security and sustainability compliance into operational governance so risk and reporting become routine, not exceptions. This sequencing aligns with benchmark evidence that leading plants scale holistic transformation rather than isolated tools [7].

For practitioners, the MOM offers a blueprint for converting external pressures into governance mechanisms and scalable execution routines. For researchers, the MOM provides a testable structure that could be operationalized into a maturity assessment and evaluated through multi-site case studies across sectors to test transferability and quantify impact pathways.

VII. CONCLUSION

Manufacturing enterprises through 2030 face interacting pressures spanning supply chain volatility, cost pressure, skills disruption, digital scaling barriers, OT cybersecurity risk, and sustainability compliance. This paper contributes a structured taxonomy of these challenges, six managerial failure modes explaining why mitigation efforts stall, and a Mitigation Operating Model (MOM) mapping pressures to governance mechanisms.

The highest-leverage interventions institutionalize ownership, KPI/incentive alignment, and measurement integrity before scaling digital/AI or compliance programs. Future work can

operationalize the MOM into a maturity assessment instrument and validate it through multi-site case studies.

REFERENCES

[1] Deloitte. (2024). 2025 Manufacturing Industry Outlook. Deloitte Insights.

[2] World Economic Forum. (2025). The Future of Jobs Report 2025 (skills outlook; 2025–2030 horizon).

[3] Stouffer, K., Falco, J., & Scarfone, K. (2023). Guide to Operational Technology (OT) Security (NIST SP 800-82 Rev. 3). National Institute of Standards and Technology.

[4] European Commission. (n.d.). Carbon Border Adjustment Mechanism (CBAM): Overview and implementation timeline.

[5] European Commission. (2026). CBAM successfully entered into force on 1 January 2026 (official update).

[6] International Energy Agency. (n.d.). Industry: Energy system and emissions tracking.

[7] World Economic Forum. (2026). Global Lighthouse Network: Rewiring operations for resilience and impact at scale.

[8] National Institute of Standards and Technology. (2025). Manufacturing Profile for the Cybersecurity Framework 2.0 (NIST IR 8183 family).

[9] Paul, A. (2025). Flexible strategies and supply chain resilience performance: A systematic literature review. Global Journal of Flexible Systems Management.

[10] Ghafoori, A., Gupta, M., Merhi, M. I., et al. (2024). Organizational culture and data-driven digital transformation. International Journal of Production Economics.

[11] Ridgway, V. F. (1956). Dysfunctional consequences of performance measurements. Administrative Science Quarterly, 1(2), 240–247.

[12] Lewis, J. M. (2015). The politics and consequences of performance measurement. Policy and Society, 34(1), 1–12.

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.11

Risk-Based Autoclave Steam Sterilization Cycle Design and Validation for Mixed Biopharmaceutical Loads: Thermodynamic Mapping Integrated with FMEA

Pujari Prasantha

Validation Engineer, Pfizer Oncology, United States

Received: 12 Feb 2026; Received in revised form: 15 Mar 2026; Accepted: 18 Mar 2026; Available online: 21 Mar 2026

Abstract *— The article addresses the development and validation of risk-based steam sterilization cycles for mixed biopharmaceutical loads by integrating thermodynamic mapping with FMEA methodology. The relevance of the work is driven by increasingly stringent regulatory expectations and the limitations of template autoclave cycles, which do not ensure reliable air removal or the uniform attainment of sterilizing conditions in complex, porous, and combined loads. The aim of the study is to establish a scientifically justified model for cycle design that ensures not only the attainment of the required lethality and SAL levels, but also their robust defensibility before regulators. Methodologically, the work is grounded in the Quality by Design paradigm and a mixed-methods research design that includes high-frequency thermal mapping using wireless data loggers in typical mixed loads, lethality calculation, and a modified FMEA that directly links cycle parameters (vacuum pulses, heating rate, equilibration time) to specific failure modes. It is shown that an unsuccessful cycle is characterized by thermodynamic instability, persistent air pockets, and failure to reach the minimum temperature of 121°C at the critical point, whereas an optimized cycle with intensified vacuuming exhibits rapid equilibration (<15 s), a narrow temperature corridor of 121–124°C, and excess lethality (F0≈29–32 min) that substantially exceeds the overkill criterion. The scientific novelty lies in treating the mixed load as a dynamic risk object and in proposing a hierarchy of acceptance criteria, placing primary emphasis on physical indicators (P/T correlation, equilibration time, absence of cold spots), with integral lethality used as a derived parameter. The article is intended for validation engineers, quality assurance specialists, and biopharmaceutical process developers involved in the design and defense of steam sterilization cycles.*

Keywords *— **steam sterilization, autoclaving, quality risk management, ISO 17665-1:2024, sterility assurance, lethality***

I. INTRODUCTION

Sterilization with saturated steam (autoclaving) has historically been the dominant method of terminal sterilization and component sterilization in the pharmaceutical industry due to its high effectiveness, economic efficiency, and absence of toxic residues (Agalloco, 2024). However, the apparent simplicity of the process, exposure to 121°C for a specified time, conceals complex thermodynamic interactions, particularly when processing heterogeneous, porous, or structurally complex loads (Feurhuber et al., 2023).

Apart from the meaningful change in regulation in the last 10 years, with the publication of Annex 1 to the EU GMP guideline and the revision of ISO 17665-1:2024, the focus is now on process understanding and Quality Risk Management (QRM), both of which can

support continuous Quality improvement (PMS, 2022; ISO, 2024). This range indicates a tight contemporary regulatory authorities, such as the FDA and EMA, which are no longer satisfied with merely demonstrating that a cycle has passed; they require scientific justification of why it passed and how its reliability is ensured under process variability (Marschall et al., 2022).

The relevance of this study stems from the critical need to revise approaches to cycle development. The traditional practice of using template parameters (e.g., three pre-vacuum pulses and 15 minutes holding time) often leads to hidden failures such as the formation of air pockets in instrument cavities or migration of cold spots, creating risks for patient safety, particularly in the manufacture of highly potent products (Chakraborty et al., 2020).

The key problem lies in the disconnection between the physics of heat and mass transfer and the formalism of validation procedures. Existing methods often treat the autoclave load as a static thermal mass, ignoring the dynamic behavior of the steam–air mixture (Feurhuber et al., 2023).

Standard protocols frequently assume that temperature and pressure are rigidly correlated via the saturation curve; however, the presence of non-condensable gases (NCGs) disrupts this relationship (Dalton's law), resulting in ineffective sterilization even when the target temperature has been reached (Miguel et al., 2022).

Using legacy parameters without considering specific load conditions (loading density, wrapping material type, presence of lumens) results in wet packs and repeated qualifications (Chen et al., 2024).

Lethality calculations assume an instantaneous inactivation effect of temperature. However, in the absence of moisture (dry heat due to an air lock), spore resistance increases by orders of magnitude, rendering the calculated lethality invalid (Cho & Chung, 2020).

The aim of this work is to develop a comprehensive risk-based model for the design and validation of steam sterilization cycles that ensures process robustness and full transparency for regulatory audit.

To achieve this aim, the following tasks are addressed:

1. To conduct an in-depth comparative analysis of thermodynamic profiles of successful and unsuccessful sterilization cycles based on real engineering data.
2. To identify the physical mechanisms underlying the formation of cold spots and barriers to steam penetration in complex mixed loads.
3. To adapt the FMEA methodology to the cycle development stage by linking each process parameter (heating rate, vacuum depth) to a specific failure mode.
4. To formulate practical acceptance criteria for engineering studies that exceed minimal regulatory requirements and thereby provide a margin of robustness (Design Space).

The study's scientific novelty lies in integrating quantitative data from high-precision thermal mapping with qualitative risk assessment tools during the pre-validation stage. For the first time, the correlation between equilibration dynamics (equilibration time) and the stability of the sterilization plateau is examined in detail as a predictor of PQ (Performance Qualification) success. The work proposes a shift from a static template to a dynamic risk object, which requires targeted engineering controls.

II. MATERIALS AND METHODS

The study was conducted within the paradigm of mixed-methods research, combining experimental data collection (engineering runs of the autoclave), quantitative analysis (statistical processing of thermograms), and qualitative synthesis (risk analysis and interpretation of regulatory requirements). Quality by Design (QbD) serves as the key methodological principle, according to which, in this context, sterility must be built into the process at the development stage rather than confirmed solely by end-product testing. The methodological logic is implemented through an iterative cycle including sequential stages of risk assessment, parameter design, empirical testing, data analysis, and optimization.

The primary data-collection tool was the wireless validation system, Ellab TrackSense Pro. Its use is justified by a combination of high measurement

accuracy (0.05°C), stable performance under vacuum and saturated steam, and the ability to place loggers inside sealed packages without compromising their integrity, which fundamentally distinguishes this approach from wired thermocouples. The configuration used an array of 18 temperature loggers (Rigid and Flexible sensors) and one integrated pressure/temperature sensor. The sampling frequency was 1 second, enabling registration of rapid transient processes characteristic of vacuum pulsations and equilibration stages.

The object of study comprised mixed loads typical for biopharmaceutical manufacturing. These included metal instruments, characterized by high thermal conductivity and low heat capacity; porous materials (e.g., gowns and filters), which exhibit heightened complexity in air removal; and items with cavities and tubing where there is a risk of air pocket formation. Additionally, containers with liquids were included, which possess pronounced thermal inertia and thus influence warming dynamics and temperature distribution throughout the cycle.

The analytical component encompassed thermodynamic analysis, lethality calculation, and risk analysis. Thermodynamic analysis was based on assessing deviations of the actual temperature from the theoretical temperature of saturated steam, calculated using the Antoine equation or steam tables derived from the measured pressure. Significant negative deviation was interpreted as evidence of air presence, since the partial pressure of air reduces the temperature of the steam–gas mixture, whereas positive deviation was considered a possible indication of superheated steam.

For lethality calculation, the standard cumulative lethality algorithm defined in ISO 17665 and pharmacopoeias was applied. To systematize cause–and–effect relationships and identify critical process vulnerabilities, a modified FMEA (Failure Mode and Effects Analysis) focused on sterilization cycle parameters was used. For each cycle stage, including air removal, heating, holding, and drying, potential failure modes, their causes, and detection methods were defined, including such scenarios as incomplete air removal from the center of a package.

III. RESULTS

This section presents a detailed analysis of the data obtained during the engineering studies. Comparing the failed and successful cycles provides a clear illustration of how process parameters influence the physics of sterilization.

Sensor placement strategy and identification of critical zones

The first stage of the study involved defining monitoring points. The effectiveness of validation directly depends on the ability to capture the worst-case scenario. Figure 1 visualizes the three-dimensional map of sensor placement.

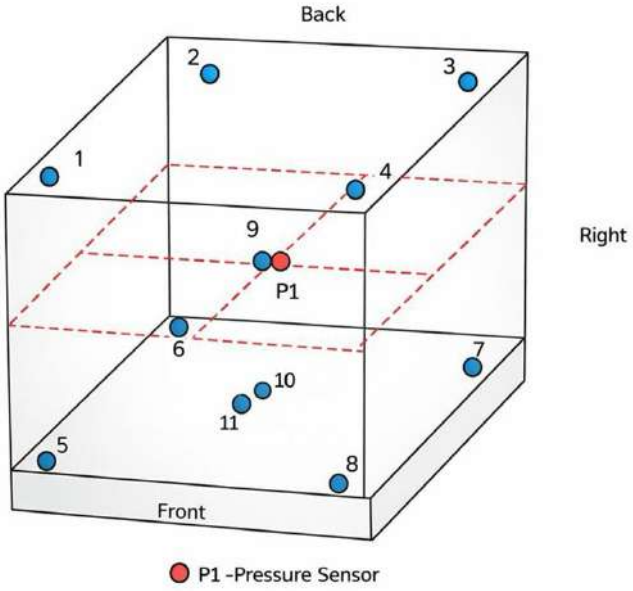


Fig. 1. Autoclave Logger placement diagram

The loggers, indicated by blue points LC 01–LC 18, are not distributed uniformly but are positioned deliberately according to a logic aimed at identifying zones with the highest probability of unfavorable heat and mass transfer conditions. This placement is oriented not toward formal symmetry but toward enhancing the measurement scheme's sensitivity to spatial temperature nonuniformity and to effects associated with residual air and condensation.

A portion of the sensors is installed at the geometric extrema of the chamber, i.e., in its corners and in the upper and lower regions. These positions serve as reference points because they reflect extreme conditions of heat distribution throughout the volume and allow detection of potential gradients arising

from specific features of steam circulation and the gravitational stratification of the gas mixture.

A separate group of loggers is placed in functional zones, primarily in the drain area. This region is interpreted as a critical point because condensate flows into it, and the coldest, heaviest gas-air mixture accumulates there. As a result, the drain zone is potentially associated with the formation of local cold regions and with the highest risk of failing to achieve target sterilization conditions.

The remaining sensors are placed directly inside the load, including the internal volumes of packages and cavities, in order to simulate the most difficult locations for steam penetration, designated as cold spots. Such positioning enables recording conditions in regions where steam access is hindered by geometry, material porosity, or the presence of cavities, thereby enabling experimental verification of the most critical scenarios for complete air removal and uniform heating.

A pressure sensor, P1, is placed at the center of the chamber. This configuration (18+1) exceeds the minimum requirements of many standards (typically 12 sensors for chambers up to 1 m^3), which allows a more detailed picture of the thermal field to be obtained (Boehler et al., 2021). Cold spots are rarely located where legacy templates suggest, so a dense sensor grid is necessary to empirically identify risk zones.

Analysis of the unsuccessful cycle (Failed Study Data)

The first iteration of cycle development was carried out using parameters based on standard templates, without thorough adaptation to the load specifics. Figure 2 shows classical signs of thermodynamic instability.

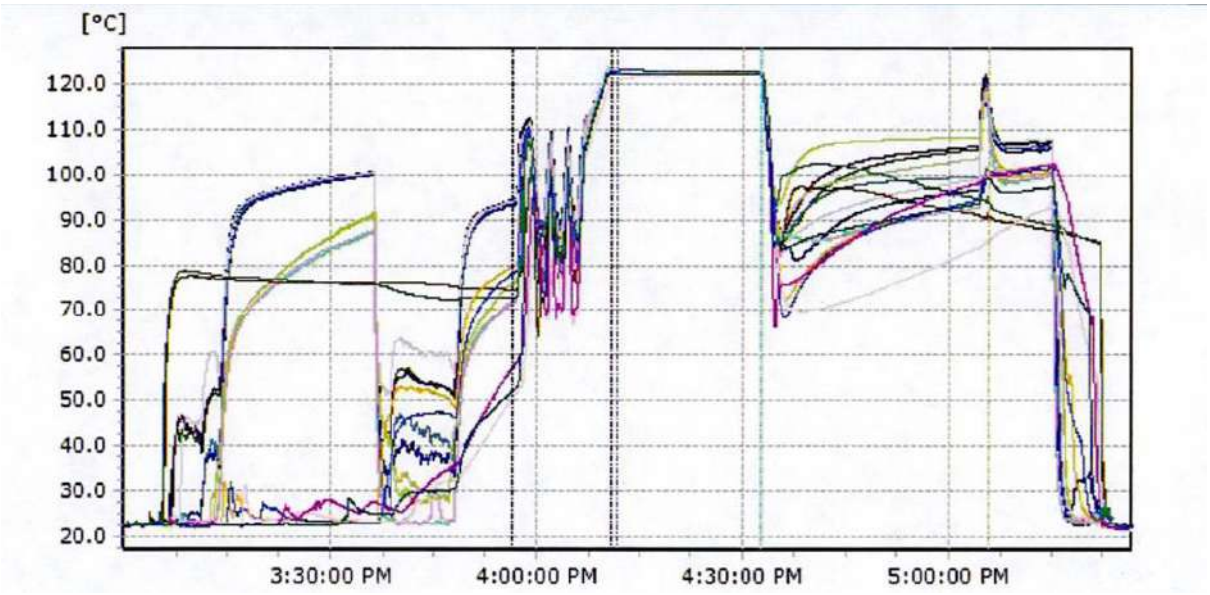


Fig. 2. Failed Study Data Autoclave Cycle Graph

The graph demonstrates classical indications of thermodynamic instability. During the come-up phase, the temperature curves of individual sensors do not form a single coherent bundle; instead, a pronounced divergence of lines and a significant spread between the hottest and coldest points are observed. This pattern indicates spatial nonuniformity in heat transfer conditions within the chamber and shows that different regions of the medium are heating at different rates.

Subsequently, slow equilibration is observed: some sensors reach the target temperature of 121°C with a marked delay relative to the control sensor positioned in the drain zone. In thermodynamic terms, this means that temperature-field equalization and the approach of the system to a quasi-stationary state do not occur synchronously throughout the volume but with substantial time lags at individual points.

During the sterilization plateau, instability persists. Temperature fluctuations are seen during the holding phase, indicating dynamic processes in the

steam–gas medium and at the load surface. Such fluctuations are interpreted as manifestations of air pockets that periodically compress and expand, or as resulting from steam condensation on relatively cold regions, which leads to local changes in the thermal regime.

Figure 3 confirms the use of 16 temperature sensors, but the absence of an independent pressure logger.

Sensor ID	Name	Validation Equipment	Description	Sample Rate 1/2 (hh:mm:ss)	Type	Sensor group
685941	LC 01	715757	Temperature	00:00:01 / N/A	LC	N/A
685943	LC 02	715739	Temperature	00:00:01 / N/A	LC	N/A
685946	LC 03	715889	Temperature	00:00:01 / N/A	LC	N/A
685948	LC 04	715753	Temperature	00:00:01 / N/A	LC	N/A
685950	LC 05	715913	Temperature	00:00:01 / N/A	LC	N/A
686043	LC 06	715892	Temperature	00:00:01 / N/A	LC	N/A
686044	LC 07	715914	Temperature	00:00:01 / N/A	LC	N/A
686045	LC 08	715920	Temperature	00:00:01 / N/A	LC	N/A
686046	LC 09	715887	Temperature	00:00:01 / N/A	LC	N/A
686049	LC 10	715882	Temperature	00:00:01 / N/A	LC	N/A
686050	LC 11	715915	Temperature	00:00:01 / N/A	LC	N/A
686052	LC 12	715905	Temperature	00:00:01 / N/A	LC	N/A
686056	LC 13	715903	Temperature	00:00:01 / N/A	LC	N/A
686057	LC 14	715902	Temperature	00:00:01 / N/A	LC	N/A
686063	LC 15	715765	Temperature	00:00:01 / N/A	LC	N/A
686067	LC 16	715780	Temperature	00:00:01 / N/A	LC	N/A

Fig. 3. 16 Ellab Data Logger used and No Pressure data logger

This is a critical study design error. Without independent pressure data, it is impossible to calculate the theoretical saturated steam temperature and thus to confirm the presence of saturation conditions.

The data in Figure 4 show catastrophic outcomes for a biopharmaceutical process.

Equilibration Time Statictics

Name:	Equilibration Time Statictics
Description:	Time difference from sterilization start (reference temp reaching 121C) and all remaining loggers reaching 121C.

Input parameters

Data Series Type:	Temperature
Data series:	LC 01, LC 02, LC 03, LC 04, LC 05, LC 06, LC 07, LC 08, LC 09, LC 10, LC 11, LC 12, LC 13, LC 14, LC 15, LC 16
Start Time:	4:10:46 PM
Stop Time:	4:10:51 PM
Output:	Global Data analysis
Functions:	Min., Max., Average

Global Data analysis

Min.	119.8 °C	4:10:46 PM	LC 14
Max.	123.9 °C	4:10:46 PM	LC 05
		4:10:47 PM	LC 05
		4:10:48 PM	LC 05
Average	122.9 °C		

Fig. 4. Equilibration Time (Failed Cycle)

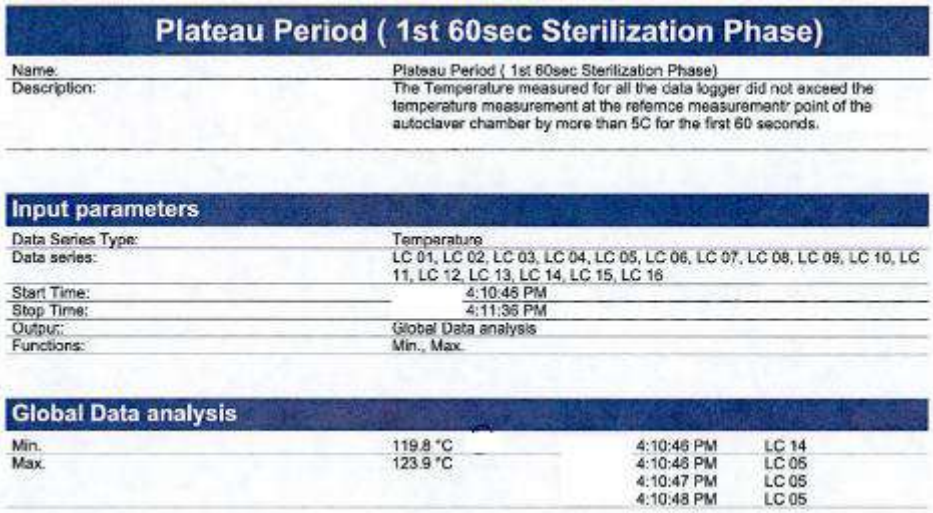
At the presumed equilibration, when the control sensor reached 121°C, the minimum temperature in the load was 119.8°C, as recorded by sensor LC 14.

Consequently, point LC 14 did not reach sterilization conditions. If a critical product component were located at this point, it would have remained non-sterile. The difference of 1.2°C may appear small. However, on a logarithmic scale of microbial inactivation, such a deviation results in a substantial reduction in lethality. Figure 5 illustrates the plateau period.

Throughout the plateau, the minimum temperature remained at 119.8°C. This demonstrates that the issue was not a transient heating lag but a systemic problem, most likely the presence of a stable air pocket acting as a thermal insulator, preventing steam from contacting the surface.

Analysis of the optimized cycle (Pass Data)

Based on the analysis of the failed cycle, modifications were introduced: parameters of vacuum pulsations were adjusted (increased depth and number), and the load configuration was altered to facilitate air removal. Figure 6 shows the autoclave cycle graph for the pass data.



Plateau Period (1st 60sec Sterilization Phase)

Name:	Plateau Period (1st 60sec Sterilization Phase)
Description:	The Temperature measured for all the data logger did not exceed the temperature measurement at the refernce measurement point of the autoclaver chamber by more than 5C for the first 60 seconds.

Input parameters

Data Series Type:	Temperature
Data series:	LC 01, LC 02, LC 03, LC 04, LC 05, LC 06, LC 07, LC 08, LC 09, LC 10, LC 11, LC 12, LC 13, LC 14, LC 15, LC 16
Start Time:	4:10:46 PM
Stop Time:	4:11:36 PM
Output:	Global Data analysis
Functions:	Min., Max.

Global Data analysis

Min.	119.8 °C	4:10:46 PM	LC 14
Max.	123.9 °C	4:10:46 PM	LC 05
		4:10:47 PM	LC 05
		4:10:48 PM	LC 05

Fig. 5. Plateau Period (Failed Cycle)

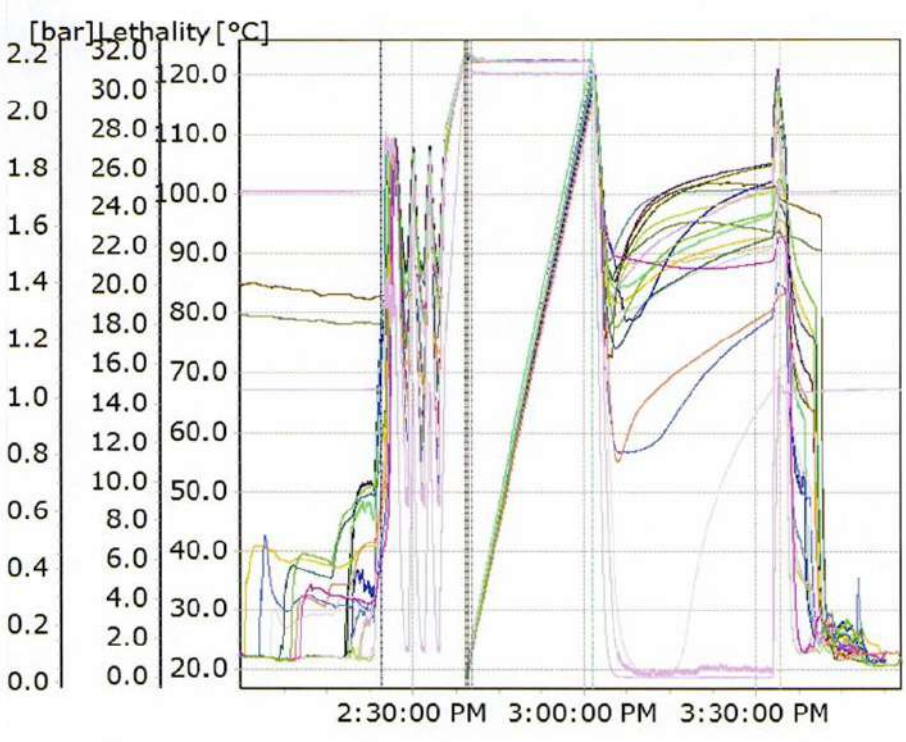


Fig. 6. Pass Data Autoclave Cycle Graph

Visually, the graph differs drastically from the previous one. Almost immediately after the start of the heating phase, convergence is observed: all 18 temperature curves merge into a narrow band, indicating coordinated behavior of measurement points and rapid transition of the system to a thermally similar state throughout the entire volume.

The cycle structure is clearly expressed on the

graph. Transitions between vacuum, heating, sterilization, and drying appear sharp and controlled, without extended segments and without indications of prolonged equilibration typical of unstable regimes. This compactness of the phase pattern reflects process controllability and the predictability of temperature dynamics.

During the holding phase, stability is maintained: the temperature of all sensors remains consistently within the range of 121–124°C. The absence of notable divergence between curves and the lack of pronounced oscillations at the plateau correspond to a state in which the temperature field remains homogeneous within the specified window and is maintained without significant disturbances. Figure 7 illustrates the use of 18 Ellab data loggers and one pressure data logger.

Sensor ID	Name	Validation Equipment	Description	Sample Rate 1/2 (hh:mm:ss)	Type	Sensor group
685941	LC 01	715757	Temperature	00:00:01 / N/A	LC	N/A
685943	LC 02	715739	Temperature	00:00:01 / N/A	LC	N/A
685946	LC 03	715889	Temperature	00:00:01 / N/A	LC	N/A
685948	LC 04	715753	Temperature	00:00:01 / N/A	LC	N/A
685950	LC 05	715913	Temperature	00:00:01 / N/A	LC	N/A
686043	LC 06	715892	Temperature	00:00:01 / N/A	LC	N/A
636557	LC 07	715909	Temperature	00:00:01 / N/A	LC	N/A
686045	LC 08	715920	Temperature	00:00:01 / N/A	LC	N/A
686046	LC 09	715887	Temperature	00:00:01 / N/A	LC	N/A
686049	LC 10	715882	Temperature	00:00:01 / N/A	LC	N/A
686050	LC 11	715915	Temperature	00:00:01 / N/A	LC	N/A
686052	LC 12	715905	Temperature	00:00:01 / N/A	LC	N/A
686056	LC 13	715903	Temperature	00:00:01 / N/A	LC	N/A
686057	LC 14	715902	Temperature	00:00:01 / N/A	LC	N/A
686063	LC 15	715765	Temperature	00:00:01 / N/A	LC	N/A
686067	LC 16	715780	Temperature	00:00:01 / N/A	LC	N/A
686069	LC 17	715799	Temperature	00:00:01 / N/A	LC	N/A
686070	LC 18	715747	Temperature	00:00:01 / N/A	LC	N/A
656186	LP 19	715899	Pressure	00:00:01 / N/A	LP	N/A
N/A	LP 19 - T-Theoretical 1	N/A	N/A	00:00:01 / N/A	TA	N/A

Fig. 7. 18 Ellab Data Logger used and 1 Pressure data logger

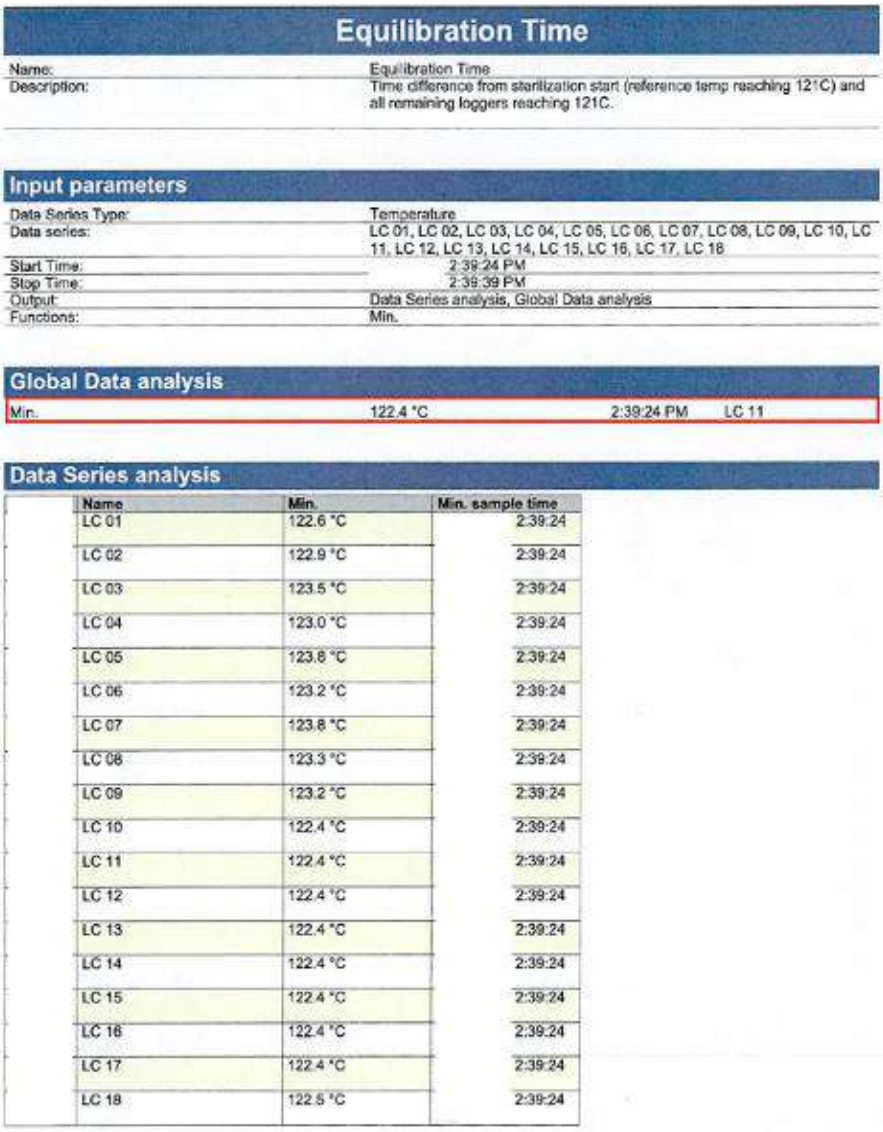
In the successful study, a pressure sensor (LP 19) was added. This allowed the report to include the line LP 19 – T-Theoretical, which shows the temperature corresponding to the current saturated steam pressure. The coincidence between actual temperatures and the theoretical curve confirms the steam quality (dryness >97%, absence of superheat). The equilibration time for the pass cycle is shown in Figure 8.

The equilibration time was less than 15 seconds. According to EN 285, the acceptable time for small loads lies within 15–30 seconds; however, the smaller this value, the better (BSI, 2021). Practically instantaneous equilibration is interpreted as an indicator of complete air removal, since the absence of delays between points indicates rapid establishment of saturated steam conditions throughout the volume.

The temperature corridor was achieved synchronously: all sensors entered the 121.0–124.0°C range simultaneously. Such simultaneous entry means that the temperature field was formed coherently, without pronounced local lags, and that the target window is maintained as a property of the system as a whole rather than a characteristic of isolated regions. Holding conditions after equilibration are shown in Figure 9.

In the summary analysis table for the holding phase, the minimum temperature was 122.2°C and was recorded by sensor LC 14, i.e., the same sensor that failed in the previous test. The maximum temperature reached 123.9°C and corresponded to sensor LC 05. This implies that the temperature distribution was very narrow, and cold spots were not strongly favored during the hold phase of the plateau.

From which follows that 1.2 degrees Celsius in excess of the 121 degree Celsius for each point will not exceed the upper limit of the safety temperature of 124 degrees Celsius, ensuring the effectiveness of the process while preserving the sterilized object in good conditions. Figure 10 shows how the lethality is calculated and what are the input values used.



Equilibration Time

Name:	Equilibration Time
Description:	Time difference from sterilization start (reference temp reaching 121C) and all remaining loggers reaching 121C.

Input parameters

Data Series Type:	Temperature
Data series:	LC 01, LC 02, LC 03, LC 04, LC 05, LC 06, LC 07, LC 08, LC 09, LC 10, LC 11, LC 12, LC 13, LC 14, LC 15, LC 16, LC 17, LC 18
Start Time:	2:39:24 PM
Stop Time:	2:39:39 PM
Output:	Data Series analysis, Global Data analysis
Functions:	Min.

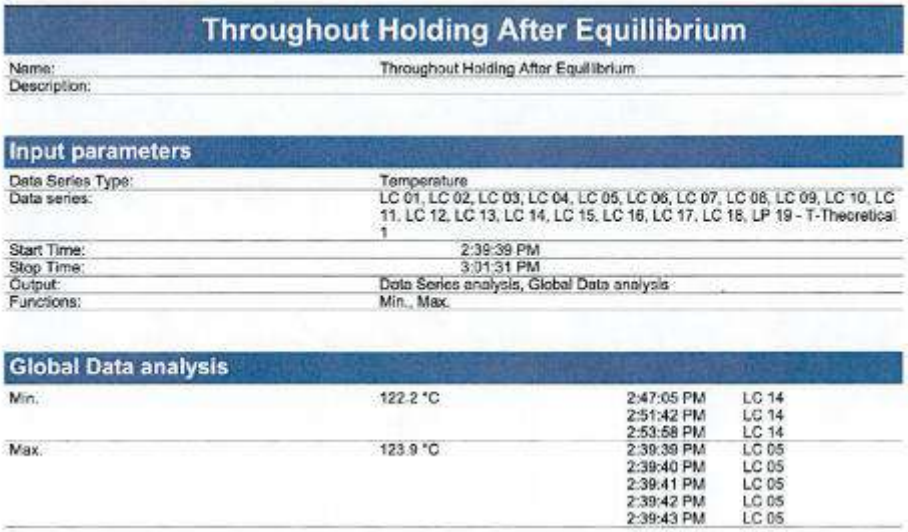
Global Data analysis

Min.	122.4 °C	2:39:24 PM	LC 11

Data Series analysis

Name	Min.	Min. sample time
LC 01	122.6 °C	2:39:24
LC 02	122.9 °C	2:39:24
LC 03	123.5 °C	2:39:24
LC 04	123.0 °C	2:39:24
LC 05	123.8 °C	2:39:24
LC 06	123.2 °C	2:39:24
LC 07	123.8 °C	2:39:24
LC 08	123.3 °C	2:39:24
LC 09	123.2 °C	2:39:24
LC 10	122.4 °C	2:39:24
LC 11	122.4 °C	2:39:24
LC 12	122.4 °C	2:39:24
LC 13	122.4 °C	2:39:24
LC 14	122.4 °C	2:39:24
LC 15	122.4 °C	2:39:24
LC 16	122.4 °C	2:39:24
LC 17	122.4 °C	2:39:24
LC 18	122.5 °C	2:39:24

Fig. 8. Equilibration Time (Pass Cycle)



Throughout Holding After Equillibrium

Name:	Throughout Holding After Equillibrium
Description:	

Input parameters

Data Series Type:	Temperature
Data series:	LC 01, LC 02, LC 03, LC 04, LC 05, LC 06, LC 07, LC 08, LC 09, LC 10, LC 11, LC 12, LC 13, LC 14, LC 15, LC 16, LC 17, LC 18, LP 19 - T-Theoretical 1
Start Time:	2:39:39 PM
Stop Time:	3:01:31 PM
Output:	Data Series analysis, Global Data analysis
Functions:	Min., Max.

Global Data analysis

Min.	122.2 °C	2:47:05 PM	LC 14
		2:51:42 PM	LC 14
		2:53:58 PM	LC 14
Max.	123.9 °C	2:39:39 PM	LC 05
		2:39:40 PM	LC 05
		2:39:41 PM	LC 05
		2:39:42 PM	LC 05
		2:39:43 PM	LC 05

Fig. 9. Throughout Holding After Equilibrium

Lethality Overview

Input parameter

Ref. Temperature:	121.1 °C
Min. Temperature :	91.1 °C
Z-Value:	10 °C
D-Value:	1

Sensor lethality overview

Name	Description	Start Time	End Time	Max. Value
LC 01 (F0) Lethality 1	L-Calculation	2:39:24 PM	3:01:31 PM	30.2
LC 02 (F0) Lethality 1	L-Calculation	2:39:24 PM	3:01:31 PM	30.8
LC 03 (F0) Lethality 1	L-Calculation	2:39:24 PM	3:01:31 PM	31.2
LC 04 (F0) Lethality 1	L-Calculation	2:39:24 PM	3:01:31 PM	30.7
LC 05 (F0) Lethality 1	L-Calculation	2:39:24 PM	3:01:31 PM	31.9
LC 06 (F0) Lethality 1	L-Calculation	2:39:24 PM	3:01:31 PM	30.8
LC 07 (F0) Lethality 1	L-Calculation	2:39:24 PM	3:01:31 PM	31.0
LC 08 (F0) Lethality 1	L-Calculation	2:39:24 PM	3:01:31 PM	31.0
LC 09 (F0) Lethality 1	L-Calculation	2:39:24 PM	3:01:31 PM	31.1
LC 10 (F0) Lethality 1	L-Calculation	2:39:24 PM	3:01:31 PM	29.5
LC 11 (F0) Lethality 1	L-Calculation	2:39:24 PM	3:01:31 PM	29.4
LC 12 (F0) Lethality 1	L-Calculation	2:39:24 PM	3:01:31 PM	29.6
LC 13 (F0) Lethality 1	L-Calculation	2:39:24 PM	3:01:31 PM	31.0
LC 14 (F0) Lethality 1	L-Calculation	2:39:24 PM	3:01:31 PM	29.3
LC 15 (F0) Lethality 1	L-Calculation	2:39:24 PM	3:01:31 PM	29.6
LC 16 (F0) Lethality 1	L-Calculation	2:39:24 PM	3:01:31 PM	29.8
LC 17 (F0) Lethality 1	L-Calculation	2:39:24 PM	3:01:31 PM	29.6
LC 18 (F0) Lethality 1	L-Calculation	2:39:24 PM	3:01:31 PM	30.7

Fig. 10. Lethality Calculation and Input Parameter Used

Theoritical Temperature on Drain Probe

Name:	Theoritical Temperature on Drain Probe
Description:	

Input parameters

Data Series Type:	Temperature
Data series:	LC 10, LC 11, LP 19 - T-Theoretical 1
Start Time:	2:39:24 PM
Stop Time:	3:01:31 PM
Output:	Data Series analysis, Global Data analysis
Functions:	Min., Max.

Global Data analysis

Min.	122.2 °C	2:53:58 PM	LC 11
Max.	122.6 °C	2:39:41 PM	LP 19 - T-Theoretica...

Data Series analysis

ID	Name	Min.	Min. sample time	Max.	Max. sample time
686049-1	LC 10	122.2 °C	2:47:05	122.5 °C	2:39:35
686050-1	LC 11	122.2 °C	2:53:58	122.5 °C	2:39:38
N/A	LP 19 - T-Theoretical 1	122.3 °C	2:51:42	122.6 °C	2:39:41

Fig. 11. Temperature at the Drain of the Autoclave

The lethality calculation demonstrates the process's integral effectiveness. The lethality range was 29–32 minutes, reflecting the accumulated sterilizing effect over the entire temperature–time profile rather than simply the attainment of a given temperature set point. In this representation, lethality serves as an aggregate measure that integrates cycle dynamics into a single indicator, allowing comparison across sensors and runs.

The values obtained exceed the standard overkill criterion by≥12 minutes, thereby providing an extremely high sterility assurance level of $<10^{-6}$. Notably, sensor LC 14, which demonstrated the lowest lethality (highlighted in red), still achieved the target values; this provides direct confirmation of cycle robustness, as even the weakest point in terms of integral effect remains within the required level. The temperature at the autoclave drain is shown in Figure 11.

The data show that the temperature in the drain (traditionally the coldest point) matched the temperature in the chamber. This eliminates the risk of a cold plug in the condensate removal system, which is often a root cause of failures (Failure Mode: Drain blockage). Figure 12 demonstrates that no temperature exceeds 124°C during the entire sterilization cycle.

Name:	No temperature exceed 124C during the entire cycle		
Description:			
Input parameters			
Data Series Type:	Temperature		
Data series:	LC 01, LC 02, LC 03, LC 04, LC 05, LC 06, LC 07, LC 08, LC 09, LC 10, LC 11, LC 12, LC 13, LC 14, LC 15, LC 16, LC 17, LC 18		
Start Time:	2:24:26 PM		
Stop Time:	3:34:18 PM		
Output:	Data Series analysis, Global Data analysis		
Functions:	Max.		
Global Data analysis			
Max.	123.9 °C	2:39:35 PM	LC 05
		2:39:37 PM	LC 05
		2:39:38 PM	LC 05
		2:39:39 PM	LC 05
		2:39:40 PM	LC 05
		2:39:41 PM	LC 05
		2:39:42 PM	LC 05
		2:39:43 PM	LC 05

Fig. 12. Temperature report during the entire sterilization cycle

The absence of overheating is evident. The maximum temperature was 123.9°C. This is critically important for preserving the physical properties of packaging materials and the product itself. Overheating may lead to embrittlement of plastics or deformation of stoppers (Zaborniak et al., 2024).

IV. DISCUSSION

This section synthesizes the empirical data obtained with theoretical concepts and regulatory requirements, forming an integrated picture of the risk-based approach.

Thermodynamics of failure: Why templates do not work

Analysis of the failed cycle (Section 3.2) provides a unique insight into sterilization physics. The fact that sensor LC 14 indicated 119.8°C while the autoclave control sensor showed 121°C is a classic manifestation of thermodynamic imbalance caused by an air pocket.

Air is an excellent thermal insulator. Its heat transfer coefficient is hundreds of times lower than that of condensing steam. Moreover, according to Dalton's law, the total pressure in the chamber is the sum of the partial pressures of steam and air (Tessarolo & Masè, 2025).

If air remains in a local region (for example, inside a tube or a dense package), the partial pressure of steam there is lower than in the rest of the chamber. Consequently, the saturation temperature at that point will also be lower. This explains why sensor LC 14 recorded a temperature lower than the setpoint.

A second-order insight is that, in the failed study, using only temperature sensors without pressure recording would essentially obscure the nature of the problem. In such a scenario, a reduced temperature would be observed, but its interpretation would

remain ambiguous: it would be impossible to determine whether the deviation was due to air in the steam–gas mixture or to condensation on cold surfaces.

The addition of a pressure sensor in the pass study fundamentally changes analytical reliability. The presence of measured pressure enables mathematical confirmation of saturated steam conditions, allowing saturation to be demonstrated not only through visual similarity of thermograms but also via calculation-based verification. Thus, the process physics is examined as the causal basis of the observed temperature regime, not merely the final result of reaching a temperature setpoint.

The role of FMEA in cycle design

The traditional approach treats validation as a confirmatory exercise, i.e., a procedure that verifies the conformity of an already-defined process to requirements. The approach proposed here, by contrast, uses FMEA as a design tool, emphasizing establishing process controllability at the development stage rather than post hoc confirmation of outcomes.

Instead of operating with abstract risks, the methodology advocates formulating engineering questions that convert uncertainty into testable hypotheses. For example, the risk of variable load density and its direct consequence, unpredictable cold-spot migration, is considered. In response, mitigation is embedded into the cycle design: the use of additional vacuum pulses is envisaged, and the equilibration time is introduced as a critical parameter that determines the start of the lethality calculation.

This FMEA adaptation enables a transition from passive observation to active control. In particular, in the pass cycle, rapid equilibration of less than 15 seconds is interpreted not as a fortuitous outcome but as the result of deliberate modification of vacuum parameters, justified through FMEA. In this sense, the cycle ceases to be a black box and becomes a transparent process in which causal relationships among settings, medium physics, and final performance indicators are subject to control.

Regulatory transformation: From Pass to Defendable

Modern regulatory requirements (ISO 17665-1:2024, Annex 1) shift the focus toward rationale. A cycle is considered validated only when the team can explain it without resorting to the report text.

ISO 17665-1:2024 requires a clear definition of the worst-case load (ISO, 2024). The graphical data presented indicate that the selection of measurement points and sensor locations, including the drain zone and package centers, was not random. Rather, it was based on an understanding of worst-case physics and an intention to record parameters precisely where the probability of deviations is highest.

EU GMP Annex 1 introduces the concept of a Contamination Control Strategy (CCS), in which autoclaving is considered a barrier (PMS, 2022). The successful cycle (Pass Data) with $F_0 \approx 30$ minutes demonstrates a substantial margin of robustness compared to the usual 12–15 minutes and can therefore serve as a strong CCS argument. This shows that the process can maintain effectiveness even in the face of an unforeseen increase in bioburden.

Critique and limitations of the lethality concept

It is important to recognize that lethality is the integral of temperature over time. This indicator accumulates lethality contributions even at 115°C because it mathematically accounts for each part of the temperature profile's contribution to the total effect. In this form, lethality is a convenient summary metric. However, it describes the integral thermal exposure but does not explain the physical nature of the medium in which it occurs.

If a temperature of 115°C is conditioned by the presence of dry air, the true biological inactivation will be negligible, since the z-value for dry heat is substantially higher than 10°C. Consequently, regimes with identical nominal lethality can exhibit fundamentally different biological performance depending on whether saturated steam is ensured or whether the process occurs in a steam–gas mixture containing air.

Therefore, acceptance criteria must be constructed hierarchically. First, strict temperature limits should be defined, including the requirement that Min ≥ 121.0°C, and equilibration time should be controlled as an indicator of complete air removal and formation of homogeneous conditions. Only after these fundamental physical conditions are met does it make sense to use lethality as the final integral indicator.

A practical rule of explainability

Validation is knowledge, not a report. If an operator or engineer cannot explain why four vacuum pulses are used instead of three, the process is at risk. A deep understanding of the physics demonstrated in the study (P/T correlation, drain analysis) builds operational resilience. In the event of future deviations (for example, vacuum pump failure), personnel will be able to diagnose the issue rapidly by understanding its root cause rather than simply following an instruction to restart the cycle.

V. CONCLUSION

The study convincingly demonstrates that the development of sterilization cycles in modern biopharmaceutical manufacturing requires a shift from empirical templates to a scientifically grounded risk-based approach. Under conditions of increasingly complex products and variable loads, process robustness can no longer rely solely on proven regimes and formal adherence to setpoints, because understanding the causal relationships among heat and mass transfer physics, cycle architecture, and actual inactivation performance becomes decisive.

The data obtained indicate pronounced physical non-uniformity of the process. Mixed and porous loads generate complex thermodynamic conditions in which standard parameters do not ensure reliable air removal and therefore do not guarantee homogeneous attainment of sterilizing conditions at all points. The results of the failed study confirm the existence of local risk zones even when the autoclave controller displays normal values, i.e., when the system-level signal masks critical local deviations.

Against this background, the effectiveness of engineering metrics that are more sensitive to the actual process physics than a simple extension of holding time is demonstrated. In particular, equilibration time and pressure–temperature correlation emerge as more reliable indicators of success, as they provide information on the completeness of deaeration and the attainment of saturated steam conditions. In the successful cycle, an equilibration time of less than 15 seconds proved a practical marker of effective air removal and rapid temperature-field equalization.

The role of FMEA as a design tool is emphasized separately. Integrating risk analysis at the cycle design stage enables proactive elimination of failure causes, such as air pocket formation, through targeted modifications to cycle parameters (e.g., vacuum pulses), rather than through trial-and-error iterations. In this way, risk is moved from the realm of post hoc detection into the realm of controlled engineering decision-making.

Finally, the proposed approach demonstrates alignment with new standards and regulatory logic. It is harmonized with ISO 17665-1:2024 and EU GMP Annex 1 requirements, ensuring not only product sterility but also the defendability of the process before auditors, grounded in transparent criteria, reproducible metrics, and traceable justifications for decisions.

The practical significance of the work lies in providing a ready-to-use framework for validation engineers that enables greater reliability in sterilization processes, reduced deviations, and the highest level of patient safety. This framework is oriented toward reproducibility, diagnostic capability, and controllability, making it applicable under conditions of manufacturing variability and rising regulatory expectations.

LIST OF ABBREVIATIONS

BSI - British Standards Institution

CCS - Contamination Control Strategy

EMA - European Medicines Agency

EN - European Norm

EU GMP - European Union Good Manufacturing Practice

FDA - Food and Drug Administration

FMEA - Failure Mode and Effects Analysis

ISO - International Organization for Standardization

LC - Temperature logger/channel ID (e.g., LC01–LC18)

LP - Pressure logger (e.g., LP19)

NCGs - Non-Condensable Gases

P/T - Pressure/Temperature correlation

PQ - Performance Qualification

QbD - Quality by Design

QRM - Quality Risk Management

SAL - Sterility Assurance Level

VI. AVAILABILITY OF DATA AND MATERIALS

The datasets generated and/or analysed during the current study are not publicly available due to confidentiality and GMP documentation restrictions, but are available from the corresponding author on reasonable request.

COMPETING INTERESTS

The author declares that they have no competing interests.

FUNDING

The author declares that this research received no external funding.

AUTHORS' CONTRIBUTIONS

The author declares that they solely conceived and designed the study, performed the data collection, conducted the analysis and interpretation, and wrote the final manuscript.

REFERENCES

[1] Agalloco J (2024) Expanding the Use of Moist Heat for Terminal Sterilization. PDA Journal of Pharmaceutical Science and Technology 79:114–121. https://doi.org/10.5731/pdajpst.2023.012918

[2] Boehler L, Daniol M, Sroka R, et al (2021) Sensors in the Autoclave-Modelling and Implementation of the IoT Steam Sterilization Procedure Counter. Sensors 21:510. https://doi.org/10.3390/s21020510

[3] BSI (2021) BS EN 285:2015+A1:2021. BSI. https://doi.org/10.3403/30278677

[4] Chakraborty S, Soman V, Dhumal V (2020) Alternate Approach for Defining Worst Case in Steam Sterilization Validation of Porous/Hard Goods Loads. PDA Journal of Pharmaceutical Science and Technology 75:91–99. https://doi.org/10.5731/pdajpst.2019.011304

[5] Chen Y, Bao L, Yi L, Hu R (2024) Analysis of Wet Pack Incidence in Steam Sterilization: A Study in a Chinese Medical Center. PubMed 30:e942601-1e942601-8. https://doi.org/10.12659/msm.942601

[6] Cho W-I, Chung M-S (2020) Bacillus spores: a review of their properties and inactivation processing technologies. Food Science and Biotechnology 29:1447–1461. https://doi.org/10.1007/s10068-020-00809-4

[7] Feurhuber M, Taupitz T, Mueller F, et al (2023) A Method to Investigate Sterilization Processes and the Bacterial Inactivation Resolved in Time and Space. PDA Journal of Pharmaceutical Science and Technology 78:331–347. https://doi.org/10.5731/pdajpst.2022.012771

[8] ISO (2024) ISO 17665:2024. In: ISO. https://www.iso.org/standard/80271.html. Accessed 3 Dec 2025

[9] Marschall L, Taylor C, Zahel T, et al (2022) Specification-driven acceptance criteria for validation of biopharmaceutical processes. Frontiers in Bioengineering and Biotechnology 10: https://doi.org/10.3389/fbioe.2022.1010583

[10] Miguel EA, Laranjeira PR, Ishii M, Pinto T de JA (2022) Analysis of water quality over non-condensable gases concentration on steam used for sterilization. PLOS ONE 17:e0274924. https://doi.org/10.1371/journal.pone.0274924

[11] PMS (2022) Review of Annex 1 2022: Environmental Monitoring Changes. In: PMS. https://br.pmeasuring.com/wp-content/uploads/2022/09/Review-of-Annex-1-2022.pdf. Accessed 2 Dec 2025

[12] Tessarolo F, Masè M (2025) Field assessment of a novel sensor for measuring noncondensable gases in steam sterilizers. Scientific Reports 15:9836. https://doi.org/10.1038/s41598-025-94798-1

[13] Zaborniak M, Kluczyński J, Stańko J, Ślęzak T (2024) The Influence of the Steam Sterilization Process on Selected Properties of Polymer Samples Produced in MEX and JMT Processes. Materials 17:5763. https://doi.org/10.3390/ma17235763

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.12

Technological Coupling and Ethical Reconstruction: A Systematic Inquiry into the Legal Paradigm Shift in the AI Era

Bingying Zhou

Associate Professor, School of Economics and Management, Guangzhou Institute of Science and Technology, Guangdong, China
chow6117@alu.ruc.edu.cn

Received: 16 Feb 2026; Received in revised form: 18 Mar 2026; Accepted: 22 Mar 2026; Available online: 28 Mar 2026

Abstract *— The leapfrog evolution of Artificial Intelligence (AI) is profoundly reshaping the operational logic of legal practice and the normative dimensions of legal ethics. Drawing upon cutting-edge scholarship from 2019 to 2025, this paper systematically examines the paradigm shift at the intersection of law and AI through three critical lenses: technological empowerment, ethical conflicts, and regulatory pathways. Technologically, Generative AI (GenAI) has transitioned from a mere auxiliary tool to a pivotal "actor" within legal practice. While the integration of Retrieval-Augmented Generation (RAG) and associated technologies has mitigated model hallucinations and enhanced algorithmic explainability, the inherent risks posed by technological boundaries remain substantial. Ethically, GenAI exerts systemic pressure on attorneys' duties of competence and confidentiality. By analyzing judicial precedents such as the UK High Court's Ayinde case, this study underscores the urgency of reconstructing professional responsibility and advocates for a recalibrated ethical framework through the lens of the "inevitability of bias." Regulatorily, the "legislative-driven" model of the EU, the "market-led" approach of the US, and the "consensus-building" strategy of international organizations exhibit distinct governance characteristics, whereas the concept of "co-regulation" provides a theoretical foundation for coupling technical standards with legal frameworks. Research indicates that the legal order in the AI era is at a critical juncture, transitioning from "heterogeneous isomorphism" toward "organic integration." Future scholarship should prioritize empirical validation, deepen cross-jurisdictional comparisons, and facilitate the collaborative design of technical solutions and ethical principles, thereby fostering a holistic paradigm update in legal education and institutional design.*

Keywords *— **Generative AI, Legal Ethics, Co-regulation, Legal Paradigm, Retrieval-Augmented Generation (RAG), Professional Responsibility.***

I. INTRODUCTION: PROBLEM STATEMENT AND RESEARCH FRAMEWORK

Over the past five years, interdisciplinary research on AI and law has undergone a paradigm shift from "marginal instrumental exploration" to "core normative agenda." This transformation is driven not only by the application potential unleashed by technological breakthroughs such as Large Language

Models (LLMs) but also by the systemic impact of AI practices on existing legal theories and ethical frameworks. From the Mata v. Avianca case in 2023, where a lawyer mistakenly submitted fictitious case law generated by ChatGPT, to the Ayinde case in 2025, which revealed the serious consequences of GenAI fabricating legal authorities, these incidents of "technological anomie" in judicial practice have evolved from isolated professional mistakes into landmark events in legal ethics discussions. These events prompt a fundamental inquiry in legal scholarship: "When AI systems become deeply embedded in the daily operations of legal practice, do traditional professional conduct norms and regulatory frameworks face a 'regulatory deficit'? In the technologically mediated legal field, how should the boundaries of human actors' responsibility be defined, and how should the legal order respond to the quasi-autonomous characteristics exhibited by technological systems?" In response, legal theory has engaged in profound normative reflection. Mei Xiaying [1] points out that AI legislation should be built upon the new types of legal objects and methods it creates, proposing "deep learning algorithms" and "human-machine relationships" as two theoretical pillars of AI law. This insight suggests that the relationship between AI and law is not a simple mechanical overlay of "technological tools + legal regulation" but a comprehensive paradigm shift involving ontology, epistemology, and methodology. Ontologically, AI systems as novel "actors" in legal processes challenge the human-centered subjectivity structure of classical jurisprudence. Epistemologically, legal practitioners urgently need to construct a cognitive framework capable of penetrating the logic of technological operations. Methodologically, traditional doctrinal legal research must achieve deep integration of knowledge resources with disciplines such as computer science and cognitive science.

Based on cutting-edge research from Chinese and international scholars between 2019 and 2025, this paper reveals key issues and theoretical frontiers in the intersection of AI and law through in-depth exploration and systematic review. By comparatively analyzing and theoretically integrating the technical aspects of GenAI, the ethical challenges it poses, and the regulatory pathways, we can grasp more profoundly its evolutionary logic and capability boundaries within the legal field, recognize more clearly its systemic impact on lawyers' ethical duties and the ethical warnings from judicial practice, and conduct more robust comparative analysis and theoretical integration of regulatory paths. Through a systematic exploration of the transformation of the legal order in the AI era, this paper not only provides an analytical framework with both theoretical depth and practical relevance for understanding the formation of the new legal ecology in this complex era but also identifies potential directions and ideas for future research.

II. TECHNOLOGICAL SOLUTIONS: FROM AUXILIARY TOOL TO LEGAL "ACTOR"

2.1 The Technological Evolution of Legal AI

Early legal AI primarily relied on symbolism and expert systems, consequently encountering the "engineering bottleneck" of knowledge acquisition and the "fragmentation" of reasoning. The technological logic of this phase was to formalize legal rules and reasoning processes into computable knowledge representations. The core difficulty lay in the fact that the openness, context-dependence, and value-laden nature of legal knowledge could not be fully encoded. For instance, rule-based expert systems required legal experts and knowledge engineers to work together to translate legal norms into IF-THEN logical rules. However, the interpretation of legal concepts often involves considerations of purpose, value balancing, and policy judgments—elements difficult to fully express within a closed rule system.

In recent years, LLMs, represented by the Transformer architecture, have overcome the limitations of traditional methods through capabilities like contextual reasoning, few-shot learning, and generative argumentation. LLMs employ statistical learning methods to "capture"

statistical regularities and semantic associations from vast amounts of text data, enabling them to generate natural language text tailored to specific contextual requirements. This approach of automatically extracting legal writing styles, argumentation patterns, and reasoning paths from data, driven by the "hard" force of data, effectively avoids the difficulty of manually encoding complex legal knowledge into formal rules and significantly reduces the rigidity and limitations inherent in "hard" legal rules. The advantage of this technological path lies in its ability to automatically learn legal writing styles, argumentation patterns, and reasoning paths through data-driven methods, eliminating the need for manual encoding of legal knowledge into formal rules.

The first comprehensive survey of legal LLMs, jointly released by multiple Chinese and international institutions [2], proposed an innovative "dual-perspective taxonomy." This survey integrates classical legal reasoning theories like Toulmin's argumentation framework, mapping structural elements of legal argument—claim, data, warrant, qualifier, rebuttal—onto the functional design of AI systems. Simultaneously, it maps the authentic workflows of professional roles such as lawyers, judges, and litigants, enabling technological development to respond to the embodied needs of legal practice. By systematically reviewing nearly 60 tools and datasets covering key technological advancements in legal text processing, knowledge integration, and reasoning formalization, the study provides a theoretical blueprint for the transition of legal AI from "laboratory tool" to "judicial infrastructure."

2.2 Application Landscape of GenAI in Legal Practice

Research by Terzidou [3] indicates that GenAI systems have become deeply embedded in the daily operations of legal practice: from electronic communication and document drafting to contract template generation, multilingual translation, and long-text summarization; LLMs are reshaping lawyers' workflows. In legal research, GenAI has been integrated into mainstream platforms like LexisNexis and Westlaw, offering legal practitioners rapid access to sources and streamlined drafting. These platforms utilize natural language processing, allowing lawyers to pose legal questions conversationally, with the system returning relevant cases, statutes, and secondary sources, and generating preliminary legal analysis suggestions. The application of GenAI is particularly significant in contract analysis and drafting. Lawyers can input contract texts into the system, requesting identification of potential risk clauses, suggestions for revisions, or generation of specific types of contract templates. Based on learning from vast amounts of contract text, AI systems can identify industry-standard clauses, common negotiation points, and ambiguous phrasing that might lead to disputes. This application not only improves the efficiency of contract review but also helps standardize contract quality and reduce human error. In litigation support, GenAI is used for evidence summarization, case fact organization, and drafting legal memoranda. Lawyers can input large volumes of evidence, asking the system to generate timelines, identify key facts, or summarize party arguments. For complex litigation requiring massive document processing, AI support can significantly reduce lawyers' workload, allowing them to focus on strategic judgment and argument innovation.

Abbasi [4] conceptualizes this trend as "Responsible Legal Augmentation," emphasizing that AI systems should be oriented towards augmenting rather than replacing human judgment. This reflects a philosophical reflection on legal augmentation and awareness of potential ethical dilemmas. He advocates orienting AI's value towards augmenting, not replacing, human judgment. The distinction between AI as "automation" versus "assistance" directly relates to the intensity of responsibility attribution and ethical constraints: if AI is seen as a "automation" tool replacing humans, responsibility might be "outsourced" to AI, reducing ethical constraints; if AI is seen as an "assistance" system augmenting human capabilities, human ultimate control and oversight obligations must be retained.

Abbasi argues that responsible legal augmentation should satisfy three core requirements: transparency (lawyers understand AI's capabilities and limitations), auditability (AI outputs can be effectively reviewed), and accountability (humans always bear ultimate responsibility for final decisions).

2.3 Technological Capability Boundaries and Core Dilemmas

Despite the promising applications of GenAI in law, core dilemmas stemming from its technological capability boundaries persist. The "hallucination" problem in LLMs' legal assertions constitutes a primary challenge. LLMs may generate entirely fictitious cases, statutes, or arguments with a high degree of confidence, and the reliance of legal reasoning on authoritative sources makes this issue particularly dangerous. The cause of model hallucinations lies in the LLM's training objective: to generate coherent text fitting the context, not to ensure factual accuracy. When the model encounters legal questions not covered in its training data, it tends to "fill in the gaps" through plausible guesswork rather than admitting ignorance. As explicitly stated in the UK High Court's Ayinde judgment, freely available LLM-based GenAI tools (like ChatGPT) cannot be relied upon for reliable legal research; the critical safeguard is "verifying any output through authoritative sources." This warning is universally applicable: no matter how advanced AI systems become, lawyers bear ultimate responsibility for the accuracy of their submissions. The court's stance can be summarized as "responsibility reinforcement under technological neutrality": not presupposing the virtue or vice of specific technologies, but consistently reinforcing the ultimate control and oversight obligations of human actors.

Second, the adaptability gap in low-resource jurisdictions is also noteworthy. LLMs' training data predominantly comes from legal texts from the English-speaking world. Model performance often significantly declines for non-English jurisdictions and legal traditions outside common law systems. This issue concerns not only technological fairness but also the normative value of legal pluralism. Profound differences exist in legal concepts, reasoning methods, and argumentation styles across jurisdictions. Applying a model trained on data from one jurisdiction directly to another may lead to systematic misunderstandings and erroneous reasoning.

Furthermore, the lack of explainability in black-box reasoning constitutes another core dilemma. The internal workings of LLMs involve hundreds of millions or even hundreds of billions of parameters, making the reasoning path leading to a specific output difficult to fully trace. For legal decisions, explainability is not merely a technical requirement but the foundation of legitimacy—litigants have the right to know the basis of decisions, and judges have the duty to state their reasoning. When AI systems participate in legal reasoning, ensuring that their contribution can be understood, reviewed, and challenged becomes a pressing methodological issue.

To address these challenges, Retrieval-Augmented Generation (RAG) technology significantly enhances the accuracy and timeliness of generated content by invoking external real-time knowledge bases rather than relying solely on static training data. The RAG architecture combines the generative capabilities of LLMs with the retrieval capabilities of external knowledge bases: when a user asks a question, the system first retrieves relevant documents from the knowledge base and then inputs these documents as context to the LLM, generating an answer based on the retrieved results. This technological path transforms the production and updating of legal knowledge from traditional "closed model training" to "open knowledge retrieval," offering new possibilities for the explainability and accountability of legal AI. To truly establish effective linkage between AI answers and human control, AI responses must have traceable authority, enabling lawyers to independently verify the "basis" relied upon by the AI, thereby establishing an effective

connection between technical assistance and human control.

III. LEGAL ETHICS: RECONSTRUCTION OF PROFESSIONAL RESPONSIBILITY

3.1 The Systemic Impact of GenAI on Lawyers' Ethical Duties

The core duties of competence and confidentiality for lawyers face systemic impacts due to the involvement of GenAI. A research report by Terzidou [3] deeply analyzes the specific manifestations and coping strategies for this impact. The research indicates that lawyers' attitudes towards GenAI show significant generational and firm-size differences: younger lawyers and those in large firms are more inclined to actively experiment with AI tools, while senior lawyers and those in smaller firms are more cautious. This disparity highlights the urgency and the need to respect diverse perspectives in reconstructing legal professional ethics.

Under the duty of competence, lawyers can effectively use GenAI systems by acquiring necessary digital literacy to enhance the efficiency, consistency, and quality of case handling. This has two aspects: positively, lawyers should understand available AI tools and their capabilities, utilizing them in appropriate scenarios to fulfill their commitment to providing competent representation to clients; negatively, lawyers must recognize the limitations of these systems and avoid relying on their outputs where their validity, accuracy, or relevance is questionable. Relevant opinions from the American Bar Association's ethics committee explicitly state that lawyers' duty to review AI-generated content is no different from their review duty in traditional legal research—ultimate responsibility always rests with the signing lawyer.

The reconstruction of competence also involves redefining the standard of "competence." In the AI era, does ignorance of available technological tools constitute professional negligence? Does failure to effectively use AI assistance mean failing to provide the best possible service to clients? These questions lack uniform answers, but it is clear that the meaning of competence is expanding from "mastery of legal knowledge" to "mastery of all practice resources, including technological tools." Lawyers need critical skills to evaluate AI outputs and balance technical assistance with independent judgment.

The duty of confidentiality faces a more direct challenge. Lawyers inputting client information directly into AI prompts or allowing AI systems to access files containing confidential data may lead to client information disclosure to third parties (including system providers), thereby violating the duty of confidentiality. Research shows that the scope of confidentiality protection should cover electronic communications between lawyer and client, including metadata in electronic documents. This means lawyers must carefully evaluate the data storage and processing policies of commercial AI tools to prevent client information from being used without authorization for model training or other commercial purposes.

Terzidou's research [3] also reveals that many lawyers lack understanding of AI systems' data processing practices, mistakenly assuming interactions with AI are as confidential as discussions with colleagues. This cognitive bias poses a significant risk: commercial AI systems often use user-input data for model optimization, and client information may become part of the training data without the lawyer's knowledge, potentially even being indirectly exposed in other users' queries. Therefore, when selecting AI tools, lawyers must review the service provider's data protection policies to ensure client confidentiality is adequately protected.

3.2 Ethical Warnings from Judicial Practice: From the Mata Case to the Ayinde Case

The 2023 Mata v. Avianca case in the US, where lawyers submitted fictitious case law generated by ChatGPT, has become a landmark event in legal ethics discussions. In that case, lawyers relied on ChatGPT for legal research but failed to verify the generated content, resulting in a brief containing multiple non-existent judicial precedents. The court's sanctions emphasized lawyers' duty of candor to the

court—regardless of technological development, lawyers are ultimately responsible for their submissions. Notably, the lawyers involved did not intend to deceive the court but lacked awareness of AI systems' capability boundaries, mistakenly believing ChatGPT could conduct reliable legal research. The core lesson is that technological reliance cannot substitute for professional judgment; lawyers must maintain skeptical scrutiny of AI outputs.

The UK High Court's Ayinde case further deepened the discussion. The judgment established an important ethical principle: lawyers cannot defend false citations by claiming the "legal principle itself is correct." Citations are not merely decorative or optional; they are the bedrock of legal reasoning and judicial trust. Even if the legal principle supported by a cited case is correct, the presence of a fictitious case still undermines the integrity of the judicial process because it prevents opposing parties and the court from verifying the basis of arguments, destroying the common foundation of legal dialogue. The UK High Court's stance in this case reflects a prudent and pragmatic approach: neither banning the use of GenAI as AI skeptics might desire nor unconditionally endorsing its reliability, but establishing a middle path by emphasizing lawyers' duty to the court—ensuring all submitted materials are accurate and not misleading. This stance can be summarized as "responsibility reinforcement under technological neutrality": courts do not presuppose the virtue or vice of specific technologies but consistently reinforce the ultimate control and oversight obligations of human actors. The guidance particularly emphasizes that lawyers must fulfill a verification duty when using AI tools; any statement relying on AI-generated content must be validated through authoritative sources.

3.3 Bias and Value: A Counterintuitive Insight

Lande [5] offers a counterintuitive theoretical perspective: "bias" in AI should not be simplistically viewed as a technical flaw to be eradicated but rather as an inherent and potentially positive feature of system operation. This view stems from the premise that technology is never neutral. The notion of "value-neutral" technology is an illusion; any system is a product of specific design choices and value judgments. From dataset selection and algorithm feature setting to the presentation of final results, every technical detail is permeated with developers' subjective presuppositions and beliefs. Based on this, Lande argues that bias reflects developers' values, and instead of trying to hide or erase these positions, they should be disclosed. This shifts the discussion from "how to eliminate bias" to "how to responsibly disclose bias." The research suggests that enabling users to understand the assumptions, priorities, and design frameworks behind a tool should be a core ethical guideline. For example, with a legal research tool, if it preferentially cites cases from specific courts or arguments with specific logic without prior explanation, users might mistakenly believe they have received a comprehensive, neutral answer. Once these preferences are clearly disclosed, users can understand that the result is from a specific perspective.

Lande further categorizes bias into two types: harmful bias and constructive bias. Harmful bias leads to discrimination against specific groups or systematically distorts factual truth. Constructive bias reflects legitimate value choices, such as designing systems to prioritize human rights, emphasize procedural justice, or pursue operational efficiency. The focus is not on pursuing the impossible goal of "zero bias" but on making these value orientations transparent, serving as the basis for users to make informed choices.

Zhou, S. [6] deepens this discussion from the perspective of virtue jurisprudence. The research points out that the term "responsible AI" carries a misleading anthropomorphic tendency that easily blurs true responsibility attribution. It argues for a strict distinction between "legal responsibility" and "moral capacity." Legally, responsibility can recognize AI as an "actor" with attributes of legal responsibility, making the legal effects it generates traceable. Morally, it insists on a "human-centered" baseline: AI cannot become a subject of moral responsibility. This shift has important practical

implications: we should not focus on whether "AI is responsible" but rather on "whether the humans using AI possess the virtues to operate technology responsibly." This redirects the ethical focus from the cold algorithm back to humans' own decision-making capabilities.

3.4 The Complex Landscape of Responsibility Attribution

With GenAI's involvement in legal practice, the issue of responsibility attribution has acquired unprecedented complexity. When an AI system generates erroneous output causing client loss, who should bear responsibility? The lawyer using the AI, the company developing the AI, the law firm deploying the AI, or the AI system itself? Answering this question involves multiple areas like contract law, tort law, and professional responsibility law, with no unified theoretical framework currently available.

From a contract law perspective, legal service contracts between lawyers and clients may implicitly or explicitly stipulate the use of AI. If a lawyer uses AI to process sensitive information without client consent, it may constitute a breach of contract; if a lawyer promises personal service but excessively relies on AI, it may violate the essence of the service contract. From a tort law perspective, a lawyer's reliance on AI outputs may constitute negligence, requiring elements like breach of duty of care, damages, and causation. From a product liability perspective, AI systems might be considered products, and developers might be liable for system defects, but the learning capabilities and autonomy of AI challenge the traditional product liability framework.

IV. REGULATORY PATHWAYS: COMPARATIVE LAW AND INSTITUTIONAL DESIGN

4.1 Three Regulatory Models in a Global Perspective

Globally, AI regulation presents a landscape of coexisting diverse models. A systematic study by Zhang Tao [7] compares three typical paths: the EU's "legislative-driven" model, the US's "market-led" model, and the international organizations' "consensus-building" model. These three models reflect fundamental differences in legal philosophies towards technology governance and embody distinct political traditions and institutional resources.

The EU model, centered on the Artificial Intelligence Act (AI Act), creatively introduces a "harmonized standards" mechanism, making compliance with technical standards an important basis for proving that AI systems meet mandatory legal requirements, thus achieving close coupling between technical standards and the legal framework. Yousefi [8] provides an in-depth interpretation of the EU AI Act from the perspective of algorithmic fairness in their doctoral thesis, emphasizing that the Act adopts a risk-based approach, classifying AI systems into four risk levels: unacceptable risk, high risk, limited risk, and minimal or no risk, with different regulatory requirements applicable to each. High-risk AI systems (e.g., used in recruitment, credit assessment, predictive policing) must meet strict compliance requirements, including risk management systems, data governance, technical documentation, transparency obligations, and human oversight. The EU model's advantage lies in providing a deterministic legal framework, offering clear behavioral expectations for businesses and users; its risk is that legislation may lag behind technological development, and rigid norms may stifle innovation. To mitigate this tension, the EU AI Act introduces "regulatory sandboxes," allowing innovators to test new technologies under regulatory supervision, seeking balance between compliance and innovation.

The US model emphasizes market leadership and industry self-regulation. The National Institute of Standards and Technology (NIST) issued the Artificial Intelligence Risk Management Framework (AI RMF), serving primarily as a guiding tool to help designers, developers, and users of AI systems manage risks associated with AI applications. The institutional logic of this model is that technological innovation outpaces legislative processes; overly rigid legal regulation may dampen industrial vitality. Through industry self-regulation and soft law

governance, risk control can be achieved while maintaining flexibility. The US model's advantage is high flexibility and adaptability, allowing quick responses to technological changes; its risk is insufficient regulatory force, potentially leading to market failures, consumer rights infringement, and accumulation of systemic risks. In recent years, discussions on AI regulation in the US have become increasingly active, with some states already enacting specific AI legislation and federal-level legislative initiatives advancing.

International standardization organizations (such as ISO/IEC JTC 1/SC 42) focus on building global technical consensus, having published dozens of international standards in areas like concepts and terminology, risk management, governance impact, and trustworthiness. The advantage of this model lies in transcending sovereign state boundaries, providing a common technical language and behavioral expectations for global industrial chains; its limitations include the democratic deficit in standard-setting processes and the weakness of enforcement mechanisms. International standard development relies primarily on consensus among technical experts, with insufficient participation from affected groups (e.g., consumers, vulnerable populations); adoption of standards depends on voluntary choices by states, lacking mandatory enforcement mechanisms.

4.2 Theoretical Construction of Co-regulation

Facing the coexistence and competition of diverse regulatory models, Zhang Tao [7] proposes that compared to purely administrative regulation models or pure self-regulation models, a "co-regulation model" is more aligned with the inherent laws of AI standardization and can more effectively address the practical dilemmas of standardization.

The theoretical foundations of the co-regulation model include embedded ethics, governance ecology, knowledge co-creation, and layered governance. Embedded ethics integrates social ethical values into the technology design process, transforming ethical considerations from external constraints into internal constituent elements. This approach requires ethicists and technologists to collaborate in the early stages of technology development, translating value demands like fairness, transparency, and explainability into specific technical requirements. Haden [9] deepens this approach from a human rights-centered perspective, advocating drawing on the institutional experience of the General Data Protection Regulation (GDPR) to construct an AI regulatory framework centered on human rights protection. She advocates extending Data Protection Impact Assessments to AI Human Rights Impact Assessments, anticipating potential human rights risks during the system design phase and taking mitigating measures.

Governance ecology emphasizes the network effects of diverse stakeholders. AI regulation involves multiple actors—developers, deployers, users, regulators, affected individuals and groups—and the knowledge and capacity of any single actor are limited. The co-regulation model integrates distributed knowledge and coordinates diverse interests by constructing a governance ecology. This approach requires establishing regular stakeholder dialogue mechanisms so that concerns of different actors can be expressed and considered in the regulatory process.

Knowledge co-creation posits that regulatory knowledge is not monopolized by legislators or regulators but is dynamically generated through the interaction of multiple actors. This perspective breaks the cognitive hierarchy of traditional "command-and-control" regulation, providing a theoretical basis for the coexistence of diverse normative forms like soft law, standards, and guidelines. In the context of rapid technological iteration, knowledge co-creation means the regulatory framework must remain open and adaptive, capable of absorbing cutting-edge experience from practice.

Layered governance distinguishes regulatory needs at different levels. The technical infrastructure layer (e.g., algorithm architecture), the application layer (e.g., specific scenarios), and the social impact layer (e.g., human rights protection) require different types of regulatory tools and institutional

arrangements; layered governance helps achieve precise matching of regulatory tools. For example, regulation at the technical infrastructure layer can focus on safety and reliability standards; application-level regulation can focus on risk assessment and mitigation for specific scenarios; social impact-level regulation needs to consider broader value balance and rights protection.

4.3 Chinese Practice and Institutional Exploration

In recent years, China has successively issued policy documents such as the New Generation Artificial Intelligence Ethics Norms and the Opinions on Strengthening Science and Technology Ethics Governance, initially constructing a normative framework for AI governance. In the field of standardization, China actively participates in international standard-setting while promoting the construction of its domestic standard system, issuing important documents like the White Paper on Artificial Intelligence Standardization. Zhang Tao [7] notes that China's regulatory path exhibits characteristics of "policy guidance + standard support." At the policy level, the state defines development goals and value orientations through strategic planning; at the standard level, technical standards provide operational tools for policy implementation. The advantage of this path lies in the combination of policy flexibility and standard professionalism, capable of absorbing the knowledge contributions of technical experts while maintaining strategic direction. However, China's AI regulation also faces unique challenges. How to ensure citizen rights while promoting technological innovation, how to enhance voice in international standard-setting, and how to coordinate regulatory functions across different government departments—these questions require ongoing institutional exploration and theoretical reflection.

4.4 Regulatory Exploration in the Judicial Sphere

Besides legislation and administrative regulation, proactive regulation of AI within the judicial sphere is also noteworthy. The guidance issued by the UK High Court in the Ayinde case reflects the judiciary's cautious approach towards AI use: neither prohibiting its use as AI skeptics might expect nor unconditionally endorsing it, but establishing a middle path by emphasizing lawyers' duty to the court to ensure all submitted materials are accurate and not misleading.

This judicial regulatory path is characterized by its focus on concrete behavioral norms in specific scenarios rather than abstract presuppositions about technology's virtue or vice; it does not create new substantive rules but activates existing ethical duties in new technological contexts; it does not rely on rigid prohibitions or permissions but guides behavior through the clarification of responsibility attribution. This model of "regulation through activating duties" offers important methodological insights for legal governance during periods of rapid technological iteration. Judicial regulation is also evident in courts' determinations of the admissibility of AI-generated evidence. When evidence generated by AI systems (e.g., deepfake detection reports, algorithmic risk assessments) is submitted to court, judges must assess its reliability, relevance, and fairness. This process is simultaneously an application of evidence rules and a judicial review of technical norms, providing a practical testing ground for technical standards for AI systems.

V. RESEARCH OUTLOOK: TOWARD A NEW JURISPRUDENCE FOR THE AI ERA

5.1 Limitations of Current Research and Future Agenda

While significant progress has been made in current interdisciplinary research on AI and law, several directions warrant further development: First, the expansion of empirical research. Existing studies are mostly normative analyses, lacking empirical evaluation of the effects of AI systems in actual judicial operations. For example, what is the actual accuracy rate of AI-assisted decision-making? Do different groups show varying levels of trust in AI decisions? Does AI use change judges' sentencing patterns? Does AI assistance truly improve the efficiency and quality of legal services? These questions require support from large-scale empirical

research, necessitating interdisciplinary collaboration combining law, computer science, sociology, psychology, and other fields. Second, the deepening of cross-jurisdictional comparisons. There are significant differences in the acceptance and regulatory paths for AI across different legal systems and jurisdictions with varying levels of development. Comparative research can provide references for global governance. Particularly noteworthy is how differences between common law and civil law systems—regarding precedent status, reasoning methods, procedural structures—affect the integration paths and regulatory needs for AI technology. Additionally, gaps between developing and developed countries in technological capacity, data resources, and regulatory capability require attention in comparative research. Third, the collaborative design of technical solutions and ethical principles. How to embed ethical principles like fairness, transparency, and explainability into technical architecture, achieving "ethics by design," is a core issue for technology-law interdisciplinary research. This requires not only enhanced ethical awareness among technologists but also basic technological literacy among legal scholars to intervene during the technology design phase rather than offering post-hoc criticism. Collaborative design necessitates establishing interdisciplinary cooperation mechanisms to translate legal insights into constraints and design goals for technical implementation. Fourth, the adaptive transformation of legal education. The innovative practice of the AI course on Legal Professional Ethics at China University of Political Science and Law demonstrates that what competency structure do legal professionals need in the AI era, and how to integrate technological literacy and ethical awareness into legal education—these are not only educational issues but also fundamental questions of professional ethics. The course design team's proposed "knowledge-problem-competence" three-part framework and their attempt to construct highly simulated case scenarios using AI agents provide valuable experience for the digital transformation of legal education. Future legal education must incorporate emerging competencies like technological literacy, data analysis, and human-AI collaboration alongside traditional legal knowledge, cultivating a new generation of legal professionals capable of practicing effectively in an AI environment.

5.2 From Heterogeneous Isomorphism to Organic Integration: The Transformation of the Legal Paradigm

Mei Xiaying [1] points out that AI possesses distinct characteristics such as systematicity, publicity, openness, and evolutionary nature, and its integration with the traditional legal system forms a relationship of "heterogeneous isomorphism." "Heterogeneous isomorphism" implies that AI and the traditional legal system are nested within each other in structural form but have fundamental differences in operational logic and value core. This judgment has important methodological significance for understanding the transformation of jurisprudence in the AI era. Moving from "heterogeneous isomorphism" to "organic integration" requires paradigm updates at several levels.

At the ontological level, the boundaries of legal subjects need re-examination. When AI systems can autonomously generate texts with legal effects, participate in legal reasoning processes, and influence judicial outcomes, the traditional human-centered theory of legal subjects faces challenges. However, this does not mean granting AI legal personality but rather acknowledging its status as an "actor" while maintaining the human-centered baseline for responsibility attribution. This position requires conceptual distinction between the "source of legal effects" and the "attribution of legal responsibility"—the former may involve AI systems, but the latter always belongs to humans.

At the epistemological level, cognitive frameworks for understanding technological systems need development. Legal professionals require basic technological literacy to understand the capability boundaries and operational logic of LLMs, avoiding mystifying or demonizing technology. Legal

scholarship needs to shift from "legal responses to technology" to "legal cognition co-evolving with technology." This means legal methodology must incorporate emerging cognitive models like computational thinking and systems thinking, enabling legal judgment to be grounded in an understanding of technology's actual operation.

At the methodological level, the integration of diverse research paradigms is necessary. Research at the intersection of AI and law is inherently interdisciplinary, requiring the integration of knowledge resources from law, computer science, ethics, cognitive science, and other fields. The methodological framework of a single discipline cannot capture the complex picture of technology-law-society interactions. Interdisciplinary research is not simple accumulation but requires establishing shared problem awareness, conceptual frameworks, and research methods, enabling knowledge from different disciplines to generate new insights through dialogue.

5.3 Conclusion: Safeguarding the Rule of Law Amidst the Technological Wave

The transformative evolution of AI technology is reshaping every aspect of legal practice, from legal research and document drafting to case prediction and adjudication assistance, from law practice to court management. This process brings both opportunities for efficiency gains and knowledge democratization and profound challenges such as the reconstruction of professional ethics, blurred responsibility attribution, and human rights protection risks.

Facing this transformation, we should neither succumb to the blind optimism of technological utopia nor fall into the pessimistic resistance of technological dystopia. The UK High Court's prudent stance in the Ayinde case perhaps offers a valuable attitude: neither presupposing the virtue or vice of specific technologies nor avoiding the challenges they bring; neither abandoning vigilance towards technological boundaries nor abandoning exploration of technological potential; neither relaxing the emphasis on human actors' responsibility nor denying the factual status of AI systems as new "actors."

Ultimately, the rule of law order in the AI era remains a human-centered order. Technology can enhance human capabilities, expand human cognition, and optimize human decisions, but it cannot replace human judgment, human responsibility, or human dignity. Safeguarding the rule of law amidst the technological wave means embracing technological possibilities while steadfastly upholding the core values of the rule of law: human autonomy, attributability of responsibility, and availability of remedy for rights. This is both the generational mission of legal professionals and the enduring topic of legal scholarship.

Research at the intersection of AI and law is in an active phase of knowledge innovation. Over the next decade, as technology continues to evolve and regulatory experience accumulates, this field will undoubtedly yield richer theoretical achievements and practical wisdom. This review is merely a starting point; more scholars are expected to join this important academic dialogue, collectively exploring the shape of the legal order in the AI era.

REFERENCES

[1] Mei, Xiaying. (2025). Fundamental Issues in AI Legislation: Algorithmic Discipline and Human-Machine Relationships in Context. Jurist (6), 16-30. (in Chinese)

[2] Shao, P., Xu, L., Wang, J., Zhou, W., & Wu, X. (2025). When large language models meet law: Dual-lens taxonomy, technical advances, and ethical governance. arXiv preprint arXiv:2507.07748.

[3] Terzidou, K. (2025). Generative ai systems in legal practice offering quality legal services while upholding legal ethics. International Journal of Law in Context, 21(3).

[4] Abbasi, Z. (2025). Responsible legal augmentation: integrating generative ai into legal practice. Law, Technology & Humans, 7(3).

[5] Lande J. (2025). RPS Coach is "Biased" - And Proud of It [R]. University of Missouri School of Law Legal

Studies Research Paper No. 2025-17, 2025. https://dx.doi.org/10.2139/ssrn.5213572

[6] Zhou, S. (2025). Deconstructing 'Responsible AI': An Examination of Legal and Ethical Accountability Through Virtue Jurisprudence. International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique, 1-22.

[7] Zhang, Tao. (2025). Regulating Artificial Intelligence through Technical Standards: A Jurisprudence Based on Co-regulation. Comparative Law Studies (4), 169-187. (in Chinese)

[8] Yousefi, Y. (2025). The quest to fairness in algorithmic decisions: ethical, legal and technical solutions. [D]. ORBilu-University of Luxembourg. https://orbilu.uni.lu/handle/10993/66101, 2025.

[9] Haden, C. (2025). Assembling the Algorithm for Human Rights Centred AI Regulation (Doctoral dissertation, University of Hertfordshire).

[10] Custers, B., & Fosch-Villaronga, E. (2022). Law and artificial intelligence. Information technology and law series, 35.

[11] DiMatteo, L. A., Poncibò, C., & Cannarsa, M. (Eds.). (2022). The Cambridge handbook of artificial intelligence: global perspectives on law and ethics. Cambridge University Press.

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.13

Environmental Disclosure in Cameroon's Industrial Sector: Insights from Case Studies

Bleck Capouell Tegofack

Institut Universitaire de La Cote
capouelltegofack@yahoo.com

Received: 20 Feb 2026; Received in revised form: 21 Mar 2026; Accepted: 25 Mar 2026; Available online: 29 Mar 2026

Abstract *— Environmental disclosure is a corporate practice in Cameroon. But very little empirical understanding about the environmental aspects of information disseminated in sustainability reports exist. Hence, the principal goal of this study is to provide an insight on the overall features of environmental information disclosed by industrial enterprises in their sustainability reports. In order to achieve this objective, descriptive qualitative research was carried out based on the content analysis involving a multiple case study of 12 sustainability reports of 4 industrial enterprises from 2018 to 2021. The content analysis reveals that 7 environmental themes are consistently disclosed by industrial enterprises, with emission (greenhouse gas) and energy (energy consumption and reduction) being the most and least disclosed environmental themes respectively. Also, the disseminated environmental information is mostly positive and narrative in nature. The non-integration of monetary information in sustainability reports highlight the necessity for the competent accounting body in the OHADA zone to design a distinct accounting reporting framework for non-financial information.*

Keywords *— **Environmental disclosure, qualitative analysis, non-financial information, sustainability reporting, content analysis.***

I. INTRODUCTION

Industrialisation turns out to be a major driver for economic development and increased standard of living in all economies. This partly explains why a good number of countries in Africa are termed developing countries due to the fact that the industrial sector is still growing. Industrial growth has devastating consequences on the environment given that industries in the course of their activities produce pollutants which are harmful to the environment and consequently acts as a threat to the existence of mankind. Greenhouse gas emissions from industrial activities also account for the emergence of global warning with its related impact on economic activities such as agriculture.

This environmental threat gives rise to the need for accountability concerning corporate impact on the environment, and call for increased transparency concerning environmental reporting (Gustafsson, 2017). Environmental reporting has been seen as a way of increasing accountability of firms related to environmental issues (Porchelvi, 2019). Environmental disclosure is a practice through which business entities demonstrate their accountability towards stakeholders by providing information related to the impact of their activities on the environment and actions undertaken to mitigate or compensate the production of negative externalities.

Historically, Porchelvi (2019) posits that environmental reporting began in the 1980s and represents the second phase of sustainability reporting. This reporting practice was adopted by organisations because of the growth of environmental challenges such as pollution, land degradation and oil spills (Deegan, 2014). Sustainability reporting incorporates three

dimensions; economic, social and environmental. Given that on global scale sustainability reporting is a voluntary activity, the Global Reporting Initiative (GRI) has proposed reporting guidelines for each of the dimensions in order to guarantee the quality and credibility of environmental reporting.

In Cameroon, industrialisation started during the colonial period as factories were created for semi processing of raw materials. These raw materials were processed in order to preserve them during the long period of exportation to the industrialised countries in Europe. After the independence of Cameroon, the trend of industrialisation started to change from export oriented to import substitution industrialisation. By this policy of import substitution; the government of Cameroon aims at reducing importation of goods that can be manufactured by home industries. Globally, industrial activities are on a rise in Cameroon and to check on the impact of industrial activities on the environment, the government created an administrative department call the Ministry of Environment, Nature Protection and Sustainable Development in the year 1992 (Tegofack and Kamdem, 2022). The practice of environmental reporting is a recent phenomenon in Cameroon and began during the second decade of 21st century. Research works on environmental reporting in Cameroon so far have tried to find out the factors explaining the presence of sustainability reporting in general (see Dongmo and Ndjetcheu, 2018) and environmental reporting in particular (see Tegofack and Kamdem, 2022). For instance, the study of Tegofack and Kamdem (2022) reveals that profitability, industry type, pressure from primary and secondary stakeholders tend out to be strong determinants of the disclosure of environmental information by industrial enterprises in the Cameroonian context.

The environmental disclosure themes, types of environmental information disclosed and methods used by industrial enterprises to disclosed environmental information have received little or no attention from researchers. This research gab takes us to find an answer to the following question; what are the inherent characteristics of environmental information disclosed by industrial enterprises in Cameroon?

The main aim of this paper is to provide an insight into the practice of environmental reporting by providing an insight into environmental reporting, practiced by industrial enterprises in Cameroon (the environmental themes, types of environmental information contained in CSR reports and method used to disclose environmental information). This paper is structure into three parts; literature review, research method, and results and discussion of findings.

II. LITERATURE REVIEW

In this section, we intend to present in brief the concept of environmental disclosure, the theoretical perspective and existing environmental disclosure studies.

2.1 The concept of environmental disclosure

Morelli (2011) associates the word "environmental" with human impact on natural systems and defines the term "environmental sustainability" by building on the most popular definition of sustainable development (meeting the needs of the present generation without compromising the ability of future generations to meet their needs). On the basis of that he considers environmental sustainability as meeting the resource and services needs of current and future generations without compromising the health of the ecosystems that provide them. Hart (1995) on his part sees corporate environmental sustainability as the firm's activities associated with protection of natural resources and efforts to preserve the environment. These efforts include reducing environmental effects and minimizing resource consumption (Gibson, 2001) through involving in environmental practices, solving pollution problem and reducing resource depletion (Henion and Kinnear, 1976).

It is worth noting that environmental disclosure is one of the dimensions of sustainability because apart from this, other dimensions of sustainability are the social, economic and governance dimensions. Environmental disclosure (ED) has been defined as the disclosure of "the impact company activities have on the physical or natural environment in which it operates" (Berthelt et al., 2003, p.2). Berthelt et al. (2003) defined it as the incorporation into annual reports of a set of clauses and information items

describing a company's past, current and future environmental management activities and performance.

With a rise in stakeholder awareness about firms' impact on the environment, some firms began to disclose environmental issues in their reports leading to voluntary reporting of environmental information. Voluntary or optional disclosure is undertaken without pressure from external parties, with the purpose of providing investors and other groups with additional information on environmental performance (Porchelvi, 2019). Beside voluntary disclosure, there is mandatory disclosure, which is undertaken according to the accounting standards issued by the relevant organisations, and the legislation and laws of the state. It commits companies to disclose the stipulated information to the users of financial statements and other annual reports (Islam et al., 2005). Therefore, firms disclosed environmental activities as a legitimate tactic to affect the decisions of their stakeholders (Deegan and Gordon, 1996).

2.2 Theoretical literature

Several theoretical perspectives have been mobilised in the social accounting literature to explain the voluntary dissemination of environmental information. According to Deegan (2002), legitimacy theory, stakeholder theory and political economy theory are the main theoretical foundation used in environmental reporting studies. Gray et al. (1996) point out that each of these theories provides explanation to the practices of environmental disclosure in a specific manner. For instance, stakeholder theory concentrates in managing the expectations of different stakeholder group by establishing a balance between primary stakeholders and secondary stakeholders. Legitimacy theory is described by Richardson (1987, p. 352) as a "means by which social values are linked to economic actions", in order to achieve harmony between corporate practices and the legitimacy of their existence. Following Guthrie and Parker (1990) political economy theory is used to explain the relationship between the political and economic and social contexts. As Ibrahim (2014, p. 47) put it, "the political, economic and social systems cannot operate in isolation from each other, given that, the economic issues cannot be investigated without considering the political, social and institutional framework in an integrated manner".

This study relies heavily on the political economy as a theoretical perspective susceptible of explaining environmental reporting practices in Cameroon. This is because in the absence of mandatory sustainability reporting standard, the interest of the vulnerable that are exposed to the negative's effects generated by the activities of industrial enterprises can only be protected by the political setting; the government. This represent the bottom line of the classical political economy theory; which is the first variant of political economy theory, the second being the bourgeois political economy. As described by Ibrahim (2014, p. 48), "in the framework of classical political economy, the possible interpretations of the social and environmental disclosure are based on the idea of maintaining the legitimacy of the system as a whole, through imposing some restrictions on organisations by the state, where the state is acting in the interests of disadvantaged groups, for example, the disabled, minority races, in order to maintain the legitimacy of the capitalist system as a whole".

2.3 Empirical Literature

Cowan and Gadenne (2005) investigated the mandatory environmental reporting as well as voluntary reporting in the Australian context. They examined the annual reports of Australian companies for three consecutive years, that is from 1998 to 2000, using content analysis. The results revealed that Australian companies have a tendency to disseminate higher levels of positive environmental information in the voluntary sections of the annual report, than in the mandatory sections. Ahmad and Sulaiman (2004) studied the extent and nature of the environmental disclosure within Malaysian industrial and construction companies. Content analysis was used to determine the amount of environmental disclosure, where the number of sentences was adopted as a unit of measurement. The environmental information was measured based on monetary and non-monetary methods, good, bad and neutral news, and environmental audit and environmental policy. The findings indicated that the level of environmental disclosure in Malaysia seems low, where most of the companies (72.46%) do not disclose any environmental information, while some of them (27.54%) have disclosed some information.

The disclosed information was mostly neutral and declarative which was located in different parts of the annual report. Sen et al. (2011) evaluated the existing status of environmental disclosure practices in Indian companies from annual reports. The findings indicate that all the environmental disclosure themes were reported in a descriptive manner, in the directors' report, while some of them were disclosed in the chairman's speech section. However, the environmental disclosure is incomplete in terms of lack of coverage of most of the environmental themes included in the study. In addition, Indian companies disclosed positive or neutral information.

In Cameroon, studies on the practice of environmental disclosure are quite limited given that existing literature focus mostly on aspects related to the explanatory factors of environmental reporting and the usage of widespread reporting materials. Tegofack and Kamdem (2022) examined the determinants of the level of voluntary disclosure of environmental information by industrial enterprises. The ordinal logistic regression revealed that profitability, industry type, pressure from primary and secondary stakeholders tend out to be strong determinants of the disclosure of environmental information by industrial enterprises in the Cameroonian context. Dongmo (2023) on the other hand investigated on the explanatory factors of the structural disparity of sustainability reporting media beyond the referential structure that is Annex Note 35 designed for the occasion. The author found out that, the presence within the company of a comity in charge of sustainable development, the quality of the informative content of CSR communication media, the sector of activity and the search for reputation among stakeholders appeared to be decisive for the decision of the managers of companies in the OHADA zone to go beyond the recommended medium of disclosure; that is note 35 in order to disclose their non-financial information through non-harmonized reporting outlets.

The studies of Tegofack and Kamdem (2022), and Dongmo (2023) show that sustainability reporting in general and environmental reporting in particular is a reality, but there exist little or no empirical understanding on the overall characteristics of the environmental information disseminated by enterprises in their CSR reports. Very little is known about the nature, types of environmental information as well as environmental disclosure themes contained in sustainability reports published by industrial enterprises operating in Cameroon. The characteristics of environmental information have already been established in other contexts (such as Asia and Europe), as discussed in this sub section but such features are lacking in the Cameroonian context even in the presence of environmental disclosure activities practiced by industrial enterprises.

2.4 Global Reporting Initiative Sustainability Reporting Guidelines

The Global Reporting Initiative has passed through several stages since the standards were first introduced. The GRI was established as a project in 1997 by the Coalition for Environmentally Responsible Economies; a non-profit organisation in Boston. The main goal of this organisation was to develop unified guidelines for voluntary sustainability reporting (Dongmo, 2017). The GRI Standards comprise four main sections; universal, economic, environmental, and social standards and each standard has sub-standards and guidelines (Buallay, 2019). The three dimensions of GRI Standards are presented in table 1.

Table 1: Global Reporting Initiative standards, 2016

GRI 100: Universal standards	**GRI 400: SOCIAL**
GRI 102 Foundation 2016	GRI 401 Employment
GRI 102 General disclosure 2016	GRI 402 Labour and management relation
GRI 103 Management approach	GRI 403 Occupational, health and safety
GRI 200: ECONOMIC	GRI 404 Training and education
GRI 201 Economic performance	GRI 405 Diversity and equal opportunity
GRI 202 Market presence	GRI 406 Non-discrimination

GRI 203 Indirect economic impacts	GRI 407 Freedom of association
GRI 203 Procurement practices	GRI 408 Child labour
GRI 205 Anti-corruption	GRI 409 Forced or compulsory labour
GRI 206 Anti-competitive behaviour	GRI 410 Security practices
GRI 300: ENVIRONMENTAL	GRI 411 Right of indigenous people
GRI 301 Material	GRI 412 Human rights assessment
GRI 302 Water	GRI 413 Local community
GRI 303 Energy	GRI 414 Supplier social assessment
GRI 304 Biodiversity	GRI 415 Public policy
GRI 305 Emissions	GRI 416 Customer health safety
GRI 306 Effluent and waste	GRI 417 Marketing and labelling
GRI 307 Environmental compliance	GRI 418 Customer privacy
GRI 308 Supplier environmental assessment	GRI 419 Socio-economic compliance

Source: Buallay (2019)

2.5 Features of environmental disclosure

The characteristics of environmental disclosure is perceived in terms of environmental disclosure themes, types of environmental information and methods of disclosing environmental information. The various features to be identified in sustainability reports are presented in this sub-section.

2.5.1 Themes of environmental disclosure

The environmental section of GRI principally consists of eight parameters or categories; material, water, energy, bio-diversity, emission, effluent and waste, environmental compliance and supplier environmental assessment as summarized in table 2.

Table 2: Environmental Disclosure Guidelines-G4

	GRI category	**Checklist or Indicators**
1	GRI-301-Materials	Materials used by weight or volume
		Recycled input materials used
		Reclaimed products and their packaging materials
2	GRI-302-Energy	Energy consumption within the organisation
		Energy consumption outside of the organisation
		Energy intensity
		Reduction of energy consumption
		Reduction in energy requirements of products and services
3	GRI-303-Water	Water withdrawal by source
		Water sources significantly affected by withdrawal of water
		Water recycled and reused
4	GRI-304- Biodiversity	Operational sites owned, leased, managed in, or adjacent to, protected areas and areas of high biodiversity value outside protected areas
		Significant impacts of activities, products, and services on biodiversity

		Habitats protected or restored
		The International Union for Conservation of Nature (IUCN) Red List species and national conservation list species with habitats in areas affected by operations
5	GRI-305- Emission	Direct (Scope 1) GHG (Greenhouse Gas) emissions
		Energy indirect (Scope 2) GHG emissions
		Other indirect (Scope 3) GHG emissions
		GHG emissions intensity
		Reduction of GHG emissions
		Emissions of ozone-depleting substances (ODS)
		Nitrogen oxides (NOX), sulfur oxides (SOX), and other significant air emissions
6	GRI-306- Effluents and Waste	Water discharge by quality and destination
		Waste by type and disposal method
		Significant spills
		Transport of hazardous waste
		Water bodies affected by water discharges and/or runoff
7	GRI-307- Environmental Compliance	Non-compliance with environmental laws and regulations
8	GRI-308- Supplier Environmental Assessment	New suppliers that were screened using environmental criteria
		Negative environmental impacts in the supply chain and actions taken

Source: Adapted from Gustafsson (2017)

2.5.2 Types of environmental information

The positive information**;** which includes any kind of environmental information that builds up the image of the company, such as: winning prizes in the field of environmental protection as well as environmental compliance or information which gives an indication on how the enterprise is working for the wellbeing of all its stakeholders. The negative information; which includes any kind of environmental information that may reflect negatively on the company, such as, accident statistics during work or even fines related to non-compliance to environmental regulations. That is information which gives an indication of the harmful effects of the enterprise's activities on the stakeholders. The neutral information**;** includes any kind of environmental information which is difficult to determine its incidence on the company. Due to the complexity involve in the identification of neutral news, we have opted to base our analysis only on good and bad information.

2.5.3 Methods of Environmental Disclosure

The disclosure of environmental information by enterprises are usually done in the form of narration (qualitative), the information is sometimes quantified (non-monetary quantitative) or monetary information. Qualitative information relates to the narration of environmental activities without quantifying the information in monetary or non-monetary terms. Non-monetary quantitative concerns the use of actual numbers of a non-financial nature used to disclose the statistics and quantities regarding the environmental issues. Monetary information consists of making use of actual numbers to disclose the amounts incurred by the company regarding the environmental issues, such as, replacement of dust filters, establishing sewage

treatment stations and judicial compensations just to name that few. From our observation of the CSR reports of the industrial enterprises that constituted our sample size, we noticed that little or no attention was paid to monetary information. Information concerning monetary value of expenditures related to environmental activities was scarcely mentioned in sustainability reports. Hence the content analysis of methods used to disseminate environmental information was limited to qualitative information and non-monetary quantitative information.

III. RESEARCH METHOD

This study adopted qualitative descriptive research strategy, in which the instrument of data collection is documentary observation or textual analysis. A multiple case study is adopted involving twelve (12) sustainability reports of four industrial enterprises that published CSR reports on a relatively consistent base in accordance to the Global Reporting Initiative guidelines. The annual sustainability reports of the industrial enterprises were retrieved from the internet. The sampling technique used to choose the enterprises whose sustainability report will be exploited is the non- probability, judgmental (purposive) sampling. Zikmund (2003) describes this as a technique in which an experienced individual selects the sample based on his or her judgement about some appropriate characteristics required of the sample members. Conceptual content analysis was used to analyse the data in order to achieve our main objective relating to the characteristics of environmental disclosure (environmental disclosure themes, types of environmental disclosure and methods of environmental disclosure respectively). The conceptual content analysis was carried out manually through eye scanning of CSR reports. This method of analysis was chosen for two reasons. Firstly, the CSR reports used for this work were written in English and French, which rendered the coding exercise difficult to be realised. Secondly, this choice was motivated by the limited number of CSR reports. In the present study, content analysis was employed to investigate the environmental themes regularly disclosed by the industrial enterprises, the type of environmental information, and the method of disclosing the environmental information.

3.1 Enterprises considered in the case study

The CSR reports of the selected enterprises were downloaded from the year 2020 to 2022 and comprised a total of 12 reports for the year 2018, 2019, 2020 and 2021; with four reports from each industry, as shown in table 3.

Table 3: Sample for the content analysis

Industrial enterprises	**Types of industry**	**Number of reports**				**Total**
		2018	**2019**	**2020**	**2021**	
Boisson du Cameroun (SABC)	Brewery (BRY)	1	1	1	1	4
SOCAPALM	Agro-food (AF)	1	1	1	1	4
CIMENCAM	Building material (BM)	0	0	1	1	2
DANGOTE CEMENT		1	1	0	0	2
	Total	3	3	3	3	12

Source: by the author

The choice of *Boisson du Cameroun*, SOCAPALM, CIMENCAM and DANGOTE CEMENT as our case studies relies on some common characteristics shared by these enterprises. Firstly, the choice for these companies is largely explained by the fact that their sustainability reports is in accordance to some international guidelines such as the Global Reporting Initiative and United Nation Global Compact. Secondly, many companies in Cameroon practice CSR but very few of them publish their

environmental policies and activities in sustainability reports. The companies selected are among the limited numbers of companies that produce sustainability reports on a relatively consistent base. Thirdly, the enterprises that have been explored in this study operate in Cameroon; hence operate under similar laws and business environment. Lastly two of the enterprises (*les Boissons du Cameroun* and CIMENCAM) are amongst the top ten of results of ASCOMT / MALARIA survey on the perception of CSR practice in Cameroon for the years 2018 and 2019.

3.2 Measurement of environmental disclosure characteristics

The measurement of environmental disclosure is done in terms of themes, types, volume and level of disclosure. This done is by means of content analysis where the unit of analysis is the number of words. The measurement techniques of environmental disclosure in previous studies are depicted in table 4.

Table 4: Measurement techniques of environmental disclosure in prior studies

Study	Country	Disclosure item examine	Techniques
(Kuasirikun, 2005)	Thailand	Corporate reports	Content analysis - the number of words
(Davey, Low, and Ratanajongkol, 2006)	Thailand	CSR disclosure	Content analysis-the number of words
(Haron, Said, and Zainuddin, 2009)	Malaysia	CSR disclosure items	Content analysis number Of sentences
(Álvarez and Custodio, 2016)	France, Portugal, Spain, the UK and the US	CSR information	Disclosure Index

Source: by our own means

The most widely used technique to measure environmental disclosure is the number of words and the disclosure index. For the sake convenience, the number of words was used to determine the volume of environmental disclosure themes. Secondly, types of environmental information contained in CSR report was determined through the count of words from paragraphs containing information favouring or being against the interest of stakeholders. Thirdly, the method of environmental disclosure was determined via the counts of words and numbers for qualitative, and non-monetary quantitative information respectively.

IV. RESULTS AND DISCUSSIONS

This section comprises presenting and discussing of environmental themes, types of environmental information and methods of disclosing environmental information.

4.1 Environmental themes disclosed by industrial enterprises

In principle, an enterprise that disseminates environmental information ought to integrate eight environmental themes in their disclosure practices. The content analysis of the CSR reports permitted the identification of the following environmental disclosure themes as shown in table 5.

Table 5: Environmental disclosure themes

Industry	Material	Water	Energy	Bio-diversity	Emission	Effluent and Wastes	Environmental compliance	Supplier environmental assessment
Brewery	1	1	1	1	0	1	0	0
Building material	1	1	1	1	1	1	1	1
Agro-food	1	1	1	1	1	1	1	0

Source: From content analysis of CSR reports

The analysis shows that the environmental disclosure themes (categories) which are consistently disseminated by industrial enterprises in Cameroon are material, water, energy, bio-diversity, emission, effluent and waste, and environmental compliance.

The information content of sustainability reports of the building and agro-food industries show that clearly as seen in the extracts of tables 6 and 7, respectively.

Table 6: Extract 1 depicting environmental disclosure theme from building material industry

*Environmental Management System (EMS) drives our products towards ISO 14001 Certification. (**environmental compliance**) These commitments are indicative of our focus on fostering a cleaner and more sustainable operational environment. Energy Management initiatives on optimal or efficient use of the two boilers for production (**energy**). Water savings Environmental Performance: initiatives which include recovery of regeneration water **(water)**. We proactively manage the various types of pollutions from our operations. At the refinery, we are replacing existing silencers in the boiler plan t(**material**). to mitigate noise pollution. Emissions – less emissions due to efficient use of gas to fire our boiler(**emission**). We also minimise our soot emission, wastewater discharge and deforestation in all our oil operations. (**bio-diversity**). Fine tuning of the boilers to fire with less emission on oil. We have a brine recovery system in place to recover minerals and water in the process house. Waste Management (**effluent and waste**) through waste segregation, less waste generation owing to process optimization……. (Page 32)*

Source: CSR reports, building material industry, 2020

Table 7: Extract 2 depicting environmental disclosure theme from agro-food industry

Material	*Graft tape; Plastic basin; Sampling bottle Germination bag…. the materials consumed leaving the central store in Douala are transferred to the operational stores of the sites. (page 42)*
Energy	*Energy produced by the local electricity company ENEO: approximately 2,166,881 KWh consumed in 2019 across all sites; Generators in the event of power cuts, inability to supply by ENEO or by the turbine: 15,012.67 liters of diesel on average per month on all sites…… (page 42)*
Water	*Total volume of water withdrawn by source in m3 (2019): Surface water and rivers 507,612 Groundwater 418,214…… (page 39)*
Bio-diversity	*Areas of lowlands and rocky outcrops, where higher biodiversity can be expected, are preserved to create pockets of biodiversity conservation. The commissioning of lagoon basins allows the preservation of surface water and the regeneration of aquatic and riparian vegetation. (page 41)*
Emission	*Employees were once again made aware of what to do with a boiler in order to avoid abnormal release of smoke into the environment... (page 41)*
Effluent and waste	*A list of companies approved and authorized to recover waste according to the categories is set up by the company…… (page 40)*
Environmental	*Two annual regulatory inspections take place within each site and are carried out by inspectors sworn*

compliance	*by the State (Ministry of the Environment, Water, Industry, Agriculture and sometimes Labour). The company did not receive any sanctions for non-compliance with environmental laws and regulations in 2019. The recommendations given are taken directly into account and, if this requires a high cost, action sheets are opened to this effect. The action is then inserted into the table of corrective actions to initiate. …… (page 44)*

Source: CSR reports, agro-food industry, 2019

Hence, for subsequent evaluation of environmental disclosure practice in terms of types of disclosure and methods used to disseminate environmental information, we shall make use of seven categories of environmental information (material, water, energy, bio-diversity, emission, and effluent and waste).

4.1.1 Volume of environmental disclosure themes

Following the identification of environmental disclosure categories which are consistently disseminated by industrial enterprises in Cameroon, the next task of our content analysis was to find out the volume (amount) of disclosure of each category. The method of evaluating the volume of disclosure was the number words used in reporting each of the categories as per industry. The content analysis realised revealed the following information as indicated in table 8.

Table 8: Volume of environmental disclosure themes of industrial enterprises

Element	Number of words							
	Material	**Water**	**Energy**	**Bio-diversity**	**Emission**	**Effluent and waste**	**Environmental compliance**	**Total**
Industry								
BRY	104	57	100	104	0	195	154	**714**
BM	849	586	160	491	885	102	341	**3414**
AF	216	132	176	128	792	532	160	**2136**
Total	**1169**	**775**	**436**	**723**	**1677**	**829**	**655**	**6264**
%	**19%**	**12%**	**7%**	**12%**	**27%**	**13%**	**10%**	

Source: From content analysis of CSR reports

From the table 8, it can be seen that the most disclosed environmental disclosure themes by industrial enterprises operating in Cameroon is emission specifically information concerning greenhouse gas emissions. This is followed suit by material used to curb the environmental impact of their activities on the local communities. Another environmental category that was prioritized by industrial enterprises is effluent and waste, given that the majority of industries described how the managed and recycled their waste in order to mitigate the effects of land and water pollution. The least disclosed environmental theme is energy due to the fact that industrial enterprises failed to provide exhaustive information concerning their energy consumption rate and how they intend to reduce the energy consumption or even how they are planning to switch to more environmentally friendly source of energy. The extent of environmental disclosure in terms of words related to each of the themes is clearly depicted by figure 1.

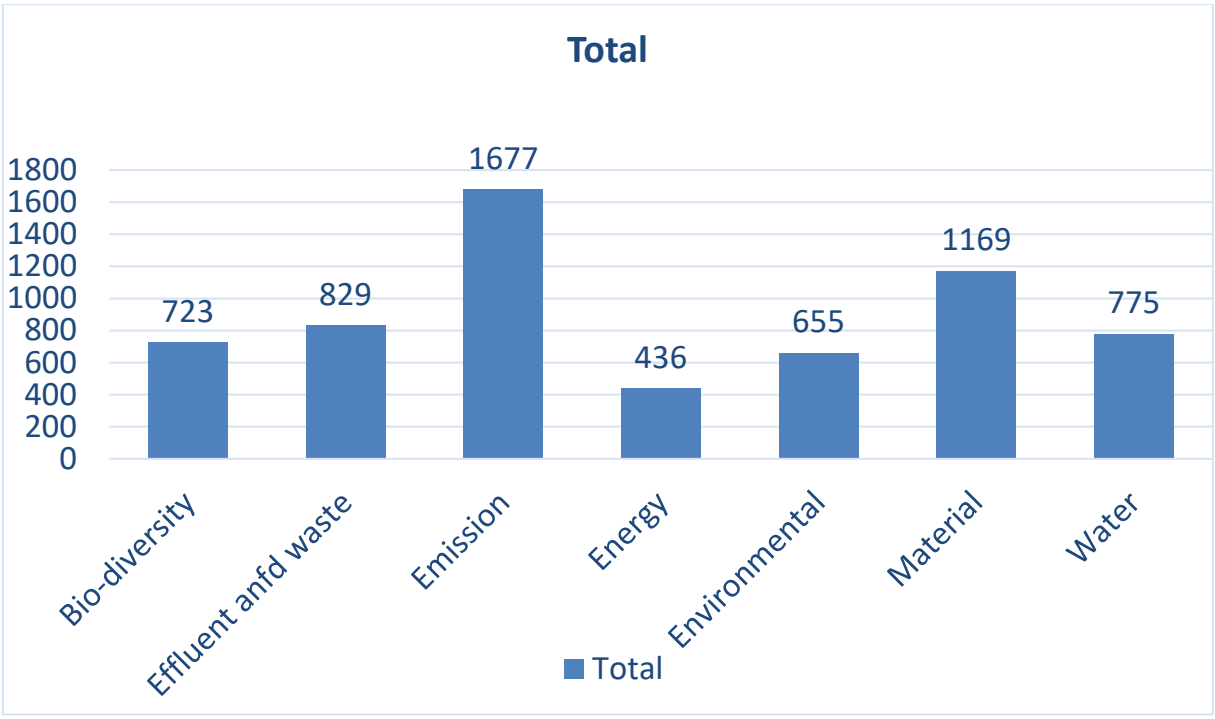


Fig.1: Environmental disclosure themes regularly disseminated

Source: by our own means

4.2 Types of environmental information disseminated by industrial enterprises

The content analysis realised on the twelve CSR reports regarding the types of environmental information disseminated by industrial enterprises operating in Cameroon is shown in table 9.

Table 9: Types of environmental disclosure themes of industrial enterprises

Element	Number of words							
Types of information	**Positive**				**Negative**			
Year	**2018**	**2019**	**2020**	**2021**	**2018**	**2019**	**2020**	**2021**
Industry								
BRY	176	157	173	208				
BM	764	764	794	812	52	80	60	66
AF	438	438	438	438	88	88	88	88
Total	**1378**	**1389**	**1405**	**1458**	**140**	**168**	**148**	**154**
Average	**5630**				**610**			
%	**90.23%**				**9.77%**			

Source: From content analysis of CSR reports

Table 9 shows clearly that, over the years the industrial enterprises in Cameroon have been disseminated more of positive information than negative information and the environmental disclosure themes which have some elements of bad

news are emission, effluent and waste as seen in the following extract:

Due to the high temperatures needed to turn limestone into clinker, the production of cement consumes a significant amount of energy and this in turn causes combustion-related CO2 emissions. (CSR report, Building material industry, 2018)

The information content of all the analysed reports is littered with good news aimed at showing their consideration for environmental issues are seen the extracts below:

We are committed to maintaining excellent standards of environmental performance. We recognise the part that we can play in improving the environment, particularly in and around our sites of operation. We apply economically sound sustainable development principles to our business and seek to maximise energy efficiency and minimise the environmental impact of our operations. (CSR report, Building material industry, 2019).

We comply with stringent environmental regulations to ensure that our raw material quarrying will not endanger local water sources, which may be used by local communities. We are committed to the implementation and maintenance of the National Industrial Standards ISO 14001:2004 Environmental Management System (EMS), which ensures a systematic approach to environmental management within the defined scope of our operations. We aim to comply with relevant legal requirements with a view to providing a sustainable environment for manufacturing, packaging, distribution and sales of cement and continuous improvement of our operations. (CSR report, Building material industry, 2019).

The proportion of the types of environmental information communicated by industrial enterprises is positively inclined as depicted by figure 2.

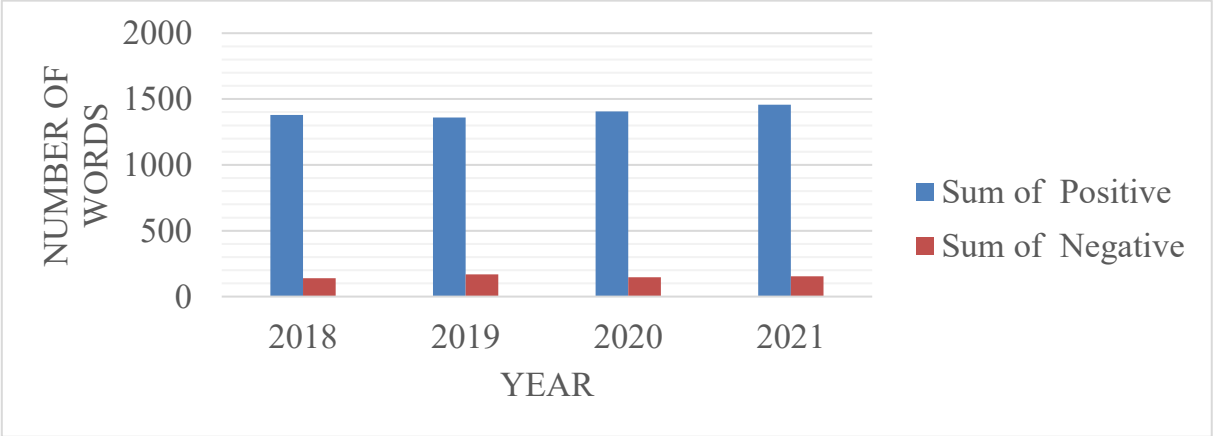


Fig.2: Types of environmental disclosure over the years

Source: by our own means

Table 9 shows that positive information is more disseminate than negative information to the proportion of 90.23% and 9.77% respectively. This is consistent with the results of previous literature investigating environmental disclosure practices for example, Deegan and Rankin (1996), Cowan and Gadenne (2005). In line with this finding, Deegan and Rankin (1996, p. 59) stated that the companies prefer to provide positive information because, "they believe there is a need to legitimize the existence of their operations". Furthermore, Yusoff and Lehman (2006) have found that Malaysian companies have disclosed just positive environmental information, in order to consolidate a good image in the community. Other researchers explain this situation by highlighting that, the major reason for companies to disclose their environmental information is to maintain their reputation and appease investors and other businesses. This is especially so now that we are seeing the emergence of the "ethical investor"; investors who prefer to invest in companies whose activities do not cause damage to the environment (Deegan and Rankin, 1997).

4.3 Methods used to disseminate environmental information

The result of the content analysis relating to the manner in which industrial enterprises communicate

to their stakeholders concerning their environmental actions is presented in table 10.

Table 10: Methods used to disclose environmental information by industrial enterprises

Element	Number of words							
Types of information	Qualitative				Non-monetary quantitative			
Year	2018	2019	2020	2021	2018	2019	2020	2021
Industry								
BRY	176	157	173	208	13	13	6	14
BM	796	824	834	858	9	30	28	37
AF	526	526	526	526	63	63	63	63
Total	**1498**	**1507**	**1533**	**1592**	**85**	**106**	**97**	**114**
Average	**1532.5**				**100.5**			
%	**93.85%**				**6.15%**			

Source: From content analysis of CSR reports

The mean number of words related to qualitative information disseminated in sustainability reports is 15 times greater than the mean number of words reserve to non-monetary quantitative information that is 1532.5 words and 100.5 words respectively. The statistics in terms of total number of words over the years is shown in the figure 3.

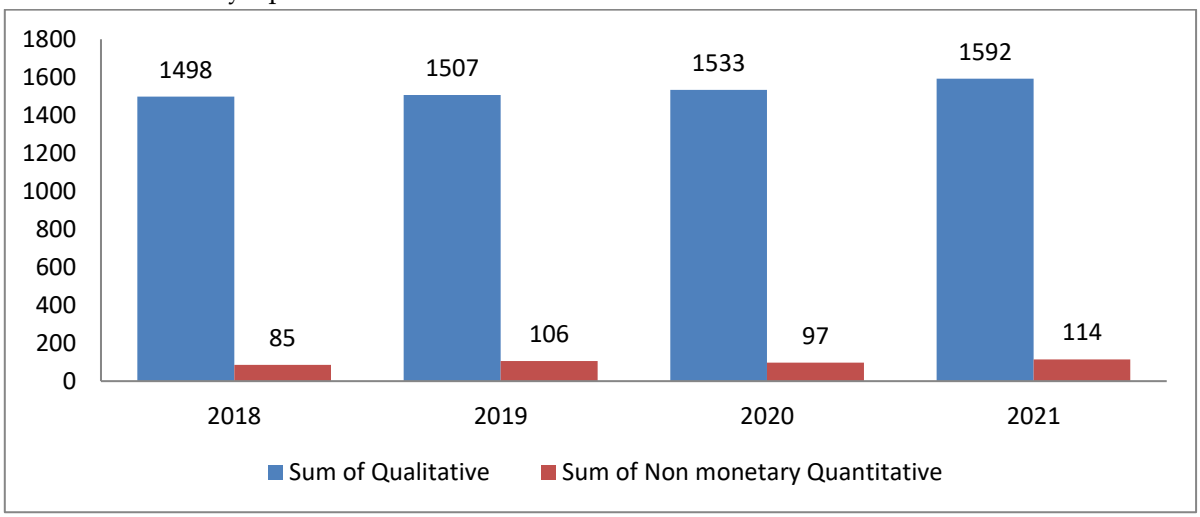


Fig.3: Methods used to disclose environmental information

Source: by our own means

It is noticed that environmental information disclosed by industrial enterprises in Cameroon is qualitative in nature and mostly descriptive. This finding is consistent with that of Sen et al. (2011), who evaluated the existing status of environmental disclosure practices in Indian companies. The findings indicate that all the environmental disclosure themes were reported in a descriptive manner, in the directors' report, while some of them were disclosed in the chairman's speech section.

V. CONCLUSION

This study also seeks to describe the practice of ED by establishing the overall characteristics of environmental information disclosed by industrial

enterprises in Cameroon. We found out that emission of greenhouse gas is the most disclosed environmental theme followed by material (material used by weight and recycled inputs material used), effluent and waste (waste by types and disposal method), water (sources of water, water recycled and reused), biodiversity (habitats protected or restored), environmental compliance, with energy (energy consumption, intensity, sources of energy used and energy reduction) as the least disclosed theme. The supplier environmental assessment is yet to be integrated by enterprises in their reporting practice. The content analysis also revealed environmental information disclosed by industrial enterprises is mostly positive because they want to maintain a good image and legitimise their existence in the society in which they operate. Industrial enterprises are reluctant to disseminate negative information in order to avoid creating a negative impression before the stakeholders regarding their environmental policy. Furthermore, the content analysis established that qualitative (descriptive), and non-monetary quantitative methods are used to diffuse environmental information in Cameroon. The monetary method is yet to be implemented due to the inherent difficulty to attach monetary value to environmental activities.

RECOMMENDATIONS

The disclosure of environmental information with respect to GRI guide lines by some industrial enterprises on one hand, and the fact that these enterprises disseminate environmental information with the help of other reporting materials such as websites and sustainability reports, rather than on note 35 of the financial statements of commercial entities as recommended by OHADA, on the other hand calls for the attention of the OHADA legislator. We therefore suggest that OHADA as an organisation should work fast towards conceiving an accounting reporting framework for social and environmental information (non-financial information), just as the existing OHADA accounting law for financial information. Alternatively, some of the dispositions of IFRS S1 & S2 could be integrated in the current accounting law of financial reporting.

The development of an accounting framework for social and environmental reporting by the OHADA legislator can contribute in facilitating the audit of environmental information contained in sustainability reports. This is because the qualitative characteristics of financial information such as comparability, relevance, timeliness, verifiability, comparability, understandability cannot be directly transpose as criteria to be used to assessed the quality of environmental information because of the disparity that may exist between financial information and environmental information, as the latter can be monetary and non-monetary and also because of the inherent difficulty of quantifying environmental information.

REFERENCES

[1] Ahmad, N. N. & Sulaiman, M. (2004). Environmental Disclosures in Malaysian Annual Reports: a Legitimacy Theory Perspective, *International Journal of Commerce & Management,* 14(1), pp. 44-58.

[2] Álvarez, I. G., & Custodio, I. A. (2016). Disclosure of corporate social responsibility information and explanatory factors. *Online Information Review*, 40(2), 1-29. doi: 10.1108/OIR-04-2015-0116

[3] Berthelt, S., Cormier, D. & Moncton, M. (2003). Environmental Disclosure Research: Review and Synthesis, *Journal of Accountancy Literature* 22, 1-44.

[4] Buallay, A. (2019). Is sustainability reporting (ESG) associated with performance? Evidence from the European banking sector. Management of Environmental Quality: *An International Journal,* 30(1), 98-115.

[5] Cowan, S. & Gadenne, D. (2005). "Australian corporate environmental reporting: a comparative analysis of disclosure practices across voluntary and mandatory disclosure systems", *Journal of Accounting & Organisational Change,* 1 (2), 165-179.

[6] Davey, H., Low, M. & Ratanajongkol, S. (2006). Corporate social reporting in Thailand: The news is all good and increasing. *Qualitative Research in Accounting & Management*, 3(1), 67-83. doi: 10.1108/11766090610659751

[7] Deegan, C. (2014). An overview of legitimacy theory as applied within the social and environmental accounting literature. *Sustainability accounting and accountability*, 2, 248-272.

[8] Deegan, C. (2002). The legitimising effect of social and environmental disclosures – a theoretical foundation. *Accounting, Auditing & Accountability Journal* 15(3), 282-311.

[9] Deegan, C. & Rankin, M. (1997). "The materiality of environmental information to users of annual reports accounting", *Auditing and Accountability Journal,* 10 (4), 562-583.

[10] Deegan, C. & Rankin, M. (1996). "Do Australian companies report environmental news objectively an analysis of environmental disclosure by firms prosecuted successfully by the environmental protection authority", *Accounting, Auditing & Accountability Journal,* 9 (2), 50-67.

[11] Deegan, C. & Gordon, B. (1996) "A study of the environmental disclosure practices of Australian corporations", Accounting and Business Research, 26 (3), 187-199.

[12] Dongmo, R. M. (2023). Analyse des determinants des disparites structurelle des supports de reporting durabilite dans un context reglementaire coecitif. *Revue Africaine de Management*, vol 9 (1), pp. 25-57.

[13] Dongmo R. & Ndjetcheu L. (2018). « Les déterminants de la communication sociétale des entreprises de l'espace OHADA: une étude en contexte camerounais », *proposition de communication, Deuxième Journée d'Etude Africaine de Comptabilité*, Dakar, Sénégal.

[14] Dongmo R. (2017). « Le reporting sociétal des entreprises de l'espace OHADA : une étude en contexte camerounais », Thèse de doctorat en Sciences de Gestion, FSEGA, Université de Douala.

[15] Gibson, R. B. (2001). Specification of sustainability-based environmental assessment decision criteria and implications for determining "significance" in environmental assessment. Ottawa: Canadian Environmental Assessment Agency.

[16] GRI (2016). Consolidated set of GRI sustainability reporting standards 2016. Amsterdam, The Netherlands.

[17] Gray, R., Owen D. & Adams, C. (1996). *Accounting and Accountability: Changes and challenges in corporate social and environmental reporting*, London: Prentice Hall.8(2), 78-101.

[18] Gustafsson, J. (2017). Single case studies vs. multiple case studies: A comparative study, Academy of Business, Engineering and Science, Halmstad University, Halmstad

[19] Guthrie, J., & Parker, L. (1990). Corporate Social Disclosure Practice: a Comparative International Analysis. *Advances in Public Interest Accounting*, 3, 159-175.

[20] Haron, H., Said, R. & Zainuddin, Y. (2009). The relationship between corporate social responsibility disclosure and corporate governance characteristics in Malaysian public listed companies. *Social Responsibility Journal*, 5(2), 212-226. doi: 10.1108/17471110910964496

[21] Hart, S. L. (1995). A natural-resource-based view of the firm. *Academy of management review*, 20(4), 986-1014.

[22] Henion, K. E., & Kinnear, T. C. (1976). Measuring the effect of ecological information and social class on selected product choice criteria importance ratings. Ecological Marketing, Chicago: American Marketing Association, pp145-156.

[23] Ibrahim, M. H. I. (2014). Corporate environmental disclosure: A case from Libyan Construction Industry. Ph. D thesis, Liverpool John Moores University.

[24] Islam, S., Hosen, A. & slam, M. (2005). An Examination of Corporate Environmental Disclosure by the Bangladeshi Public Limited Companies, *Journal of Social Sciences* 3 (9), 1095-1102.

[25] Kuasirikun, N. (2005). "Attitudes to the development and implementation of social and environmental accounting in Thailand", *Critical Perspectives on Accounting*, 16, 1035–1057.

[26] Morelli, J. (2011). Environmental sustainability: A definition for environmental professionals. *Journal of environmental sustainability*, 1(1), 2.

[27] Porchelvi, A. (2019). Environmental reporting practices: a study of selected companies PhD Thesis, department of commerce delhi school of economics, University of Delhi.

[28] Richardson, A. J. (1987). Accounting as a legitimating institution, *Accounting, Organisations & Society*, 12, 341-355

[29] Sen, M., Mukherjee, K. & Pattanayak, J. K. (2011). "Corporate environmental disclosure practices in India", *Journal of Applied Accounting Research*, 12 (2), 139-156.

[30] Tegofack, B. C. & Kamdem D. (2022). The determinants of corporate environmental disclosure in sub-Sahara Africa: an empirical study of industrial enterprises from Cameroon. International Journal of Management Studies and Social Science Research, 4 (5), 211-225. DOI: https://doi.org/10.56293/IJMSSSR.2022.4519

[31] Yusoff, H. & Lehman, G. (2006). "International differences on corporate environmental disclosure practices: a comparison between Malaysia and Australia", 4th International conference on Accounting and Finance in Transition. University of South Australia.

[32] Zikmund, W. G. (2003). *Business research methods* (7th ed.). OHIO: South- W

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.14

Methodological Aspects of the Transition from Accuracy Metrics to Risk Modelling in the Design of Hybrid Intelligent Systems

Bezhentsev Yurii

Software Developer, Houston, Texas, United States

Received: 20 Feb 2026; Received in revised form: 22 Mar 2026; Accepted: 26 Mar 2026; Available online: 01 Apr 2026

Abstract *— The article examines the methodological transition from the use of standard accuracy metrics to risk-oriented modelling in the design of hybrid intelligent systems embedded in a management and control loop. The relevance of the approach stems from the fact that, in real processes, the quality of a solution is determined by the probability of undesired events, the magnitude of their consequences, and the speed of error detection and reversibility in the operational environment. The aim of the work is to translate the evaluation of intelligent models from the plane of numerical indicators into the language of systemic risk associated with people, infrastructure, and organisational procedures. The scientific novelty lies in integrating cost-sensitive error assessment, class imbalance analysis, and data shifts with an architectural description of hybrid (neuro-symbolic) systems, in which risk is distributed along the entire chain: data – model – rules – human – action. Accuracy metrics are proposed to be treated as particular input characteristics within a more general scheme for managing undesired events, defined by loss functions, barrier architecture, traceability, explainability, and controlled degradation modes. It is shown that, under class imbalance, label defects, and drifting data, the choice of metric and thresholds becomes a methodological decision that directly influences actual damage rather than a technical detail of the experiment. A conclusion is formulated on the necessity of shifting acceptance criteria from maximising aggregated metrics to constraining expected loss and ensuring risk controllability at the level of the hybrid system as a whole. The article is intended for researchers and engineers developing and deploying risk-sensitive intelligent systems in safety-critical and regulated domains.*

Keywords *— **hybrid intelligent systems, risk-oriented evaluation, accuracy metrics, risk modelling, cost-sensitive classification***

I. INTRODUCTION

The transition from evaluating intelligent solutions via accuracy metrics to risk modelling is associated with changes in the tasks for which such solutions are deployed. An intelligent system increasingly functions as part of a management and control loop, where the outcome is determined by how that prediction is translated into action, how quickly an error is detected, and the consequences for people, infrastructure, and the organisation [1]. Under these conditions, a risk-oriented methodological frame defines a different object of analysis. The focus shifts to undesired events and their consequences, as well as to prevention, containment, and recovery mechanisms that can be designed and empirically verified.

Accuracy metrics remain useful; however, their interpretation in real processes is unstable and depends on factors that are rarely captured in a laboratory formulation of the task. In practice, metric values may change substantially with shifts in class prevalence and with errors in the reference

annotation, leading to systematic misjudgements of a model's suitability for decision-making. This effect has been demonstrated in modelling studies, where even metrics commonly considered robust to imbalance vary significantly with changes in prevalence and ground-truth quality, and where high aggregate scores can arise alongside low actual utility of a classifier in the applied process [2].

In practice, asymmetry in consequences further amplifies this effect, since the costs of a false alarm and a miss rarely coincide. The target criterion is then naturally displaced towards expected losses and resource savings, which is formalised through error cost functions and enables comparison of solutions in terms of their impact on damage [3].

Hybrid intelligent systems integrate statistical models, formal rules, expert knowledge, and human-in-the-loop procedures; as a result, risk is distributed across the entire transformation chain from data to action. An error may be induced by data, algorithms, rule sets, integration logic, or usage protocols, and then amplified by organisational delays and by improper distribution of trust between automation and human operator [4]. Within such an architecture, a risk-oriented perspective enables responsibility to be decomposed across components and controlled operating regimes to be specified, including constraints, input-condition checks, routing of difficult cases, and traceability requirements. A survey of research on neuro-symbolic approaches emphasises that combining data-driven learning with symbolic reasoning improves interpretability and reliability in decision-making tasks, where the robustness of system behaviour in complex contexts is of primary importance [5].

II. MATERIALS AND METHODOLOGY

The materials are based on seven sources that present complementary lines of argument for why the evaluation of intelligent solutions should rest on risk and systemic consequences rather than solely on agreement with a reference annotation. The theoretical framework is provided by work on risk-oriented AI and ML, which emphasises that quality manifests through the probability of undesired events, the severity of damage, and the controllability of system behaviour within the operational loop, including error detectability and the effectiveness of response procedures [1].

The empirical basis for the challenges of interpreting accuracy metrics under real-world conditions comprises studies demonstrating the sensitivity of metrics to class prevalence and to defects in reference annotation, as well as the possibility of high aggregate scores coexisting with a classifier's low factual usefulness in the applied process [2]. For the formalisation of applied utility and the transition to criteria of expected losses, results on cost-sensitive classification and error cost functions are employed, enabling the linkage of threshold choice to the profile of consequences and resource constraints [3].

Materials on the architectural specificity of hybrid intelligent systems include studies of neuro-symbolic models in risk-sensitive domains and a survey of dependable neuro-symbolic approaches, highlighting interpretability, traceability, and robustness as engineering properties spanning the entire data–decision–action chain [4, 5]. Additionally, data are used on how class imbalance affects the consistency of conclusions when comparing models under different metrics, which is important for the methodological choice of criteria in tasks characterised by rare events and asymmetry of damage [6]. Practice-oriented measurement of the gap between out-of-deployment evaluation and in-deployment behaviour is supported by findings on harmful data shifts in clinical applications, where limited availability of true labels amplifies the importance of monitoring and protective mechanisms [7].

The research methodology is constructed as an analytical integration of several interrelated procedures aimed at translating model quality into the language of systemic risk for hybrid intelligent architectures.

First, a conceptual analysis of evaluation objects is performed, in which accuracy metrics are interpreted as partial descriptors of agreement with annotation under a fixed decision-making logic, while the central object becomes the risk of undesired events and its controllability within the application loop [1, 2]. Second, a methodological reconstruction of the relationship between metrics and consequences is carried out through a cost model of errors, where the criterion for comparing solutions is posed in terms of

expected losses and allows for threshold tuning that accounts for asymmetry of damage, contextual constraints, and instance-dependent costs [3].

Third, a comparative analysis of uncertainty regimes in hybrid systems is undertaken, in which sources of risk are distributed across components and integration interfaces, including data, model, rules, organisational delays, and human participation, while requirements for barriers, traceability, and robust degradation are identified in a manner consistent with the neuro-symbolic logic of reliability enhancement [4, 5]. Fourth, an empirically motivated analysis examines the influence of imbalance and distributional shifts on the validity of conclusions about model suitability. This analysis substantiates the need for risk segmentation, prevalence control, and monitoring of degradation indicators in operation [6, 7].

As a result, the chosen methodology fixes the transition from evaluation by numbers to a designed scheme of undesired-event management, in which accuracy metrics serve as input characteristics, while acceptance and validation criteria are formulated in terms of damage, detectability, reversibility, and response procedures at the level of the hybrid system as a whole [1–3, 5, 7].

III. RESULTS AND DISCUSSION

Basic accuracy metrics describe agreement between predictions and reference annotations under a fixed decision-making procedure [6]. The Accuracy indicator reflects the proportion of correct responses among all instances and depends on the selected threshold when the system outputs probabilities or scores. Precision and Recall capture different aspects of quality for a rare target event and are based on the structure of type I and type II errors in the confusion matrix. F1 aggregates precision and recall into a single value, which is convenient for comparing alternatives, but it obscures the underlying trade-off between error types and conveys limited information about system behaviour under threshold variation. ROC-AUC evaluates ranking performance across the entire threshold range and is often used to compare models; however, the resulting scalar value may have different practical interpretations depending on event prevalence and error-consequence profiles.

These differences are critical for hybrid intelligent systems because, further along the decision loop, the model score is converted into action via rules, constraints, and control procedures; consequently, the meaning of a metric is determined by how exactly the system will be used in the process.

Error asymmetry and the costs of false-positive and false-negative decisions shift evaluation from the domain of universal indicators into the domain of applied utility. When the cost of missing a high-risk event exceeds the cost of a false alarm, the optimal decision threshold shifts, and maximising standard metrics no longer coincides with minimising expected loss. Under class imbalance, the problem of average values is exacerbated: high Accuracy may be obtained by systematically ignoring a rare class, while F1 and related metrics yield different model rankings under identical base data and noise levels, making metric choice a component of methodology rather than a technical detail. Empirical studies demonstrate that, as the imbalance grows, the agreement between popular metrics decreases, and the selection of an indicator begins to alter which model is deemed best [6].

The gap between pre-deployment evaluation and in-deployment behaviour arises from changes in the context of use, shifts in data flows, and the influence of the procedures into which the solution is embedded. Data may drift over time, across sites, and across user groups, as well as through changes in measurement and recording practices. This leads to quality degradation and to shifts in probabilistic estimates. In clinical applications, situations arise in which data shifts can degrade performance and increase the risk of harmful interventions, while timely evaluation against true labels is often infeasible; consequently, monitoring shifts and designing protective mechanisms at the system level are required [7]. Model Evaluation Metrics are summarised in Table 1.

Table 1. Model Evaluation Metrics

Metric/idea	Meaning	Pitfall/limit	What matters in a system
Accuracy	share correct	class imbalance, threshold	don't rely on it alone
Precision	alert quality	threshold/prevalence sensitive	when FP are costly
Recall	don't miss	can inflate FP	when FN are costly
F1	P+R in one	hides trade-off/threshold behavior	add threshold-based view
ROC-AUC	ranking ability	interpretation shifts with prevalence/costs	tie to how the score is used
FP/FN cost	utility	metrics ≠ expected loss	set threshold by expected loss
Imbalance	rare class	metrics disagree	metric choice is methodological
Offline≠Online	deployment behavior	data shift, labels unavailable	monitoring + safeguards

The risk-oriented paradigm for evaluating intelligent systems rests on the notion that solution quality manifests in the probability of undesired events, the severity of their consequences, and the system's ability to detect hazardous regimes and keep them under control. Risk is conveniently conceptualised as the product of propensity to error, the cost of that error, and the degree of controllability, which is determined by detectability, reversibility, and the presence of response procedures. This definition renders the evaluation application-oriented and enables the linking of modelling results to safety, reliability, and resilience requirements in real operation. In hybrid intelligent systems, risk is not localised in a single component, so its assessment requires a description of the entire architecture and of how information and decisions propagate through rules, constraints, and human interaction.

A typology of risks is based on sources of uncertainty and mechanisms that amplify them. Model risks include behavioural instability at distributional boundaries, erroneous overconfidence, and degradation under changing conditions. Data-related risks include incomplete or biased data collection, distributional drift (shift in distribution), and data transformations before it is input to the model. Rules- and knowledge-related risks include contradictory rules, incomplete case coverage, and changes in regulations that such systems must comply with. Risks include operational risks of delays, failures of infrastructure, and breaks in integration and version incompatibilities; and human-factor risks of improper forms of trust allocation, ignoring processes, and misinterpretation of recommendations and assumptions. Regulatory risks manifest as non-compliance with established procedures, insufficient decision traceability, and a lack of audit readiness.

Risk as a systemic characteristic requires end-to-end consideration of the decision-making pipeline. At the input, measurements and data are collected; these are followed by cleaning and transformations, then by state estimation via the model, then by rule-based logic and constraints, after which the decision is converted into action and generates new feedback data. Errors may arise and accumulate at any stage, and their effect depends on the application context. A critical feature is that an identical model error occurring at different points of the pipeline leads to different consequences. Therefore, evaluation must account for escalation paths, stopping conditions, manual review procedures, and recovery speed, since these factors determine controllability and detectability of risk.

The methodology for transitioning to risk management begins with identifying undesired events at the process level, where an event is described in terms of the disrupted objective, context, and potential damage. Causal models are then constructed that connect events to combinations of conditions, data defects, model errors, rule conflicts, and

execution failures. Such analysis defines control points at which risk can be reduced through input-condition checks, action constraints, and switching to safe modes.

Subsequently, the loss function is formalised via an error cost matrix and an outcome utility, so that comparisons of alternatives reflect expected damage in process terms. Accuracy metrics are embedded into this frame through probability calibration, threshold tuning, and stratified evaluation across segments, since risk is unevenly distributed and concentrated in specific regimes. Demonstrable risk reduction is achieved through a barrier architecture that includes constraints, routing of difficult cases, abstention from decision-making under low reliability, control over action execution, and monitoring of degradation signals and feedback mechanisms that keep the system within an acceptable behavioural region. Figure 1 illustrates Risk Management in Intelligent Systems.

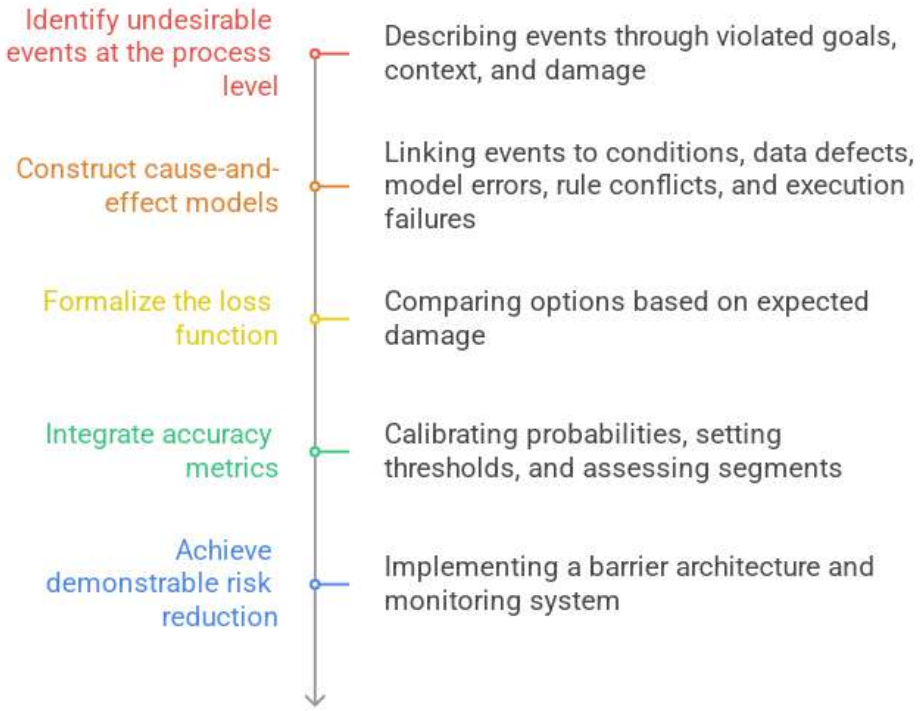


Fig. 1. Risk Management in Intelligent Systems

In risk modelling, a hybrid intelligent system is a linkage of heterogeneous components, each contributing its own uncertainty and mechanisms for constraining it. The statistical model generates estimates and rankings; rules define admissible decision regions and ensure compliance with regulations; expert knowledge determines the semantics of features and the prioritisation of consequences. The human operator makes decisions in situations where automatic action is hazardous or poorly specified. The distribution of responsibility becomes both a technical and an organisational object of design, specifying which component makes the decision, which component constrains the action, which component explains the reasons, and which component bears the duty to stop and escalate. Such decomposition links risk to architecture and prevents situations in which responsibility is blurred among model, rules, and operator.

Hybridisation mechanisms determine where exactly a control signal arises and how it is transformed into action. In a model-to-rule scheme, the model outputs an estimate, after which a set of constraints determines whether action is permissible and in what mode. Sometimes rules specify which cases are admissible, so the model can assume a smaller, more stable domain. For example, an ensemble or cascade may split risk across levels, with simple classifiers filtering out simple cases and more complex classifiers activating only in the presence of more complex cases. In this configuration, risk is shifted to interfaces, where alignment of confidence scales, a shared

interpretation of context, and predictable behaviour in the face of signal conflicts become crucial. In a hybrid system, value is provided by controlled degradation: failure of one layer should trigger a safe mode rather than uncontrolled action.

The human-in-the-loop contour enhances reliability when it is designed as an engineering mechanism rather than merely a formal signature in documentation. Escalation zones are defined according to uncertainty level, potential damage magnitude, and decision type. Trust control includes operator training, interface design, comprehensible rationales for recommendations, and explicit constraints that prevent blindly following system advice. Automation bias arises when an operator comes to perceive automated outputs as guaranteed correct and ceases to verify applicability conditions. Its prevention requires modes that allow the system to communicate its limitations, offer alternative actions, and demand confirmation in high-risk situations. In such settings, the human acts as part of the barrier architecture and assumes the role of active diagnostician rather than passive executor.

Traceability and explainability become instruments of risk management because they support verifiability of decisions and accelerate recovery after incidents. Traceability links a decision to input data, component versions, applied rules, context, and human actions. This enables localising the cause of an undesired event and updating a specific segment of the pipeline without dismantling the entire system. Explainability is important to ensure that escalation rules operate correctly and that operators understand which features and constraints led to an action. In hybrid systems, explanations must remain consistent with rules and data. Otherwise, they become decorative and increase risk by fostering a false sense of control. Risk management requires explanations that clearly delineate applicability boundaries and reveal the reasons for doubt.

Risk-oriented design relies on practices that keep the system within controlled regimes throughout its life cycle. Handling uncertainty includes confidence calibration, abstaining from decision-making when reliability is insufficient, and detecting cases that do not resemble known patterns. Scenario-based testing is built around stress conditions, boundary feature combinations, and rare events that account for the bulk of damage. Risk segmentation leads to differentiated thresholds and policies for different contexts, because identical errors incur different costs in different segments. Operational monitoring tracks data drift, performance degradation, incident growth, and changes in confidence distributions, after which feedback and component-update procedures are triggered. Risk management is reinforced through audits, decision logs, response procedures, and periodic revisions to the registry of undesired events, which transform acceptance criteria and the development life cycle itself. Requirements are formulated in terms of damage constraints and escalation rules. Design establishes barriers and control points. Validation examines risk scenarios and the system’s capacity for safe degradation. The operation comprises continuous monitoring, incident analysis, and policy and model updates to keep risk controllable in a changing environment. Figure 2 illustrates Hybrid Intelligent Systems.

When implementing a risk-oriented approach, a common distortion of meaning arises, treating risk as a cosmetic complication of familiar metrics. A team may introduce additional indicators, apply class weighting, and complicate averaging schemes, thereby creating an appearance of progress, while the connection to process-level consequences remains unspecified. As a result, the solution is optimised with respect to formal numbers, whereas undesired events continue to occur because their generating mechanism lies in the application's logic, in escalation rules, and in the limits of admissible actions. The absence of a formalised list of undesired events aggravates the problem, as discussion shifts toward convenient quality indicators while error costs remain implicit and each project participant interprets them idiosyncratically. Consequently, system requirements become blurred, and acceptance degrades into comparing models using metrics that are only weakly related to damage and controllability.

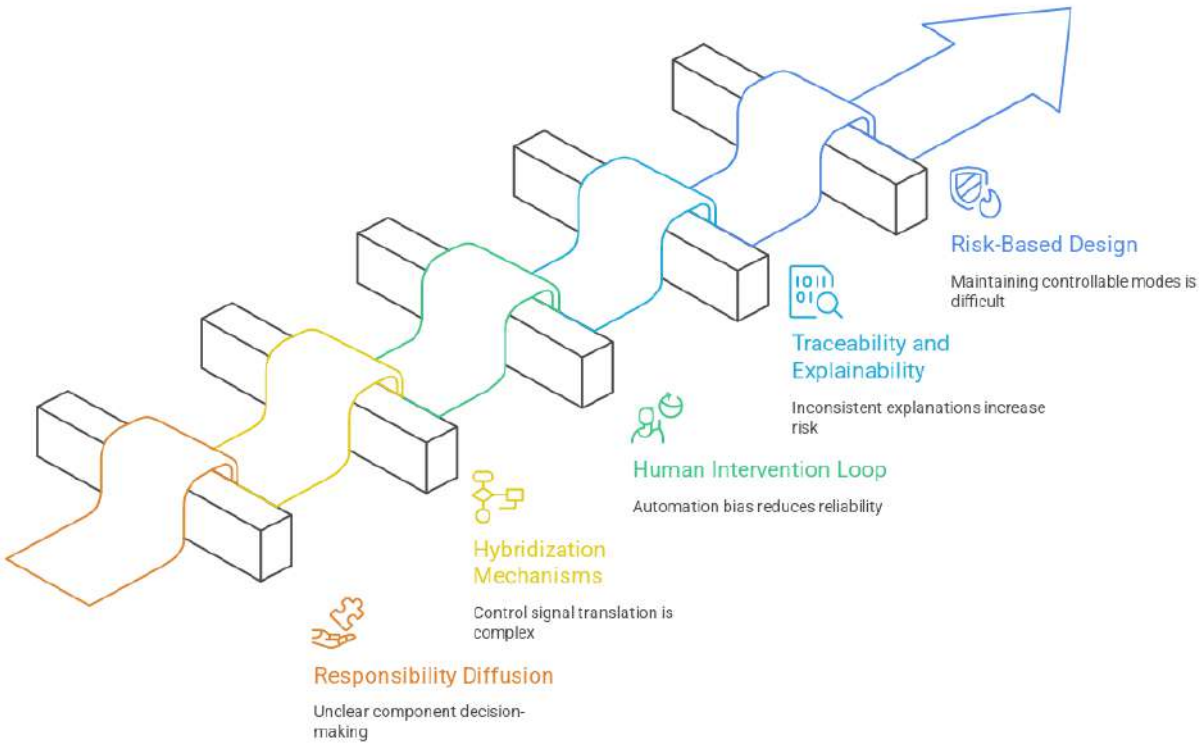


Fig. 2. Hybrid Intelligent Systems

Another persistent antipattern is associated with uncalibrated probabilities and arbitrary threshold management. Probabilistic estimates are perceived as reliable measures of confidence, although they may be systematically over- or underestimated, so thresholds begin to act as generators of latent risks. Threshold errors are particularly dangerous in hybrid architectures, where a threshold can trigger rule cascades, switch service modes, and influence human decisions. Neglect of the human and organisational loop entrenches these errors, because operators receive signals without clear applicability boundaries, and protocols lack procedures for stopping and verification. Over-complex modelling of damage completes this picture, when, instead of a minimally sufficient loss scheme, a cumbersome construct is introduced with numerous parameters that cannot be reliably estimated from data and cannot be maintained in a changing environment. In such circumstances, risk modelling loses practical footing, and the system reverts to intuitive decisions that are poorly reproducible and poorly controlled.

IV. CONCLUSION

The transition from evaluation via accuracy metrics to risk modelling naturally follows from the growing integration of intelligent systems into management and control loops, where significance attaches not only to agreement with a reference standard but also to how a prediction is converted into action, how rapidly an error is detected, and how extensive the consequences are for people, infrastructure, and the organisation. In this formulation, the object of analysis shifts to undesired events, their consequences, and designed mechanisms of prevention, containment, and recovery. Accuracy metrics retain their role, but their interpretation becomes context-dependent. In practice, metric values change substantially with shifts in class prevalence and with errors in reference annotation. High aggregate scores may coexist with low actual utility. Asymmetry of consequences renders threshold and criterion selection fundamental, because the cost of a false alarm and the cost of a miss generally differ. This shifts evaluation into the realm of expected losses and error cost functions, where comparisons of solutions are linked to impacts on damage and resource savings.

The article demonstrates that classical accuracy metrics describe agreement between predictions and

reference annotations under a fixed decision logic and therefore lose determinacy when the model score is used as an input to rules, constraints, and control procedures. Accuracy is sensitive to imbalance and threshold. Precision and Recall capture different facets of quality and depend directly on prevalence and the chosen threshold. F1 is convenient for comparison but obscures the trade-off between error types and provides limited insight into behaviour under threshold variation. ROC-AUC reflects ranking ability, yet the practical interpretation of its scalar value changes with event prevalence and consequence profiles. Against this background, the gap between pre-deployment evaluation and in-deployment behaviour becomes central, exacerbated by temporal and cross-site data shifts, changes in measurement practices, and limited availability of true labels. In clinical scenarios, it is particularly evident that drift can lead to degradation and to the risk of harmful interventions, necessitating shift monitoring and protective mechanisms at the system level. These conclusions are reinforced by the observation that, as imbalance increases, agreement among popular metrics diminishes, and the choice of indicator begins to change model-comparison outcomes.

A risk-oriented framework formalises quality in terms of the probability of undesired events, the severity of consequences, and controllability, which is determined by detectability, reversibility, and the presence of response procedures. In hybrid intelligent systems, risk emerges along the entire chain from data to action, as errors may arise in data, models, rules, integration logic, and usage protocols, and may be amplified by organisational delays and misallocation of trust between automation and human operators. Consequently, evaluation requires architectural description and mapping of decision paths. It must account for control points, stopping conditions, manual review, recovery speed, and the system's barrier organisation, in which constraints, routing of difficult cases, abstention from decisions under low reliability, and degradation monitoring keep behaviour within acceptable bounds. Traceability and explainability play a key role, as they provide verifiability, accelerate incident analysis, and help operators understand applicability limits.

Antipatterns are explicitly identified as: risk reduced to mere complication of familiar metrics without linkage to consequences; thresholds set arbitrarily under uncalibrated probabilities; the human and organisational loop remaining outside design; and damage models excessively parameterised and losing practical grounding. Against this backdrop, the methodological conclusion is that acceptance criteria for hybrid intelligent systems should shift from maximising aggregated accuracy metrics to constraining expected damage and ensuring risk controllability at the system level.

REFERENCES

[1] X. Zhang, F. T. S. Chan, C. Yan, and I. Bose, "Towards risk-aware artificial intelligence and machine learning systems: An overview," *Decision Support Systems*, vol. 159, p. 113800, May 2022, doi: https://doi.org/10.1016/j.dss.2022.113800.

[2] G. M. Foody, "Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient," *PLOS ONE*, vol. 18, no. 10, p. e0291908, Oct. 2023, doi: https://doi.org/10.1371/journal.pone.0291908.

[3] S. De Vos, T. Vanderschueren, T. Verdonck, and W. Verbeke, "Robust instance-dependent cost-sensitive classification," *Advances in Data Analysis and Classification*, vol. 17, pp. 1057–1079, Jan. 2023, doi: https://doi.org/10.1007/s11634-022-00533-3.

[4] C. K. Kolli, "Hybrid Neuro-Symbolic Models for Ethical AI in Risk-Sensitive Domains," *arXiv*, 2025, doi: https://doi.org/10.48550/arXiv.2511.17644.

[5] Z. Lu, I. Afridi, H. J. Kang, I. Ruchkin, and X. Zheng, "Surveying neuro-symbolic approaches for reliable artificial intelligence of things," *Journal of Reliable Intelligent Environments*, vol. 10, pp. 257–279, Jul. 2024, doi: https://doi.org/10.1007/s40860-024-00231-1.

[6] J.-G. Gaudreault and P. Branco, "Empirical analysis of performance assessment for imbalanced classification," *Machine Learning*, vol. 113, pp. 5533–5575, Jan. 2024, doi: https://doi.org/10.1007/s10994-023-06497-5.

[7] V. Subasri *et al.*, "Detecting and Remediating Harmful Data Shifts for the Responsible Deployment of Clinical AI Models," *JAMA Network Open*, vol. 8, no. 6, p. e2513685, Jun. 2025, doi: https://doi.org/10.1001/jamanetworkopen.2025.13685.

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.15

Integration of Artificial Intelligence Algorithms into On-Board Vehicle Diagnostic and Control Systems

Madugula Abhiram

Software Developer, Saanavi Technologies, Bloomington, Illinois, US

Received: 22 Feb 2026; Received in revised form: 21 Mar 2026; Accepted: 28 Mar 2026; Available online: 02 Apr 2026

Abstract *— The article examines architectural and engineering approaches for integrating artificial intelligence algorithms into on-board vehicle diagnostic and control systems. Relevance follows from software-defined vehicles, Electronic Control Unit (ECU) density, and the need to convert On-Board Diagnostics II (OBD-II) and Controller Area Network (CAN) telemetry into dependable decisions under real-time and safety constraints. Novelty consists of a synthesis that links data acquisition and preprocessing, model families for fault prediction, multi-fault diagnosis, and intrusion detection, and deployment patterns spanning in-vehicle compute with cloud/edge services. The objective is to systematize design decisions that yield robust diagnostics, interpretable outputs for service workflows, and controlled mitigation actions. Methods comprise comparative analysis of recent publications, taxonomy building across task types, and architectural decomposition of pipelines. The source base covers OBD-II machine learning, deep-learning predictive maintenance, attention-based fault prediction, explainable multi-fault diagnosis, CAN intrusion detection, connected-vehicle platforms, deterministic AUTOSAR Adaptive communication, and ML-aware functional-safety extensions. The article targets automotive software engineers, data scientists, and platform architects.*

Keywords *— **on-board diagnostics, OBD-II, CAN bus, ECU, fault diagnosis, predictive maintenance, intrusion detection, AUTOSAR Adaptive, functional safety, edge–cloud deployment.***

I. INTRODUCTION

Modern vehicles incorporate software functions that were traditionally implemented in dedicated hardware. This shift intensifies diagnostic complexity because faults emerge not only from component degradation but also from interactions among Electronic Control Units (ECU), network traffic, calibration states, and software updates. On-Board Diagnostics II (OBD-II) and Controller Area Network (CAN) provide standardized access to operational signals and diagnostic trouble codes, yet conventional rule-based diagnostics struggle with correlated or cascading failures and with adversarial conditions on in-vehicle networks. Recent studies show strong performance of deep learning for predictive maintenance from OBD streams, attention-based models for fault forecasting, and neural approaches for intrusion detection in CAN traffic, which motivates an integrated view that unifies diagnostic inference with control-oriented mitigation and with fleet-scale feedback loops for model governance and updates.

The purpose of the article is to develop an analytic synthesis of how AI algorithms can be embedded into on-board diagnostic and control pipelines without violating real-time, determinism, and safety expectations. The study solves three tasks:

1) to classify data sources and inference targets for on-board diagnostics, spanning Diagnostic Trouble

Codes (DTC) centric reasoning, time-series condition monitoring, and security anomaly detection;

2) to compare algorithm families and deployment patterns suitable for ECU- and vehicle-level constraints, including hybrid in-vehicle and cloud/edge arrangements;

3) to derive design recommendations that connect model outputs to actionable control or mitigation steps, while aligning the workflow with determinism and ML-related safety engineering practices.

Novelty is provided by a pipeline-centric consolidation that explicitly ties model choice, runtime placement, and safety-oriented lifecycle controls into one engineering narrative rather than treating them as separate topics.

II. MATERIALS AND METHODS

The material base for this article consists of recent peer-reviewed publications used to cover complementary layers of the problem—from embedded determinism to AI inference and connected-vehicle data infrastructure. D. Bellassai et al. analyzed deterministic publish/subscribe communication for AUTOSAR Adaptive, supporting real-time reasoning about service-oriented in-vehicle software stacks [1]. A. Errezgouny et al. studied predictive maintenance from OBD-derived time series using a hybrid Long Short-Term Memory (LSTM) and clustering strategy, informing model selection under sparse labeling [2]. M. N. Hossain et al. reviewed AI-driven vehicle fault diagnosis across subsystems, supporting the formation of a taxonomy for diagnostic targets [3]. P. Iyenghar et al. proposed ML-specific lifecycle phases and testing methods to extend ISO 26262-aligned assurance, shaping governance requirements for safety-relevant AI [4]. H. Jia et al. proposed an attention-based fault prediction model that captures fault correlations, enabling multi-fault reasoning beyond single-code diagnostics [5]. M. J. Khan et al. evaluated trade-offs in vehicle–edge–cloud deployments for deep models, informing placement decisions under latency constraints [6]. W. Lo et al. developed a hybrid deep intrusion detection approach for in-vehicle CAN traffic, supporting security integration within diagnostics [7]. E. T. Michailidis et al. reviewed OBD-II-based ML applications, consolidating signal types and use cases relevant to on-board diagnostic integration [8]. A. Mostefaoui et al. reported an in-production connected-vehicle big-data platform that informs fleet analytics and a feedback-loop architecture [9]. R. Rai et al. evaluated deep learning methods for CAN bus intrusion detection, supporting robustness considerations for cyber-physical threat detection [10].

For writing the article, comparative analysis, structured literature analysis, and architectural decomposition were applied. The approach emphasizes cross-source synthesis of constraints (latency, compute, determinism, safety assurance), mapping them to algorithm choices and to deployment topologies.

III. RESULTS

A consolidated interpretation of the literature supports viewing on-board AI diagnostics as a closed pipeline rather than as an isolated model embedded into an ECU. OBD-II and CAN telemetry jointly provide the observable layer: OBD exposes standardized parameters and DTCs, while CAN frames expose fine-grained inter-Electronic Control Unit (ECU) communication patterns. The OBD-focused review literature indicates that ML systems built on OBD signals are routinely used for anomaly detection, efficiency optimization, predictive maintenance, and broader driving support, with task feasibility depending on feature availability, sampling regimes, and data quality constraints typical for consumer-grade and service-grade telemetry streams [8]. From an integration standpoint, this implies that model design is inseparable from preprocessing choices such as synchronization, denoising, windowing, and handling of missing values, because these steps define what the embedded inference engine treats as a stable “state” for prediction.

Evidence from predictive maintenance research supports a pragmatic strategy for situations where labeled failure data are scarce. A hybrid LSTM-plus-clustering method applied to unlabeled OBD time series illustrates how temporal sequence learning can be coupled with unsupervised structure discovery to separate operational regimes and then predict condition signals with high reported accuracy, directly relevant to fleet settings where maintenance outcomes are sporadically recorded, and ground truth

arrives with a delay [2]. Attention-based fault prediction extends this idea by explicitly modeling relationships among fault events rather than treating faults as independent targets; the reported approach uses attention mechanisms (including graph attention) to capture inter-fault dependencies and forecast fault evolution from historical records [5]. When translated into on-board integration terms, these methods motivate a shift from "single DTC interpretation" to "fault graph evolution," where the AI component outputs ranked hypotheses about fault propagation and the expected following faults, enabling service workflows and on-board mitigations to prioritize interventions.

Multi-fault diagnosis in highly electronic vehicles introduces an additional integration pressure: diagnostic outputs are helpful only if they remain interpretable to technicians and consistent across operational conditions. An explainable hybrid deep model for multiple-fault diagnosis in automotive electronic systems explicitly positions interpretability as a practical requirement, pairing a high-performing fusion architecture with explainability mechanisms that render diagnostic logic in a human-consumable form [5]. This supports a design rule for on-board software teams: if diagnostic inference participates in control decisions or triggers maintenance actions, the output interface should expose not only a label, but supporting evidence (salient signals, temporal segments, confidence bands), because downstream modules—whether a service advisor interface or a safety supervisor—operate on justifications, not only predictions.

In-vehicle cybersecurity has matured into an inseparable part of diagnostics because compromised CAN traffic can masquerade as component failure and distort control loops. Deep intrusion-detection research on CAN traffic repeatedly highlights that CAN lacks built-in authentication and that attacks such as DoS, fuzzing, impersonation, and spoofing are feasible through physical or wireless entry points. A hybrid CNN–LSTM intrusion detection approach based on spatial–temporal representations reports very high detection results on benchmark car-hacking datasets, emphasizing automatic feature learning over manual engineering [7]. A complementary open-access study evaluates recurrent models (LSTM/GRU and variants). It demonstrates high detection performance across multiple datasets, reinforcing the feasibility of embedding anomaly-detection logic at the vehicle network edge [10]. For integration into diagnostic and control systems, these findings justify treating intrusion detection as a first-class diagnostic signal: a security classifier can gate, rate-limit, or quarantine specific message patterns before they contaminate control decisions or inflate false fault codes.

The deployment dimension links algorithm capacity to real-time constraints. Vehicle–edge–cloud analyses indicate that execution-time and accuracy trade-offs become deployment-defining variables, especially as model complexity increases and hardware diversity spans ECU microcontrollers, domain controllers, and external edge nodes [6]. In practice, the literature supports a layered placement strategy: inference steps that must act within tight latency budgets (e.g., detecting gross anomalies affecting safety-relevant functions) remain on-vehicle, while heavier analytics (fleet trend mining, model retraining, drift detection, and root-cause correlation across vehicles) sit in cloud/edge services [6, 9]. The connected-vehicle platform report, based on an in-production automotive big data stack, reinforces that large-scale ingestion, storage, and analytics can be organized into an integrated pipeline for telematics services, making fleet-scale feedback loops technically feasible rather than purely conceptual [9]. For an automotive software profile centered on Java and Google Cloud Platform, this architecture naturally maps to microservice-based ingestion, stream processing, and model serving patterns, while keeping the on-board component minimal and deterministic.

Determinism and safety assurance constrain how diagnostic outputs can influence control. Service-oriented in-vehicle platforms, especially AUTOSAR Adaptive, broaden software flexibility but introduce challenges for deterministic communication. Work on deterministic communication for AUTOSAR Adaptive proposes mechanisms to ensure determinism in publish/subscribe interactions, offering a concrete direction for integrating diagnostic messaging into safety-relevant software paths without unpredictable timing [1]. Safety engineering literature that extends ISO 26262 practices for ML emphasizes additional lifecycle phases—data preparation, ML training, and ML deployment—and

ties them to properties such as robustness, uncertainty handling, and interpretability, reflecting an assurance logic that complements classical code-centric safety processes [4]. The integration consequence is direct: an on-board AI diagnostic module cannot be treated as a static "component"; it becomes a managed asset with controlled updates, monitored drift, and evidence artifacts supporting its operational domain boundaries.

Figure 1 consolidates these results into a reference pipeline that connects in-vehicle data acquisition to AI inference, and then to control/mitigation actions and fleet-scale services. The figure is constructed as an author synthesis grounded in OBD-II ML application patterns, predictive maintenance modeling, intrusion detection in CAN, layered deployment practices, determinism in AUTOSAR Adaptive, and ML-aware safety lifecycle extensions [1, 2, 4, 6–10].

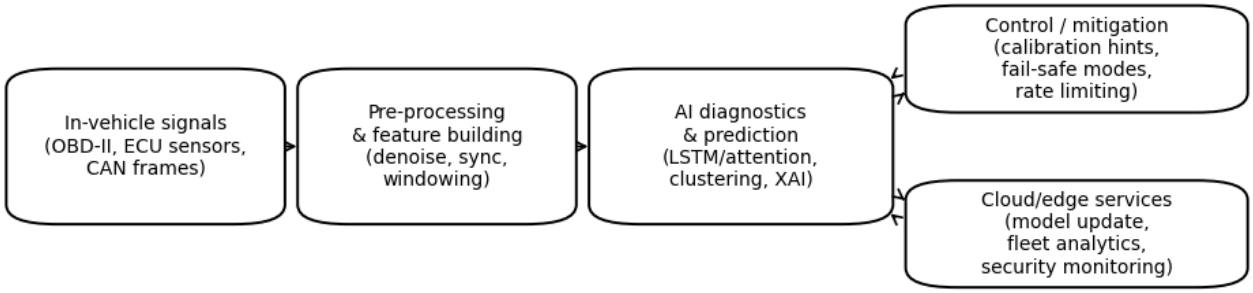


Fig.1. Reference pipeline for integrating AI diagnostics into on-board control loops

The synthesized pipeline supports a unified interpretation of AI-enabled on-board diagnostics as a closed technical loop that starts with standardized vehicle telemetry (OBD-II parameters, ECU sensor streams, CAN frames), proceeds through bounded preprocessing and feature construction, and terminates in two tightly coupled output channels: on-board mitigation and control-aligned recommendations constrained by deterministic execution and message timing, and fleet-scale cloud/edge services responsible for monitoring, model lifecycle operations, and security surveillance. This arrangement is consistent with recent evidence that deep sequential models and hybrid learning strategies can extract predictive maintenance value from OBD-derived time series under sparse labeling, that CAN-bus intrusion detection benefits from spatial–temporal deep representations suitable for near-real-time screening, and that safety-critical integration requires ML-specific lifecycle phases and testing practices layered onto functional-safety engineering, while deterministic communication patterns in AUTOSAR Adaptive provide an engineering basis for integrating diagnostic inference into service-oriented in-vehicle stacks without timing instability.

IV. DISCUSSION

The reviewed evidence supports treating "on-board diagnostics plus AI" as a system-of-systems problem: the diagnostic function is shaped by telemetry interfaces (OBD/CAN), by algorithm families tuned to data scarcity and temporal dependence, by runtime placement across vehicle and cloud/edge, and by governance for determinism and safety assurance [1, 4, 6, 8, 9]. A practical implication for automotive software teams working primarily with Java and Google Cloud Platform is that the most stable engineering boundary is not between "embedded" and "cloud," but between "latency-critical decisions" and "fleet-scale learning and governance." The former favors compact models, bounded execution, and deterministic messaging, while the latter favors scalable data platforms and controlled model lifecycle operations [1, 6, 9].

Table 1 organizes algorithm families by diagnostic objective and integration implications, highlighting where interpretability and data labeling pressure engineering choices.

Table 1. *AI algorithm families and integration implications for on-board diagnostics and control*

Objective	Representative approaches	Typical vehicle data	Integration implication
Predictive maintenance from OBD time series	Sequence models (LSTM) combined with unsupervised clustering [2]	OBD sensor time series, derived features [8]	Supports weak-label regimes; requires stable preprocessing and drift monitoring.
Fault prediction with correlated faults	Attention-based and graph-attention formulations [5]	Fault records, event sequences	Encourages modeling fault dependencies; output interfaces should expose ranked hypotheses and their confidence levels.
Multi-fault diagnosis with interpretability	Hybrid deep fusion with explainability mechanisms [5]	Electronic subsystem signals, fault samples	Enables technician-facing evidence; reduces ambiguity when multiple simultaneous anomalies occur.
CAN intrusion detection	Hybrid CNN–LSTM spatial–temporal detectors [7]; deep learning IDS variants [10]	CAN frames and traffic patterns	Security classification serves as a gating signal for diagnostics and control, requiring low-latency inference placement.

The platform layer benefits from separating the "model execution surface" into on-vehicle inference and cloud/edge management. Vehicle–edge–cloud studies explicitly frame execution time as a deployment driver and motivate layered pipelines that shift expensive operations away from the car when safety timing allows [6]. The connected-vehicle big-data platform experience indicates that production architectures already exist for large-scale telemetry ingestion and analytics, which can be repurposed for diagnostic model monitoring, retraining triggers, and fleet-level root-cause analysis [9]. Deterministic communication mechanisms for AUTOSAR Adaptive strengthen the feasibility of routing diagnostic outputs through service-oriented stacks without compromising timing predictability in safety-sensitive paths [1]. ML-aware safety lifecycle extensions support the governance side by treating model updates and validation as explicit engineering processes rather than ad hoc operational activities [4].

Table 2 translates these discussion points into an integration matrix spanning in-vehicle, edge, and cloud responsibilities, emphasizing where Java/GCP-aligned services naturally fit.

Two engineering tensions recur across the source base. First, model accuracy gains are not automatically convertible into operational value unless the output is interpretable and action-linked; explainable multi-fault diagnosis supports technician workflows, while predictive maintenance outputs gain value only when they map to maintenance scheduling or conservative control adjustments. Second, security and fault diagnosis intersect at the signal level: a compromised network can generate symptoms that resemble mechanical or electronic faults, making intrusion detection a diagnostic prerequisite rather than a separate security feature. Both tensions reinforce the use of a pipeline architecture that treats preprocessing, inference, and mitigation as a single controlled chain with explicit governance and deterministic communication where safety constraints apply.

Table 2. Integration matrix for AI-enabled diagnostics across vehicle, edge, and cloud layers

Layer	Primary responsibilities	Technical emphasis	Evidence basis
On-board (ECU/domain controller)	Low-latency anomaly screening; basic fault prediction; message gating/mitigation hooks	Bounded runtime, deterministic messaging, minimal model footprint	Deterministic AUTOSAR Adaptive communication [1]; CAN IDS feasibility on vehicle data [7,10]
Edge (near-vehicle compute)	Aggregation across short horizons; privacy-preserving buffering; selective offloading	Latency relief with localized compute; controlled bandwidth use	Vehicle–edge–cloud execution trade-offs [6]
Cloud (fleet platform)	Fleet analytics, drift detection, retraining pipelines, update orchestration, and historical correlation	Scalable ingestion and analytics services; lifecycle governance	Connected-vehicle significant data platform patterns [9]; ML lifecycle phases and testing methods for safety alignment [4]

V. CONCLUSION

The analysis supports three conclusions aligned with the stated tasks. First, OBD-II and CAN telemetry enable complementary diagnostic observability: OBD-centric signals and DTCs support condition monitoring and maintenance forecasting. In contrast, CAN traffic analysis supports the detection of abnormal inter-ECU communication that can corrupt diagnostic reasoning or control behavior. Second, the literature indicates that data realities and temporal structure drive algorithm selection: hybrid sequence learning with unsupervised components fits weak-label OBD regimes, attention-based modeling improves forecasting under correlated faults, and deep spatial–temporal detectors achieve strong CAN intrusion-detection performance on benchmark datasets. Third, robust integration depends on deployment and governance, not only on model choice: layered vehicle–edge–cloud placement aligns execution time and compute limits with fleet-scale learning, deterministic messaging mechanisms in AUTOSAR Adaptive support predictable diagnostic communication, and ML-aware safety lifecycle phases formalize data preparation, training, deployment, and testing activities required for safety-relevant AI functions.

REFERENCES

[1] Bellassai, D., Scordino, C., Casini, D., & Biondi, A. (2025). AP-LET: Enabling deterministic Pub/Sub communication in AUTOSAR Adaptive. *Journal of Systems Architecture, 162*, 103390. https://doi.org/10.1016/j.sysarc.2025.103390

[2] Errezgouny, A., Chater, Y., Barranco González, C. D., & Cherkaoui, A. (2025). An integrated deep learning approach for predictive vehicle maintenance. *Decision Analytics Journal, 16*, 100597. https://doi.org/10.1016/j.dajour.2025.100597

[3] Hossain, M. N., Rahman, M. M., & Ramasamy, D. (2024). Artificial intelligence-driven vehicle fault diagnosis to revolutionize automotive maintenance: A review. *CMES - Computer Modeling in Engineering and Sciences, 141*(2), 951–996. https://doi.org/10.32604/cmes.2024.056022

[4] Iyenghar, P., Gracic, E., & Pawelke, G. (2024). A systematic approach to enhancing ISO 26262 with machine learning-specific life cycle phases and testing methods. *IEEE Access, 12*, 179600–179627.

[5] Jia, H., Qian, D., Chen, F., & Zhou, W. (2025). Collaborative fusion attention mechanism for vehicle fault prediction. *Future Internet, 17*(9), 428. https://doi.org/10.3390/fi17090428

[6] Khan, M. J., Khan, M. A., Turaev, S., Malik, S., El-Sayed, H., & Ullah, F. (2024). A vehicle-edge-cloud framework for computational analysis of a fine-tuned deep learning model. *Sensors, 24*(7), 2080. https://doi.org/10.3390/s24072080

[7] Lo, W., Alqahtani, H., Thakur, K., Almadhor, A., Chander, S., & Kumar, G. (2022). A hybrid deep learning based intrusion detection system using spatial-temporal representation of in-vehicle network traffic. *Vehicular Communications, 35*, 100471. https://doi.org/10.1016/j.vehcom.2022.100471

[8] Michailidis, E. T., Panagiotopoulou, A., & Papadakis, A. (2025). A review of OBD-II-based machine learning applications for sustainable, efficient, secure, and safe

vehicle driving. *Sensors, 25*(13), 4057. https://doi.org/10.3390/s25134057

[9] Mostefaoui, A., Merzoug, M. A., Haroun, A., Nassar, A., & Dessables, F. (2022). Big data architecture for connected vehicles: Feedback and application examples from an automotive group. *Future Generation Computer Systems, 134*, 374–387. https://doi.org/10.1016/j.future.2022.04.020

[10] Rai, R., Grover, J., Sharma, P., et al. (2025). Securing the CAN bus using deep learning for intrusion detection in vehicles. *Scientific Reports, 15*, 13820. https://doi.org/10.1038/s41598-025-98433-x

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.16

Modern Architectural Patterns for Ensuring Security in Cloud-Native Applications

Praveen Ravula

Software Engineer at Amazon, Arlington, VA - USA

Received: 17 Mar 2026; Received in revised form: 15 Apr 2026; Accepted: 18 Apr 2026; Available online: 23 Apr 2026

Abstract *— The article presents a theoretical and methodological analysis of architectural patterns for ensuring security in distributed applications operating in cloud environments. The study integrates microservice protection models, access control mechanisms, methods for identifying architectural vulnerabilities, and configuration-analysis approaches into a unified framework for designing secure cloud-native systems. It is shown that the structure of modern applications is defined by dynamic component orchestration, distributed trust boundaries, and a highly variable environment, which together form a specific set of threats and constraints. The paper proposes a conceptual model of multilayer protection that includes network segmentation, privilege separation, configuration-integrity control, cryptographically reinforced trust mechanisms, and automated anomaly detection. This model is considered as a foundation for aligning the behavior of services, the container runtime, and security policies. The study establishes that insufficient formalization of architectural decisions leads to risk-assessment bias and reduced system predictability. Promising directions for advancing architectural security are substantiated, including the integration of telemetry-driven analysis, adaptive recovery mechanisms, enhanced privilege-management models, and formalized maturity metrics. The article may be useful for architects, engineers, and researchers involved in designing secure distributed applications under conditions of high automation and regulatory pressure.*

Keywords *— **architectural security, microservices, access control, configuration vulnerabilities, multilayer protection.***

I. INTRODUCTION

The widespread adoption of cloud environments, containerization, and modular architectures has led to a transition toward systems built from a multitude of small components. This structure enhances flexibility through scaling and automated recovery, but simultaneously expands the attack surface due to the growth of interaction points and the increasing complexity of network connections. Under these conditions, security becomes an intrinsic property of the architecture.

As distribution increases, ensuring trust between components becomes paramount: access control, protection of transmitted data, secrets management, and adherence to the principle of least privilege [3]. Frequent configuration changes and automated deployment increase the likelihood of errors capable of leading to leaks and integrity violations.

Practice is shifting toward formalized architectural solutions that ensure protection at the design level: unified entry points, multilayer privilege separation, verified cryptographic methods, network rule control, automated configuration analysis, and the identification of architectural defects.

The aim of the study is to systematize modern architectural solutions for ensuring the security of distributed applications, identify their role in forming a resilient trust model, and analyze how containerization, a dynamic runtime environment, and a modular structure alter the requirements for building secure systems.

The scientific novelty of the research lies in the integration of disparate approaches to security into a unified architectural framework covering infrastructure, business logic, and component interaction. It is shown that the application of verified cryptographic solutions, the formalization of entry points, distributed authorization, strict network rules, automated setting analysis, and the identification of architectural errors form a complementary complex of measures capable of maintaining system stability even under a high pace of change.

The research hypothesis posits that the coordinated use of architectural approaches to security ensures higher predictability, manageability, and resilience of distributed applications compared to individual, fragmented measures.

The scope of the research covers distributed software complexes deployed in cloud environments and built from numerous interacting components. Particular attention is paid to highly automated systems utilizing containerization, a modular structure, and centralized runtime management tools, which allows security to be viewed as a result of architectural decisions rather than a collection of separate technical techniques.

II. MATERIALS AND METHODS

The methodological basis of the study is formed by the systematization of architectural security models covering the protection of cloud and container environments, microservice applications, and access control mechanisms. Works from 2021–2025 devoted to the evolution of architectural patterns, risk analysis of distributed systems, and methods for increasing application resilience through formalized trust mechanisms serve as the source corpus.

In the study by Arif et al. [1], key directions for protecting distributed applications are examined – enhancing confidentiality, trust management, and service interaction control – which form the basis of architectural security in cloud systems. Rahaman et al. [8] summarize methods for the static analysis of configurations and components, allowing for the early detection of vulnerabilities and the assessment of the correctness of design decisions. In the work of Theodoropoulos et al. [9], characteristic threats to cloud services and countermeasure mechanisms are systematized, setting a conceptual framework for risk analysis. The research by El Akhdar et al. [3] complements the methodology with aspects of microservice protection in environments with high dynamics and complex topology. A significant element of the methodological base is the analysis of architectural defects. Dell'Immagine et al. [2] present a tool for identifying insecure configuration patterns, and the review by Ponce et al. [7] offers a typology of architectural errors and measures for their elimination. These works allowed for the formation of a classification of "smells" and their linkage to security policy violations. The comparison of access control models is based on the results of Venčkauskas et al. [10], where the differences between centralized and decentralized authorization schemes and the characteristics of token mechanisms are experimentally demonstrated. An important block of the methodology is related to research on migration to cloud architectures. Nascimento et al. [5] reveal the impact of architectural stack changes on the availability, scalability, and security of services. The work of Falcão et al. [4] provides an analysis of secure deployment models and their influence on resilience. Early detection of deviations and attacking actions is included in the methodology thanks to the AIDS model by Park et al. [6], which allowed for the combination of architectural mechanisms with anomaly detection methods.

The research methodology is based on a combination of structural analysis of architectural models, a comparative review of access control mechanisms, the typologization of architectural defects, and the synthesis of conclusions regarding the relationship between system configuration and resilience. This approach provides the opportunity to view security as a property of the architecture, rather than as a collection of disjointed technical measures, and forms the basis for analyzing modern protection patterns for cloud applications.

III. RESULTS

An analysis of the infrastructure characteristics of containerized applications revealed that the transition from monolithic deployments to modular orchestration ensures higher system

resilience and behavioral predictability under load. The study by Nascimento et al. [5] experimentally confirmed the ability of Kubernetes-based clusters to maintain the stability of computational and network resources during changes in request intensity, which makes such architectures pivotal for ensuring continuous service availability. The structure of the observed indicators is illustrated in Figure 1, which depicts the multilayer interaction of infrastructure, orchestration, and security mechanisms in the applied cloud-native configuration.

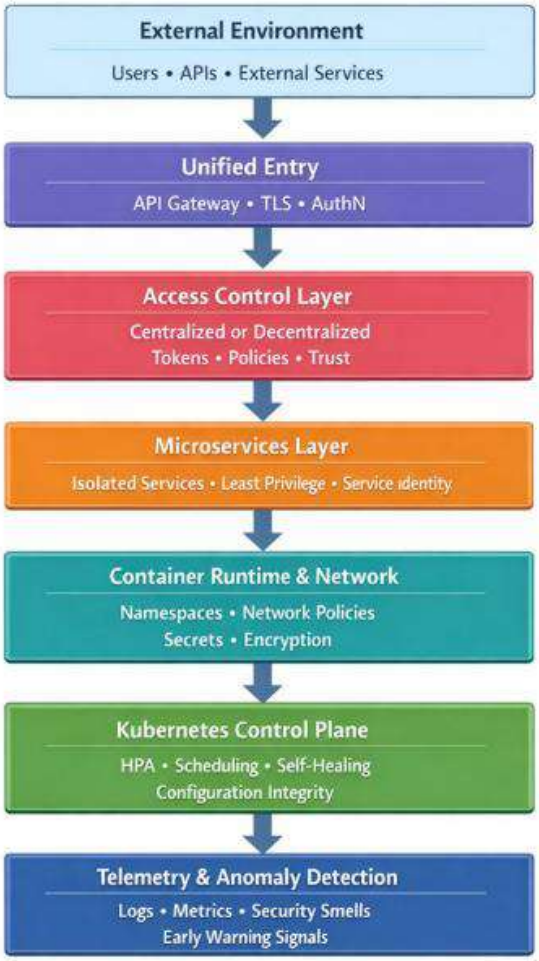


Fig.1 – Multilayer security architecture illustrating infrastructure mechanisms of availability, scalability, and network load management for containerized applications during Kubernetes migration (Compiled by the author based on sources: [3, 5, 6]).

Observations illustrated in Figure 1 reveal fundamental features of the behavior of distributed applications in a cloud environment. The uniformity of node loading and the absence of sharp memory fluctuations confirm the effectiveness of built-in scheduling mechanisms, which ensure the smoothing of local overloads. This effect corresponds to the general conclusions of Rahaman et al. [8], linking architectural resilience to the correct distribution of computing resources. The automatic recovery of cluster components confirms the presence of built-in self-healing mechanisms in Kubernetes, which fundamentally distinguishes it from traditional monolithic environments.

Network dynamics demonstrate that the growth of inbound traffic does not lead to a violation of service availability, which aligns with the results of Arif et al. [1], where the role of network isolation and managed namespaces as basic means of increasing resilience is emphasized. Simultaneously, the controlled nature of outbound traffic indicates the correct integration of platform optimization tools, as described in the study by Falcão et al. [4] regarding secure deployment models.

Automatic horizontal scaling confirms the architecture's ability to adapt to load changes without operator intervention. As shown in the work of El Akhdar et al. [3], it is precisely such self-organizing mechanisms that form the foundation for the further complication of microservice systems and the expansion of their functionality.

From a security perspective, the obtained results allow for the assertion that the combination of namespaces, access control mechanisms, and secret management structures creates a stable basic trust perimeter. This aligns with the conclusions of Theodoropoulos et al. [9], who systematized the requirements for service security in distributed cloud environments. The aggregate of the presented observations confirms that the transition to container orchestration forms a qualitatively different level of infrastructure manageability, where resilience, scalability, and security act not as isolated properties, but as interconnected elements of the architectural solution.

An analysis of distributed application architectures indicates that the methods of organizing access control determine the technical parameters of the system and its resilience to failures and attacks. In the study by Venčkauskas et al. [10], it is experimentally demonstrated that differences between centralized and decentralized authorization

schemes manifest in load profiles, throughput, and the nature of vulnerabilities. These data allowed for the comparison of the architectural distribution of identification and authority confirmation functions with real performance indicators. Table 1 systematizes the key indicators of two access control models based on experimental measurements.

Table 1 – Comparison of centralized and decentralized access control in microservice (Compiled by the author based on sources: [4, 7, 10])

Criterion	Centralized model	Decentralized model
CPU (low load)	Almost 2× lower	Significantly higher
CPU (medium/high load)	~5% lower	~5% higher
RAM (non-optimized)	Lower	Higher
Requests to Authorization Service	Much fewer	+100,000 (≈40% of internal)
Final score (low load)	8.27	–
Final score (medium load)	6.94	–
Security score	–	7.77
Throughput	Higher	Lower
Resilience	SPOF possible	Higher resilience

Architectural comparison shows that the centralized access control scheme provides a more predictable load profile on computing resources, which is particularly noticeable at low request intensity. With this approach, authority confirmation is performed in a single component, which reduces the number of calls between services and decreases the total volume of internal traffic. Experimental values confirm a reduction in processor load by approximately half during low system utilization [7], which indicates the efficiency of centralization from a resource consumption perspective. However, the concentration of the authorization function in a single node forms a potential single point of failure, which reduces resilience.

The decentralized scheme demonstrates opposite properties. The growth in the number of requests to the authorization service leads to a noticeable increase in computational costs. Nevertheless, the logic of distributed verification itself contributes to strengthening architectural trust: each service independently validates requests, which eliminates dependence on a single node and increases resilience to failures. The integral resilience indicator highlighted in the study confirms the increase in the security level precisely due to the distribution of responsibility [10].

A comparison of the models allows for the assertion that access control in microservice architectures represents an applied protection task and a fundamental element of distributed system design. Centralized control ensures resource economy and behavioral predictability but amplifies architectural risks. The decentralized approach increases the load but creates a more robust trust model, reducing the impact of failures of individual components.

Thus, the obtained differences allow for the formulation of architectural recommendations taking into account the requirements of a specific environment. Prioritizing performance justifies centralized solutions, while elevated requirements for resilience and security reasonably shift the choice in favor of decentralized schemes.

IV. DISCUSSION

An analysis of experimental and theoretical data indicates that the choice of access control architecture inevitably forms its own structural limitations, which must be considered when designing secure cloud systems. The problem of crossing trust boundaries comes to the fore. In centralized models, the entire flow of requests relies on a single decision-making point, which creates a concentration of risks and substantially amplifies the consequences of any configuration error or short-term unavailability of the authorization service. The study by Venčkauskas et al. [10] shows that a central element is capable of ensuring high throughput; however, it is this element that becomes the convergence point of all

trust connections, and a disruption in its correct operation leads to the degradation of the entire system.

With the transition to distributed authorization, the situation changes: responsibility is dispersed across a multitude of components, which reduces the probability of a simultaneous failure and creates a more resilient trust model. However, this leads to a growth in network activity and an increase in computational load, as each microservice is forced to independently contact the rights confirmation system.

Systems oriented toward high performance prove to be more sensitive to additional latencies, while architectures designed for multi-stage authority verification are less vulnerable to local failures. Under conditions of diverse service load models, a single universal option cannot be identified. If predictable and uniform requests dominate, the central model can ensure a balance of speed and control. If the structure of requests is dynamic, and fault tolerance is a priority, the distributed architecture proves to be more aligned with security requirements.

The influence of the network structure deserves separate attention. In centralized schemes, the crossing of trust boundaries often contracts to a single route where the rights verification logic is concentrated. In distributed configurations, these boundaries spread throughout the entire system, which facilitates incident localization but complicates observation and analysis. Practical observations presented in the works of Arif et al. [1] and El Akhdar et al. [3] confirm that the complication of topology reinforces the importance of correct service classification, strict separation of responsibility zones, and control over interaction channels.

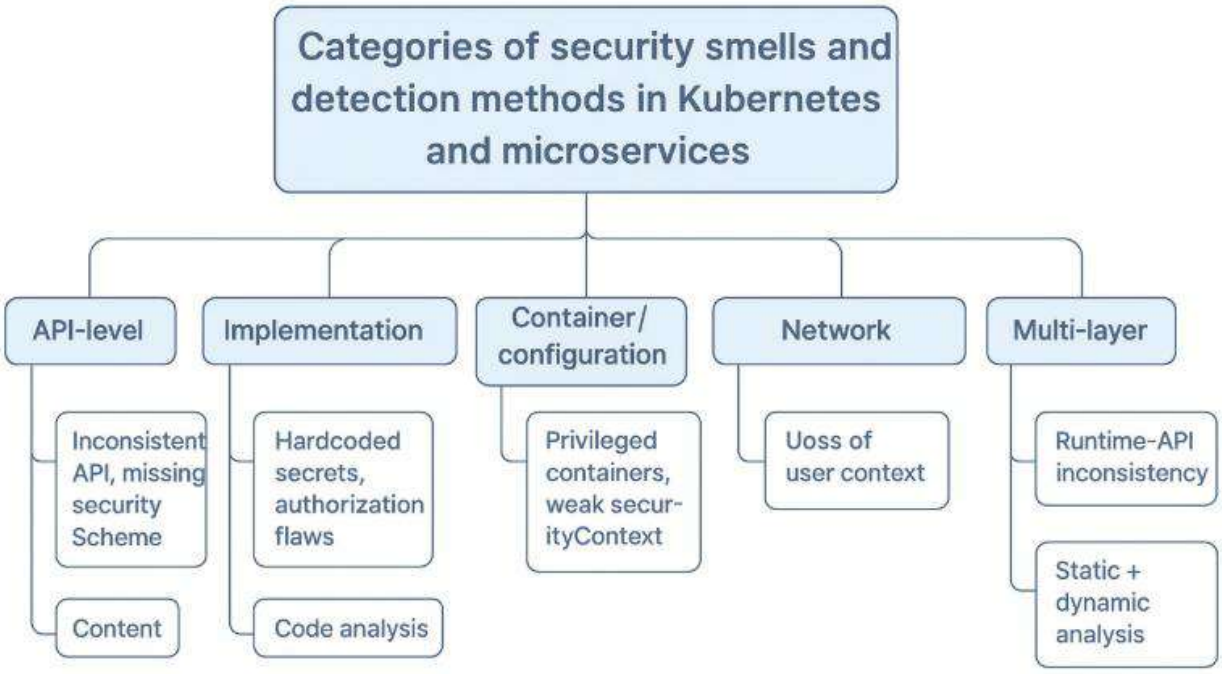


Fig.2 – Multilayer security smells and detection mechanisms in Kubernetes-based microservice architectures (Compiled by the author based on sources: [2, 6, 9])

At the design approach level, the key becomes not the choice of a specific scheme per se, but the understanding that the access control architecture determines the entire system configuration. Its reaction to errors, the nature of the network load, requirements for computing resources, and the degree of manageability of inter-service interaction. Without accounting for these factors, it is impossible to form a resilient security model oriented toward the long-term development of distributed infrastructure. The decision to apply a centralized or distributed model must be based on the load profile, the anticipated intensity of changes, permissible risks, and strategic architectural goals – only with such a formulation is it possible to achieve a balanced level of protection,

confirmed by empirical data and consistent with general patterns reflected in scientific research.

The design error categories presented in the study by Dell'Immagine et al. [2] demonstrate that vulnerabilities in distributed environments form not as isolated defects, but as a consequence of accumulated architectural inconsistencies. Such inconsistencies manifest at the levels of interfaces, implementation, container environment, network policy, and observability, forming a collection of signals indicating a hidden weakness in the architecture. Before proceeding to the analysis, Figure 2 illustrates the propagation of security smells across architectural layers and highlights the corresponding detection mechanisms.

The architectural interpretation of the listed categories shows that vulnerability signals are not reducible to singular configuration defects. Their appearance reflects a divergence between the design intent and the actual behavior of the system. Thus, the identification of inconsistency in application programming interfaces indicates a gap between specification and implementation. Such a gap complicates the verification of access rights correctness and renders the boundaries of component responsibility undefined. The detection of hardcoded secrets or errors in authority verification demonstrates the absence of mature development processes and weak control over the data lifecycle, which correlates with the threat analysis presented by Theodoropoulos et al. [9].

Vulnerability signals associated with the container and configuration environment are interpreted as the result of insufficient control over execution parameters. Privileged containers or weakened launch parameters show that the actual distribution of rights exceeds the limits of the principle of least privilege. In a distributed computing environment, such deviation creates prerequisites for the vertical and horizontal propagation of attacks. This link aligns with the conclusions of Arif et al. [1], where the importance of strict regulation of interactions between services and control of their authorities is emphasized.

Network vulnerability signals demonstrate the tension between infrastructure flexibility and the requirement for clear data flow isolation. The absence of encryption or the presence of weakened traffic processing policies leads to uncertainty in network behavior. Such uncertainty reduces the accuracy of matching the actual configuration with the architectural model and decreases the reliability of deviation detection mechanisms. This corresponds to the conclusions of Rahaman et al. [8] that the correct interpretation of configurations is the foundation of static security analysis.

Particular interest is drawn to multi-layer vulnerability signals, as they indicate systemic inconsistencies across several levels simultaneously – the interface, execution environment, access policy, and network interaction rules may diverge from one another. The diagnostic value of such signals lies in the fact that they reveal precisely architectural contradictions that remain unnoticeable during the analysis of individual levels. In this sense, they act as precursors to complex vulnerabilities and necessitate a correction of the system's design logic, rather than just changes to individual parameters. The review by Ponce et al. [7] confirms that such inconsistencies are early indicators of future security breaches and require a systemic revision of architectural decisions.

Thus, matching these signals with threat analysis models shows that they can serve as empirical markers of those zones in the system where initial design assumptions diverge from real interaction mechanisms. Such a reading allows them to be used as a diagnostic tool and a basis for revising the principles of building distributed systems. In aggregate, the indicated categories of signal signs form an analytical framework allowing for the identification of hidden dependencies between system levels and the explanation of the origin of complex vulnerabilities characteristic of modern cloud architectures.

V. CONCLUSION

The obtained results confirm that security in modern distributed applications is formed not by a set of separate technical tools, but by a coordinated architecture in which infrastructure, access control, and the identification of design errors are linked. Within this framework, trust between components becomes a conscious object of design, rather than a side effect of platform choice.

The analysis of the transition from monolithic solutions to the managed launch of numerous small services showed that mechanisms of automatic scaling, recovery, and strict environmental separation fulfill the role of the main supporting contour of resilience. However, this contour remains reliable only when requirements for data protection and rights management are embedded in the architecture from the outset, rather than added retroactively after incidents.

A comparison of centralized and distributed access control schemes showed that the choice of model determines the volume of resources, latencies, and the nature of risks. The centralized approach is appropriate when performance and controlled single-point-of-failure threats are prioritized, whereas distributed authority verification is justified given elevated requirements for the survivability and independence of individual parts of the system.

The classification of characteristic signs of insecure architecture at the levels of interfaces, implementation, execution environment, network, and observability demonstrates that such signals can be used as a tool for the early diagnosis of discrepancies between intent and actual configuration. The inclusion of these signals in design and maintenance processes allows for a shift from one-time fixes to the purposeful correction of architectural decisions, cementing security as a sustainable property of the developing system, rather than a sum of disjointed protection measures.

REFERENCES

[1] Arif, T., Jo, B., & Park, J. H. (2025). A comprehensive survey of privacy-enhancing and trust-centric cloud-native security techniques against cyber threats. Sensors, 25(8), 2350. https://doi.org/10.3390/s25082350

[2] Dell'Immagine, G., Soldani, J., & Brogi, A. (2023). KubeHound: Detecting microservices' security smells in Kubernetes deployments. Future Internet, 15(7), 228. https://doi.org/10.3390/fi15070228

[3] El Akhdar, A., Baidada, C., Kartit, A., Hanine, M., García, C. O., Lara, R. G., & Ashraf, I. (2024). Exploring the potential of microservices in Internet of Things: A systematic review of security and prospects. Sensors, 24(20), 6771. https://doi.org/10.3390/s24206771

[4] Falcão, E., Silva, F., Pamplona, C., Melo, A., Asadujjaman, A. S. M., & Brito, A. (2025). Confidential Kubernetes deployment models: Architecture, security, and performance trade-offs. Applied Sciences, 15(18), 10160. https://doi.org/10.3390/app151810160

[5] Nascimento, B., Santos, R., Henriques, J., Bernardo, M. V., & Caldeira, F. (2024). Availability, scalability, and security in the migration from container-based to cloud-native applications. Computers, 13(8), 192. https://doi.org/10.3390/computers13080192

[6] Park, H., EL Azzaoui, A., & Park, J. H. (2025). AIDS-based cyber threat detection framework for secure cloud-native microservices. Electronics, 14(2), 229. https://doi.org/10.3390/electronics14020229

[7] Ponce, F., Soldani, J., Astudillo, H., & Brogi, A. (2021). Smells and refactorings for microservices security: A multivocal literature review. arXiv. https://doi.org/10.48550/arXiv.2104.13303

[8] Rahaman, M. S., Islam, A., Cerny, T., & Hutton, S. (2023). Static-analysis-based solutions to security challenges in cloud-native systems: Systematic mapping study. Sensors, 23(4), 1755. https://doi.org/10.3390/s23041755

[9] Theodoropoulos, T., Rosa, L., Benzaid, C., Gray, P., Marin, E., Makris, A., Cordeiro, L., Diego, F., Sorokin, P., Di Girolamo, M., Barone, P., Taleb, T., & Tserpes, K. (2023). Security in cloud-native services: A survey. Journal of Cybersecurity and Privacy, 3(4), 758–793. https://doi.org/10.3390/jcp3040034

[10] Venčkauskas, A., Kukta, D., Grigaliūnas, Š., & Brūzgienė, R. (2023). Enhancing microservices security with token-based access control method. Sensors, 23(6), 3363. https://doi.org/10.3390/s23063363

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.17

Integration of ISO 14971 Requirements into the Design Processes of Class III Life-Sustaining Medical Devices

Sandeep Reddy Koppula

Senior Quality Engineer-Senseonics Holdings, Germantown, Maryland
ksandeepreddy3@gmail.com

Received: 20 Mar 2026; Received in revised form: 14 Apr 2026; Accepted: 19 Apr 2026; Available online: 23 Apr 2026

Abstract *— This article examines how ISO 14971 requirements can be embedded into the design processes of Class III life-sustaining medical devices as an operating design discipline. The subject is relevant because implantable and life-sustaining systems combine severe clinical consequences of failure with complex interactions among materials, software, mechanical components, and use conditions. The aim is to develop an analytical model for integrating risk management into design planning, evidence generation, and lifecycle control. The study relies on recent scientific and regulatory publications focused on high-risk devices, implantable systems, quality system regulation, post-market surveillance, cybersecurity, and clinical evaluation. Comparative analysis, conceptual synthesis, and analytical generalization were used. The analytical part shows that effective integration requires early hazard framing, traceability between risk controls and verification evidence, and continuous transfer of post-market signals into design change logic. The proposed interpretation has practical value for quality engineering, design reviews, and regulatory readiness of Class III medical technologies.*

Keywords *— **ISO 14971, Class III medical devices, life-sustaining devices, design controls, risk management, implantable devices, regulatory compliance, verification and validation, post-market surveillance, reliability engineering.***

I. INTRODUCTION

The design of Class III life-sustaining medical devices cannot be reduced to technical optimization alone. In this product category, every design decision has safety, verification, and regulatory consequences. The tighter the physiological coupling between device and patient, the less tolerance remains for fragmented development logic. For implantable and life-sustaining technologies, risk management must be integrated into the design process at the point when intended use, user need, and system architecture are first defined.

This article aims to develop an analytical framework for integrating ISO 14971 requirements into the design processes of Class III life-sustaining medical devices.

The first research objective is to determine where ISO 14971 requirements should be inserted into the design process so that risk management shapes design planning, inputs, outputs, and design review activity. The second research objective is to identify how the specific properties of high-risk implantable and life-sustaining devices change the content of hazard analysis and the structure of verification and validation evidence. The third research objective is to formulate an implementation logic that connects premarket risk controls with post-market signal capture and design change management.

The study develops an original analytical model for embedding ISO 14971 into the full design logic of Class III life-sustaining medical devices. Its novelty lies in treating risk management as a design-

driving and evidence-forming mechanism that connects pre-market controls, verification and validation evidence, post-market surveillance findings, and design change decisions within a single lifecycle structure.

II. MATERIALS AND METHODS

The source base combined recent systematic reviews, narrative reviews, retrospective cohort data, regulatory analyses, and current FDA (Food and Drug Administration) regulatory texts relevant to Class III and other high-risk medical devices [14]. Screening focused on publications from 2023 to 2026 that addressed one or more of the following domains: integration of risk management into quality systems and design controls, clinical evaluation and regulatory oversight of high-risk devices, implantable-device engineering constraints, reliability and long-term stability, cybersecurity in benefit-risk evaluation, and post-market surveillance for complex medical technologies. The selected literature covers three interconnected layers of inquiry: regulatory harmonization and evidence expectations [3; 4; 6; 7; 11], risk expansion in implantable and life-sustaining technologies [2; 5; 8; 10; 12; 13], and lifecycle feedback from clinical evidence and post-market observation [1; 9; 12].

Methods. The article uses comparative analysis to align regulatory and engineering interpretations, source analysis to extract recurring design obligations, conceptual synthesis to connect risk management with design control checkpoints, typologization to structure major integration points, and analytical generalization to formulate an implementation model suitable for Class III life-sustaining devices.

III. RESULTS

The recent regulatory trajectory makes one point difficult to ignore that risk management is increasingly built into the quality architecture that governs design itself. The regulatory change that became operational on February 2, 2026, was significant because the FDA's Quality Management System Regulation incorporated ISO 13485:2016 by reference and explicitly positioned risk management within the governing quality framework for device design and development [3; 4]. For Class III life-sustaining devices, this changes the practical meaning of design control. Risk analysis is no longer an accessory to design planning. It becomes one of the mechanisms through which planning decisions are justified, staged, reviewed, and translated into objective evidence.

That regulatory shift matters because the design process for a life-sustaining implant does not fail at a single point. Failure can arise from material degradation, energy instability, interface drift, software behavior, infection pathways, thrombogenic or inflammatory responses, use errors, or weak clinical evidence regarding long-term performance [2; 8; 10; 12; 13]. When ISO 14971 is inserted too late, these problems are discovered as testing surprises or review deficiencies [15]. When it is inserted early, it works differently. It helps define which design inputs are safety-critical, which interfaces require tighter tolerances, which verification methods need physiological realism, and which residual risks must remain visible during review and labeling decisions [2; 6; 8; 13].

A consistent pattern across recent publications is that implantable and life-sustaining devices expand the hazard field beyond conventional component failure. Reviews on vascular implants, implantable sensors, and long-term bioelectronic reliability converge on the same engineering reality: device safety in vivo depends on interactions between hardware, biological environment, time, and clinical use conditions [2; 10; 13]. Material choice influences corrosion behavior, encapsulation stability, inflammatory response, and signal integrity [8; 10]. Mechanical architecture influences fatigue, sealing, lead durability, and interface strain [2; 8]. Miniaturization and sensing sophistication create additional exposure to drift, power limitations, packaging defects, and failure modes that may not appear under simplified bench assumptions [10; 13]. In such systems, hazard analysis cannot be limited to component catalogs. It has to model functional degradation pathways over time.

This point becomes even sharper when evidence from different high-risk domains is compared. The systematic review of CE-marked high-risk glucose-management devices found clinically meaningful benefits, yet it simultaneously highlighted short study duration and methodological

heterogeneity in the evidence base [1]. The quantitative study on Medical Device Regulation clinical evaluation burden reached a related conclusion from another angle, showing that manufacturers face substantial pressure when stronger high-risk evidence expectations meet limited internal readiness for expanded evaluation requirements [7]. The review of medical device trials in the European Union notes that a risk management plan aligned with ISO 14971 is expected within the trial oversight logic, which places design assumptions, clinical investigation strategy, and benefit-risk framing into a single evidentiary chain [11]. Taken together, these publications suggest that for Class III life-sustaining devices, ISO 14971 integration is inseparable from evidence strategy. A weak connection between hazard definition and clinical investigation design does not stay confined to documentation. It weakens the credibility of the entire safety case for the device [1; 7; 11].

A second cross-source pattern concerns the movement from static hazard lists to dynamic risk architecture. The cybersecurity review demonstrates that regulators increasingly expect connected-device risks to be incorporated into benefit-risk evaluations, yet practical guidance on how to do so remains uneven across jurisdictions [5]. That observation matters far beyond software-only products. Many contemporary life-sustaining devices operate within broader digital ecosystems that involve telemetry, wireless configuration, or data-dependent clinical decision pathways. In these circumstances, cybersecurity is not external to patient safety. It becomes a route through which safety, performance, and trustworthiness can degrade [5]. ISO 14971 integration into design has to account for threat-driven failure initiation.

The same logic applies to post-market information. The empirical analysis of high-risk device surveillance across regulatory data sources shows that hardware modifications are more effective at reducing recurrence than software-only corrective actions. At the same time, software-driven issues remain persistent in some product groups [9]. The retrospective cohort study on cardiac implantable electronic device infections reinforces the value of real-world signal capture by showing that knowledge of risk factors can inform decisions about device type, upgrades, replacements, and prophylactic measures [12]. From a design perspective, post-market surveillance should be treated as a continuing source of evidence that tests whether validated safety assumptions remain stable under real clinical use. It reveals whether the original risk controls were robust enough under actual clinical variability, implant duration, maintenance patterns, and patient selection [9; 12].

At the design stage, this leads to a practical conclusion. ISO 14971 should be mapped onto design controls as a sequence of decision gates, not as a file created in parallel. During design planning, risk management defines the scope of critical interfaces, foreseeable misuse, and essential performance boundaries. During design input development, it forces safety claims into measurable requirements. During design output generation, it makes visible which outputs are true risk controls and which are merely enabling features. During verification, it tests whether risk controls operate as intended under relevant technical and environmental conditions. During validation, it examines whether user interaction, clinical workflow, and real-use conditions preserve the intended safety margin. During design transfer and change control, it prevents production adaptation or late modification from undermining earlier control assumptions [3; 4; 6; 11].

The literature on implantable technologies helps explain why this gate-based interpretation is preferable to linear compliance mapping. Reviews of implantable sensors and bioelectronic medicine repeatedly emphasize that long-term performance depends on stability over time [10; 13]. A design that passes early verification can still fail clinically if packaging permeability, biofouling, material fatigue, signal degradation, or tissue response were underestimated during hazard framing [8; 10; 13]. For Class III life-sustaining devices, residual risk evaluation has to be anchored in temporal realism. It should ask whether the control remains effective after implantation-related stress, prolonged exposure to body fluids, repeated use cycles, or clinically plausible maintenance scenarios.

Figure 1 summarizes this integration logic by adapting the challenge structure reported for implantable sensors and translating it into a design-

process view relevant to Class III life-sustaining devices [13].

Fig. 1. Integration logic for embedding ISO 14971 into the design process of Class III life-sustaining devices (author's analytical reconstruction adapted from the challenge typology in [13])

A further implication follows from the convergence of regulatory and engineering literature. ISO 14971 integration works best when traceability is bidirectional. Design inputs should point to hazards, but hazards should also point forward to verification evidence, validation rationale, and post-market monitoring targets. Without that structure, risk files become archival. With it, they become operational. For Class III life-sustaining devices, that distinction affects review readiness, investigation design, and patient safety simultaneously.

IV. DISCUSSION

The literature supports a stricter interpretation of ISO 14971 integration than the one often seen in routine development practice. For Class III life-sustaining devices, risk management should not be positioned as a review package compiled around milestones. It should function as the design grammar that determines what the team treats as a requirement, what it accepts as evidence, and what it escalates for redesign. A useful implementation model begins with one decision: every major design artifact must answer a risk question. If a document, review, or test does not change the understanding of hazardous situations, control effectiveness, residual risk, or monitoring need, its place in a high-risk development program should be reconsidered.

The first operational step is to connect the intended use to a structured hazard map before architecture freeze. At this stage, teams need more than a conventional failure list. They need a layered description of clinical-harm pathways, component-level initiating events, interface vulnerabilities, foreseeable use deviations, and degradation mechanisms that emerge only over time. The purpose of Table 1 is to show how ISO 14971 can be distributed across design control checkpoints so that risk activity changes the content of development, not merely its documentation.

The comparison in Table 1 shows that integration depends less on the number of documents and more on timing, ownership, and test relevance. The common weakness in high-risk programs is not the absence of a risk file. It is the delayed conversion of hazards into design-driving requirements. Once that delay appears, verification becomes generic, validation becomes defensive, and design reviews lose their discriminating function.

A second implementation issue concerns what should be monitored after the design freeze. Life-sustaining implants operate in changing biological and clinical environments. That means residual risk cannot be managed solely through complaint handling. Monitoring has to be built around leading indicators that reveal weakening control effectiveness before catastrophic harm accumulates. Table 2 presents a monitoring structure suited to this task.

Table 1. Operational integration of ISO 14971 within the design controls of Class III life-sustaining devices [1–10]

Design control stage	ISO 14971 activity	Primary output	Main decision question	Escalation trigger
Design planning	Risk management planning and initial hazard framing	Risk management plan linked to the development plan	Which functions, interfaces, and clinical scenarios require heightened control?	Unclear intended use, uncertain essential performance, unresolved hazard ownership
Design inputs	Translation of hazards into measurable requirements	Safety-critical input set	Are the requirements specific enough to control identified harms?	Inputs describe features but not risk boundaries
Architecture and design outputs	Allocation of risk controls to hardware, software, materials, labeling, and workflow	Risk-linked outputs matrix	Is each relevant hazard controlled by a defined design element or justified information control?	Controls depend on vague future testing or unspecified user behavior
Verification	Confirmation of control effectiveness	Risk control verification evidence	Did the control work under technically relevant stress conditions?	Testing omits physiological, environmental, or interface stressors
Validation	Confirmation of intended use conditions	Use and clinical validation rationale	Does the integrated device remain safe and effective in realistic use?	Simulated use or clinical logic fails to reflect actual care conditions
Design transfer and change control	Reassessment of modified risk profile	Updated traceability and residual risk review	Did manufacturing translation or design change alter prior safety assumptions?	Process changes introduce new interfaces, tolerances, or variability
Production and post-production	Signal collection and feedback	Surveillance-linked risk review	Which real-world signals require corrective action or redesign?	Recurrent incidents, drift patterns, infection signals, and software anomalies

Table 2. Post-freeze monitoring architecture for residual risk in Class III life-sustaining devices [1–10]

Signal domain	Leading indicator	Review cadence	Design implication	Responsible function
Mechanical integrity	Fatigue trend, seal degradation, abnormal wear	Periodic and event-triggered	Reassess durability margins and packaging design	Reliability engineering

Biological interface	Infection pattern, inflammatory response, thrombogenic events	Continuous trend review	Reevaluate materials, coatings, and implant procedure assumptions	Clinical and biocompatibility team
Functional performance	Signal drift, output instability, and power irregularity	Periodic trending with threshold alarms	Reopen control limits and performance specifications	Systems engineering
Software and connectivity	Recurring anomaly class, failed update behavior, cybersecurity-relevant deviation.	Continuous surveillance	Revise software controls, update architecture, and strengthen safety fallback logic	Software quality and cybersecurity
Use-related safety	Recurrent user deviation, maintenance error, interpretation error	Scheduled human factors review	Refine interface, instructions, training, and service process	Human factors and field support
Corrective action effectiveness	Recurrence after Corrective and Preventive Action or field action	Milestone-based review	Decide whether containment is sufficient or redesign is needed	Quality and regulatory affairs

The value of this monitoring architecture lies in its ability to transform surveillance into design intelligence. A mature Class III program does not wait for a recall-level pattern before reinterpreting residual risk. It looks for shifts in signal quality, complication clustering, replacement patterns, and failure recurrence that indicate erosion of prior assumptions. Once those signals are visible, the manufacturer can decide whether the appropriate response is a documentation update, a process correction, a design modification, or a renewed clinical evaluation.

From an applied standpoint, the most defensible integration sequence is straightforward. Risk planning starts before design inputs are finalized. Hazard framing then shapes the safety-critical input set. Architecture decisions are reviewed against control allocation. Verification is built around control effectiveness, not around generic requirement coverage. Validation tests the complete system in realistic use conditions. Post-market data are preassigned to risk review pathways before launch. In this sequence, ISO 14971 ceases to be an isolated standard. It becomes the logic that binds engineering, clinical evidence, and regulatory readiness into one continuous development system.

V. CONCLUSION

The research showed that ISO 14971 requirements are most effective when embedded as decision logic in design planning, design input definition, control allocation, verification, validation, and change management.

Class III life-sustaining and implantable devices expand hazard analysis beyond ordinary component failure. Long-term in vivo stability, biological interaction, software behavior, material degradation, and evidence uncertainty all reshape the structure of verification and validation.

Sustainable integration requires a lifecycle model in which premarket controls, residual risk evaluation, and post-market signal capture remain connected through traceability and structured review. In this form, ISO 14971 supports safer design execution, stronger evidence architecture, and greater regulatory readiness for high-risk medical technologies.

REFERENCES

[1] Bano, A., Künzler, J., Wehrli, F., Kastrati, L., Rivero, T., Llane, A., Valz Gris, A., Fraser, A. G., Stettler, C.,

Hovorka, R., Laimer, M., Bally, L., & CORE-MD investigators. (2024). Clinical evidence for high-risk CE-marked medical devices for glucose management: A systematic review and meta-analysis. *Diabetes, Obesity and Metabolism, 26*(10), 4753–4766. https://doi.org/10.1111/dom.15849

[2] Babboni, S., Sicari, R., Russo, L., Mattoli, V., Basta, G., & Del Turco, S. (2025). *Small Science, 5*, e202500379. https://doi.org/10.1002/smsc.202500379

[3] Food and Drug Administration. (2024, February 2). Medical devices, quality system regulation amendments. *Federal Register*. https://www.federalregister.gov/documents/2024/02/02/2024-01709/medical-devices-quality-system-regulation-amendments

[4] Food and Drug Administration. (2026, February 2). *Quality management system regulation (QMSR)*. https://www.fda.gov/medical-devices/postmarket-requirements-devices/quality-management-system-regulation-qmsr

[5] Freyer, O., Jahed, F., Ostermann, M., Rosenzweig, C., Werner, P., & Gilbert, S. (2024). Consideration of cybersecurity risks in the benefit-risk analysis of medical devices: Scoping review. *Journal of Medical Internet Research, 26*, e65528. https://doi.org/10.2196/65528

[6] Jurczak, K. M., van der Boon, T. A. B., Devia-Rodriguez, R., et al. (2025). Recent regulatory developments in EU Medical Device Regulation and their impact on biomaterials translation. *Bioengineering & Translational Medicine, 10*(2), e10721. https://doi.org/10.1002/btm2.10721

[7] Kearney, B., & McDermott, O. (2023). The challenges for manufacturers of the increased clinical evaluation in the European Medical Device Regulations: A quantitative study. *Therapeutic Innovation & Regulatory Science, 57*, 783–796. https://doi.org/10.1007/s43441-023-00527-z

[8] Khan, A. A., & Kim, J.-H. (2024). Recent advances in materials and manufacturing of implantable devices for continuous health monitoring. *Biosensors and Bioelectronics, 261*, 116461. https://doi.org/10.1016/j.bios.2024.116461

[9] Manimuthukumar, O., Manjuladevi, M., & Arunkumar, T. (2025). Global trends in post-market surveillance of high-risk medical devices: An empirical analysis based on regulatory data. *Indian Journal of Medical Research, 162*(2), 163–168. https://doi.org/10.25259/IJMR_86_2025

[10] Mariello, M. (2025). Reliability and stability of bioelectronic medicine: A critical and pedagogical perspective. *Bioelectronic Medicine, 11*, Article 16. https://doi.org/10.1186/s42234-025-00179-4

[11] Pannonhalmi, Á., Sipos, B., Kurucz, R. I., Katona, G., Kemény, L., & Csóka, I. (2025). Advancing regulatory oversight of medical device trials to align with clinical drug standards in the European Union. *Pharmaceuticals, 18*(6), 876. https://doi.org/10.3390/ph18060876

[12] Shawon, M. S. R., Sotade, O. T., Li, J., Hill, M. D., Strachan, L., Challis, G., King, K., Ooi, S.-Y., & Jorm, L. (2024). Factors associated with cardiac implantable electronic device-related infections, New South Wales, 2016–21: A retrospective cohort study. *Medical Journal of Australia, 220*, 510–516. https://doi.org/10.5694/mja2.52302

[13] Yogev, D., Goldberg, T., Arami, A., Tejman-Yarden, S., Winkler, T. E., & Maoz, B. M. (2023). Current state of the art and future directions for implantable sensors in medical technology: Clinical needs and engineering challenges. *APL Bioengineering, 7*(3), 031506. https://doi.org/10.1063/5.0152290

[14] Food and Drug Administration. (1997, March 11). Design Control Guidance for Medical Device Manufacturers: Guidance for Industry (archived). https://www.fda.gov/medical-devices/guidance-documents-medical-devices-and-radiation-emitting-products/withdrawn-or-expired-guidance

[15] International Organization for Standardization. (2019). ISO 14971:2019 Medical devices — Application of risk management to medical devices. https://www.iso.org/standard/72704.html

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.18

Real-Time Underground Cable Fault Detection Using Arduino Platform

Kautilya Jangid*, Aditya Chouhan, Ankit Loyal*, Aman Yadav, Vikas Ranveer Singh Mahala

Department of Electrical Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan Jaipur, India
*kautilyajangid1628@gmail.com, ankitloyal.1999@gmail.com

Received: 19 Mar 2026; Received in revised form: 16 Apr 2026; Accepted: 20 Apr 2026; Available online: 24 Apr 2026

Abstract *— In today's world, underground power cables are an increasingly popular method of distributing electricity for power distribution networks because they provide increased safety and reliability and they are less susceptible to being affected by environmental disruptions. Locating faults in underground cables can be very challenging and time-consuming due to the fact that the cables are typically not easily accessed for inspection purposes. The proposed project will utilize an Arduino microcontroller to develop an underground cable fault detection system based on the microcontroller and voltage and current sensors and a display device to continuously monitor the cable's electrical properties. The fault detection system will use Ohm's Law to detect fault conditions such as line-to-ground, line-to-line, open-circuit and three-phase using the properties of electricity. The proposed project is cost effective and provides a user-friendly and reliable method for detecting faults within underground cable systems. The time required to detect faults will be significantly reduced, with maintenance costs decreasing and the average length of time to remain without being powered-up due to a fault also decreasing. The proposed method is suitable for use in Smart Grids, Industrial, and Community types of applications.*

Keywords *— **Underground cable fault detection, Arduino microcontroller, Power distribution systems, Voltage and current sensors, Ohm's law, Smart grid applications.***

I. INTRODUCTION

In recent years, the use of underground cables for power distribution has increased steadily, especially in cities where space is limited and reliability is important. Unlike overhead lines, underground systems are less affected by weather con-ditions and reduce the risk of accidental damage. Because of these advantages, many utilities are shifting towards under-ground cabling as a more dependable solution for electricity supply. Underground cable networks have become an impor-tant part of modern power distribution systems, especially in urban and industrial areas. Compared with overhead trans-mission lines, underground cables provide several advantages such as improved safety, better protection from environmental disturbances, and a cleaner visual appearance in cities and residential areas [1]. They are also less vulnerable to weather-related damage, which helps improve the overall reliability of electrical power distribution.

Despite these advantages, detecting faults in underground cables remains a major challenge for power utilities. Since the cables are buried beneath the ground, faults cannot be easily observed or inspected visually. Locating the exact point of failure often requires specialized testing equipment or excavation along the cable route, which can be both time-consuming and costly [2]. Conventional techniques used for underground fault detection usually involve manual testing procedures and

require trained personnel to operate complex diagnostic equipment. These methods may also result in longer power outages while the fault location is being identified [3]. With the increasing demand for reliable power supply and the expansion of underground cable networks, there is a growing need for faster and more efficient fault detection methods. Automated monitoring systems based on embedded platforms provide a practical solution to this problem, as they can continuously observe electrical parameters and detect abnormal conditions at an early stage.

In this work, a real-time underground cable fault detection system based on an Arduino microcontroller platform is pro-posed. The system monitors voltage and current values of the cable using sensors and processes the data through the microcontroller to identify abnormal conditions. When a fault occurs, the system estimates the approximate distance of the fault along the cable and displays the information immediately. This helps reduce the time required for fault detection and makes maintenance work easier.

The remainder of this paper is organized as follows. The next section describes the experimental setup used for im-plementing the proposed system [4]. This is followed by the working principle of the underground cable fault detec-tion method. The simulation results and discussion are then presented to evaluate system performance. Finally, the paper concludes with a summary of the work and possible directions for future improvements.

II. EXPERIMENTAL SETUP

The experimental setup for the real-time underground cable fault detection system was designed to replicate practical fault conditions in a controlled laboratory environment [5], [6]. The objective was to create a simple and flexible arrangement where different types of underground cable faults could be introduced and studied without using an actual buried cable system. The overall circuit diagram of the proposed system is shown in Fig. 3. It illustrates the interconnection of the Arduino controller, sensing circuits, relay modules, and the cable simulation network.

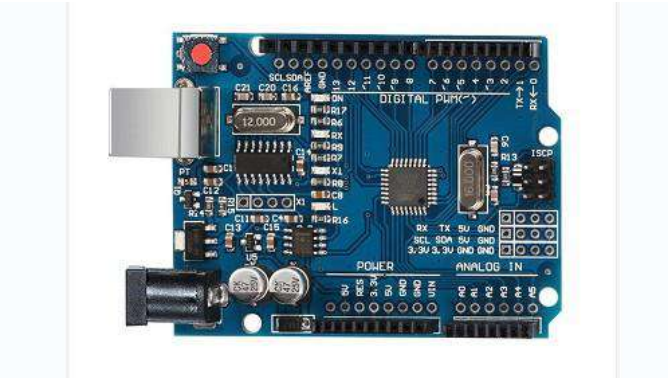


Fig. 1. Arduino ESP

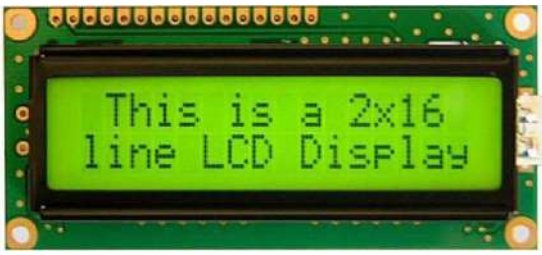


Fig. 2. 16x2 LCD Display

The objective was to create a simple and flexible arrange-ment where different types of underground cable faults could be introduced and studied without using an actual buried cable system. The complete experimental setup of the project is shown in Fig. 4.

The system is built around a microcontroller-based platform, where an Arduino or ESP board is used as the main processing unit, as shown in Fig.1. The setup also includes a regu-lated power supply, a resistive cable model, fault simulation switches, sensing circuits, and a display unit Fig. 2. Since real underground cables are not easily accessible for testing, a resistor network is used to represent the cable, where each resistor corresponds to a fixed length. This makes it possible to simulate faults at different distances along the cable.

Fault conditions are introduced manually using switches connected at various points in the resistive network. When a fault is created, it results in a change in voltage and current values. These variations are detected by the sensors and sent to the microcontroller through analog input pins. The controller processes the data continuously and identifies abnormal conditions. Using Ohm's law, the system estimates the approximate distance of the fault from the source end [9], [10]. The fault type

and location are then displayed on the LCD screen, allowing quick identification without manual inspection.

This setup also allows repeated testing under different fault scenarios, which helps in evaluating the performance and response time of the system. The arrangement is simple, low-cost, and suitable for demonstration as well as practical applications [11], [12].

A. Components

The proposed underground cable fault detection system is developed using a combination of hardware modules that work together to monitor electrical parameters and identify fault conditions. The main components used in the system are described as follows:

1) Microcontroller Unit (Arduino/ESP Board): As shown in Fig. 1, the microcontroller acts as the central processing unit of the system. It continuously receives input signals from the voltage and current sensors, processes the data, and identifies abnormal conditions. Based on the observed variations, the controller determines the type of fault and estimates its approx-imate location using Ohm's law. The processed information is then sent to the display unit. The Arduino/ESP platform is selected due to its low cost, ease of programming, and flexibility for future upgrades such as IoT-based monitoring.

2) Underground Cable Simulation Network (Switch Array): The underground cable is simulated using a switch-based network, as shown in Fig. 3. Each switch represents a possible fault point along the cable length. By operating these switches, different fault conditions such as line-to-ground, line-to-line, and open-circuit faults can be created. This arrangement provides a safe and controlled environment for testing the system without requiring an actual underground cable.

3) Resistor Network (Cable Model): A series combination of resistors is used to model the resistance of the underground cable, as depicted in Fig. 3. Each resistor represents a fixed segment of the cable length. This arrangement allows the system to estimate fault distance based on resistance variation. The resistor network plays an important role in simulating realistic cable behavior under different fault conditions.

4) Voltage Sensor Circuit: The voltage sensor circuit, inte-grated into the setup (Fig. 3), is used to measure the voltage across the cable. It converts the measured voltage into a suitable analog signal that can be read by the microcontroller. Any sudden drop or variation in voltage is treated as an indication of a possible fault condition.

5) Current Sensor Circuit: The current sensor monitors the flow of current through the cable. It detects abnormal increases or decreases in current, which are common during fault condi-tions. The sensed data is transmitted to the microcontroller for further processing. This component is essential for identifying short-circuit and open-circuit faults.

6) Power Supply Unit: The power supply unit provides a stable DC voltage to all components of the system. It typically consists of a transformer, rectifier, and voltage regulator (buck converter), as shown in Fig. 3. The unit ensures proper operation of the microcontroller and sensors by maintaining a constant voltage level and protecting the circuit from fluc-tuations.

7) Relay Modules: Relay modules are used as electrically controlled switches within the system (Fig. 3). They help in isolating faulted sections and controlling the flow of current in different parts of the circuit. Relays also provide electrical iso-lation between high-power and low-power sections, ensuring safe operation of the microcontroller.

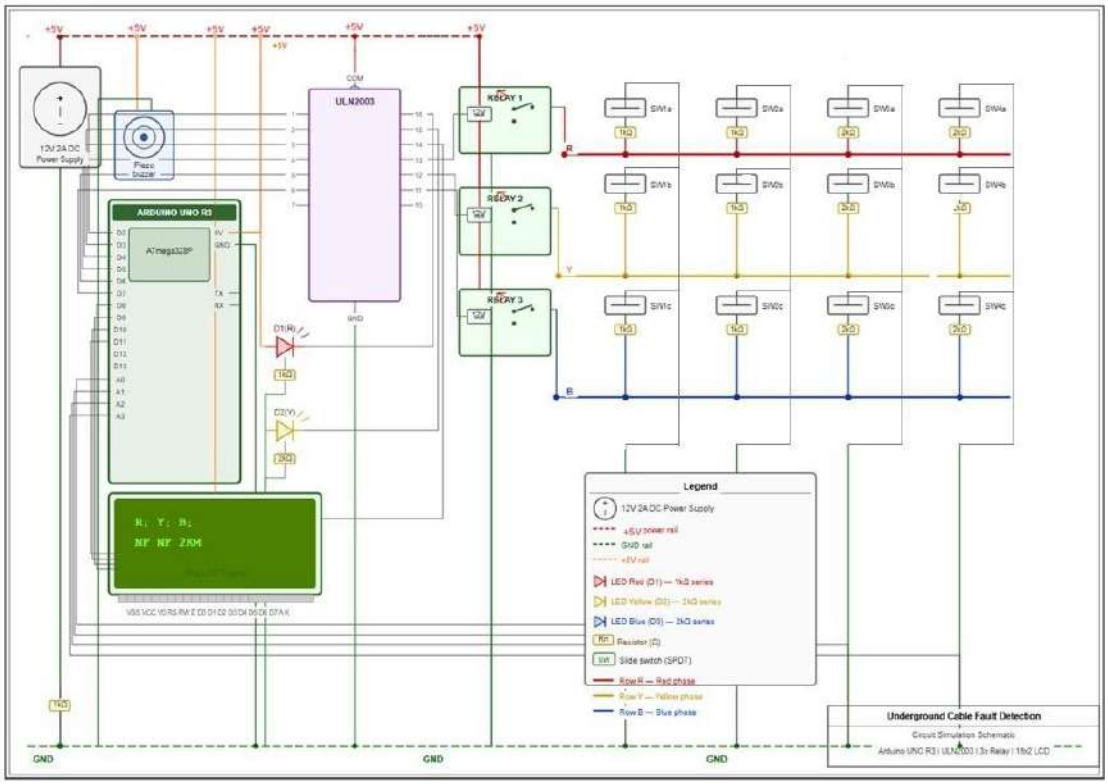


Fig. 3. Circuit diagram of the proposed underground cable fault detection system

8) LCD Display (16×2): The LCD display, shown in Fig. 2, is used to present real-time information about the system. It displays the detected fault type, estimated fault distance, and overall system status. This allows the user to quickly under-stand the condition of the cable without additional equipment.

9) Indicator LED: An indicator LED is used to provide a simple visual indication of system operation. It glows during fault conditions or when the system is active, which helps during testing and debugging of the setup.

10) Connecting Wires and PCB: All components are inter-connected using connecting wires and mounted on a printed circuit board (PCB), as shown in Fig. 3. These elements ensure proper signal transmission, structural stability, and organized layout of the system.

III. WORKING PRINCIPLE

The proposed underground cable fault detection system mainly works on a simple electrical concept, which is Ohm's law ($V = IR$). In a normal condition, an underground cable behaves in a predictable way, where the resistance along its length remains almost uniform. As a result, the voltage and current remain stable during normal operation [13].

In the developed setup, as shown in Fig. 4, the underground cable is not used directly. Instead, it is represented using a se-ries of resistors connected together. Each resistor corresponds to a small portion of the cable length. This makes it easier to study the behavior of the system in a laboratory environment. Voltage and current sensors are connected at the input side of the setup to continuously measure electrical values. These measured signals are then given to the Arduino microcontroller for processing [14], [15].

When a fault is introduced in the system, such as a line-to-ground, line-to-line, or open-circuit fault, the normal flow of current gets disturbed. This disturbance causes a change in voltage and current values. In most cases, faults either increase or decrease the current suddenly, and this change can be clearly observed through the sensors. The microcontroller keeps checking these values and compares them with normal operating conditions. If a significant deviation is found, the system recognizes that a fault has occurred [16].

To find the location of the fault, the system uses the idea that resistance is related to the length of the cable. Since the resistance per unit length is already known from the resistor model, any change in resistance can be used to estimate how far the fault is from the starting point. The Arduino performs this calculation and gives an approximate distance of the fault. At the same time, it also tries to identify the type of fault by looking at how the voltage and current have changed [17].

Once the fault is detected and analyzed, the result is shown on the LCD display. The display provides simple information such as fault type and its approximate distance. This makes it easier for the user to understand the situation without doing any manual calculations or testing.

Overall, the working of the system is straightforward and easy to follow. It uses basic electrical principles along with a microcontroller to detect faults quickly. The main advantage is that the system reduces the need for manual inspection and helps in locating faults in less time.

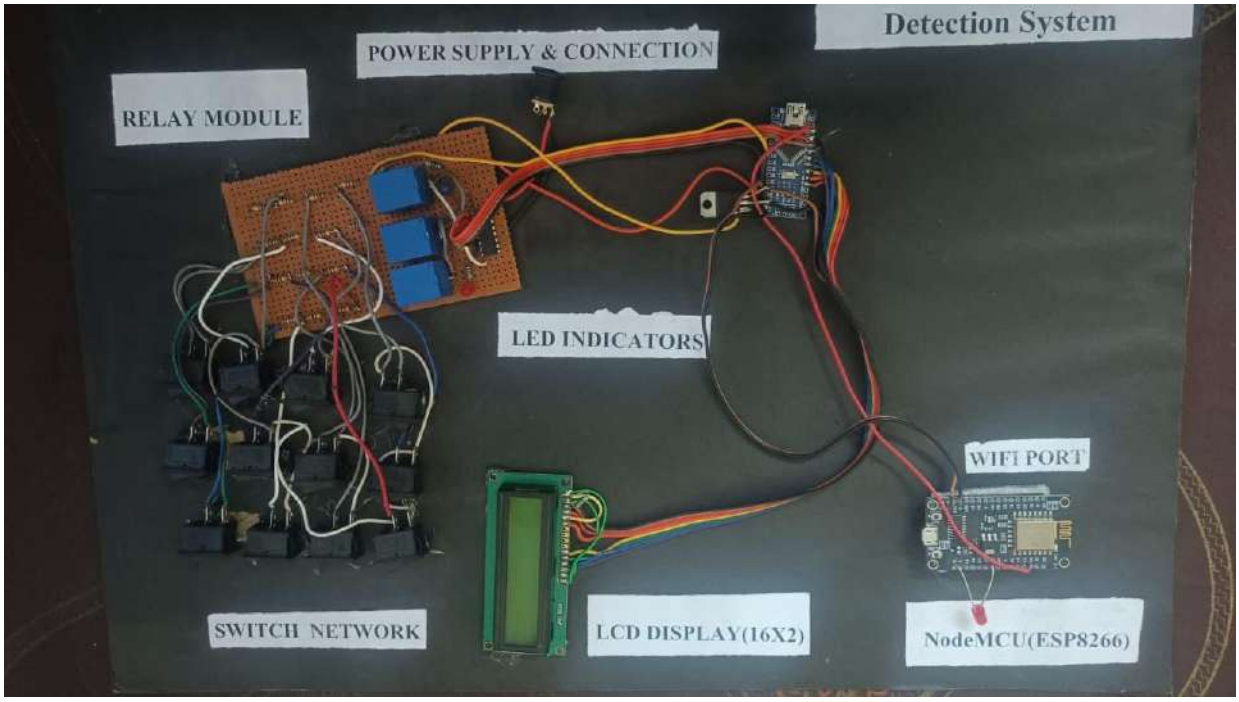


Fig. 4. Experimental setup of the proposed underground cable fault detection system

IV. SIMULATION RESULTS AND DISCUSSION

The proposed underground cable fault detection system was evaluated through simulation to understand how it behaves under different operating conditions. The idea was to check whether the system can correctly detect faults and give a reasonable estimate of their location along the cable. For this purpose, different scenarios were created in the simulation model, including normal operation as well as various fault conditions. These situations were generated by changing re-sistance values and current flow in the cable model.

Under normal operating conditions, the system showed stable voltage and current values along the entire cable. The sensors continuously monitored these parameters and passed the data to the microcontroller. Since there were no disturbances, the readings remained steady, indicating healthy operation. This condition was useful as a reference to compare how the system behaves when faults occur.

When fault conditions were introduced, clear changes in electrical parameters were observed. In the case of a line-to-ground fault, the resistance of the affected section reduced, which caused an increase in current. This change was quickly detected by the system. The microcontroller processed the sensor data and identified the condition as a fault. It then estimated the approximate location by relating the resistance change to the cable length.

A similar observation was made for line-to-line

faults. In this case, the current increased suddenly due to direct contact between conductors. The sensors captured this abnormal rise, and the system recognized it as a fault condition. By looking at the pattern of voltage and current variation, the controller was able to distinguish this fault from others and calculate its approximate position along the cable.

Open-circuit faults were also tested in the simulation. Here, the current dropped sharply because the circuit path was broken. The system detected this drop without delay and classified it as an open-circuit condition. The microcontroller then estimated the fault distance using the same resistance-based approach.

One of the noticeable outcomes from the simulation was the fast response of the system. As soon as a fault was introduced, the change in voltage and current was detected almost immediately. The processing time was very short, and the result was displayed without any significant delay. This quick response is important in practical situations, as it helps reduce the time required to locate faults and restore power.

The simulation also included waveform analysis to study the behaviour of different circuit components. Parameters such as input current, output voltage, and stresses on switches, inductors, capacitors, and diodes were observed. During nor-mal conditions, the waveforms appeared smooth and stable. However, when faults were introduced, disturbances were clearly visible in the waveforms. These changes matched the expected behaviour of the system, which supports the correctness of the detection method.

Another point worth noting is the performance of the resistance-based fault location method. While the system may not always give the exact physical location of the fault, it provides a close approximation. This is usually sufficient in practice, as it helps narrow down the faulty section of the cable. Instead of checking the entire cable, maintenance work can be focused on a smaller region, saving both time and effort. Overall, the simulation results show that the proposed system can detect different types of faults and respond in a reliable manner. It is able to monitor electrical parameters, identify abnormal conditions, and estimate fault locations with reasonable accuracy. These observations suggest that the system can be useful for real-world underground cable monitoring, especially where quick fault detection is required.

V. CONCLUSION

This work presents a simple underground cable fault de-tection system using an Arduino-based setup. The system monitors voltage and current values and uses these changes to identify when a fault occurs in the cable. From the simulation study, it can be seen that the system is able to detect common fault conditions such as line-to-ground, line-to-line, and open-circuit faults. It also gives an approximate idea of where the fault is located by using resistance variation, which makes the process of inspection easier and faster.

One of the main advantages of this approach is that it does not require complex equipment. The setup is built using low-cost components and can be implemented without much difficulty. The response of the system is also quite fast, which helps in reducing delay in fault detection and improving overall reliability.

There is still scope for improvement in future work. Features like wireless communication, IoT-based monitoring, or GPS support can be added to make the system more practical for real-time applications and smart grid environments.

REFERENCES

[1] M. H. Rashid *et al.*, "Power Electronics: Circuits, Devices and Appli-cations, Pearson Education. review," *IEEE Trans. Power Electron.*, vol. 32, no. 12, pp. 9143–9178, 2017.

[2] R. W. Erickson and D. Maksimovic´, "Fundamentals of Power Electron-ics, Springer.

[3] S. Ghosh and M. Chakraborty *et al.*, "Fault detection and location in underground cable system using microcontroller," International Journal of Engineering Research, 2019.

[4] J. Arrillaga, N. R. Watson, "Power System Harmonics, Wiley.

[5] A. R. Gupta, *et al.*, "Microcontroller-based underground cable fault detector," IJEEE, 2020.

[6] J. M. Kharade and H. P. Patil *et al.*, " Fault Detection in Underground Cable by using Arduino, Applied Science and Engineering Journal for Advanced Research, vol. 4, no. 2, pp. 63–66, Mar. 2025.

[7] S. Gouda, S. Sudhakar, V. J. Vinaykumarswamy, R.

Singadi, and Navaprakash, *et al.*, "Underground Cable Fault Detection System, IJRASET, Apr. 29, 2023.

[8] M. T. HOD, N. Nagabhushan, and P. C. H. Pushpanjali, *et al.*, "Smart System for Detection of Underground Cable Faults Using Arduino and Ohm's Law, ShodhKosh: Journal of Visual and Performing Arts, vol. 5, no. 1, 2024

[9] A. Jangir, D. Jain, K. Nayak, and V. R. S. Mahala, "Density Based Traffic Control System and Emergency Vehicle Detection Using Arduino," *In-ternational Journal of Advanced Engineering, Management and Science*, vol. 11, no. 2, p. 609970, 2025.

[10] S. A. Danole, A. S. Mane, and V. M. Heralge *et al.*, "Underground Cable Fault Detection, IARJSET, 2023.

[11] S. Inampudi, R. T. B. Rajitha, K. Sawant, and L. Gaur *et al.*, "Under-ground Cable Fault Detection Device Using Microcontroller, IJRASET, 2022.

[12] S. Kesavan and V. S. Srinivasan *et al.*, "Underground Cable Fault Analysis System Using Arduino with Bluetooth Controller, Int. J. Sci. Res. Sci. Eng. Technol., vol. 11, no. 4, pp. 169–175, Jul. 2024.

[13] M. Dudi, M. Bhati, N. Sharma, N. Choudhary, V. Mahala, and J. Vijay, "Optimizing solar powered charging stations for electric vehicles: Integration of fast and slow charging with renewable energy sources," 2024.

[14] Abhishek Pandey and N. H. Younan, *et al.*, "Underground cable fault detection and identification via Fourier analysis," IEEE, 2010.

[15] Xia Yang *et al.*, "Fault location for underground power cable using distributed parameter approach, IEEE Trans. Power Syst., 2008.

[16] V. Mahala, R. Saini, S. Sharma, R. Verma, and R. Jangid, "High step-up modified four phase interleaved boost converter with coupled inductor topology," in *Proc. 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 129–134, 2021.

[17] Md. Fakhrul Islam et al., *et al.*, "Locating underground cable faults: A review and guideline for new development, IEEE, 2013.

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.19

Seismic Analysis of G+10 Storey Building using Structural Lightweight and Normal Weight: A Review

Er. Shubham Rathod[1,*], Prof. Satish Sahebrao Manal[2]

[1]MTech Structural student, CSMSS, Chh. Shahu College of Engineering, Chh. Sambhajinagar. – 431011, Maharashtra,, India
[2]Assistant Professor, CSMSS, Chh. Shahu College of Engineering, Chh. Sambhajinagar. - 431011, Maharashtra,, India
*Corresponding Author

Received: 20 Mar 2026; Received in revised form: 18 Apr 2026; Accepted: 22 Apr 2026; Available online: 26 Apr 2026

Abstract *— Concrete is the most important material used in many construction applications. The multistorey buildings are constructed of ordinary concrete, steel and other materials and are subjected to heavy loads requiring heavy construction and may not be cost effective. But the structural lightweight concrete produced using lightweight aggregates may reduce the dead load of the structure, so now a days it is used in construction of multistorey buildings. Concrete is considered as light weight concrete which has density of less than 2000 kg/m³. In this research, a G+10 multistorey plan irregular building is analysed with structural lightweight concrete using perlite as a fine aggregate and normal weight concrete using the Response Spectrum Method under different seismic zones and Time History Method for Bhuj earthquake. The parameters like storey displacement, storey drift, storey shear and overturning moments are considered and the results of NWC and SLWC buildings are compared. From the results obtained it is observed that seismic damages are considerably reduced in structural lightweight concrete buildings as compared to normal weight concrete buildings.*

Keywords *— **Structural lightweight concrete, Perlite aggregate, Seismic analysis, Response spectrum method, Time history analysis, ETABS***

I. INTRODUCTION

Concrete plays an important role in modern construction due to its strength, durability, fire resistance, and low cost. However, one of the major drawbacks of conventional concrete is its high self-weight. During earthquakes, higher mass leads to larger inertial forces, which increases damage to structures. Structural lightweight concrete helps in reducing the dead weight of structures, thereby lowering seismic forces.

Lightweight concrete can be produced by using lightweight aggregates, introducing air voids, or removing fine aggregates. Among these methods, the use of lightweight aggregates is the most suitable for structural applications. Natural perlite is a volcanic material that expands when heated and becomes very light in weight. Due to this property, it can be used as a lightweight aggregate in concrete. This paper presents a review of studies related to lightweight concrete and the use of perlite aggregate, especially for improving seismic performance of buildings.

Structural Lightweight Concrete

Structural lightweight concrete generally has a density between 1440 and 1850 kg/m³ and sufficient strength for load-bearing applications. The main advantage of SLWC is the reduction in dead load, which results in lower seismic forces, smaller member sizes, and reduced foundation cost.

Although SLWC has lower modulus of elasticity compared to normal weight concrete, it can still perform well under seismic loading if properly designed. The reduced stiffness may increase

displacement, but overall seismic demand is lowered due to reduced mass.

Perlite as a Lightweight Aggregate

Perlite is a naturally occurring volcanic glass that contains moisture. When heated, the moisture inside turns into steam and causes the material to expand, forming a porous structure with very low density. Expanded perlite is widely used in construction due to its fire resistance and insulating properties. When used in concrete, perlite reduces density and improves thermal performance. However, higher porosity can lead to a reduction in compressive strength. Many studies show that partial replacement of conventional aggregates with perlite, along with the use of mineral admixtures, can produce concrete suitable for structural use. Several researchers have studied the mechanical behaviour of lightweight concrete.

History of Structural Lightweight Concrete

Structural lightweight concrete has been used in engineering applications for more than a century, particularly where reduction in self-weight and improved structural efficiency are required. One of the earliest modern applications was the construction of a lightweight concrete ship between 1917 and 1920 using concrete with a compressive strength of about 35 MPa. This project marked the first significant use of high-performance lightweight concrete in structural design. Another notable example is the Bank of America Corporate Center in Charlotte, North Carolina, a 60-storey building with a height of 265 m. The structure was built using lightweight concrete with compressive strength ranging from 43 to 51 MPa and an average density of approximately 1890 kg/m³. The use of lightweight concrete helped reduce dead load while maintaining structural performance. The Wellington Regional Stadium in New Zealand, with a seating capacity of 40,000, was the country's first major structure built using lightweight concrete. All precast components were constructed using lightweight concrete made with expanded slate aggregate imported from the United States.

In Sweden, the first lightweight concrete bridge was completed in 1975. The prestressed slab bridge, constructed using sintered fly ash aggregate, achieved a compressive strength of 35 MPa and a density of about 1800 kg/m³. Similarly, Norway has constructed several bridges using high-strength structural lightweight concrete since the late 1980s, with compressive strengths up to 60 MPa and densities around 1900 kg/m³. In the United Kingdom, the use of lightweight concrete began with a reinforced concrete office building in 1958. This was followed by major projects such as aircraft hangar expansions and high-rise hospital towers constructed using sintered fly ash lightweight aggregate concrete.

Natural Perlite

Perlite is a naturally occurring volcanic glass formed through the hydration of obsidian and contains a relatively high amount of chemically bound water. When subjected to rapid heating, perlite undergoes a significant volumetric expansion due to the vaporization of this water, resulting in a lightweight cellular material. The expanded form of perlite can increase in volume by 15 to 20 times its original size while maintaining a very low bulk density. Expanded perlite is widely used in construction due to its lightweight, fire-resistant, and thermal insulation properties. When used as a fine aggregate in concrete, perlite produces lightweight concrete with reduced density compared to conventional mixes. Although the compressive strength of perlite-based concrete is generally lower than that of normal weight concrete, its reduced density plays a crucial role in minimizing dead load and improving seismic performance.

Perlite fine aggregate is obtained from pumice stone, a glassy volcanic material containing trapped moisture. During heating, the moisture converts into steam, forming numerous internal pores that give perlite its lightweight and insulating characteristics. Due to these properties, expanded perlite is increasingly being considered as an alternative fine aggregate in structural lightweight concrete applications.

Aim and Objectives

Aim:

To evaluate and compare the seismic performance of a G+10 storey plan-irregular reinforced concrete building constructed using structural lightweight concrete with perlite as a fine aggregate and conventional normal weight concrete.

Objectives:

1. To perform seismic analysis of a G+10 storey plan-irregular building using structural lightweight concrete and normal weight concrete through the Response Spectrum Method and Time History Method.
2. To compare key seismic response parameters such as overturning moment, storey shear, storey drift, and storey displacement for both structural systems using ETABS software.

II. LITERATURE SURVEY

Brindha Sathiaseelan & M. Hannah Angelin (2025)

Brindha Sathiaseelan and M. Hannah Angelin (2025) evaluated the durability and mechanical properties of sustainable lightweight concrete prepared with perlite and vermiculite as partial replacements for traditional aggregates. Concrete mixes with 5%, 7%, and 10% replacement levels were tested to examine their effect on density, compressive strength, and durability indicators. The study found that incorporating perlite and vermiculite significantly reduced concrete density while influencing mechanical behaviour and long-term durability. The research demonstrated the potential of mineral aggregate modifications to improve sustainability and performance of structural lightweight concrete.

V. Navin Ganesh, N. Divyah & R. Rajkumar (2024)

V. Navin Ganesh et al. (2024) investigated fiber reinforcement in sintered flyash lightweight aggregate concrete to enhance structural performance, energy absorption, and impact resistance. Basalt fibers were added to concrete mixtures containing sintered flyash aggregates to study improvements in toughness and dynamic behaviour. Results showed that fiber additions markedly increased impact resistance and crack-bridging capacity, providing improved mechanical performance after cracking. The research highlighted the benefits of combining recycled lightweight aggregates with fiber reinforcement for more resilient structural lightweight concrete designs.

B. Prasanna Kumar et al. (2024)

Prasanna Kumar, S. Sai Charan, A. Hema Venkata Suresh, and M. Chandra Kanth (2024) conducted an experimental study on the mechanical properties of lightweight aggregate concrete using expanded clay aggregates. This work focused on M25 grade concrete with partial replacement of coarse aggregates by lightweight expanded clay. Fresh and hardened concrete properties were evaluated, including compressive strength and splitting tensile strength. The findings revealed a reduction in density and mechanical properties compared to conventional mixes, but the trends indicated that lightweight aggregate substitution can be tailored to achieve structural requirements with optimized mix proportions.

Abhishek Kumar Singh et al. (2022)

Abhishek Kumar Singh and colleagues (2022) studied the mechanical properties of lightweight concrete using lightweight expanded clay aggregate (LECA) as partial replacement for conventional coarse aggregate. The research demonstrated that LECA incorporation significantly lowered concrete self-weight, improving thermal insulation and reducing structural mass. The study also reported effects on compressive strength and workability, showing that lightweight mixes with LECA present practical advantages for structural applications where reduced density and improved sustainability are desired.

Singh, A. K., Verma, R., and Sharma, R. (2022)

Singh et al. (2022) examined the mechanical properties of lightweight concrete prepared using lightweight expanded clay aggregate as a replacement for conventional coarse aggregates. The experimental program included compressive strength, split tensile strength, flexural strength, and modulus of elasticity tests. The results showed that lightweight concrete achieved sufficient strength for structural use while offering a significant reduction in unit weight. It was observed that aggregate strength and porosity had a direct influence on stiffness and load-carrying capacity.

Although the modulus of elasticity was lower than that of normal weight concrete, the material exhibited better deformation capacity. The authors highlighted that reduced self-weight leads to lower seismic forces, which is beneficial for multistorey buildings. The study concluded that lightweight expanded clay aggregate concrete can be effectively used in structural applications where dead load

reduction and improved seismic response are required.

Patel, J., and Desai, M. (2023)

Patel and Desai (2023) conducted a detailed analytical study to evaluate the seismic performance of reinforced concrete buildings constructed using lightweight aggregate concrete. The study compared normal weight concrete and lightweight concrete frames using response spectrum analysis as per Indian seismic design provisions. Parameters such as base shear, storey displacement, storey drift, and natural time period were analyzed. Results indicated that lightweight concrete buildings experienced reduced base shear and overturning moments due to lower seismic mass. Although storey displacements were marginally higher, they remained within the allowable limits prescribed by IS 1893. The authors emphasized that reduced member forces can lead to more economical structural design. The study concluded that lightweight concrete is a suitable alternative for earthquake-resistant construction, especially in high seismic zones.

Ramesh, K. (2021)

Ramesh and Prakash (2021) investigated the mechanical and durability characteristics of perlite-based lightweight concrete. The experimental work focused on compressive strength, density, water absorption, and resistance to chloride penetration. The results showed that increasing perlite content significantly reduced the density of concrete, making it suitable for lightweight structural applications. However, higher perlite replacement levels led to increased porosity, resulting in reduced compressive strength. The authors noted that the use of supplementary cementitious materials helped improve strength and durability. Despite the reduction in strength, the concrete achieved acceptable performance for structural and non-structural elements. The study concluded that perlite-based lightweight concrete is beneficial for reducing dead load and improving durability when properly designed and optimized.

Kumar, V., and Rao, P. S. (2024)

Kumar and Rao (2024) studied the effect of partial replacement of fine aggregate with perlite on the strength and stress–strain behavior of lightweight concrete. Experimental investigations included compressive strength tests and evaluation of stress–strain characteristics. The results indicated that perlite improved workability and reduced density, leading to lighter concrete mixes. Although a reduction in stiffness was observed, the concrete showed improved strain capacity and gradual failure behavior. Stress–strain curves revealed lower peak stress but enhanced ductility compared to normal concrete. The authors highlighted that improved deformation capacity is advantageous under seismic loading. The study concluded that perlite-based lightweight concrete can be effectively used in seismic-resistant structures where controlled strength reduction is acceptable in exchange for improved energy absorption.

Chaudhari, S., and Kulkarni, P. (2023)

Chaudhari and Kulkarni (2023) analyzed the seismic behavior of multistorey reinforced concrete buildings constructed using lightweight concrete. Both response spectrum and time history analyses were carried out to assess structural performance under earthquake loading. Parameters such as base shear, storey drift, overturning moment, and axial forces in columns were examined. The results showed a significant reduction in base shear and overturning moments due to decreased structural mass. Although storey displacements increased slightly, they were within permissible limits specified by seismic design codes. The authors concluded that lightweight concrete helps reduce seismic demand and improves overall structural efficiency. The study recommended the use of lightweight concrete in high-rise buildings located in seismic-prone regions.

Mohammed Ibrahim et al. (2020)

Mohammed Ibrahim et al. (2020) focused on developing durable structural lightweight concrete using expanded perlite aggregate (EPA). EPA was used in proportions ranging from 0% to 20%, while cement was partially replaced with 50% GGBFS and 7% silica fume. Mechanical and durability properties were evaluated along with seismic behaviour using finite element modelling. Results showed a 20–30% reduction in unit weight compared to normal concrete. Concrete containing 10–15% EPA achieved adequate strength and durability. Improved thermal insulation and better seismic performance were also observed.

Annie Sweetlin et al. (2020)

Annie Sweetlin et al. (2020) studied structural lightweight concrete using natural perlite as a complete replacement for coarse aggregate. Concrete mixes were prepared with different water–cement ratios and tested for compressive and split tensile strength. The results showed a significant reduction in self-weight compared to conventional concrete. However, compressive and tensile strengths were reduced by approximately 46% and 57%, respectively. Despite lower strength, the reduced density makes perlite concrete suitable for seismic-resistant construction where reduced dead load is beneficial.

Bing Han et al. (2017)

Bing Han et al. (2017) investigated the axial compressive stress–strain behaviour and Poisson's effect of lightweight concrete made using expanded slate aggregates. Four different aggregate volume fractions were studied. Results showed that compressive strength did not change linearly with aggregate content. Minimum strength occurred at 20% replacement, after which strength increased. At 60% replacement, compressive strength was comparable to normal concrete. Lightweight aggregates were found to be more deformable, resulting in lower elastic modulus. Aggregate type significantly influenced mechanical properties.

Bhuvaneshwari et al. (2017)

Bhuvaneshwari et al. (2017) examined the use of natural perlite as a partial replacement for sand in lightweight concrete. Perlite replaced sand at levels of 5%, 10%, 15%, 20%, and 25%. Compressive, split tensile, and flexural strength tests were conducted. The study found that 10% perlite replacement produced optimum results, with increases in compressive, tensile, and flexural strength at 28 days. Beyond this level, strength decreased due to increased porosity in the concrete matrix.

Kwang-Soo Youm et al. (2016)

Youm et al. (2016) conducted an experimental study on the strength and durability of high-performance lightweight concrete containing silica fume. Concrete mixes were designed to achieve 60 MPa compressive strength with oven-dry density below 1900 kg/m³. Three lightweight aggregates and silica fume contents of 0%, 3.5%, and 7% were investigated. Silica fume improved compressive strength and chloride resistance depending on aggregate type. However, split tensile strength and elastic modulus were not significantly affected. Improved durability was attributed to refined cement paste microstructure.

Yingwu Zhou et al. (2016)

Zhou et al. (2016) investigated the axial compressive behaviour of fibre-reinforced polymer (FRP) confined lightweight aggregate concrete. Three different lightweight aggregates were used, including shale ceramics and hollow steel balls. Axial compression tests showed that FRP confinement significantly enhanced load-carrying capacity and ductility. Analytical models for ultimate strength and strain closely matched experimental results. Strength enhancement depended on FRP content, with higher efficiency observed in normal concrete at larger FRP ratios.

Murat Emre Dilli et al. (2015)

Dilli et al. (2015) compared the strength and elastic properties of conventional concrete and lightweight structural concrete produced using expanded clay aggregates. Properties such as compressive strength, modulus of elasticity, Poisson's ratio, and ductility were studied. The dry density of lightweight concrete ranged from 1640 to 2050 kg/m³. Results showed that compressive strength alone is insufficient for predicting elastic modulus. Lightweight concrete exhibited lower stiffness and more brittle behaviour than normal concrete, while Poisson's ratios remained similar.

H.Z. Cui et al. (2012)

Cui et al. (2012) carried out experimental investigations and developed an analytical model for the pre-peak stress–strain behaviour of structural lightweight aggregate concrete. Five different lightweight aggregates with varying volume fractions were studied. Results showed that peak stress and Young's modulus decreased with increasing lightweight aggregate content. Lightweight aggregate was identified as the weakest component in LWAC. Higher aggregate density and crushing strength improved mechanical performance, while increased aggregate content led to more brittle failure behaviour.

Tommy Y. Lo et al. (2006)

Tommy Y. Lo et al. (2006) studied the influence of aggregate properties on the strength of lightweight concrete. The effects of aggregate strength, water–cement ratio, and porosity within the interfacial transition zone were examined. Concrete mixes with water–cement ratios of 0.40, 0.44, and 0.48 were tested. Results indicated that concrete strength mainly depends on aggregate crushing strength. Higher water–cement ratios increased pore volume and reduced strength, particularly at the aggregate–paste interface.

J.M. Chi et al. (2002)

Chi et al. (2002) investigated the effect of aggregate properties on the strength and stiffness of lightweight concrete using cold-bonded pelletized aggregates. Three aggregate types with different fly ash contents were produced. The study showed that water–binder ratio and aggregate volume fraction strongly influence compressive strength and elastic modulus. When the aggregate volume fraction was limited to 18%, concrete strength and stiffness were governed mainly by the cement paste, rather than aggregate type.

T.H. Almusallam and S.H. Alsayed (1995)

Almusallam and Alsayed (1995) proposed a simple mathematical model to represent the complete stress–strain relationship of normal, high-strength, and lightweight concrete. The model was developed using average experimental results from specimens tested under various conditions. It accurately described both ascending and descending portions of the stress–strain curve. The authors concluded that the model is equally applicable to different concrete types and effectively accounts for factors influencing stress–strain behaviour, showing good agreement with experimental data.

III. LITERATURE GAP

- Most existing studies focus on material-level properties of lightweight concrete, such as compressive strength, stress–strain behavior, modulus of elasticity, and durability, rather than overall structural performance.
- Research on perlite-based lightweight concrete mainly addresses density reduction and strength characteristics, with limited attention to its application in structural systems.
- Very few studies have evaluated the seismic behavior of multistorey reinforced concrete buildings constructed using structural lightweight concrete.
- Comparative studies between structural lightweight concrete with perlite replacement of fine aggregates and normal weight concrete for multistorey buildings are scarce.
- Limited literature is available on the combined use of response spectrum method and time history method for seismic analysis of buildings using lightweight concrete.
- The influence of reduced self-weight on base shear, storey drift, overturning moment, and member forces in multistorey buildings using perlite-based lightweight concrete is not well documented.
- There is a lack of studies addressing the practical feasibility and seismic efficiency of using structural lightweight concrete with perlite replacement in G+10 or higher storey buildings.
- Hence, a clear research gap exists in conducting a comprehensive comparative seismic analysis of multistorey buildings using structural lightweight concrete with perlite replacement and conventional normal weight concrete.

IV. SUMMARY OF LITERATURE

The reviewed research studies provide a comprehensive understanding of the behaviour, strength, and structural performance of lightweight concrete, particularly with the use of perlite and other lightweight aggregates. These studies collectively highlight the advantages and limitations of lightweight concrete in terms of strength, density, durability, and seismic performance. The findings from previous investigations form a strong foundation for identifying research gaps and defining the scope of the present study.

- The reviewed literature confirms that incorporating perlite in concrete significantly reduces the self-weight of structural members, which helps in minimizing dead loads and optimizing member sizes.

- Due to its low density, perlite-based lightweight concrete contributes to reduced seismic forces acting on structures, making it suitable for earthquake-resistant design.
- Several studies indicate that an increase in perlite content leads to a reduction in density as well as compressive strength of concrete.
- To compensate for strength loss while maintaining reduced density, researchers have incorporated supplementary materials such as silica fume, GGBFS, and other mineral admixtures in the concrete mix.
- Experimental findings show that the peak stress and Young's modulus of lightweight aggregate concrete generally decrease when lightweight aggregates replace conventional aggregates.
- The mechanical performance of lightweight concrete is strongly influenced by the strength, density, and crushing resistance of the lightweight aggregates used.
- The literature clearly establishes that aggregate properties play a dominant role in governing the stress–strain behaviour and overall structural response of lightweight concrete.
- After a detailed review, it is observed that limited research has been carried out on the comparative seismic behaviour of multistorey buildings using structural lightweight concrete with perlite as fine aggregate replacement.
- No comprehensive study was found comparing the seismic performance of a multistorey building constructed with perlite-based lightweight concrete and normal weight concrete.

Therefore, the present research addresses this gap by performing seismic analysis of a G+10 storey building using structural lightweight concrete with perlite as a replacement for fine aggregates and comparing it with normal weight concrete using Response Spectrum Method and Time History Method.

V. FUTURE SCOPE

This research area has a vast scope for the future studies. In this subject, large scale analysis can be conducted with a wide range of variable and characteristics. After finishing the analysis and scrutiny of data about structure design with SLWC and NWC under different seismic zone, there are still some avenues that can be studied in details which are listed below

1. Introduction of other lateral load resisting like bracing, belt, truss and outrigger system can be done and their results can be compared.
2. Seismic analysis of multistorey building design with SLWC with design position & location of shear wall, with or without shear wall can be done and their results can be compared with structure design with NWC.

REFERENCES

[1] Brindha Sathiaseelan, M., and Hannah Angelin, M., "Mechanical and durability performance of lightweight concrete using perlite and vermiculite," *International Journal of Engineering Science and Technology*, 2025.

[2] Navin Ganesh, V., Divyah, N., and Rajkumar, R., "Enhancing structural performance of sintered fly ash lightweight concrete using fiber reinforcement," *Indian Journal of Science and Technology*, 2024.

[3] Prasanna Kumar, B., Sai Charan, S., Hema Venkata Suresh, A., and Chandra Kanth, M., "Experimental investigation on mechanical properties of lightweight aggregate concrete using expanded clay," *International Journal of Research and Analytical Reviews*, 2024.

[4] Singh, A. K., et al., "Mechanical properties of lightweight concrete using lightweight expanded clay aggregate," *International Journal for Research in Applied Science and Engineering Technology*, 2022.

[5] Ibrahim, M., Ahmad, A., Barry, M. S., Alhems, L. M., and Mohamed Suhoothi, A. C., "Durability of structural lightweight concrete containing expanded perlite aggregate," *International Journal of Concrete Structures and Materials*, vol. 14, 2020.

[6] Sweetlin, J. P. A., Boologanageswari, S., Madhu, S., Priyanka, B., and Rekha, S., "Experimental study of lightweight concrete using perlite," *International Research Journal of Engineering and Technology*, vol. 7, no. 5, pp. 1–6, 2020.

[7] Han, B., and Xiang, T. Y., "Axial compressive stress–strain relation and Poisson effect of structural lightweight aggregate concrete," *Construction and Building Materials*, vol. 146, pp. 338–343, 2017.

[8] Bhuvaneshwari, K., Dhanalakshmi, G., and Kaleeswari, G., "Experimental study on lightweight concrete using perlite," *International Research Journal of Engineering and Technology*, vol. 4, no. 4, pp. 1–5, 2017.

[9] Youm, K. S., Moon, J., Cho, J. Y., and Kim, J. J., "Experimental study on strength and durability of

lightweight aggregate concrete containing silica fume," *Construction and Building Materials*, vol. 114, pp. 517–527, 2016.

[10] Zhou, Y., Liu, X., Xing, F., Cui, H., and Sui, L., "Axial compressive behavior of FRP-confined lightweight aggregate concrete: Experimental study and stress–strain model," *Construction and Building Materials*, vol. 119, pp. 1–15, 2016.

[11] Dilli, M. E., Atahan, H. N., and Şengül, C., "Comparison of strength and elastic properties of conventional and lightweight structural concretes using expanded clay aggregates," *Construction and Building Materials*, vol. 101, pp. 260–267, 2015.

[12] Cui, H. Z., Lo, T. Y., Memon, S. A., Xing, F., and Shi, X., "Experimental investigation and analytical modeling of pre-peak stress–strain behavior of structural lightweight aggregate concrete," *Construction and Building Materials*, vol. 36, pp. 845–859, 2012.

[13] Lo, T. Y., Tang, W. C., and Cui, H. Z., "Effects of aggregate properties on lightweight concrete," *Building and Environment*, vol. 42, pp. 3025–3029, 2007.

[14] Chi, J. M., Huang, R., Yang, C. C., and Chang, J. J., "Effect of aggregate properties on the strength and stiffness of lightweight concrete," *Cement and Concrete Composites*, vol. 25, pp. 197–205, 2003.

[15] Almusallam, T. H., and Alsayed, S. H., "Stress–strain relationship of normal, high-strength and lightweight concrete," *Magazine of Concrete Research*, vol. 47, no. 170, pp. 39–44, 1995.

[16] Bureau of Indian Standards, *IS 456:2000 – Plain and Reinforced Concrete: Code of Practice*, New Delhi, India.

[17] Bureau of Indian Standards, *IS 875 (Part 1):1987 – Code of Practice for Design Loads (Dead Loads)*, New Delhi, India.

[18] Bureau of Indian Standards, *IS 1893 (Part 1):2016 – Criteria for Earthquake Resistant Design of Structures*, New Delhi, India.

[19] Bureau of Indian Standards, *IS 13920:2016 – Ductile Detailing of Reinforced Concrete Structures Subjected to Seismic Forces*, New Delhi, India.

[20] American Concrete Institute, *ACI 318R – Building Code Requirements for Structural Concrete and Commentary*, USA.

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.20

Six Sigma Dimensional Control Frameworks for Safety-Critical Assemblies in Emerging Aviation Technologies

Srinivasrao Balaga

Sr. Manager, Dimensional Engineering, Archer Aviation, CA, USA
ORCID: 0009-0001-6240-9798

Received: 19 Mar 2026; Received in revised form: 17 Apr 2026; Accepted: 22 Apr 2026; Available online: 28 Apr 2026

Abstract *— Safety-critical aerospace assemblies are produced under a narrower dimensional margin than conventional industrial systems. At the same time, digital manufacturing, composite structures, and hybrid joining routes increase the number of deviation pathways that can affect fit, load transfer, sealing, and long-term reliability. This article develops an analytical model of Six Sigma dimensional control for emerging aviation technologies. The study examines how recent research links tolerance allocation, metrology, digital twins, deviation-source modeling, and capability assurance in aerospace and adjacent high-precision assembly environments. The source base includes ten recent peer-reviewed studies published between 2022 and 2024. Comparative analysis, source analysis, conceptual synthesis, typologization, and analytical generalization were applied. The analytical section reconstructs a closed-loop control architecture, identifies the most stable implementation sequence, and formulates decision rules for process monitoring and intervention. The proposed framework can be used for early design screening, fixture planning, measurement planning, digital model calibration, and production-stage capability governance in safety-critical assembly programs.*

Keywords *— **six sigma, dimensional control, aerospace assembly, digital twin, tolerance allocation, geometry assurance, metrology, assembly variation, capability control, safety-critical manufacturing.***

I. INTRODUCTION

Safety-critical aerospace assemblies are exposed to a form of dimensional risk distinct from conventional tolerance management. A local deviation in a locating feature, a fixture point, or a compliant mating surface does not stay local for long. It moves through the assembly chain, changes the contact condition, alters the fit at downstream interfaces, and in severe cases affects sealing, aerodynamic contour, vibration behavior, or service reliability. Recent aerospace studies describe this shift directly. Current aerospace research treats dimensional control as a problem of deviation propagation, in-process observability, and closed-loop correction within digital production environments.

This study develops an analytical framework for Six Sigma dimensional control in safety-critical assemblies used in emerging aviation technologies. To achieve this goal, three objectives are being solved. The first identifies the main deviation sources and control layers that structure dimensional governance in aerospace assembly. The second determines how metrology and digital twin environments convert static tolerance planning into a closed control cycle. The third formulates an implementation logic for selective tolerance tightening, capability verification,

and production-stage monitoring. The novelty of the article lies in combining recent studies of aerospace deviations, metrology-driven assembly simulation, and tolerance-cost research within a single six sigma framework for aviation use cases.

The working hypothesis states that Six Sigma dimensional control in aerospace assembly achieves practical value only when tolerance allocation, deviation prediction, measurement feedback, and intervention rules operate as a single closed architecture.

II. MATERIALS AND METHODS

The source corpus comprises 10 peer-reviewed studies published between 2022 and 2024. The selection logic followed three filters. First, each source had to address one of the following: aerospace assembly, aerospace-oriented digital twins, metrology-supported geometry assurance, or tolerance-cost control in high-precision mechanical assemblies. Second, the source had to contribute to at least one of the three research objectives fixed in the Introduction. Third, the set as a whole had to cover the full control chain, from deviation-source modeling to monitoring and capability governance. The final corpus covers five connected clusters: deviation accumulation in aircraft structures, metrology-supported virtual assembly, digital twin control for aerospace or precision assembly, tolerance-cost optimization, and geometry assurance under individualized or data-driven production conditions [1–10].

The study applies comparative analysis to align heterogeneous engineering studies within one control problem, source analysis to isolate transferable technical propositions, conceptual synthesis to reconstruct a closed-loop dimensional architecture, typologization to separate design, measurement, digital-model, and decision layers, and analytical generalization to translate literature findings into an implementation-ready framework for safety-critical aerospace assembly.

III. RESULTS

The first recurring finding concerns the source structure of dimensional risk. Recent aerospace studies do not treat deviation as an isolated geometric error confined to a single part. Research on multi-level assembly of an aircraft rear fuselage frame models deviation as the combined result of manufacturing deviation, fixture positioning deviation, and assembly contact deviation, all of which interact during station-to-station transfer [9]. Work on fabricated aerospace assemblies with digital twins reaches a similar position from a different path. In-line manufacturing data and non-nominal welding simulation are used to individualize assembly operations, thereby reconstructing the effective geometry from actual part conditions [3]. The same logic applies in aero-engine casing research, where tolerance analysis is linked to performance instability [4]. The authors show that the control problem extends beyond isolated geometric variation. Engineers manage coupled deviation pathways.

That shift has immediate consequences for Six Sigma thinking. In aerospace assembly, sigma performance has to be interpreted through structural sensitivity, fixture influence, contact closure, and performance-critical interfaces. In the rear fuselage frame model, rigid and compliant deviations are treated differently because weakly rigid parts do not accumulate error in the same manner as rigid components [9]. In the aero-engine casing study, assembly tolerance is evaluated through its influence on dynamic performance instability, thereby tying dimensional governance to product behavior under operating conditions [4]. The practical implication is direct. A six sigma framework for aviation assemblies cannot be reduced to a late-stage capability calculation. It needs a prior map of structural sensitivity, fixture influence, contact closure, and performance-critical interfaces.

A second line of findings concerns measurement and virtual representation. The recent literature moves away from tolerance allocation as a one-time design act. A metrology-aided virtual approach for aeronautical product assembly uses measurement data to drive assembly simulation before final mating decisions are locked in [5]. A digital twin method for complex product assembly extends this logic by integrating virtual and physical information into a progressively deduced tolerance-allocation scheme designed for high-precision aerospace components [10]. A related study on digital twin-based quality management in aerospace assembly links process

disturbances, abnormality prediction, and cause identification within a single quality management environment [8]. The authors reach the same conclusion. Dimensional control becomes more reliable when the digital model is updated by the measured state of the physical assembly.

This transition from static planning to closed control is reinforced by research outside the narrow aerospace label but still inside precision assembly. Online geometry assurance through digital twin calibration treats control and calibration as connected tasks, with the virtual model corrected during the assembly process [6]. A broader review of digital twins in aircraft production and MRO argues for fit-for-purpose twins, which means that the twin gains industrial value when it is tied to a sharply bounded operational task such as assembly planning, quality assurance, or maintenance support [7]. A perspective on digital twins in mechanical and aerospace engineering reaches a similar conclusion and frames them as capability-driven constructs [2]. Read together, these studies clarify the architecture needed for six sigma dimensional control. The digital twin is the medium through which measurement, prediction, and intervention are connected.

The literature also clarifies where cost enters the control chain. Tolerance tightening remains expensive, and recent work does not treat that expense as an afterthought. One 2024 study estimates cost reduction by tolerance optimization and shows that allocation decisions shape manufacturing choices in measurable ways [1]. A 2023 cost-based optimization study in mechanical assemblies builds the same argument in a more general design setting and frames tolerance development as a balance between manufacturing cost, non-quality cost, and functional constraints [8]. In aerospace assembly, that balance becomes stricter because some constraints are non-negotiable. The aero-engine casing study links tolerance choices to final dynamic performance [4]. In such cases, it moves into a ranked decision structure, where performance-critical interfaces receive tighter control while secondary contributors are managed with wider bands.

A comparison of four studies sharpens this point. The aircraft rear fuselage study identifies multiple deviation sources and their transfer rules [9]. The metrology-aided aeronautical assembly study brings measured geometry into the simulation loop [5]. The online geometry assurance study adds calibration and feedback control [6]. The 2024 data-driven tolerance-allocation study turns that closed loop into an explicit deduction mechanism for later-stage tolerance decisions [10]. The progression across these sources occupies a distinct layer of the same architecture. The first explains where deviations form and accumulate. The second captures the physical state before irreversible assembly steps. The third updates the twin and supports intervention. The fourth reallocates dimensional control based on real production evidence. Six sigma fits this sequence because sigma performance depends on all four layers.

Figure 1 condenses this result into a control chain reconstructed from the literature. It should be read as a closed architecture instead of a linear checklist.

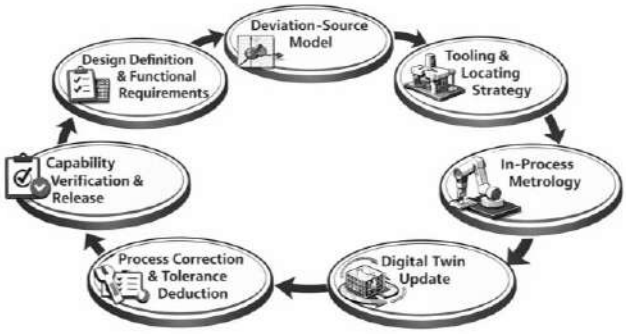


Fig. 1: Closed-loop architecture of six sigma dimensional control for safety-critical assemblies, adapted from [3; 5–7; 9; 10]

Recent studies call for early identification of dominant contributors, online observation of real geometry, calibration of the virtual model, and selective intervention in the small subset of variables that control assembly precision. In applied engineering language, that logic is close to the one hard-point sensitivity, selective tightening, and cost-aware robustness screening. The FISITA paper describes six sigma achievement through the identification of critical hard points, tolerance optimization in 3DCS, and controlled cost impact, reporting a 34% cost reduction under optimized tolerance tightening and noting that further reductions follow from relaxing non-critical hard points.

IV. DISCUSSION

The literature supports a closed-loop model of dimensional governance, but deployment fails when aerospace programs treat the loop as a stack of disconnected specialist tasks. Design engineers define tolerances. Tooling teams define locators. Metrology teams collect data. Quality teams calculate capability. Digital teams maintain the twin. This division of labor is manageable only as long as deviations remain small and local. Safety-critical assemblies operate under a tighter margin. The same geometric disturbance touches fit, load path, surface continuity, sealing, and inspection planning. A working model has to assign each decision to a specific stage and bind it to a clear escalation rule.

Table 1 structures that sequence. The purpose of the table lies in separating the design-stage questions from the production-stage questions. Without that separation, organizations tend to tighten tolerances too late, or measure the wrong feature too early.

Table 1. Implementation logic for six sigma dimensional control in safety-critical aerospace assembly

Lifecycle stage	Primary decision	Dominant evidence	Release criterion	Main output
Concept definition	Identify safety-critical interfaces and failure-sensitive dimensions	Functional architecture, load paths, sealing and alignment requirements	Agreement on critical interface map	Initial control boundary
Detailed design	Rank deviation contributors and define tolerance priorities	GD&T scheme, interface sensitivity, compliance assumptions	Sensitivity-ranked tolerance plan	Targeted tolerance architecture
Tooling planning	Select locating and clamping logic with minimum amplification	Locator layout, access constraints, deformation risk, fixture repeatability	Approved locating strategy	Robust fixture concept
Pre-production validation	Check agreement between predicted and measured deviation behavior	Trial metrology, virtual assembly, early capability estimates	Acceptable prediction fidelity	Calibrated digital model
Serial production	Maintain capability at critical interfaces	In-process measurements, drift patterns, and correction history	Stable capability inside the release band	Controlled production state
Change management	Decide whether drift requires compensation or redesign	Recurring escapes, tooling wear, engineering changes, and twin mismatch	Escalation by source severity	Updated control baseline

The table points to a practical conclusion. Six sigma in aerospace assembly starts long before Cp or Cpk enters a dashboard. The decisive move occurs earlier, when the program identifies which dimensions govern function and which ones merely consume engineering attention. Once that distinction is made, measurement planning becomes sharper and intervention becomes cheaper.

A second implementation problem concerns metrics. Many-dimensional dashboards collect everything that can be measured and still fail to support action. A usable control system separates symptom metrics from source metrics, and it distinguishes model

quality from process quality. Table 2 translates that requirement into a compact monitoring set.

Table 2. Monitoring metrics for closed-loop dimensional governance

Metric family	What is monitored	Operational meaning	Typical trigger
Interface metrics	Gap, flush, coaxiality, positional error, clearance spread	Direct evidence of assembly acceptability	Repeated proximity to the specification edge
Source metrics	Locator drift, fixture offset, clamp instability, local deformation	Probable origin of downstream error growth	Persistent source dominance
Model-fidelity metrics	Difference between predicted and measured deviation, update latency, and residual pattern	Trust level of the digital twin	Stable mismatch between model and shop-floor state
Capability metrics	Sigma level, spread stability, tail behavior, defect recurrence	Readiness for release and continuity of control	Capability decline or unstable dispersion
Economic metrics	Rework hours, scrap, inspection burden, maintenance effort	Cost of the current control regime	Rising cost without capability gain
Learning metrics	Repeated escapes, time to isolate the root cause, recurrence of known patterns	Organizational retention of dimensional knowledge	Return of previously corrected failure modes

Table 2 organizes monitoring by the mechanisms that generate deviation, not solely by defect totals. Interface metrics answer whether the build passes. Source metrics answer where to intervene. Model-fidelity metrics answer whether the twin remains trustworthy enough for closed-loop decisions. Economic metrics protect the program from indiscriminate tightening. Learning metrics reveal whether the organization is carrying dimensional knowledge forward or solving the same problem again under a new label.

Aviation programs gain the highest return when they adopt a ranked intervention logic. The program benefits from ranking interfaces by their contribution to functional risk. The upper tier requires tighter tolerance governance, richer metrology, and faster digital updates. Dimensions with lower influence remain under a lighter monitoring regime. This ranked-logic pattern aligns with the engineering approach, where critical hard points were tightened to achieve six sigma, while non-critical hard points were designated for later cost relief. In an aerospace setting, the same logic supports cleaner trade-offs between reliability, inspection load, and tooling cost.

The broader implication concerns governance culture inside engineering teams. A closed-loop dimensional system requires disciplined alignment between four elements: the deviation source map, the measurement strategy, the twin update rule, and the intervention threshold. Programs that keep those four elements synchronized are more likely to detect amplification early, intervene with precision, and preserve manufacturability without turning every tolerance into a premium requirement.

V. CONCLUSION

Safety-critical aerospace assembly demands a wider understanding of dimensional control than nominal tolerance assignment can provide. Recent studies identify deviation propagation, fixture influence, measurement feedback, and digital synchronization as central to the problem. Local geometric error acquires system significance only through transfer,

amplification, and interaction with real assembly conditions.

Metrology-supported virtual assembly and digital twin calibration form the operational core of Six Sigma control. Once the geometry is measured, the virtual model is updated during production, and tolerance planning stops being a static design exercise and becomes part of a closed decision cycle. That cycle supports earlier diagnosis, sharper intervention, and better alignment between design intent and build reality.

Selective tightening offers the most credible path for implementation. The strongest gains come from ranking critical contributors, tightening them with discipline, and preserving wider tolerance space where contributions to risk remain low. The hypothesis stated in the Introduction is therefore confirmed. Six sigma dimensional control in emerging aviation technologies attains practical engineering value when tolerance allocation, deviation prediction, measurement feedback, and intervention rules are integrated into a single closed-loop framework.

REFERENCES

[1] Armillotta, A. (2024). Estimation of cost reduction by tolerance optimization. *The International Journal of Advanced Manufacturing Technology, 134*(3–4), 1379–1393. https://doi.org/10.1007/s00170-024-14227-x

[2] Ferrari, A., & Willcox, K. (2024). Digital twins in mechanical and aerospace engineering. *Nature Computational Science, 4*(3), 178–183. https://doi.org/10.1038/s43588-024-00613-8

[3] Hultman, H., Cedergren, S., Wärmefjord, K., & Söderberg, R. (2022). Predicting geometrical variation in fabricated assemblies using a digital twin approach, including a novel non-nominal welding simulation. *Aerospace, 9*(9), 512. https://doi.org/10.3390/aerospace9090512

[4] Kang, H., Li, Z.-M., Liu, T., Yuan, W., & Wu, Y. (2022). A novel method to design tolerance of aero-engine casing by integrating 3-D assembly tolerance with performance instability. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 236*(8), 1052–1070. https://doi.org/10.1177/09544054211060917

[5] Kortaberria, G., Mutilba, U., Eguskiza, J., & Martins, J. (2022). Simulation of an aeronautical product assembly process driven by a metrology aided virtual approach. *Metrology, 2*(4), 427–445. https://doi.org/10.3390/metrology2040026

[6] Moenck, K., Rath, J.-E., Koch, J., Wendt, A., Kalscheuer, F., Schüppstuhl, T., & Schoepflin, D. (2024). Digital twins in aircraft production and MRO: Challenges and opportunities. *CEAS Aeronautical Journal, 15*(4), 1051–1067. https://doi.org/10.1007/s13272-024-00740-y

[7] Sjöberg, A., Önnheim, M., Frost, O., Cronrath, C., Gustavsson, E., Lennartson, B., & Jirstrand, M. (2023). Online geometry assurance in individualized production by feedback control and model calibration of digital twins. *Journal of Manufacturing Systems, 66*, 71–81. https://doi.org/10.1016/j.jmsy.2022.11.011

[8] Umaras, E., Barari, A., Horikawa, O., & Tsuzuki, M. S. G. (2023). Dimensional tolerances in mechanical assemblies: A cost-based optimization approach. *Applied Sciences, 13*(16), 9202. https://doi.org/10.3390/app13169202

[9] Zhang, H., Li, Y., Xue, D., Tong, X., Gao, B., & Yu, J. (2024). Digital twin technology facilitates precision improvement in complex product assembly: A progressive deduction method of data-driven tolerance allocation. *Advanced Engineering Informatics, 62*, 102790. https://doi.org/10.1016/j.aei.2024.102790

[10] Zheng, Y., Huang, X., Wang, M., & Hu, P. (2023). Source and accumulation analysis of deviation during multi-level assembly of an aircraft rear fuselage frame. *Applied Sciences, 13*(17), 9914. https://doi.org/10.3390/app13179914

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.21

Problems of Scaling Large Language Model Inference During Simultaneous Operation of Multiple Autonomous Agents

Kapil Verma

Software Engineer, Google, Mountain View, CA, USA

Received: 21 Mar 2026; Received in revised form: 22 Apr 2026; Accepted: 25 Apr 2026; Available online: 29 Apr 2026

Abstract *– The article examines the transformation of large language models from a single-query, single-response regime to a multi-agent configuration, in which a single external stimulus generates a tree of dependent calls to the model, and analyzes the specific constraints on scaling inference. The relevance of the study is determined by the proliferation of LLM-based agents and the growing share of workloads in which not the total number of tokens but the cadence of short iterations is decisive. The objective is to identify causal bottlenecks that determine throughput and tail latencies during the concurrent operation of multiple autonomous executors. On the basis of an analytic–synthetic review of 11 sources, a framework is proposed that shifts the unit of analysis from an individual response to a chain of dependent micro-steps, interpreted as competing job classes. The scientific contribution consists of systematizing the role of the KV cache as a dynamic scarce resource, introducing the phenomenon of contextual inflation, and linking these effects to batching policies, service fairness, distributed inference, and step routing across models of different sizes. It is shown that bottlenecks in multi-agent systems shift from arithmetic performance to memory, attention-state management, stopping discipline, and context engineering, while tail and network latencies take on the character of cascading lockups; the necessity of role-dependent token budgets and carefully designed eviction and state-folding strategies is substantiated. The article is intended for researchers and engineers developing LLM-based multi-agent systems and the infrastructure for their operation.*

Keywords *–* ***large language models, multi-agent systems, inference scaling, KV cache.***

I. INTRODUCTION

In recent years, large language models have ceased to function merely as one interlocutor per query and increasingly act as the computational core of systems in which multiple autonomous software executors jointly solve a task, partitioning planning, search, verification, and answer construction into specialized roles [1]. This transition has become practical for two reasons. First, the same model can support heterogeneous cognitive functions if the input–output format is rigidly fixed and the role is imposed via context, reducing development costs compared with training separate models for each module [2]. Second, collective organization enhances robustness: errors of one executor are partially compensated by a critic, tool-based verification, and repeated refinement cycles, particularly evident in complex multi-step tasks where the priority is not elegant text but consistent decision-making [3]. This evolution from single executors to multi-agent configurations and communities is systematized in recent survey works, which emphasize that interaction among specialized roles has become the canonical engineering pattern for improving the quality and controllability of model behavior [4].

The load induced by the simultaneous operation of many autonomous executors differs qualitatively from that of an ordinary chat, where each

user query is typically completed by a single, relatively long response. In a multi-agent system, one external query triggers a cascade of internal calls to the model: short decision-making commands are interleaved with intermediate plan generation, clarifying tool calls, and repeated checks, and these chains are launched in parallel and often synchronize on shared resources [5]. As a result, not only does the average number of calls increase, but the variance in context and response length grows: alongside long inputs (action logs, tool outputs, replanning) appear short one- or two-step replies, followed again by extended text fragments. This heterogeneity makes tail latencies critical and amplifies queue-blocking effects, forcing serving systems for generative models to abandon naïve run-to-completion schemes and to seek more flexible computation-scheduling strategies [6].

But there is also a subtler difference: in chat mode, the user is mostly interested in the final result. In multi-agent mode, on the other hand, there is an individual cost per step, and since each step depends on the previous one, any one-step slowdown affects all steps. As a consequence, since the bottleneck is not in how many words are generated but in how many small steps with different contexts can be taken, high performance and low latency cannot be achieved simultaneously. This change in the unit of work, from a response to a sequence of interdependent steps, provides the motivation for further analysis of inference-scaling challenges under concurrent operation of a multitude of autonomous executors [7].

II. MATERIALS AND METHODOLOGY

The empirical material for the study comprises 11 peer-reviewed sources selected as a representative cross-section of two interrelated lines of work: (a) survey and architectural studies of multi-agent LLM configurations and their typical operating loops, and (b) systems papers on inference serving, where latency, batching, and memory management are treated as primary scaling constraints. As the theoretical foundation, surveys on LLM-based agents and multi-agent systems were used that document the shift from single dialogue to orchestration of specialized roles and iterative plan–act–observe–reflect cycles (Li et al., 2024; Xi et al., 2025; Chowa et al., 2026), as well as works that demonstrate the practical effectiveness of distributing cognitive functions via context and prompt-engineering strategies (Gozzi & Di Maio, 2024) and the robustness effects of collective verification/error smoothing in LLM-driven agents (Hu et al., 2025). Applied multi-agent case studies were not treated as benchmark sources, but as material for reconstructing a characteristic load profile: a cascade of internal calls, heterogeneity of context and response lengths, and parallel execution branches with synchronization on shared resources (Lv et al., 2025; Kim et al., 2025; Jimenez-Romero et al., 2025).

The methodology had an analytic–synthetic character and was constructed as a linkage of three procedures aimed at identifying causal bottlenecks in the problem of inference scaling under concurrent operation of multiple autonomous agents. First, a conceptual analysis of the unit of work in a multi-agent system was performed: instead of the single query, single response metric, a chain of dependent micro-steps with different urgency and cost was examined, which accentuates the importance of tail latencies and makes queue planning a central object of study (Li et al., 2024; Xi et al., 2025). Second, a comparison of systems-level mechanisms for serving LLMs was conducted from the perspective of their behavior under heterogeneous agent load, with flexible scheduling/speculative execution strategies and their potential incompatibility with short, format-strict agent iterations considered (Miao et al., 2024; Zhao & Wang, 2024). Third, a structural analysis of resource constraints was conducted through the lens of memory and attention state: the KV cache and its allocations were treated as a dynamic scarce resource that determines the upper bound on parallelism and the predictability of latencies, with the principles of PagedAttention used as a canonical model of how memory management becomes a system-forming factor in inference serving (Kwon et al., 2023).

III. RESULTS AND DISCUSSION

Multi-agent systems based on large language models are typically constructed as a composition of specialized executors that sequentially or in parallel perform different cognitive functions: one forms a plan and decomposes the goal into steps, another

executes steps and invokes external tools, a third acts as a critic and attempts to refute intermediate conclusions, a fourth retrieves relevant information from external memory or a search environment, and a fifth monitors process state and enforces constraints on time, cost, and safety. Such functional decomposition is described in recent survey studies as a stable architectural technique that replaces a universal interlocutor with a system of interacting roles, where quality is achieved not only through model parameters but also through the structure of interaction and feedback [8]. In practical scenarios, this manifests in software engineering, research workflows, simulations, and tool-use tasks, where agent behavior must be reproducible and auditable rather than merely persuasive [9].

The orchestration of such executors can be centralized, with a single dispatcher allocating steps and budgets and managing queues of calls to the model, or distributed, with executors exchanging messages and making decisions locally, coordinating actions through an interaction protocol. In both variants, an iterative plan–act–observe–reflect loop dominates, in which observation includes the results of external calls, and reflection involves replanning in light of new data and criticism. A survey on autonomous agents emphasizes that the presence of a closed loop with tools and memory turns the language model into a component that functions as a control device: it must not only generate text, but also maintain state, respond to environmental signals, and adjust strategy [8]. This is important for scaling inference because the loop generates many short control calls interleaved with rare but long episodes of generation and summarization, and also creates long-lived dialogues among executors.

The load profile under parallel agent operation is characterized by a single external query unfolding into an internal call tree: a planner may spawn several execution branches, each branch may initiate a series of tool clarifications, and the critic may generate additional verification and alternative-hypothesis queries. As a consequence, not only does the aggregate call frequency grow, but the dispersion of input and output lengths increases: short commands and probing utterances are interspersed with long contexts containing action logs and embedded tool results [10]. At the computational level, this splits into two asymmetric phases: input-context processing, in which the model iterates over all input tokens, and step-wise generation, in which each subsequent token is computed sequentially. The paradox is that multi-agent organizations often make the first phase expensive due to context growth, while simultaneously increasing sensitivity to the second phase, because latency at every short step stretches the entire decision-making loop and worsens system-level tail latency.

The primary constraints quickly shift from the raw arithmetic performance of the accelerator to memory and its bandwidth, since accelerating step-wise generation requires storing intermediate attention states in a key–value cache. In an experimental configuration on an accelerator with 40 gigabytes of video memory, it is shown that this cache can occupy more than one-third of the device memory even in typical serving regimes, and inefficient management leads to severe fragmentation and loss of usable capacity: measurements for existing systems revealed that the actually utilized memory for token states amounted to only about one-fifth to two-fifths of the allocated cache volume, which directly limits the degree of parallel processing and, accordingly, the service throughput [11]. In multi-agent mode, the effect is exacerbated by long-lived sessions and unpredictable generation length: memory grows and breathes over time, allocations become irregular, and memory bandwidth increasingly becomes the limiting factor, because step-wise generation forces the system to access large state structures repeatedly, and any unfortunate queue-scheduling decision amplifies lockups and latency spikes. The multi-agent LLM system challenges are systematized in Figure 1.

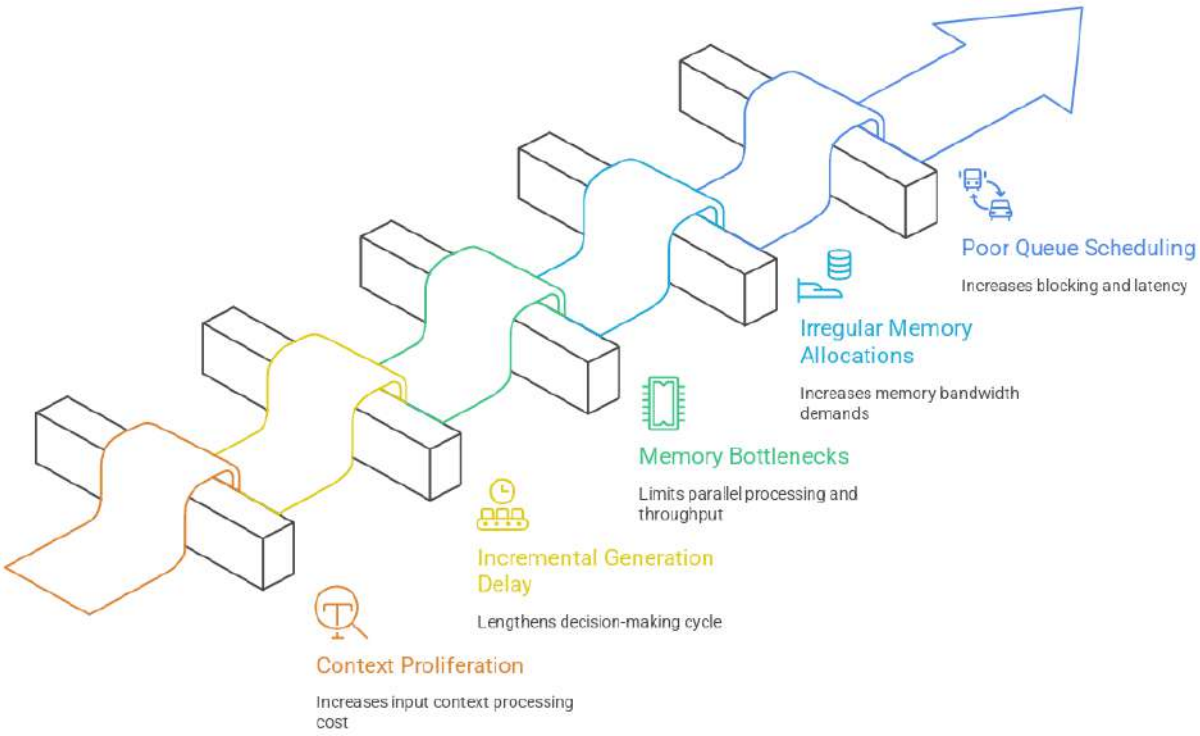


Fig. 1. Multi-Agent LLM System Challenges

With the parallel execution of numerous executors, model inference becomes a dispatching problem in which heterogeneous requests compete on a single accelerator, while the objective is twofold: to maximize aggregate throughput and to maintain step latency within tight bounds. Batching requests increases the efficiency of compute-unit utilization, but the price of this efficiency manifests itself as waiting time: short steps are forced to accumulate companions, thereby elongating the closed decision-making cycles of agents. As a result, the same batching mechanism can improve average performance while degrading perceived system reactivity, especially when the agent logic is structured as a chain of dependent steps, where each subsequent step is launched only after the previous response has been received.

A critical subtlety is that input-context processing and step-wise generation behave as distinct job classes and require different batching rules. Input processing benefits from large batches because computations proceed as dense matrix operations that fully utilize the accelerator, whereas step-wise generation is sensitive to waiting, since it consists of many short iterations that accumulate latency across steps. If these modes are mixed without differentiation, queues form in which heavy requests with long inputs block the flow of small iterations, and the system experiences a head-of-line blocking effect: fast jobs idle behind slow ones, even though the resource could have served them in a different order. In the limiting case, some executors begin to starve because their small requests are constantly displaced by heavier neighbors or by more active agents that generate a higher volume of calls.

A separate issue is fairness, because in a multi-agent environment, requests are incomparable in both cost and importance. One executor may generate long contexts and long responses, consuming memory and time; another may issue short control commands that are critical to the entire system; a third may periodically initiate monitoring or verification steps in which the cost of error exceeds the cost of delay. Therefore, simple fairness by number of requests fails: policies are required that account for token budgets, the cost of memory for state, step urgency, and the agent's role in the overall process. Without such policies, the system either becomes unaffordable or loses robustness as critical decision branches begin to experience systematic delays.

Growth in context length amplifies all of the aforementioned effects and introduces the

phenomenon of contextual inflation, which is particularly characteristic of autonomous executors. During operation, action logs, results of external calls, intermediate plan versions, and self-critiques accumulate, and there is a tendency to feed all of this back into the model for reliability. However, every redundant input fragment increases processing costs and inflates the state required for generation, while also bringing the system closer to the context-window limit. Once the window overflows, the developer must either discard data and risk losing important decision justifications or compress them and risk semantic distortion; in addition, an overly saturated input degrades quality because the model diffuses its attention and more frequently confuses current facts with obsolete traces of previous steps. Consequently, compression and deduplication become prerequisites for scaling rather than cosmetic refinements: the system must be able to transform a long history into a compact state, store knowledge outside the primary input, and retrieve it only when necessary; otherwise, computational resources will inevitably be consumed by textual noise.

The use of external tools disrupts the computational flow and violates the assumption of a smooth queue. Pauses for external services, which exhibit their own latencies and failures, appear between agent steps; upon completion of such calls, many executors often return to the model simultaneously, creating load spikes. At the infrastructure level, a state-management dilemma arises: it is possible to retain intermediate generation state to resume the dialogue quickly, but then memory requirements grow, and long-lived session management becomes more complex; alternatively, state can be released and the input reprocessed from scratch, at the cost of increased compute and context-reconstruction latency. In multi-agent systems, this dilemma becomes especially acute because pauses and spikes are the norm, and the choice of strategy directly determines whether the system remains resilient to peak loads and can preserve reactivity without uncontrolled cost growth. Challenges in parallel model execution are shown in Figure 2.

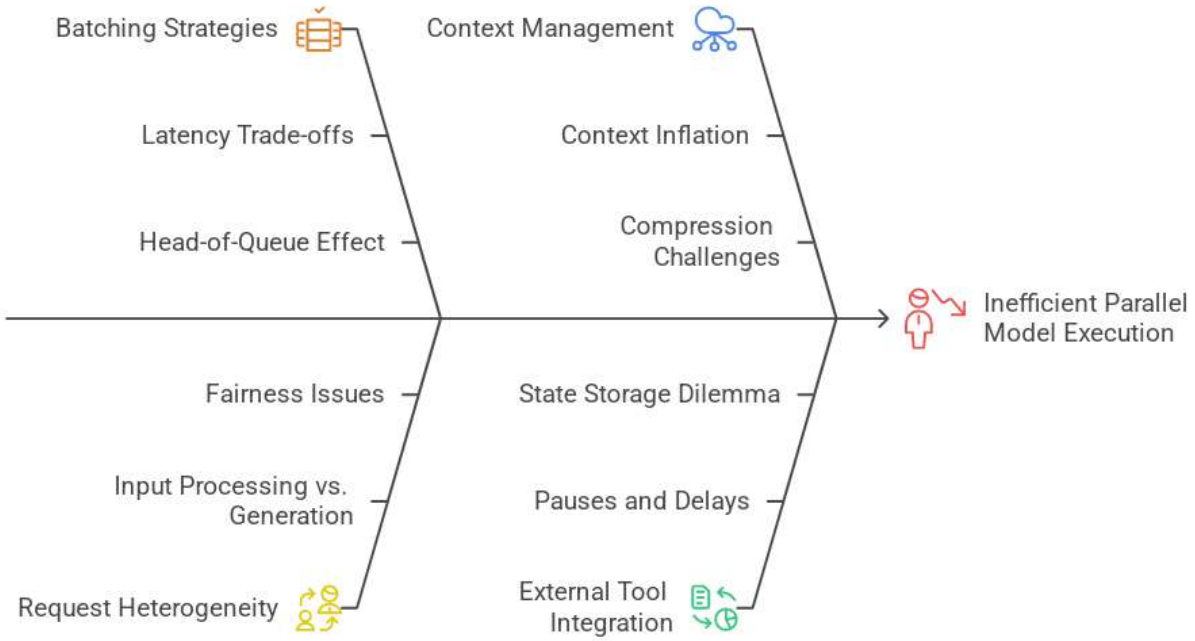


Fig. 2. Challenges in Parallel Model Execution

When the parallel operation of executors saturates memory and the queue on a single accelerator, a natural step is to distribute computation across multiple devices and nodes. However, in distributed mode, inference ceases to be a local operation and becomes an alternation of compute and communication phases, in which intermediate representations must be repeatedly transmitted among participants. Under layer-wise partitioning, some nodes are forced to wait for neighbors to complete, and

under parameter partitioning, each step requires coordination of partial results, with the cost of this coordination growing as the number of participants increases. For multi-agent systems, this is particularly problematic because they generate many short steps: the finer the step granularity, the higher the fraction of time spent on coordination rather than useful computation, and the harder it becomes to hide communication behind computation.

The network contributes not only to increased average latency but also to increased variability, which manifests as tail latencies and introduces instability into closed-loop control cycles. Even if the mean response time is acceptable, rare but long delays break synchronization among executors: one agent waits for the critic's confirmation, the critic waits for a tool result, and the dispatcher holds resources, unable to determine whether a branch will complete. In such a system, tail latency becomes not a mere statistic but a mechanism of cascading slowdown, in which an individual stall escalates into a series of lockups, reinforcing queueing effects and increasing cost due to retries and context expansion. The problem is exacerbated by load spikes: agent systems often exhibit staged behavior, and when several branches simultaneously return from external calls, the distributed infrastructure experiences a synchronous influx of requests that is poorly absorbed because scaling computation requires warm-up, state distribution, and queue stabilization.

Against this backdrop, model-level optimizations are treated as a means to alleviate pressure on compute and memory, but they inevitably embed a trade-off between speed and quality. Reducing parameter precision frees memory and increases parallel-processing density; however, in multi-step chains, even minor distortions can accumulate: an early inaccuracy in the plan leads to an incorrect tool choice, after which an erroneous result is recorded in the context, and the critic then evaluates an already distorted picture and reinforces false confidence. Thus, a gain in speed may turn into an increase in the number of iterations and in total cost, because the agent system compensates for quality degradation with additional refinement and self-verification cycles.

A separate class of methods attempts to accelerate generation by predicting fragments of the answer more quickly and then confirming them via more precise computation. This yields benefits when responses are long and predictable, but becomes less meaningful when agent steps are short, frequently interrupted by tool calls, and require strict adherence to format. In a multi-agent environment, many responses are control commands or compact conclusions, where the savings from accelerated generation are minimal, while the overhead of verification and correction becomes noticeable. Moreover, such methods complicate queue planning because they introduce yet another source of heterogeneity: some requests are accelerated, others are not, and the dispatcher must anticipate this in advance. Otherwise, there will be no net benefit.

For this reason, routing requests among models of different sizes is increasingly employed, with simple steps handled by lighter models and complex, high-stakes decisions delegated to more powerful ones. At the architectural level, this resembles the division of labor among executors but is shifted within the computational layer: the system attempts not only to distribute agent roles but also to assign each step an appropriate computational depth. Nonetheless, this strategy requires a careful complexity criterion; otherwise, routing errors become systematic: a light model may take over a step that requires precision and thereby launch a correction cascade, or, conversely, a heavy model may expend resources on trivial actions and worsen latency for others. Ultimately, model-level optimizations cannot be considered in isolation from agent dynamics: their effect is determined by how they alter the number of iterations, the shape of the query tree, and the system's robustness to tail latencies in a distributed environment. The cycle of distributed agent system optimization is illustrated in Figure 3.

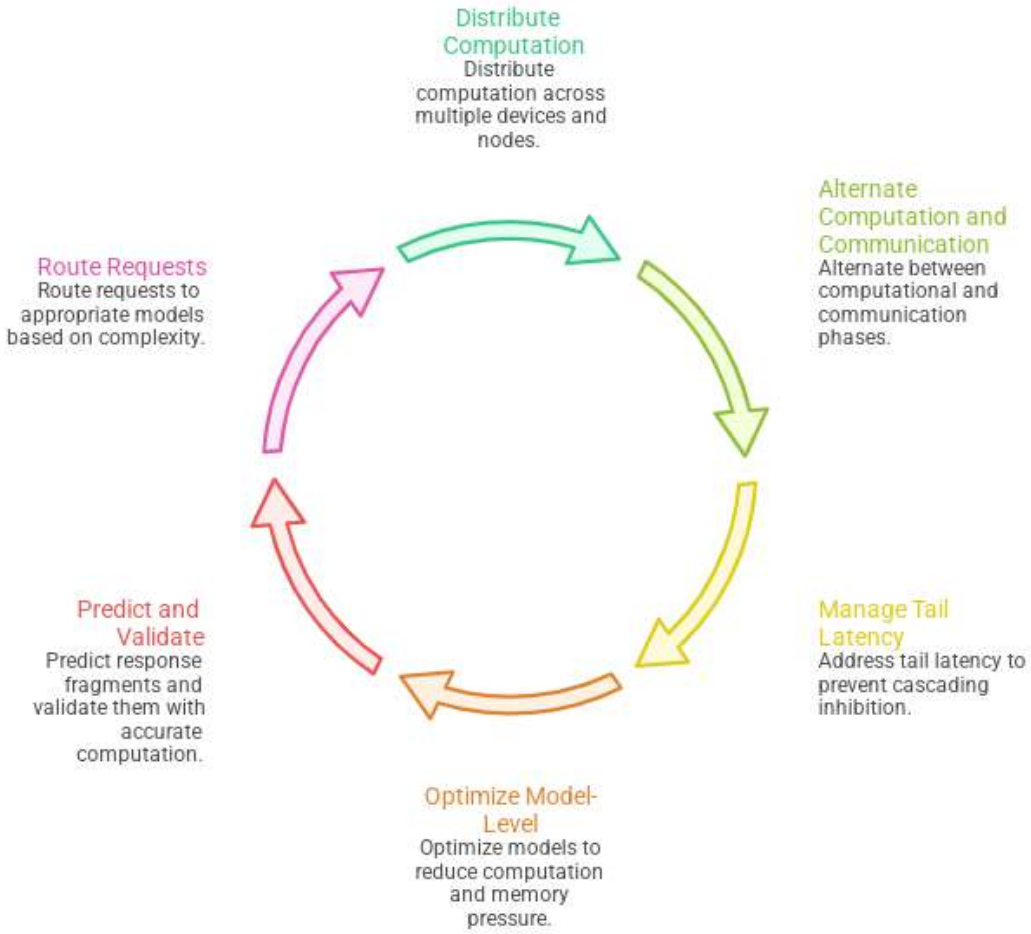


Fig. 3. Cycle of Distributed Agent System Optimization

Practice shows that scaling a multi-agent system begins not with accelerators but with stopping discipline. If executors are allowed to refine plans indefinitely, re-check themselves, and spawn new branches, any infrastructure improvement will be consumed by growth in the number of steps. It is therefore useful to predefine limits on the number of iterations and on the volume of text that each executor may add to the shared state, and these limits should be role-dependent: planners and critics are dangerous precisely because they can inflate discussion, whereas action executors more often require short commands. It is important that the constraints be not only hard but also semantic: once its budget is exhausted, an agent must switch to result-folding mode or request an external signal. Otherwise, the system will simulate activity, accumulating context and latency without commensurate quality gains.

Memory policies must take into account that attention state is a resource no less scarce than compute and, in a multi-agent environment, also temporally unstable. It is impossible to keep all sessions warm and simultaneously expect predictable latency: some states must inevitably be treated as cold and evicted according to rules that minimize damage to control chains. Good eviction is not reducible to least-recently-used because an agent may return after a pause in an external tool and prove critical for completing an entire branch; therefore, priority should be determined by proximity to a decision point, expected continuation length, and the cost of reconstruction. Ideally, the system should distinguish in advance between sessions for which it is advantageous to preserve state and sessions that are cheaper to reconstruct from a compact description; otherwise, memory will be occupied by history that no longer influences future steps.

Context engineering in multi-agent systems must be organized as knowledge management rather than transcript accumulation. However, for action

logs, tool output, and intermediate plans, it can pay off to only keep information useful for the next decision step. Thus, deduplicating and keeping only canonical facts, constraints, and verifiable facts can prevent the remainder of the search tree from becoming overwhelming. The context for a module includes the current goal, current plan, key observations, tool status, and a brief decision history. Detailed traces should be stored only for loading on request. This reduces the effort required to process input, reduces the risk of corrupting the signal of interest with noise, and makes the system's behavior more repeatable, as each step is based on a concise representation.

The most typical errors arise from a desire to hedge with text. Overly long logs turn the context into a landfill, where the model begins to confuse causes and effects and to devote attention to details unrelated to the current step. Another antipattern is excessive reflexive cycling, where the critic and planner repeatedly discuss the quality of their reasoning, creating an illusion of reliability, but incurring high latency and cost, and degrading the final result through the accumulation of confluent formulations. Finally, loose formats allow agents to write whatever they want. Because there is no regularity to the content, the system cannot collapse state, extract facts, or queue requests. Any infrastructure-level optimization has to navigate the erratic growth of tokens and requests.

IV. CONCLUSION

Multi-agent deployment of large language models crystallizes a shift from the single-query, single-response paradigm to a regime in which a single external stimulus unfolds into an internal call tree, heterogeneous in context and response length, and coupled by stepwise dependencies. In this configuration, performance is no longer measured solely by the total number of tokens: the decisive factor becomes the rate of short iterations under high input variability, because the latency of each micro-step stretches the entire plan–act–observe–reflect loop, and tail latencies transform from a statistical artifact into a systemic blocking mechanism. From this follows the central engineering challenge of scaling: the need to maintain high compute utilization and predictable reactivity simultaneously, despite the fact that batching, advantageous for dense input processing, can undermine step-wise generation latency and provoke starvation of lightweight but critical requests in the vicinity of heavy contexts.

At the resource level, the primary bottleneck quickly becomes not only compute units but also memory, primarily due to the storage of attention states in a KV cache, which, in typical regimes, can occupy a significant share of video memory and suffer from fragmentation, directly limiting parallelism and service throughput. Attempts to escape local limitations by distributing workloads across devices introduce communication phases and network variability, and in multi-agent dynamics with numerous short steps, coordination overhead and rare but long delays begin to disrupt branch synchronization in a cascading fashion. Consequently, model- and infrastructure-level optimizations cannot be evaluated apart from agent dynamics: acceleration or compression inevitably collides with error accumulation along the chain and the risk of increasing iteration counts, while robustness hinges on stopping discipline, meaningful budgets, state-eviction policies, and context engineering in which memory is managed as knowledge rather than as an unbounded transcript.

REFERENCES

[1] X. Li, S. Wang, S. Zeng, Y. Wu, and Y. Yang, "A survey on LLM-based multi-agent systems: workflow, infrastructure, and challenges," *Vicinagearth*, vol. 1, no. 9, Oct. 2024, doi: https://doi.org/10.1007/s44336-024-00009-2.

[2] M. Gozzi and F. Di Maio, "Comparative Analysis of Prompt Strategies for Large Language Models: Single-Task vs. Multitask Prompts," *Electronics*, vol. 13, no. 23, p. 4712, Nov. 2024, doi: https://doi.org/10.3390/electronics13234712.

[3] J. Hu, Y. Dong, Z. Ding, and X. Huang, "Enhancing robustness of LLM-driven multi-agent systems through randomized smoothing," *Chinese Journal of Aeronautics*, p. 103779, Aug. 2025, doi: https://doi.org/10.1016/j.cja.2025.103779.

[4] Z. Xi *et al.*, "The rise and potential of large language model-based agents: a survey," *Science China Information Sciences*, vol. 68, no. 2, p. 121101, Jan. 2025, doi: https://doi.org/10.1007/s11432-024-4222-0.

[5] W. Lv, Q. Cao, and X. Liu, "A multi-agent classical Chinese translation method based on large language models," *Scientific Reports*, vol. 15, p. 40160, Nov. 2025, doi: https://doi.org/10.1038/s41598-025-23904-0.

[6] X. Miao *et al.*, "SpotServe: Serving Generative Large Language Models on Preemptible Instances," *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*, pp. 1112–1127, Apr. 2024, doi: https://doi.org/10.1145/3620665.3640411.

[7] Y. Zhao and J. Wang, "ALISE: Accelerating Large Language Model Serving with Speculative Scheduling," *Proceedings of the 43rd IEEE/ACM International Conference on Computer-Aided Design*, pp. 1–9, Oct. 2024, doi: https://doi.org/10.1145/3676536.3676659.

[8] S. S. Chowa *et al.*, "From language to action: a review of large language models as autonomous agents and tool users," *Artificial Intelligence Review*, vol. 59, no. 2, Jan. 2026, doi: https://doi.org/10.1007/s10462-025-11471-9.

[9] C. Jimenez-Romero, A. Yegenoglu, and C. Blum, "Multi-agent systems powered by large language models: applications in swarm intelligence," *Frontiers in Artificial Intelligence*, vol. 8, May 2025, doi: https://doi.org/10.3389/frai.2025.1593017.

[10] S. Kim, J. Ha, H. Lee, S. Park, and S. Cho, "Human-guided collective LLM intelligence for strategic planning via two-stage information retrieval," *Information Processing & Management*, vol. 63, no. 1, p. 104288, Jul. 2025, doi: https://doi.org/10.1016/j.ipm.2025.104288.

[11] W. Kwon *et al.*, "Efficient Memory Management for Large Language Model Serving with PagedAttention," *Proceedings of SOSP '23*, pp. 611–626, Oct. 2023, doi: https://doi.org/10.1145/3600006.3613165.

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.22

Study of Ligand-Assisted Co-precipitation and Structural Passivation in Semiconducting SnO_2 Nanostructures

Dr. Surender Kumar

Assistant Professor, Department of Physics, Govt. P.G. College for Women, Rohtak, Haryana, India.
E-mail: iitatmotivation@gmail.com

Received: 29 Mar 2026; Received in revised form: 24 Apr 2026; Accepted: 27 Apr 2026; Available online: 30 Apr 2026

Abstract *— This study investigates the synthesis, thermal evolution, and doping of pure, aluminum-doped (Al-doped), and silver-doped (Ag-doped) tin dioxide (SnO_2) nanoparticles. As a highly versatile wide-bandgap semiconductor, SnO_2 offers exceptional properties for advanced electronic and optical applications; however, its functional efficacy relies heavily on precise morphological control at the nanoscale. To achieve this, a bottom-up chemical co-precipitation method was employed using stannic chloride and sodium carbonate. A central focus of this research is the systematic evaluation of four structurally diverse capping agents, which were integrated to govern nucleation kinetics, confine crystal growth, and prevent nanoparticle agglomeration. Furthermore, targeted doping was performed by introducing aluminum hydroxide and silver nitrate during the precipitation phase to produce Al-doped and Ag-doped SnO_2 nanostructures, respectively. The thermal degradation profiles of the synthesized carbonate precursors were comprehensively evaluated utilizing Thermogravimetric and Differential Thermal Analysis in a nitrogen atmosphere from 28°C to 800°C. Thermal analysis identified the exact thresholds for complete volatile mass loss, which is critical for inducing porosity while preventing thermal sintering. Consequently, precise calcination parameters were established: pure and Al-doped precursors were annealed at 400°C, whereas the thermodynamic stability of the Ag-doped matrix necessitated a calcination temperature of 450°C. All samples were processed with a careful heating ramp of 5°C/min and held for an optimized duration of 3 hours. The low-temperature exothermic decomposition of the carbonaceous material successfully yielded phase-pure, highly crystalline semiconducting oxides. Ultimately, this work demonstrates that pairing scalable co-precipitation techniques with specific organic surface passivation and metal doping provides a robust framework for tailoring the structural and physical properties of advanced nanomaterials.*

Keywords *— **Morphological, Doping, Nanostructures, Degradation, Precursors.***

I. INTRODUCTION

The frontier of modern condensed matter physics and materials science lies in the deliberate manipulation of matter at the nanoscale. Tailoring the physical, optical, and electronic properties of semiconductors by manipulating their composition, morphology, and particle size has unlocked entirely new paradigms in device functionality. Transitioning a material from the bulk to the nano-regime intrinsically increases the surface-to-volume ratio, resulting in a high density of surface states that drastically alter the overall physical characteristics. Among these advanced materials, Tin Dioxide stands out as a paramount wide-bandgap, n-type semiconductor. By engineering pure and doped SnO_2 nanostructures, scientists can fine-tune its electrical conductivity, optical transparency, and catalytic reactivity, paving the way for next-

generation gas sensors, optoelectronic devices, and advanced antimicrobial coatings.

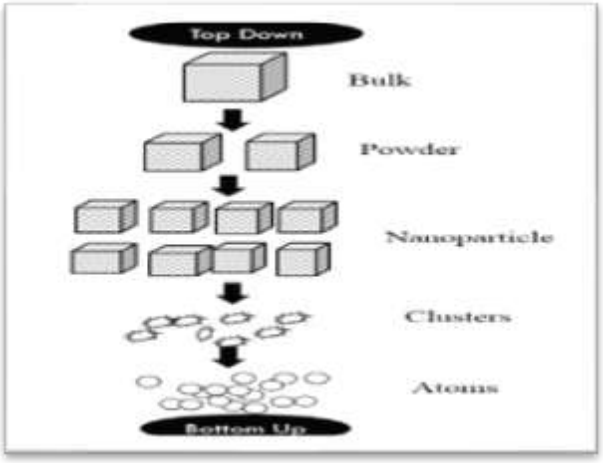
1.1. The Dichotomy of Nanofabrication

The synthesis of nanocrystals directly governs the manifestation of quantum confinement effects within the semiconductor lattice. The architectural strategies for nanomaterial synthesis are broadly bifurcated into two foundational methodologies:

1.1.1. The Top-Down Paradigm

The top-down approach relies on the systematic dismantling of bulk materials into nano-dimensional fragments. Attrition, milling, and lithographic slicing are quintessential examples. While industrially scalable, this approach inherently introduces substantial crystallographic damage and surface imperfections. These structural defects create rapid diffusion paths and localized electron traps, presenting formidable challenges for high-precision optoelectronic device fabrication.

1.1.2. The Bottom-Up Paradigm



Conversely, the bottom-up approach mimics natural molecular assembly, constructing nanostructures atom by atom, molecule by molecule, or cluster by cluster. Colloidal dispersion and vapor-phase nucleation are classic examples. This methodology is heavily favored in advanced physics and materials science because it yields highly homogeneous chemical compositions, pristine crystallographic surfaces, and a drastically reduced defect density.

Feature	Top-Down Synthesis	Bottom-Up Synthesis
Starting State	Bulk solid material	Atoms, ions, or molecules
Primary Mechanism	Mechanical or energy-driven cleavage	Chemical reactions or phase nucleation
Defect Density	Typically high (lattice strain, dislocations)	Generally low (highly crystalline)
Control over Shape	Limited	Excellent (via templates and capping agents)

II. SYNTHESIS TECHNIQUES

Several sophisticated bottom-up and hybrid schemes are deployed to synthesize nanometal oxides. Understanding the physical mechanisms behind these techniques is crucial for optimizing the resultant nanostructures.

- **Inert Gas Evaporation-Condensation:** A cornerstone technique since the 1930s, IGC involves the thermal evaporation of a metallic source within a highly controlled vacuum chamber (evacuated to 10^{-7} torr) backfilled with a low-pressure inert gas. Evaporated metal atoms migrate into the cooler gas via convective flow and diffusion. Kinetic energy is dissipated through collisions with inert gas atoms, leading to homogeneous nucleation. The resultant nanoparticles are harvested on a cold substrate in a highly sterile environment to prevent unwanted oxidation.
- **Mechanical Alloying:** A high-energy milling protocol where powder particles undergo successive cold welding, fracturing, and re-welding. This mechanical energy transfer introduces severe lattice strain and generates a high density of dislocations. The continuous refinement of grain sizes reduces diffusion distances, promoting solid-state alloying even at room temperature. MA is exceptional for producing supersaturated solid solutions and metastable quasi-crystalline phases.
- **Magnetron and Ion-Beam Sputtering:** Operating typically below 10^{-3} mbar, sputtering utilizes high-velocity inert gas ions to bombard a solid target, ejecting atomic clusters via momentum transfer. Differential pumping systems allow for reactions in specific atmospheres (e.g.,

introducing N_2 to form nitrides). The primary advantage of this physical vapor deposition method is the exact stoichiometric replication of the target material in the resulting nanoscale thin film or powder.

- **Pulsed Laser Ablation:** This highly non-equilibrium technique utilizes short, intense laser pulses (e.g., Nd:YAG at 532 nm or excimers at 248 nm) to irradiate a substrate, generating a localized high-energy plasma plume. The extremely brief heating duration (10-50 ns) allows for the synthesis of complex semiconductor nanoparticles that would otherwise thermally decompose under conventional vaporization.
- **Spray Pyrolysis:** An intermediate liquid-to-gas phase technique where solute-rich droplets are atomized and subjected to rapid thermal decomposition. As the solvent evaporates, the droplet shrinks, forcing supersaturation and subsequent nanoparticle precipitation within the micro-reactor of the droplet itself, often yielding unique hollow spherical morphologies. For precise control over nanoparticle morphology without the need for extreme vacuum or laser systems, wet-chemical synthesis routes are globally preferred in both research and industry.
- **Solvent Vaporization & Aerosol Techniques:** By utilizing spray arrays to create isolated microscopic droplets, researchers can circumvent the macroscopic segregation that often plagues slow-evaporating multi-component systems. Aerosol methods further utilize molecular clusters (micelles, liposomes) as nanoreactors to confine growth, followed by thermophoretic collection.
- **Hydrothermal Processing:** Conducted in specialized autoclaves, hydrothermal synthesis leverages the unique thermodynamic properties of water at elevated temperatures and high pressures (often near or past the critical point). This process bypasses the need for extreme high-temperature calcination, facilitating crystallization and oxidation simultaneously.

III. THE CO-PRECIPITATION THERMODYNAMICS

Among wet-chemical techniques, Co-precipitation remains the most versatile, scalable, and economical method for synthesizing pure and doped SnO_2 nanomaterials. The process relies on achieving supersaturation, leading to the simultaneous precipitation of multiple soluble species into an insoluble solid matrix.

3.1. Crystal Growth and Ostwald Ripening

The synthesis hinges on mastering the kinetics of nucleation versus growth. To isolate monodisperse quantum dots or nanoparticles, a brief, rapid burst of nucleation must be followed by a slow, diffusion-controlled growth phase. If growth kinetics are not strictly moderated, the system will naturally undergo Ostwald Ripening. This thermodynamic phenomenon is driven by the minimization of total surface free energy. . Because atoms on the surface of smaller nanoparticles are highly energetic and less stable, they tend to dissolve back into the solvent and re-deposit onto the surfaces of larger particles, which possess a more stable, lower surface-to-volume ratio.

3.2. Optimizing the Co-precipitation Matrix: Successful co-precipitation necessitates

> Complete insolubility of the resultant compound in the mother liquor.

> Rapid precipitation kinetics to ensure homogeneous nucleation.

> The use of highly volatile precipitating agents (e.g., organic ammonium carbonates or oxalates) that leave no contaminating residue during the subsequent thermal decomposition phase.

3.3. Surface Passivation: The Physics of Capping Agents

To counteract agglomeration and precisely dictate the final crystallographic architecture, capping agents are introduced during synthesis. These agents function via steric hindrance and electrostatic stabilization, selectively adsorbing onto specific crystal facets to retard their growth rate, thereby shaping the nanocrystal.

Chemical Classification	Structural Influence in SnO2 Synthesis
Amino-poly-carboxylic Acid	Acts as a strong hexadentate chelating ligand. It forms dense protective layers, tightly restricting particle size and ensuring high stability against agglomeration.
Cationic Surfactant	Forms micellar templates. The positively charged head group and long hydrophobic tail guide the anisotropic growth of the semiconductor.
Biopolymer (Nucleic Acid)	Serves as a highly complex biological soft template. The double-helix structure and phosphate backbone provide unique nucleation sites.
Polysaccharide	Acts as an eco-friendly, green stabilizing matrix. Its complex carbohydrate chains create a sterically hindered web that confines nanoparticle expansion.

The thermodynamic balance of the binding and unbinding kinetics of these agents is the absolute prerequisite for maximizing the quantum yield and engineering the density of states in the resultant nanostructure.

IV. DOPING SnO2 NANOPARTICLES

Pristine SnO_2 is highly effective, but intentional doping introduces new energy levels within the bandgap, radically expanding its functional utility.

Al-Doped SnO_2: Introducing Aluminum into the Tin lattice generates oxygen vacancies to maintain charge neutrality, or acts as a dopant that modifies charge carrier mobility and electrical conductivity.

Ag-Doped SnO_2: Silver doping profoundly influences the localized surface plasmon resonance and catalytic behavior of the semiconductor. Furthermore, silver acts as an electron sink, which significantly amplifies the material's ability to generate reactive oxygen species (ROS), making Ag-doped SnO_2 an exceptional candidate for advanced antibacterial and biomedical applications.

V. SYNTHESIS OF PURE AND DOPED SnO2

The rigorous experimental methodology for preparing these nanoscale semiconductor lattices utilizes analytical grade (AR) stannic chloride and sodium carbonate. The precipitation is modulated using a highly controlled burette drip system (10 ml/hr).

5.1. Synthesis of Pure SnO_2

A 0.1 M solution of stannic chloride and a 0.1 M solution of sodium carbonate are co-titrated into a 0.01 M solution of the chosen capping agent (EDTA, CTAB, DNA, or Starch) under vigorous, continuous magnetic stirring. The resultant white tin carbonate precipitate undergoes rigorous washing to eliminate residual ionic species. Upon natural drying, the precursor is pulverized into a fine powder, ready for thermal activation. Samples are classified as SNE, SNC, SND, and SNS corresponding to their respective capping agents.

5.2. Synthesis of Ag-Doped SnO_2

To achieve silver intercalation, a 0.1 M silver nitrate solution is introduced simultaneously with the stannic chloride and sodium carbonate into the capping agent

matrix (EDTA or DNA). The co-precipitation ensures homogeneous distribution of the ions throughout the carbonate host lattice. The resulting precursors are classified as AGE and AGD.

5.3. Synthesis of Al-Doped SnO_2

Similarly, aluminum integration is achieved by co-titrating a 0.1 M aluminum hydroxide solution alongside the primary reactants into the EDTA or DNA capping environment. The resulting homogeneously mixed aluminum-tin carbonate precursors are classified as ALE and ALD.

5.4. Thermal Evolution: Thermogravimetric Analysis (TGA/DTA)

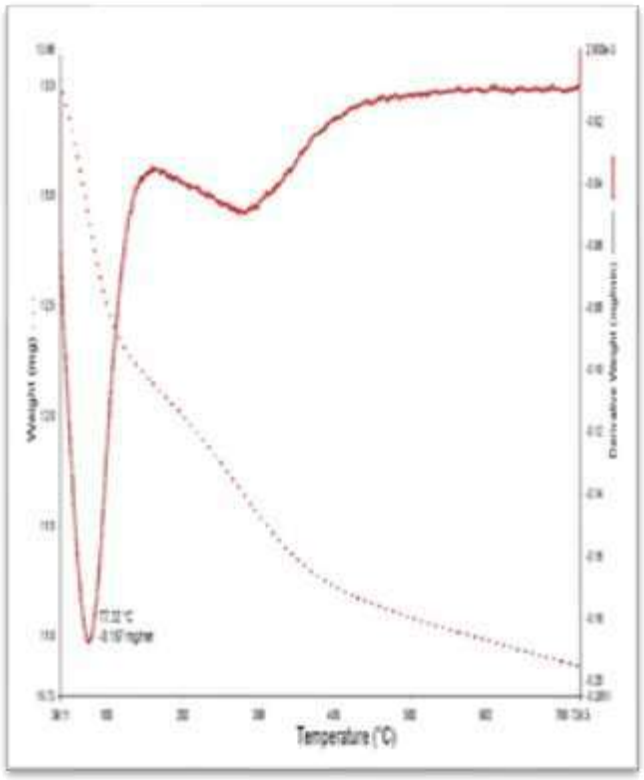
To transition the precursor carbonates into highly crystalline semiconducting oxides, precise thermal decomposition is required. Thermogravimetric Analysis (TGA) coupled with Differential Thermal Analysis (DTA) is employed to monitor mass variance and endothermic/exothermic transitions as a function of temperature. TGA is typically conducted from room temperature up to 800°C at a heating rate of 15°C/min in an inert nitrogen atmosphere. As the precursor is heated, non-covalent bonds rupture and structural rearrangements occur. The evolution of volatile gases from the breakdown of the carbonate matrix serves a dual physical purpose:

Porosity Generation: The escaping gases create micro-pores, yielding a highly fluffy, fine-particulate material.

Sintering Inhibition: The endothermic release of these gases absorbs excess thermal energy, effectively quenching localized hot spots and preventing the nanoparticles from sintering and agglomerating into bulk masses.

VI. RESULTS AND DISCUSSION

Thermal characterization of the synthesized samples was conducted utilizing a Perkin-Elmer Diamond TGA/DTA analyzer. Measurements were recorded in a continuous nitrogen atmosphere, scanning from an initial temperature of 28°C up to 800°C at a steady heating rate of 15°C/min. During this controlled heating phase, both the pristine and metal-incorporated tin carbonate precursors underwent complete thermal decomposition, yielding the desired pure and doped tin oxide nanocrystals. Because variables such as the calcination atmosphere and the temperature ramp rate critically influence the structural morphology and defect density of the resulting nanomaterials, these conditions were strictly regulated. For the annealing process, the precursors were placed inside a furnace at ambient temperature and gradually heated at a rate of 5°C/min. The peak calcination temperature was established at 400°C for the pure and Al-doped specimens, whereas the Ag-doped samples required a slightly elevated temperature of 450°C. Based on iterative experimental trials, the optimal holding time for this calcination phase was standardized to 3 hours.



6.1. Based on the derivative mass loss curves (TGA thermograms):

Pure and Al-Doped Precursors: Exhibit complete volatile mass loss by 350°C. Consequently, the optimal calcination temperature is established at 400°C.

Ag-Doped Precursors: Due to the distinct thermodynamic stability of the silver-integrated carbonate matrix, the calcination set-point is elevated to 450°C.

Following ambient-to-target temperature ramping at 5°C/minute, the precursors are annealed for precisely 3 hours. This carefully calibrated thermal budget

ensures complete oxidation, phase purity, and the crystallization of the semiconducting nanolattice.

VII. CONCLUSION

The deliberate synthesis and structural manipulation of SnO2, Al-doped SnO2, and Ag-doped SnO2 nanoparticles represent a highly sophisticated branch of semiconductor physics. Utilizing the chemical co-precipitation method, combined with structurally diverse capping agents (EDTA, CTAB, DNA, and Starch), allows for the precise architectural control of the nanolattice.

By utilizing rigorous TGA/DTA analytics to determine exact thermal degradation thresholds—specifically 400°C for pure and Al-doped variants, and 450°C for Ag-doped composites—the successful transition from amorphous carbonate precursors to highly functional, crystalline semiconducting nanoparticles is achieved. This thermodynamically controlled exothermic processing ensures the preservation of the nano-regime, ultimately empowering these materials for advanced electronic, optical, and catalytic device integration.

REFERENCES

[1] Ahmed, A. S., Azam, A., Chaman, M., & Naqvi, A. H. (2012). Structural and optical properties of Al-doped SnO2 nanoparticles. Journal of Luminescence, 132(10), 2566–2571.

[2] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2014). Molecular biology of the cell (6th ed.). Garland Science.

[3] Alivisatos, A. P. (1996). Semiconductor clusters, nanocrystals, and quantum dots. Science, 271(5251), 933–937.

[4] Amutha, T., Sahaya Seetha, S., Vidhya, M., & Prabha, K. (2016). Pure and Ni-doped tin oxide nanoparticles prepared by co-precipitation technique. International Journal of Latest Research in Engineering and Technology (IJLRET), 104–107.

[5] Anandan, K., & Rajendran, V. (2010). Solvents effect of quantum sized SnO2 nanoparticles via solvo thermal process and optical properties. Material Science Research India, 7(2), 389–397.

[6] Anbuvannan, M., Ramesh, M., Viruthagiri, G., Shanmugam, N., & Kannadasan, N. (2015). Synthesis, characterization and photo catalytic activity of ZnO nanoparticles prepared by biological method. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 143(15), 304–308.

[7] Anglin, E. J., Cheng, L., Freeman, W. R., & Sailor, M. J. (2008). Porous silicon in drug delivery devices and materials. Advanced Drug Delivery Reviews, 60(11), 1266–1277.

[8] Anitha, R., Ramesh, K. V., & Sudheer Kumar, K. H. (2017). Photo catalytic activity of CeO2 nanoparticles: Synthesis using artocarpus gomezianus fruit mediated facile green combustion method. International Journal of Pharma and Bio Sciences, 8(2), 933–936.

[9] Ansari, S. G., Dar, M. A., Dhage, M. S., Kim, Y. S., Ansari, Z. A., Al-Hajryans, A., & Shin, H. S. (2009). Binding of chloroquine–conjugated gold nanoparticles with bovine serum albumin. Review of Scientific Instruments, 80(4), 045112.

[10] Aragon, F. H., Coaquira, J. A. H., Hidalgo, P., da Silva, S. W., Brito, S. L. M., Gouvea, D., & Morais, P. C. (2010). Evidences of the evolution from solid solution to surface segregation in Ni-doped SnO2 nanoparticles using Raman spectroscopy. Journal of Raman Spectroscopy, 42(5), 1081–1087.

[11] Aruldhas, G. (2005). Molecular structure and spectroscopy. Prentice-Hall of India.

[12] Batzill, M., & Diebold, U. (2005). The surface and materials science of tin oxide. Progress in Surface Science, 79(2-4), 47–154.

[13] Bhattacharjee, A., & Ahmaruzzaman, M. (2015). A novel and green process for the production of tin oxide quantum dots and its application as a photo catalyst for the degradation of dyes from aqueous phase. Journal of Colloid and Interface Science, 448, 130–139.

[14] Carmona-Ribeiro, A. M., & Carrasco, L. D. (2013). Cationic antimicrobial polymers and their assemblies. International Journal of Molecular Sciences, 14(5), 9906–9946.

[15] Das, S., & Srivastava, V. C. (2016). Synthesis of SnO2 nanoparticles by co-precipitation method and their application in photocatalytic degradation of dye. Journal of Molecular Catalysis A: Chemical, 420, 243–252.

[16] Krishnakumar, T., Jayaprakash, R., Parthibavarman, M., Phani, A. R., Nayak, V. S., & Deraman, M. (2009). Microwave-assisted synthesis and characterization of SnO2 nanoparticles. Materials Letters, 63(21), 1898–1901.

International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.23

The Impact of Place and Promotion Strategies on Consumers' Home Purchase Intention: A Case Study of Construction Companies in the Greater Tainan Area

Huang Hsiao-Chun

School of International Businesses, Fuzhou University of International Studies and Trade, Fuzhou 350202, China
Email: Sandy.huang825@gmail.com

Received: 17 Jan 2026; Received in revised form: 20 Feb 2026; Accepted: 23 Mar 2026; Available online: 22 Apr 2026

Abstract— *In an increasingly competitive real estate market, how construction companies effectively utilize place layout and promotional tactics to accurately convey product information to target customers and stimulate their purchase intention has become a critical issue. This study focuses on the Place and Promotion dimensions of the 4P marketing theory to explore their influence on consumers' willingness to purchase houses from construction companies. Based on a questionnaire survey of 122 consumers in the Greater Tainan area, empirical analysis was conducted using the Linear Probability Model (LPM). The research results indicate that, within the Place dimension, the location of the sales center has a significant positive impact on home purchase intention. However, entrusting real estate agencies for sales and establishing online virtual storefronts show a significant negative impact, contradicting traditional expectations. Within the Promotion dimension, offering free interior decoration and low down payment with low interest rates have significant positive impacts, while lucky draw and grab bag events show no significant effect. This study provides empirically based strategic guidance for construction companies in planning place selection and promotional activities, along with concrete suggestions for future research.*

Keywords— ***Greater Tainan Area, Home Purchase Decision, Linear Probability Model, Marketing 4P, Place Strategy, Promotion Strategy.***

I. INTRODUCTION

1.1 Research Background and Motivation

The domestic real estate industry has long been regarded as a bellwether of economic development. Its rise and fall not only affect related industrial chains but also directly impact the national economy and people's livelihoods. In the past, construction companies primarily relied on traditional sales centers and word-of-mouth for promotion. However, with the rapid advancement of digital technology, the diversification of consumer information acquisition channels, and the drastic fluctuations in the real estate market cycle, traditional marketing models are facing unprecedented challenges. Particularly in the age of information overload, consumers can obtain property information through various channels such as online platforms, social media, and real estate apps, making the home purchase decision process more complex and information-intensive. Concurrently, to stand out in a fiercely competitive market, construction companies have launched a wide array of promotional activities, ranging from "buy a house, get free interior decoration," "low down payment and low interest rates," to lucky draws and celebrity lectures. However, what is the true effectiveness of these place and promotion strategies? Which strategies can truly move consumers to make a purchase decision? Which strategies merely increase costs with limited

effectiveness? These questions urgently require clarification through rigorous empirical research.

Specifically in the Greater Tainan area, recent major developments such as the expansion of the Southern Taiwan Science Park, TSMC's establishment of new fabs, and MRT planning have driven rapid growth in the real estate market, followed by a unique phase of consolidation. According to 2025 statistics from the Tainan City Government Land Administration Bureau, the primary homebuyer demographic in Greater Tainan consists of owner-occupiers (approximately over 80%). These consumers exhibit high price sensitivity and a strong sense of identity with their local living circles. Such market characteristics make the localization of place and promotion strategies particularly important. For instance, a project relying solely on online advertising without a physical reception center may struggle to gain the trust of local consumers. Similarly, a simple lucky draw event without substantial purchase incentives might only attract crowds of onlookers rather than converting into actual sales.

Given this context, this study focuses on the Place and Promotion dimensions of the 4P marketing theory. Through a questionnaire survey of consumers in the Greater Tainan area and rigorous econometric analysis, it seeks to answer the following core questions:

a. Among different place types—traditional sales centers, entrusting real estate agencies, and online virtual storefronts—which is most effective in enhancing consumers' home purchase intention?

b. Among promotional activities—free interior decoration, low down payment/low interest rates, and lucky draw events—which holds the most substantial influence?

c. By answering these questions, this study aims to provide construction companies with scientific, data-driven marketing decision-making references under resource constraints.

1.2 Research Objectives

Based on the research background and motivation stated above, the specific objectives of this study are as follows:

a. Analyze the impact of place strategy on consumers' home purchase intention: Explore the direction and magnitude of the impact of three place types—sales center location, entrusting real estate agencies, and online virtual storefronts—on consumers' home purchase intention.

b. Analyze the impact of promotion strategy on consumers' home purchase intention: Explore the effects of three promotional methods—free interior decoration, low down payment/low interest rates, and lucky draw events—on consumers' home purchase intention.

c. Compare the relative importance of different place and promotion strategies: Compare the influence of each strategy based on the magnitude and significance of regression coefficients to provide construction companies with recommendations for resource allocation priorities.

d. Provide practical recommendations: Based on the empirical analysis results and considering the market characteristics of the Greater Tainan area, propose concrete and feasible suggestions for construction companies regarding place layout and promotional activity planning.

1.3 Research Scope and Subjects

This study focuses on the Place and Promotion dimensions of the 4P marketing theory and adopts a cross-sectional quantitative research design. The research subjects are general consumers in the Greater Tainan area (including all 37 administrative districts of Tainan City) who have home buying experience or potential home purchase needs, regardless of industry, education level, age, or gender. Convenience sampling was employed, with data collected through two channels: online questionnaires (distributed via Line groups, PTT Real Estate Forum) and paper questionnaires (distributed at major shopping malls and project reception centers in Tainan City). A total of 130 questionnaires were distributed, with 122 valid responses collected, yielding an effective response rate of 93.8%. The sample structure consisted of 85% owner-occupier consumers and 15% investors, consistent with the actual market structure of the Greater Tainan area.

1.4 Research Process

This study first established the research theme and motivation, followed by a review and synthesis of

relevant literature on place and promotion strategies to develop the research framework and hypotheses[1]. Subsequently, a questionnaire was designed and administered. After collection, reliability and validity analyses were conducted to confirm the quality of the instrument. Finally, regression analysis was performed using the Linear Probability Model (LPM), and conclusions and recommendations were drawn based on the empirical results. The overall research process is illustrated in Table 1-1.

Table 1-1 Overall Research Process

Step	Content	Phase
1	Research Theme and Motivation	Preparation Phase
2	Literature Review and Synthesis (Place & Promotion)	Preparation Phase
3	Establishment of Research Framework and Hypotheses	Design Phase
4	Questionnaire Design and Survey	Implementation Phase
5	Reliability and Validity Analysis	Implementation Phase
6	LPM Regression Analysis	Implementation Phase
Conclusion	Conclusions and Recommendations	Conclusion Phase

II. LITERATURE REVIEW

2.1 Characteristics of Place and Promotion for Construction Companies

The place and promotion strategies in the construction industry differ significantly from those for general consumer goods, primarily due to the unique characteristics of real estate products: high unit price, high heterogeneity, low transaction frequency, and severe information asymmetry[2]. In terms of place, the main options for construction companies include:

a. Self-built Sales Centers: The construction company establishes its own sales team and sets up a temporary reception center near the construction site to sell directly to consumers.

b. Commissioning Sales Agencies: Outsourcing the entire sales process to professional real estate marketing agencies responsible for planning, advertising, and sales.

c. Entrusting Real Estate Brokerages: Delegating the sale of projects to real estate brokerage firms, leveraging their existing storefronts and agent networks.

d. Online Channels: Establishing official websites, listing on real estate platforms (e.g., 591 Housing Network), and managing social media accounts.

In terms of promotion, common tactics used by construction companies include:

a. Price Promotions: Early bird discounts, limited-time offers, waived management fees, etc.

b. Gift Promotions: Free interior decoration with purchase, free appliances, free parking spaces, etc.

c. Financial Promotions: Collaborating with banks to offer low down payment, low interest rate, and long-term loan packages.

d. Event Promotions: Hosting seminars, celebrity lectures, lucky draw events, family carnivals, etc.

The cost-effectiveness of different promotional tactics varies significantly, and their effects may differ based on consumer characteristics (e.g., first-time buyers vs. existing homeowners, owner-occupiers vs. investors).

2.2 Impact of Place Strategy on Home Purchase Decisions

2.2.1 Sales Center Location

The sales center serves as the primary physical interface between the construction company and consumers. The choice of its location directly impacts consumer visitation willingness and experience quality. Prior research indicates that sales center site selection should consider the following factors[3]:

a. Transportation Convenience: Proximity to major roads and public transit stations for easy access.

b. Visibility: Prominent exterior and clear signage to attract the attention of passing traffic and pedestrians.

c. Parking Convenience: Provision of ample parking spaces or partnerships with nearby parking facilities to prevent loss of potential clients due to parking difficulties.

d. Pedestrian Traffic: Selection of high-traffic areas (e.g., commercial districts, transportation hubs) to increase natural footfall.

e. Display Functionality: Sufficient space to display project models, sample units, floor plans, etc. Based on this, the study proposes Hypothesis H3a: The sales center location has a significant positive impact on consumers' home purchase intention.

2.2.2 Entrusting Real Estate Agencies for Sales

Entrusting real estate agencies involves construction companies handing over project sales to brokerage firms (e.g., Sinyi Realty, Yungching Realty). The advantages of this model include: brokers possess extensive client databases, professional agent teams, and mature sales networks, allowing for rapid expansion of project exposure and reach. However, this model also has notable drawbacks: consumers generally perceive purchasing through an agent as incurring additional costs ("being skimmed"), as extra commission fees increase the overall purchase expense. Furthermore, agents may simultaneously handle multiple projects[4], potentially diluting their focus on any single development. Therefore, the empirical direction of the impact of entrusting real estate agencies on purchase intention is not definitively clear. This study proposes Hypothesis H3b: Entrusting real estate agencies for sales has a significant positive impact on consumers' home purchase intention, but its effect is expected to be weaker than that of self-built sales centers.

2.2.3 Online Virtual Storefronts

With the proliferation of the internet, a growing number of construction companies are establishing online virtual storefronts, including official websites, listings on platforms like 591, Facebook fan pages, Instagram project accounts, and even VR online property viewings. The advantages of online channels include fast information dissemination, wide reach, relatively low cost, and 24/7 availability. However, given that real estate is a high-value, high-involvement product, consumers typically prefer to see and feel the property in person before making a decision. Online photos, videos, and 3D renderings may deviate from the actual condition of the house and cannot fully replace the physical viewing experience[5]. Consequently, the role of online virtual storefronts in the home purchase decision may be more of an auxiliary tool for information gathering rather than a decisive factor prompting purchase. This study proposes Hypothesis H3c: Online virtual storefronts have a significant positive impact on consumers' home purchase intention, but their influence is expected to be smaller than that of physical sales centers.

2.3 Impact of Promotion Strategy on Home Purchase Decisions

2.3.1 Free Interior Decoration with Purchase

Offering free interior decoration is a very common promotional tactic in Taiwan's real estate market[6]. The core logic is that after purchasing a pre-sale or newly completed house, consumers usually face significant additional decoration costs (approximately NT$ 30,000-80,000 per ping) and a time-consuming, labor-intensive process. If the construction company provides the decoration service directly, it not only saves consumers time and money but also ensures the home is move-in ready upon handover, significantly reducing post-purchase hassles. Furthermore, through bulk purchasing of materials and consolidated contracting, construction companies can often achieve better quality at a lower cost. Therefore, free interior decoration is considered a win-win promotional strategy. This study proposes Hypothesis H4a: Offering free interior decoration has a significant positive impact on consumers' home purchase intention.

2.3.2 Low Down Payment and Low Interest Rates

For most consumers, the biggest barrier to homeownership is not the property price itself, but the burden of the down payment and subsequent mortgage payments[7]. Particularly in an era of high housing prices, insufficient savings for the down payment is often the primary reason young people cannot afford a home. Consequently, construction companies collaborating with financial institutions to offer low down payment and low interest rate loan packages has become an effective means of stimulating housing demand. Examples include offering zero payment during the construction period, deferred down payment until after handover,

or negotiating mortgage rates below market average with extended loan terms. These schemes directly lower the financial threshold for home buying and reduce monthly repayment pressure, thereby enhancing purchase intention. This study proposes Hypothesis H4b: Low down payment and low interest rate offers have a significant positive impact on consumers' home purchase intention.

2.3.3 Lucky Draw and Grab Bag Events

Lucky draw and grab bag events are common methods used by construction companies to generate excitement and foot traffic at sales centers. Common practices include weekend raffles with prizes ranging from small appliances to cars, or limited-edition grab bags containing coupons and small gifts during holidays (e.g., Lunar New Year, Mid-Autumn Festival). The primary purpose of such events is to attract crowds, create buzz, and increase visitor numbers. However, whether this type of promotion can genuinely translate into purchase behavior remains debated academically. On one hand, crowds can indeed bring more potential clients. On the other hand, attendees may simply be bargain hunters with no actual purchase intention, potentially consuming the time and energy of sales staff without contributing to sales. Thus, the effectiveness of lucky draw events requires empirical testing. This study proposes Hypothesis H4c: Lucky draw events have a significant positive impact on consumers' home purchase intention, but their effect is expected to be smaller than that of substantive promotional schemes.

2.4 Summary of Research Hypotheses

Based on the literature review above, this study establishes the following six research hypotheses:

Table 2-1 Research Hypotheses

Code	Hypothesis Content	Expected Direction
H3a	Sales center location has a significant positive impact on home purchase intention.	Positive
H3b	Entrusting real estate agencies for sales has a significant positive impact.	Positive
H3c	Online virtual storefronts have a significant positive impact.	Positive
H4a	Offering free interior decoration has a significant positive impact.	Positive
H4b	Low down payment and low interest rates have a significant positive impact.	Positive
H4c	Lucky draw events have a significant positive impact.	Positive

III. RESEARCH DESIGN AND METHODOLOGY

3.1 Research Framework

This study utilizes the Place dimension[8] (sales center location, entrusting real estate agencies, online virtual storefronts) and Promotion dimension (free interior decoration, low down payment/low interest rates, lucky draw events) of the 4P marketing theory as independent variables. Consumers' home purchase intention serves as the dependent variable. Basic consumer demographic data (average annual income, down payment amount, purchase purpose, distance between home and workplace) are included as control variables. The research framework is illustrated in Fig. 3-1.

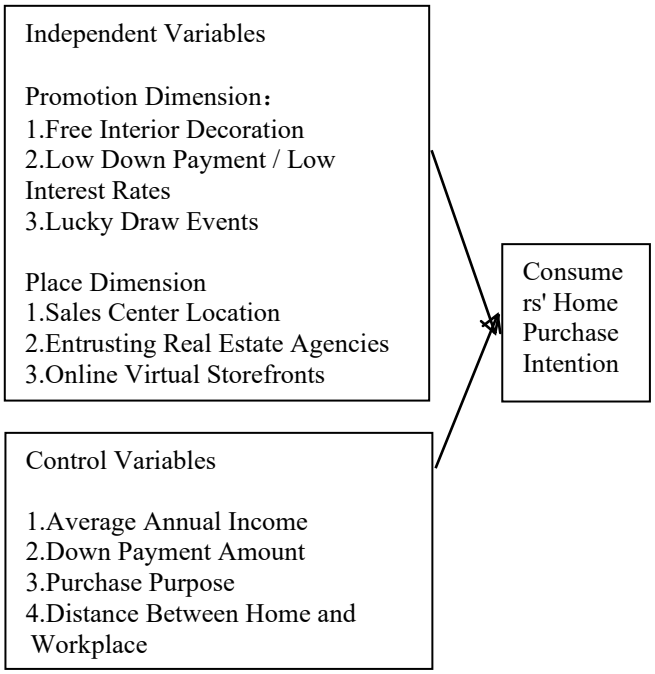


Fig. 3-1 Research Framework

3.2 Research Subjects and Sampling Design

This study conducted a questionnaire survey targeting consumers in the Greater Tainan area,

yielding 122 valid responses. The sample structure, with 85% owner-occupiers, aligns with the predominantly owner-occupier driven market in Greater Tainan.

3.3 Operational Definitions and Measurement of Variables

3.3.1 Dependent Variable

Home Purchase Intention (Y) is defined as the consumer's willingness to enter into a purchase relationship with the construction company. It is measured as a dummy variable: 1 indicates willingness to purchase, 0 indicates unwillingness.

3.3.2 Independent Variables

X_7 Sales Center Location: Refers to the site selection of the construction company's sales reception center, including transportation convenience, visibility, parking convenience, etc. Measured on a 5-point Likert scale (1=Very Unimportant, 5=Very Important).

X_8 Entrusting Real Estate Agencies for Sales: Refers to the construction company delegating project sales to real estate brokerage firms. Measured on a 5-point Likert scale regarding consumer agreement with this channel.

X_9 Online Virtual Storefront: Refers to the construction company utilizing official websites, real estate platforms, social media, etc., for sales. Measured on a 5-point Likert scale.

X_{10} Free Interior Decoration with Purchase: Refers to the promotion scheme where the construction company provides complimentary interior decoration upon home purchase. Measured on a 5-point Likert scale.

X_{11} Low Down Payment and Low Interest Rates: Refers to preferential loan packages offered in collaboration with financial institutions. Measured on a 5-point Likert scale.

X_{12} Lucky Draw Events: Refers to raffles, sweepstakes, and similar activities held at the sales center. Measured on a 5-point Likert scale.

3.3.3 Control Variables

D_1: Average Annual Income (1 = NT$ 1 million and above, 0 = Below NT$ 1 million)

D_2: Down Payment Amount (1 = NT$ 2 million and above, 0 = Below NT$ 2 million)

D_3: Purchase Purpose (1 = Owner-occupier, 0 = Investor)

D_4: Distance between Home and Workplace ($D_{4\text{-}1}$ = 3-5 km, $D_{4\text{-}2}$ = Over 5 km; Reference group = Within 3 km)

3.4 Statistical Analysis Methods

3.4.1 Descriptive Statistics

Calculation of mean, standard deviation, and ranking for each variable to understand the importance consumers place on various place and promotion strategies.

3.4.2 Reliability Analysis

Cronbach's α coefficient was used to test the internal consistency of the questionnaire. An α value greater than 0.7 indicates good reliability.

3.4.3 Validity Analysis

The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity were employed. A KMO value > 0.5 and a significant Bartlett's test ($p < 0.05$) indicate the data are suitable for factor analysis. Communalities greater than 0.5 were used as the criterion for retaining variables.

3.4.4 Linear Probability Model (LPM)

Since the dependent variable is binary (purchase/not purchase), this study employs the Linear Probability Model for regression analysis. The model is specified as in Equation 1:

$$Y=\alpha+\gamma_1D_1+\gamma_2D_2+\gamma_3D_3+\gamma_{4\text{-}1}D_{4\text{-}1}+\gamma_{4\text{-}2}D_{4\text{-}2}+\beta_1X_1+\beta_2X_2+\beta X_3+\beta_4X_4+\beta_5X_5+\beta_6X_6+\beta_7X_7+\beta_8X_8+\beta_9X_9+\beta_{10}X_{10}+\beta_{11}X_{11}+\beta_{12}X_{12} \quad (1)$$

Where Y represents consumers' home purchase intention, α is the constant term, γ represents the coefficients for control variables, β represents the coefficients for independent variables, and ε is the error term.

IV. EMPIRICAL RESULTS ANALYSIS

4.1 Sample Structure Analysis

The demographic distribution of the 122 valid samples is as follows:

Table 4-1 Demographic Data

Variable	Option	Count	Percentage	Cumulative Percentage
Average Annual Income	NT$ 1 Million and Above	76	62%	62%
	Below NT$ 1 Million	46	38%	100%
Down Payment	NT$ 2 Million and Above	58	48%	48%
	Below NT$ 2 Million	64	52%	100%
Purchase Purpose	Owner-occupier	104	85%	85%
	Investor	18	15%	100%
Distance (Home to Workplace)	Within 3 km	58	48%	48%
	3-5 km	47	39%	86%
	Over 5 km	17	14%	100%
Total	NA	122	100.0%	NA

Average Annual Income: 62% above NT$1M, 38% below.

Down Payment: 48% above NT$2M, 52% below.

Purchase Purpose: 85% Owner-occupier, 15% Investor.

Distance (Home-Workplace): 48% within 3km, 39% within 3-5km, 14% over 5km.

The sample structure indicates that respondents are primarily middle-to-high income earners with owner-occupier needs, and they tend to prefer purchasing homes closer to their workplace.

4.2 Descriptive Statistics

4.2.1 Descriptive Statistics for Place Dimension

Table 4-2 Place Dimension Statistics

Place Dimension Variable	N	Mean	Std. Deviation	Rank
Entrusting Real Estate Agencies for Sale	122	2.96	1.708	1
Online Virtual Storefront	122	2.85	1.694	2
Sales Center Location	122	2.71	1.593	3

Based on the mean scores, consumers attach the highest level of importance to entrusting real estate agencies (2.96), followed by online virtual storefronts (2.85), and finally sales center location (2.71). However, it is crucial to note that these means reflect perceived importance, not the actual impact on purchase intention (which requires regression analysis to ascertain). Notably, all three mean scores are below 3 (the midpoint of the 5-point scale), suggesting that overall, consumers do not place high importance on these place factors. This might be attributed to the high value of properties and consumers' tendency to personally gather information.

4.2.2 Descriptive Statistics for Promotion Dimension

Table 4-3 Promotion Dimension Statistics

Promotion Dimension Variable	N	Mean	Std. Deviation	Rank
Free Interior Decoration	122	3.11	1.719	1
Low Down Payment & Low Interest Rate	122	3.09	1.749	2
Lucky Draw Events	122	2.38	1.501	3

The mean scores indicate that consumers value "Free Interior Decoration" (3.11) and "Low Down Payment & Low Interest Rate" (3.09) almost equally, both slightly above the midpoint, indicating favorable

reception. In contrast, the mean score for "Lucky Draw Events" is only 2.38, significantly below the midpoint, suggesting consumers find such activities generally unappealing. This aligns with practical observations: lucky draws may draw crowds but have minimal impact on actual purchase decisions.

4.3 Reliability Analysis

4.3.1 Reliability of Place Dimension

The Cronbach's α for the Place dimension is 0.956, well above the threshold of 0.7, indicating excellent reliability. The item-to-total correlations are: Sales Center Location (0.901), Entrusting Real Estate Agencies (0.881), and Online Virtual Storefront (0.943), all exceeding 0.5, demonstrating good internal consistency for each item.

4.3.2 Reliability of Promotion Dimension

The Cronbach's α for the Promotion dimension is 0.912, also exceeding 0.7, indicating good reliability. The item-to-total correlations are: Free Interior Decoration (0.869), Low Down Payment & Low Interest Rate (0.890), and Lucky Draw Events (0.727), all exceeding 0.5. While the correlation for Lucky Draw Events is relatively lower (0.727), it still meets the reliability requirement.

4.4 Validity Analysis

4.4.1 Validity of Place Dimension

For the Place dimension, the KMO value is 0.741 (>0.7), and Bartlett's test of sphericity yields an approximate Chi-Square of 408.280 (df=3, p=0.000<0.05), confirming the data's suitability for factor analysis. The communalities for the variables are: Sales Center Location (0.915), Entrusting Real Estate Agencies (0.894), and Online Virtual Storefront (0.952), all exceeding 0.5. The total variance explained is 92.054%, surpassing the 60% benchmark. Therefore, all variables are retained.

4.4.2 Validity of Promotion Dimension

For the Promotion dimension, the KMO value is 0.704 (>0.7), and Bartlett's test of sphericity yields an approximate Chi-Square of 293.201 (df=3, p=0.000<0.05), indicating suitability for factor analysis. The communalities are: Free Interior Decoration (0.891), Low Down Payment & Low Interest Rate (0.910), and Lucky Draw Events (0.752), all exceeding 0.5. The total variance explained is 85.082%, above 60%. Thus, all variables are retained.

4.5 Regression Analysis (Linear Probability Model)

4.5.1 LPM Analysis without Control Variables

First, a regression analysis was conducted using only the six independent variables from the Place and Promotion dimensions (X_7~X_{12}) on the dependent variable Y, excluding control variables (D_1~D_4). The results are shown in Table 4-4:

Table 4-4 Regression Analysis Results

Variable	Coefficient	Std. Error	t-value	p-value	Result
X_7	0.0307	0.01315	2.333	0.0215	Positive Significant
X_8	-0.0261	0.01062	-2.463	0.0154	Negative Significant
X_9	-0.0896	0.01280	-6.998	0.0000	Negative Significant
X_{10}	0.0289	0.01073	2.693	0.0082	Positive Significant
X_{11}	0.0512	0.01097	4.666	0.0000	Positive Significant
X_{12}	0.0060	0.00545	1.107	0.2709	Not Significant

Adjusted R^2 = 0.9858, Durbin-Watson statistic = 2.286.

Key Findings:

a. Place Dimension: X_7 (Sales Center Location) is significantly positive, supporting H3a. X_8 (Entrusting Real Estate Agencies) is significantly negative, not supporting H3b. X_9 (Online Virtual Storefront) is significantly negative, not supporting H3c.

b. Promotion Dimension: X_{10} (Free Interior Decoration) and X_{11} (Low Down Payment/Low Rate) are significantly positive, supporting H4a and H4b. X_{12} (Lucky Draw Events) is not significant, not supporting H4c.

c. Comparison of Coefficient Magnitudes: Among the significant variables, X_{11} (Low Down Payment/Low Rate) has the largest coefficient (0.0512), followed by X_7 (0.0307) and X_{10} (0.0289). This indicates that low down payment/low interest rate offers are the most potent promotion strategy influencing purchase intention, while sales center location is the only effective positive factor among the place strategies.

4.5.2 LPM Analysis with Control Variables

After including control variables (D_1~D_4) in the regression model, the results remain largely consistent with the previous analysis, albeit with slight changes in coefficient magnitudes. X_7, X_{10}, and X_{11} remain significantly positive; X_8 and X_9 remain significantly negative; and X_{12} remains non-significant. None of the control variables (D_1: Income, D_2: Down Payment, D_3: Purchase Purpose, D_4: Distance) reached statistical significance, suggesting these demographic factors do not significantly impact purchase intention in this sample. The adjusted R^2 remained high at 0.9859, and the Durbin-Watson statistic was 2.224, indicating excellent model explanatory power and no autocorrelation issues.

4.5.3 In-depth Discussion of Negative Findings

The significant negative impacts observed for X_8 (Entrusting Real Estate Agencies) and X_9 (Online Virtual Storefront) are counterintuitive findings[9]. Possible explanations are proposed below:

Entrusting Real Estate Agencies (X_8): Consumers widely believe that purchasing through an agent incurs additional commission fees (typically 1-2% of the transaction price), which they perceive as ultimately being added to their purchase cost. Even if the developer covers the commission in some cases, the perception of an extra layer of cost persists. Moreover, as agents may handle multiple projects concurrently, consumers may doubt receiving comprehensive and unbiased information. Consequently, this channel may inadvertently deter purchase intention.

Online Virtual Storefront (X_9): Real estate is a high-involvement, high-value product requiring physical inspection. Online photos and VR tours, while convenient, cannot fully convey the nuances of ventilation, natural light, material texture, and community ambiance. Furthermore, concerns about digitally enhanced or misleading representations online may heighten consumer skepticism. Therefore, relying solely on a virtual storefront might create unease and reduce purchase intention. This underscores that in real estate, online platforms are best positioned as information-gathering aids rather than standalone sales endpoints.

V. CONCLUSIONS AND RECOMMENDATIONS

5.1 Research Conclusions

Through a questionnaire survey of 122 consumers in the Greater Tainan area and LPM regression analysis, this study reveals the unique operational logic of place and promotion strategies in the real estate market. The main conclusions are as follows:

5.1.1 Empirical Findings on Place Strategy

Within the Place dimension, sales center location has a significant positive impact, while entrusting real estate agencies and online virtual storefronts show significant negative impacts. This finding challenges the conventional wisdom that "more channels are better," highlighting that in real estate, the quality of the place is far more important than quantity.

Specifically, the physical sales center remains the most trusted channel for consumers because it provides tangible product experience (sample units, material displays, community models), face-to-face professional consultation, and an opportunity to build trust. Conversely, entrusting agencies may trigger cost-related concerns, and virtual storefronts, burdened by high information asymmetry and inability to replicate physical experience, may actually deter potential buyers. This does not negate the value of agencies or online platforms but serves as a reminder to construction companies: these channels should function as supplementary tools, not replacements for the core physical sales center.

5.1.2 Empirical Findings on Promotion Strategy

In the Promotion dimension, offering free interior decoration and low down payment/low interest rates both exert a significant positive impact, while lucky draw events show no significant effect. This finding aligns closely with the concept of perceived value in consumer behavior theory: promotional schemes that directly address consumer pain points, lower the barrier to entry, and save on post-purchase

costs are most effective at enhancing perceived value and driving purchase intent.

Free interior decoration directly resolves the onerous and costly post-purchase issue of fitting out a new home, enabling move-in readiness. Low down payment and low interest rates directly mitigate financial hurdles and monthly repayment burdens, particularly appealing to young buyers or those with limited savings. In contrast, lucky draw events, though capable of generating a festive atmosphere, are loosely connected to the core values of a home (living quality, price fairness). Consumers largely view them as incidental perks rather than decisive factors, limiting their impact on actual purchase decisions.

5.2 Theoretical Contributions

This study makes three primary theoretical contributions:

First, it supplements the empirical foundation of the 4P marketing theory in the real estate domain. Prior research on the 4Ps has predominantly focused on consumer goods or manufacturing, with less attention paid to unique, high-involvement goods like real estate. This study confirms that Place and Promotion strategies play a crucial role even in real estate, though the direction and magnitude of their effects diverge notably from typical consumer goods.

Second, it reveals the potential negative effects of certain place strategies. While existing literature often assumes that more diverse and convenient channels are inherently better, this study finds that entrusting agencies and relying on virtual storefronts can decrease purchase intention. This underscores the need for academia to consider product characteristics (e.g., credence attributes) and consumer decision psychology (e.g., risk aversion) when evaluating place strategies, rather than solely focusing on channel quantity or convenience.

Third, it differentiates the effects of distinct promotional tactics. The study demonstrates that substantive promotions (decoration, financing) have a significantly greater impact than atmospheric promotions (lucky draws)[10]. This aligns with Kotler's classification of promotional tools and provides empirical support for it within the real estate context.

5.3 Practical Recommendations

Based on the findings, the following practical recommendations are proposed for construction companies:

5.3.1 Recommendations for Place Strategy

Continue investing in and optimizing physical sales centers: Sales centers are indispensable trust-building environments. Developers should focus on:

a. Site Selection: Prioritize locations with convenient transportation, ample parking, and high visibility.

b. Hardware: Create high-quality sample units, material display zones, and project models for an immersive experience.

c. Software: Train professional, sincere sales staff to provide complete, transparent information and foster trust.

Carefully evaluate the necessity of entrusting real estate agencies: If using this channel is unavoidable, implement the following:

d. Transparency: Clearly communicate the arrangement regarding commission fees to alleviate cost concerns.

e. Consistency: Ensure pricing and service conditions are identical to those offered directly by the developer.

f. Exclusivity: Consider exclusive partnerships with select brokerage brands to ensure agents are knowledgeable and committed to the project.

Position online platforms as information and filtering tools, not final sales endpoints: The primary functions of a virtual storefront should be:

a. Providing comprehensive, accurate, and up-to-date project information.

b. Utilizing VR/360-degree tours for preliminary screening.

c. Guiding consumers towards the physical sales center for a deeper, in-person experience. Avoid overly embellished online content that could lead to mismatched expectations and erode trust.

5.3.2 Recommendations for Promotion Strategy

Prioritize low down payment and low interest rate schemes: Given its strongest effect size, developers should actively collaborate with financial institutions to develop diverse loan packages, such as:

a. 90% financing with interest-only payments for the first three years for first-time buyers.

b. Bridge loans for existing homeowners buying before selling.

c. Tailored programs for self-employed individuals with non-traditional income verification. These directly lower the entry barrier.

Effectively utilize the "free interior decoration" strategy: This tactic is also highly effective. Developers should:

a. Offer multiple design style options for flexibility.

b. Clearly specify the scope, specifications, and brands of included finishes to manage expectations and prevent disputes.

c. Consider offering a decoration allowance option, providing a subsidy for buyers to engage their own contractors.

Cautiously evaluate the ROI of lucky draw events: Given their insignificant impact on purchase decisions, these activities are best positioned as supplementary crowd-pulling tactics, not core promotional strategies. If held, budgets should be controlled, and participation should be linked to concrete purchase actions (e.g., contract signing) to target resources toward serious buyers.

5.3.3 Recommendations for Localized Operations in Tainan

The Greater Tainan real estate market is characterized by:

a.Predominantly owner-occupier demand (~85%).

b.High consumer price sensitivity.

c.Strong local identity.

Therefore, construction companies should:

In Place Strategy: Strengthen local physical sales centers and incorporate Tainan's cultural elements (e.g., revitalized old house aesthetics, temple square motifs) to enhance familiarity and appeal.

In Promotion Design: Tailor offers to local preferences (e.g., emphasis on living amenities, lower total price points). Implement combined packages of low down payment and basic interior decoration to meet rigid demand effectively.

5.4 Research Limitations and Future Research Directions

This study has several limitations that suggest avenues for future research:

Sample Representativeness: The sample is limited to the Greater Tainan area and predominantly consists of owner-occupiers (85%). Generalizability of the findings is thus geographically and demographically constrained. Future research should expand the survey scope to other regions in Taiwan (North, Central, South) for cross-regional comparisons and specifically target investor-type consumers.

Cross-sectional Design: The cross-sectional nature of this study captures consumer attitudes at a single point in time. The effectiveness of place and promotion strategies may vary with economic cycles and market competition. Future research could employ longitudinal designs (e.g., panel studies) to explore these variations across different market phases.

Measurement of Variables: Strategy variables were measured using subjective consumer perceptions on a 5-point scale rather than objective behavioral data. Future research could attempt to obtain actual sales data from developers (e.g., foot traffic per channel, conversion rates) or employ experimental designs (e.g., A/B testing promotional offers) for more objective causal inference.

Mechanisms Behind Negative Place Effects: While this study identified negative impacts for agency sales and virtual storefronts, the underlying psychological mechanisms (e.g., cost concerns, lack of trust, information asymmetry) require further elucidation. Future studies could employ in-depth interviews or experimental designs to investigate the specific causes of these negative effects.

Moderating Effects of Property Type: This study did not differentiate between property types (pre-sale, newly completed, existing homes). Future research could explore whether the effectiveness of place and promotion strategies varies by product type (e.g., pre-sale projects might rely more heavily on sales centers and show flats, while existing homes might depend more on brokerage channels).

REFERENCES

[1] Kotler, P., & Keller, K. L. (2016). *Marketing management (15th ed.)*. Pearson Education.

[2] Kotler, P. (2003). *Marketing insights from A to Z: 80 concepts every manager needs to know*. John Wiley & Sons.

[3] Kotler, P. (2005). *This is marketing: The essence of Kotler (S. M. Hong, Trans.)*. Sun Moon Publishing. (Original work published 2005)

[4] Kotler, P., & Armstrong, G. (2018). *Principles of marketing (17th ed.)*. Pearson.

[5] Yang, Z. L. (1997). *A study on market segmentation and promotion activity preferences in housing purchase behavior: A case study of Hsinchu City (Unpublished master's thesis)*. Chung Hua Polytechnic Institute, Department of Industrial Engineering and Management.

[6] Yan, Z. H. (2023). *Analysis of the impact of physical and virtual sales models of housing marketing on consumers' purchase intention*. Feng Chia University.

[7] Hou, X. Q. (2004). *A study on creative marketing strategies of real estate in Taiwan*. National Taipei University of Technology, Graduate Institute of Architecture and Urban Design.

[8] Zhang, H. S. (2023). *Exploring the influence of homebuyers' preferences and marketing knowledge management capabilities on purchase intention*. Asia University, Department of Business Administration.

[9] Wu, C. L. (2012). *A study on regional real estate marketing strategies: A case study of housing projects in Taoyuan area*. Yuan Ze University, Master Program in Information Sociology.

[10] Zhang, Z. H. (2023). *A study on marketing strategy planning for achieving expected results in the real estate consignment industry: A case study of C New Village, Xindian District*.