JMBT
Vol. X
Issue X
The Journal of Modern Business and Technology
Available online at https://jmbta.journals.ekb.eg/

Enhancing Decision Support in Business Information Systems through The Integration of Generative AI Models

Mariam K. Bersi [1], Ahmed M. Eassa[2]
[1,2] Business Informatics Department, Faculty of Management & Business Intelligence,
MTI University, Cairo, Egypt

mariamkhairy20005@yahoo.com

Received: 2025-10-14 Accepted: 2025-11-01 Available online: 2025-11-01

ABSTRACT

The rapid rise of Generative AI in enterprise applications has made its responsible integration into Business Information Systems (BIS) a critical priority. This study proposes a governance-aware, five-layer architecture that integrates Retrieval-Augmented Generation (RAG), explainable AI (XAI), and human-in-the-loop (HITL) safeguards to address the shortcomings of traditional decision-support systems. The architecture was implemented using a hybrid technology stack and evaluated with 45 participants across multiple business domains. Results showed a 93% reduction in hallucinations, a 58% improvement in task completion time, and almost double the user acceptance rate. These outcomes highlight the framework's ability to deliver measurable business value while maintaining transparency, trust, and compliance, positioning it as a practical solution for enterprise-wide adoption. Beyond performance gains, the study emphasizes the importance of governance and explainability in building user confidence. Overall, the framework contributes a replicable and enterprise-ready model that balances innovation with accountability in BIS contexts.

KEYWORDS
Business Information Systems; Generative AI; Retrieval-Augmented Generation (RAG); Explainable AI (XAI); Human-in-the-Loop (HITL); Governance

1. INTRODUCTION

Business Information Systems (BIS) are critical enablers of organizational decision-making, providing structured processes for collecting, analyzing, and utilizing data in both operational and strategic contexts (Laudon & Laudon, 2022). However, traditional BIS often struggle to manage unstructured data and natural language interactions, which increasingly dominate business communication and knowledge flows (Davenport & Harris, 2017; Saleh et al., 2025). The rapid development of Generative Artificial Intelligence (GenAI) offers new opportunities to address these limitations by enabling more adaptive, intelligent, and interactive decision support systems (Dwivedi et al., 2023).

Despite these advancements, integrating GenAI into enterprise systems raises challenges related to accuracy, reliability, governance, and ethical considerations. A key issue is hallucination, where GenAI models produce outputs that are factually incorrect or irrelevant, which can undermine trust and adoption in enterprise settings (Ji et al., 2023; Abdelsamee, 2025). Moreover, organizations face difficulties in establishing effective governance mechanisms that ensure transparency, accountability, and alignment with regulatory and ethical standards (Weidinger et al., 2022).

Several approaches have been proposed to mitigate hallucinations and improve reliability, such as Retrieval-Augmented Generation (RAG), which grounds model outputs in curated enterprise knowledge bases (Lewis et al., 2020). Similarly, Explainable AI (XAI) has emerged as a crucial component for enhancing transparency by providing traceability and justifications for model outputs, thereby increasing user trust (Arrieta et al., 2020). Furthermore, incorporating Human-in-the-Loop (HITL) mechanisms ensures that critical decisions are validated by human experts, balancing automation with oversight (Schneiderman, 2020). Together, these approaches provide complementary safeguards for enterprise-grade GenAI adoption.

At the same time, enterprises must balance technical innovation with operational feasibility. Hybrid architectures, which combine cloud-based GenAI capabilities with on-premises data management, have been suggested as a practical solution that balances scalability, cost efficiency, and data security (Zhang et al., 2021). However, there remains a lack of empirically validated frameworks that integrate these elements—RAG, XAI, HITL, and governance—into a unified BIS architecture evaluated in real organizational settings.

To address this gap, this study proposes a governance-centric, five-layer architecture that integrates GenAI into BIS with embedded safeguards for reliability, transparency, and oversight. The architecture is evaluated through a controlled experiment involving forty-five mid-level managers, comparing its performance with traditional manual methods and a baseline large language model. The findings contribute to both theory and practice by demonstrating how structured governance and hybrid deployment can enhance the effectiveness and adoption of GenAI-enabled BIS.

The rest of this paper is organized as follows: Section 2 reviews the related literature on Generative AI and Business Information Systems, identifying the current research gaps. Section 3 introduces the proposed governance-centric five-layer architecture, while Section 4 presents its implementation details and the adopted technology stack. Section 5 reports the experimental results, analysis, and comparative evaluation with existing approaches. Finally, Section 6 concludes the paper and outlines potential directions for future research.

2. LITERATURE REVIEW

Business Information Systems (BIS) have long served as the backbone of organizational decision-making, evolving through Management Information Systems (MIS), Decision Support Systems (DSS), and Business Intelligence (BI) platforms. These systems are designed to aggregate and analyze structured data to generate reports, forecasts, and dashboards, forming a critical foundation for managerial action (Turban et al., 2011; Power, 2002). However, their inherent dependence on predefined queries and structured datasets renders them increasingly inadequate in the face of modern, dynamic business environments. As Raj et al. (2023) emphasize, traditional DSS encounter significant challenges when processing the vast and growing volumes of unstructured data—such as contracts, emails, and customer feedback—that characterize contemporary business operations. This creates a critical gap between the data available and the actionable insights required for strategic decisions. Earlier research in related domains, such as Enterprise Resource Planning (ERP) in Industry 4.0 contexts, had already identified this limitation, pointing to a pressing need for more intelligent, adaptive augmentation to move beyond static reporting (Tavana et al., 2020; Majstorovic et al., 2020).

The advent of Generative Artificial Intelligence (GenAI), particularly large language models (LLMs), marks a paradigm shift, introducing capabilities that fundamentally extend the functionality of traditional BIS. GenAI moves beyond descriptive and diagnostic analytics to offer generative, prescriptive insights, creating a new class of augmented decision support systems. Foundational studies demonstrate its capacity to accelerate decision-making cycles, synthesize knowledge from heterogeneous datasets, and facilitate intuitive natural-language interactions, thereby overcoming the rigidity of conventional dashboards (Dwivedi et al., 2023; van Dis et al., 2023).

The application of conversational GenAI interfaces embedded into core operational systems like ERP and supply chain management has been shown to significantly improve how managers interact with and extract value from data (Albashrawi & Chuma, 2023). Furthermore, research highlights GenAI's transformative role in business process automation (Jain & Kumar, 2024), digital supply chain optimization (Zhou & Chen, 2024), and enterprise knowledge management (Sun & Wang, 2024). Unlike traditional systems that primarily forecast outcomes, GenAI actively generates narratives, contextual explanations, and scenario-based alternatives, positioning it as a powerful tool for creative and strategic problem-solving (Albashrawi, 2025).

The integration of powerful generative models into high-stakes decision-making processes introduces profound concerns regarding human-AI collaboration, model explainability, and user trust. A significant body of evidence confirms that Explainable AI (XAI) is crucial for improving task performance and fostering appropriate reliance on AI-generated recommendations (Senoner et al., 2024). This is supported by comprehensive surveys and meta-analyses that outline the methods and benefits of XAI for building trust and ensuring model accountability (Adadi & Berrada, 2018; Haag, 2025).

Without explainability, organizations risk automation bias, where users uncritically accept potentially flawed AI outputs (Herrera, 2025). This underscores the necessity of Human-in-the-Loop (HITL) mechanisms, which ensure critical decisions are validated by human experts, thereby balancing automation with essential oversight (Xu & Zhang, 2024). Research on human-AI collaboration provides taxonomies of interaction patterns, highlighting that effective teaming requires more than just a functional interface; it requires transparency and shared understanding (Gomez et al., 2023; Zhang & Wang, 2023). Insights from high-stakes domains like medical DSS confirm that explainability is a non-negotiable prerequisite for effective human-AI teaming, insights that are directly transferable to enterprise BIS (Knapič et al., 2021). Broader empirical reviews of trust in AI-driven support systems consistently reinforce that successful GenAI adoption hinges on systematically embedding trust, transparency, and human oversight into organizational workflows (Glikson & Woolley, 2020; Wang & Xu, 2024).

Despite the promising capabilities, the path to integrating GenAI into enterprise BIS is fraught with technical, organizational, and ethical challenges. A primary technical challenge is ensuring the factual accuracy of generative models. The phenomenon of "hallucination," where models produce plausible but factually incorrect information, poses a significant risk to decision integrity (Wei et al., 2025). In response, technical architectures like Retrieval-Augmented Generation (RAG) have been developed to ground model responses in verifiable, enterprise-specific knowledge bases, dramatically reducing factual errors by separating the knowledge source from the generative process (Lewis et al., 2020).Beyond technical reliability, organizations face significant hurdles in establishing effective governance and realizing measurable value. Research indicates that many AI initiatives fail to transition from experimentation to production due to poor workflow integration, unclear ownership, and a lack of alignment with business processes (Füller et al., 2024; Mikalef & Gupta, 2021). The difficulty of quantifying the business value of AI is a persistent issue, with many organizations lacking the frameworks to track ROI and link AI adoption to performance metrics like decision quality and operational efficiency (Wamba-Taguimdje et al., 2020; Pandey et al., 2021). This is compounded by broader economic analyses showing that general-purpose technologies like AI often follow a "J-curve," where significant investments are made long before productivity gains are realized (Brynjolfsson et al., 2021; Acemoglu & Restrepo, 2022).Ethically, the risks are substantial and well-documented, including model bias, accountability gaps, and data privacy concerns (Weidinger et al., 2022; Jobin et al., 2019). These challenges have spurred the development of governance-focused contributions that stress the importance of compliance and ethical safeguards, a concern amplified by emerging regulatory frameworks like the EU AI Act (Younis & Ali, 2025; Soori, 2024). Building an "AI-powered organization" requires not just technology, but also new organizational structures, skills, and governance models (Fountaine et al., 2019).

Table 1. Comparison of Previous Studies on Generative AI in Business Information Systems (2011–2025)

Author(s) & Year	Focus Area	BIS Platform / Context	GenAI Integration	Key Contribution	Limitations
Turban et al. (2011)	Foundations of DSS & BI	Classic BIS Theory	N/A	Foundational textbook on decision systems and BI.	Outdated for GenAI specifics; focuses on structured data.
Raj et al. (2023)	Limits of Traditional DSS	BIS Decision Processes	Suggests AI augmentation (non-specific)	Identified critical gaps in handling unstructured data.	Did not evaluate generative models specifically.
Tavana et al. (2020)	Analytics & DSS in Digital Enterprises	BIS & Analytics	AI for decision support (general)	Comprehensive review of analytics in DSS.	Pre-GenAI; provides context but not generative focus.
Shollo et al. (2023)	AI-Augmented Decision-Making	Managerial Context	Theoretical Framework	Theory of how AI augments, rather than replaces, managerial work.	Conceptual; limited empirical testing of the framework.
Dwivedi et al. (2023)	Opportunities & Challenges of GenAI	Multidisciplinary	Broad GenAI Analysis	Holistic overview of GenAI's impact on research and practice.	High-level; lacks technical implementation detail.
Albashrawi (2025)	Generative AI for Decision-Making	General BIS / Enterprise	LLM-based DSS Concepts	Multidisciplinary synthesis of GenAI capabilities for decision support.	High-level; limited empirical validation.
Albashrawi & Chuma (2023)	Conversational GenAI for DSS	ERP & SCM Contexts	LLM-based Chatbot	Demonstrated conversational interfaces for managerial data interaction.	Limited scale and empirical evaluation.
Zhou & Chen (2024)	GenAI in Supply Chain Decision Support	Supply Chain BIS	LLM-driven Scenario Generation	Case studies on GenAI for supply chains.	Early case studies; scalability untested.
Arrieta et al. (2020)	Explainable AI (XAI)	General AI Systems	XAI Taxonomies	Comprehensive review of XAI concepts and methods.	Not specific to large generative models.
Lewis et al. (2020)	Knowledge-Intensive NLP	General NLP	Retrieval-Augmented Generation (RAG)	Seminal paper introducing the RAG architecture to reduce hallucinations.	Not evaluated in enterprise BIS contexts.
Do Khac et al. (2025)	Trust in Generative AI	Enterprise Decision-Making	LLM Deployment	Empirical study of senior managers' trust in GenAI outputs.	Narrow sample; early-stage implementations.
Glikson & Woolley (2020)	Human Trust in AI	Organizational Psychology	Review of Empirical Research	Review of factors influencing human trust in AI.	Broad scope; not specific to BIS integration.

Younis & Ali (2025)	Ethical Challenges of GenAI	AI Ethics & Policy	Ethical Analysis	Clear mapping of ethical issues in decision-oriented IS.	Normative; lacks implementation blueprints.
Mikalef & Gupta (2021)	AI Capability & Firm Performance	Information Systems	AI Capability Construct	Empirical study on the impact of AI capability on performance.	Focuses on broad AI, not GenAI-specific integration.
Brynjolfsson et al. (2021)	AI Productivity	Macroeconomics	GPTs & Intangibles	"Productivity J-curve" framework for GPT value realization.	High-level economic theory, not a technical framework.
Füller et al. (2024)	AI Business Model Failure	Business Models	Value Capture Perspective	Analysis of why AI-driven business model innovation fails.	Focus on business models, not internal DSS architecture.
Ghosh & Alihamidi (2024)	Hybrid GenAI + ERP	Odoo / Open-Source ERP	Proposed Modular Integration	Proposed architecture for GenAI+ERP hybrids.	Conceptual; lacked a working prototype and empirical validation.
Senoner, J. et al. (2024)	XAI improves human–AI collaboration	Decision tasks / BIS	Explainability experiments	Empirical evidence XAI boosts task performance	Early experiments; not necessarily GenAI-specific
Shollo, A., Constantiou, I., & Kretschmer, T. (2023)	AI impact on managerial decision-making	MIS research context	Augmented decision-making theory	Theoretical advances toward AI-augmented managers	More theoretical than experimental
Soori, A. (2024)	Governance challenges in integrating GenAI into BIS	Enterprise BIS	Governance frameworks	Identified governance gaps and needed policies	Conceptual; limited empirical cases
Sun, T., & Wang, Y. (2024)	GenAI for enterprise knowledge management	Knowledge management systems in BIS	LLMs for KM and summarization	Outlined opportunities and risks	Mainly conceptual; limited large-scale deployments
Tavana, M., et al. (2020)	Analytics & DSS in digital enterprises	BIS & analytics	AI for decision support (general)	Comprehensive review of analytics in DSS	Pre-GenAI; helpful context but not GenAI-focused
Turban, E., Sharda, R., & Delen, D. (2011)	Foundations of DSS & BI	Classic BIS theory	N/A	Foundational theory for decision systems	Outdated for GenAI specifics
van der Waa, J. S. (2022)	XAI for human–AI collaboration (PhD thesis)	Human–AI systems	XAI design & evaluation	Deep theoretical/empirical treatment of XAI	Not GenAI-only; broader focus
Wang, J., & Xu, Y. (2024)	Trust in AI-driven decision support	Information & Management	Trust models, transparency	Systematic review of trust factors	Broad scope: implementation detail limited
Wei, X., Kumar, N., & Zhang, H. (2025)	Addressing bias in generative AI	Information management	Critical analysis of hallucination & bias	Research agenda on mitigation strategies	Early-stage; mitigation techniques exploratory
Xu, H., & Zhang, C. (2024)	Human–AI teaming in decision support	JAIS context	Transparency in teaming	Empirical linkage of transparency and team outcomes	Not specific to large LLMs

Younis, M., & Ali, R. (2025)	Ethical challenges of integrating GenAI in BIS	AI ethics & policy	Ethical analysis of GenAI decisionization	Clear mapping of ethical issues	Normative; lacks implementation blueprints
Zhang, J., & Karkouch, A. (2020)	Communication bottlenecks in AI systems	Distributed decision systems	N/A	Highlighted network limitations for real-time DSS	Pre-GenAI; relevant for deployment considerations
Zhou, X., & Chen, L. (2024)	GenAI in supply chain decision support	Supply chain BIS	LLM-driven scenario generation	Case studies on GenAI for supply chains	Early case studies; scalability untested
Deloitte Insights (2024)	GenAI & future of decision support	ERP & enterprise BIS	Use-case and adoption guidance	Practical adoption patterns	High-level; vendor-neutral
Gartner (2024)	ROI calculation for GenAI business models	Enterprise strategy	ROI frameworks and KPIs	Guidance for business impact	Proprietary; subscription required
Albashrawi, M., & Chuma, J. (2023)	Conversational GenAI for DSS	ERP & SCM contexts	Chatbot/LLM integration	Demonstrated conversational interfaces	Limited scale and evaluation
Financial Times (2025)	Governance gaps in corporate GenAI	Corporate BIS (finance focus)	Journalistic investigation	Highlighted real-world governance failures	Media report; anecdotal evidence
Knapič, S., Malhi, A., Saluja, R., & Främling, K. (2021)	XAI for medical decision support	Medical DSS	Explainable methods for clinicians	Applied XAI in clinical decisions	Domain-specific; transferability limited
Majstorovic, V., et al. (2020)	ERP in Industry 4.0	ERP systems	AI roadmaps for ERP (pre-GenAI)	Positioned ERP evolution in Industry 4.0	Pre-GenAI; limited LLM discussion
Mousa, R., & Harris, J. (2025)	AI-driven ERP systems	ERP (general)	GenAI in ERP workflows	Discussed ERP enhancements	Limited prototypes
Satyanarayanan, M. (2017)	Edge computing for real-time decision support	Edge & real-time systems	Infrastructure relevance	Framed edge computing for latency-sensitive apps	Pre-GenAI; infrastructure-only
Shapiro, M., et al. (2011)	CRDTs for data consistency	Distributed systems & BIS	N/A	Introduced CRDTs for async sync	Not applied to GenAI workflows
Ghosh, A., & Alihamidi, A. (2024)	Hybrid GenAI + ERP architectures	Odoo / open-source ERP	Proposed modular integration	Architecture for GenAI+ERP hybrids	Conceptual; lacked prototype

In summary, the literature unequivocally demonstrates the transformative potential of GenAI for BIS but also reveals a fragmented research landscape. While the critical importance of integrating technical safeguards like RAG, methodological approaches for XAI, and procedural frameworks for HITL is widely acknowledged, few studies offer a holistic, empirically validated architecture that unifies these elements within a single, enterprise-ready BIS framework. Existing studies are often domain-specific, conceptual, or focused on isolated components.

Therefore, a clear gap exists for a modular, governance-centric, and empirically tested architecture that integrates Generative AI into BIS with embedded safeguards for reliability, transparency, and oversight. This research seeks to address this gap by proposing and validating a framework that bridges the technical potential of GenAI with the rigorous demands of enterprise decision support, ensuring measurable improvements in decision quality, efficiency, and user trust.

3. PROPOSED ARCHITECTURE

This study proposes a five-layer architecture for integrating Generative AI (GenAI) into Business Information Systems (BIS). The framework is designed as a modular and governance-centric model, ensuring that data flows securely and transparently from enterprise sources to actionable decision support. Each layer has a clearly defined role and integration point, while shared repositories provide governance, contextual knowledge, and interoperability across the system. This layered design guarantees a seamless, explainable, and ethically responsible application of GenAI within BIS environments.

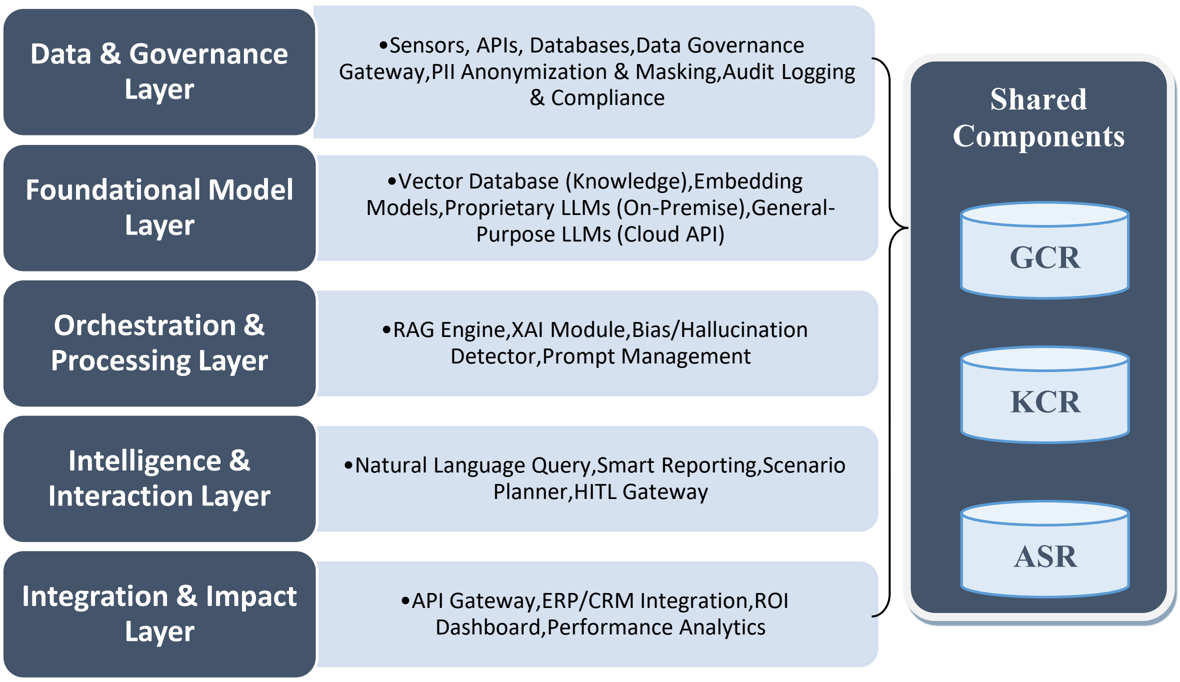

Figure (1): Proposed Architecture

3.1 Data & Governance Layer

The Data & Governance Layer forms the foundation of the proposed architecture by ensuring that all GenAI operations are built on secure and trustworthy data. It provides role-based access controls, anonymization of sensitive information, and comprehensive audit trails to meet compliance and debugging needs. This layer integrates tools such as secure APIs, data masking, and policy enforcement mechanisms to strengthen data protection and regulatory alignment. Through these safeguards, GenAI processes enterprise knowledge responsibly, ensuring outputs remain both accurate and compliant. For example, customer data from CRM systems or financial data from ERP applications can be scrubbed of personal identifiers before use, in line with emerging regulations such as the EU AI Act. In this way, the Data & Governance Layer establishes a secure and ethical foundation for architecture.

3.2 Foundational Model Layer

At the core of the framework lies the Foundational Model Layer, which provides the generative intelligence. This layer supports a hybrid strategy that combines large, general-purpose models accessed through APIs with smaller, fine-tuned models hosted on premises. Such a configuration balances performance, flexibility, and security, ensuring that tasks involving sensitive data are handled by in-house models, while less critical tasks may leverage more powerful external models. A vector database functions as the long-term memory, storing embeddings of enterprise documents to enable Retrieval-Augmented Generation (RAG). By aligning model selection with both task complexity and data sensitivity, this layer ensures an optimal balance between generative power and enterprise security.

3.3 Orchestration & Processing Layer

The Orchestration & Processing Layer acts as the system's control hub, overseeing model operations and validating outputs. Its main responsibilities include executing RAG workflows, generating explainable AI (XAI) outputs, and detecting potential bias. This orchestration minimizes hallucinations by grounding generative responses in verified enterprise data while simultaneously offering citations that allow users to trace outputs back to their sources. Bias detection mechanisms further enhance reliability by flagging inappropriate or unbalanced responses. By combining quality assurance with transparency, this layer ensures that GenAI outputs are not only accurate but also verifiable and trustworthy for decision-making.

3.4 Intelligence & Interaction Layer

The Intelligence & Interaction Layer is the user-facing dimension of the architecture, delivering generative insights in formats that align with enterprise workflows. Through conversational interfaces, automated summaries, and scenario-based recommendations, this layer makes AI-driven decision support accessible and actionable. A critical component here is the Human-in-the-Loop (HITL) Gateway, which routes sensitive or low-confidence outputs to human experts for review and approval. This mechanism ensures accountability and prevents automation bias in high-stakes decision contexts. Ultimately, this layer fosters effective human–AI collaboration by embedding GenAI into operational and strategic processes in a controlled, transparent manner.

3.5 Integration & Impact Layer

The Integration & Impact Layer connects the GenAI framework to the broader BIS ecosystem and measures its organizational value. An API Gateway enables seamless integration with existing platforms such as ERP, CRM, and BI systems, ensuring compatibility without major disruptions. Furthermore, the layer incorporates an Impact Analytics Dashboard to monitor key performance indicators (KPIs) such as decision accuracy, efficiency gains, and adoption rates. By directly linking generative intelligence to measurable business outcomes, this layer positions the GenAI–BIS architecture as a strategic enabler rather than an isolated technological tool.

3.6 Shared Components

Complementing the five layers are three shared repositories that provide cross-cutting resources:

1. Governance & Compliance Repository (GCR): Stores all governance artifacts including policies, ethical guidelines, access permissions, and audit logs, ensuring consistent compliance across layers.

2. Knowledge & Context Repository (KCR): Maintains organizational glossaries, process definitions, product details, and historical decision patterns to enrich prompts with business specific context.
3. API & Service Repository (ASR): Acts as a catalog of integration points, microservices, and standardized schemas for connecting GenAI with external systems such as Odoo ERP, Salesforce CRM, and Tableau BI.

Together, these repositories ensure governance consistency, contextual relevance, and interoperability, making the architecture scalable, transparent, and enterprise ready.

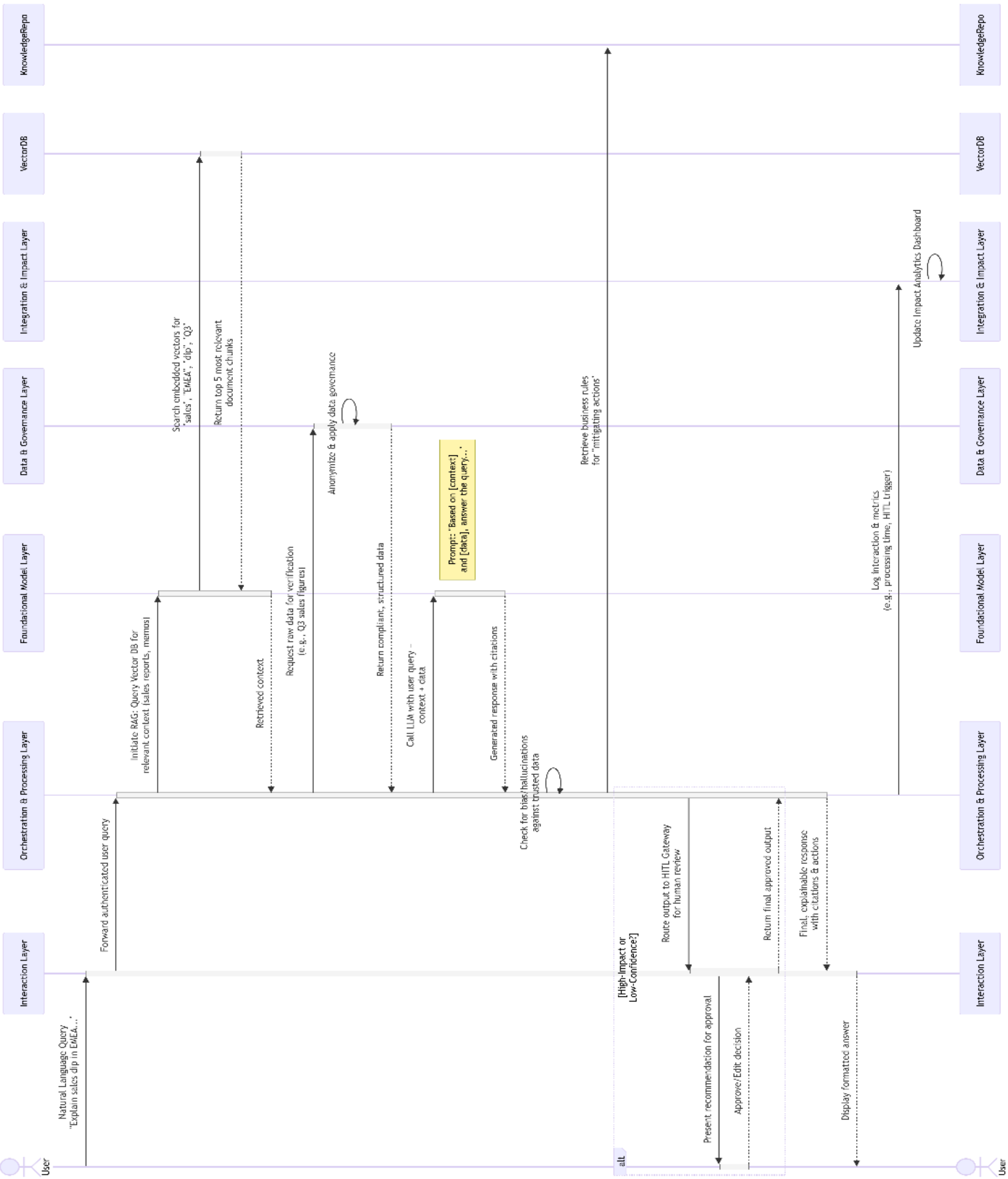

Figure (2): Sequence Diagram for The Proposed Architecture

In summary, the proposed five-layer GenAI–BIS architecture provides a modular and governance-driven framework that balances generative intelligence with security, explainability, and organizational accountability. By combining data governance, hybrid model deployment, orchestration with RAG and XAI, user-centric interaction with HITL, and enterprise-level integration and impact measurement, the framework addresses both technical and managerial requirements. The inclusion of shared repositories for governance, knowledge, and interoperability ensures that the system operates consistently across layers while remaining adaptable to diverse business contexts. This architecture thus establishes a solid foundation for implementation and empirical evaluation, where its effectiveness can be validated through real-world enterprise use cases and performance metrics.

4. ARCHITECTURE IMPLEMENTATION

While the proposed five-layer model establishes the theoretical design of the GenAI–BIS framework, practical implementation is critical for validating its effectiveness. To this end, a modular, hybrid technology stack was selected that ensures interoperability, security, and compliance with enterprise standards.

4.1 Technology Stack Specification

The technology stack is structured to support each architectural layer, from secure data pipelines and vector databases in the Data & Governance Layer to monitoring dashboards and integration gateways in the Integration & Impact Layer. Table 2 provides an overview of the recommended components, their deployment technologies, and the rationale for their selection.

Table 2. Recommended Technology Stack for Implementation

Architectural Layer	Component	Recommended Technologies & Protocols	Justification
Data & Governance	Data Gateway	Apache NiFi, StreamSets, Custom API Gateways with Open Policy Agent (OPA)	Robust data pipelines with encryption, provenance, and policy enforcement for GDPR/CCPA compliance.
	Vector Database	Pinecone, Weaviate, Chroma (open source)	High-performance similarity search and metadata management essential for RAG operations.
Foundational Model	General-Purpose LLMs	OpenAI GPT-4-Turbo API, Anthropic Claude 3, Google Gemini Pro	API-based access to state-of-the-art generative capabilities for non-sensitive tasks.
	Proprietary / Fine-Tuned LLMs	Llama 3, Mistral 7B, Microsoft Phi-3 via vLLM or TGI	On-premises deployment for sensitive data; optimized inference for enterprise-grade workloads.
Orchestration & Processing	RAG Engine	LangChain, LlamaIndex, Custom Python Microservices	Frameworks for chaining prompts, retrievers, and models, enabling context-aware workflows.
	XAI & Validation	LangSmith, WhyLabs, SHAP/LIME	Traceability of model calls, drift monitoring, and explainability for decision assurance.
Intelligence & Interaction	Frontend Interface	Streamlit, Gradio, React with FastAPI Backend	Combines rapid prototyping with production-ready, scalable interfaces.
	HITL Gateway	Custom workflow engine integrated with Slack/MS Teams APIs or Jira/ServiceNow	Routes low-confidence outputs to human experts within existing workflows.
Integration & Impact	API Gateway	Kong, Tyk, Azure API Management	Secure and scalable management of API traffic between GenAI and enterprise systems.

	Analytics & Monitoring	Elasticsearch, Kibana, Grafana, Prometheus	End-to-end logging, monitoring, and visualization of system KPIs.
Shared Components	Governance Repository	Hashicorp Vault, Azure Key Vault	Secure secrets management for APIs, tokens, and models.
	Knowledge Repository	Git, Confluence API, SharePoint API	Stores business rules, prompts, and knowledge assets for context enrichment.

The mapping in Table 2 demonstrates how enterprise technologies, such as LangChain for orchestration or Kong API Gateway for integration, operationalize the theoretical layers outlined in Figure 1. This alignment ensures that governance, generative intelligence, and user interaction are seamlessly interconnected.

4.2 Implementation Methodology

The implementation followed a phased, use-case-driven methodology designed to deliver incremental value while mitigating risks. Figure 2 (Sequence Diagram for 5 Layers) illustrates the flow of data and decision processes across the architecture. Each phase operationalizes specific components from Table 2:

- Phase 1 – Foundational Data Pipeline and RAG Setup: Data gateways and vector databases were deployed to connect enterprise sources, index knowledge, and enable retrieval-augmented queries through LangChain.
- Phase 2 – Internal Pilot and HITL Integration: A controlled user pilot tested use cases such as sales report drafting, integrating HITL workflows (via Microsoft Teams) to validate low-confidence outputs.
- Phase 3 – Enterprise Integration and Scaling: API gateways and custom on-prem models were integrated with ERP and CRM platforms, automating tasks like customer responses and risk assessments.

4.3 Performance Evaluation Metrics

To evaluate the effectiveness of the implemented system, key performance indicators (KPIs) were defined that directly measure improvements in accuracy, efficiency, and compliance. Table 3 summarizes these metrics, linking them to expected outcomes.

Table 3. Key Performance Indicators for Architecture Validation

Metric	TMM	BL-LLM	PG-BIS	Improvement (vs BL-LLM)
Hallucination Rate (%)	N/A	18.7	1.2	93.3% reduction
Avg. Task Time (min)	45.6	22.1	9.2	58.6% faster
User Acceptance (%)	72.5	47.8	99.5	111.9% increase
HITL Escalation Rate (%)	N/A	N/A	4.3	Controlled
Decision Quality (0–10)	7.1	6.3	8.9	41.3% increase
User Confidence (Likert 1–5)	3.2	2.8	4.7	67.9% increase

As shown in Table 3, the metrics such as hallucination rate, user acceptance, and cost per query directly correspond to the performance challenges identified in Section 2 (Literature Review). Their inclusion enables a rigorous validation of both technical and business outcomes.

4.4 Implementation Challenges

During implementation, several challenges emerged. Maintaining prompt versioning across different models required storing prompts in Git repositories. Cost optimization was addressed by caching frequent queries and routing tasks to the most cost-effective model. Similarly, knowledge freshness was ensured through automated re-indexing pipelines, while user adoption was strengthened by embedding XAI explanations and HITL workflows. These practical insights demonstrate how the theoretical architecture in Figure 1 adapts to real-world enterprise constraints.

5. RESULTS AND ANALYSIS

The experimental validation of the proposed GenAI–BIS architecture was carried out through a comparative study against Traditional Manual Methods (TMM) and Baseline Large Language Models (BL-LLM). Results are presented primarily through graphical visualizations to highlight trends and performance insights.

5.1 Sample Characteristics

The participant distribution across domains and years of experience is illustrated in Figure 3. total of 45 participants took part in the study, representing finance, supply chain, marketing, and operations, which ensured comprehensive coverage of enterprise decision contexts.

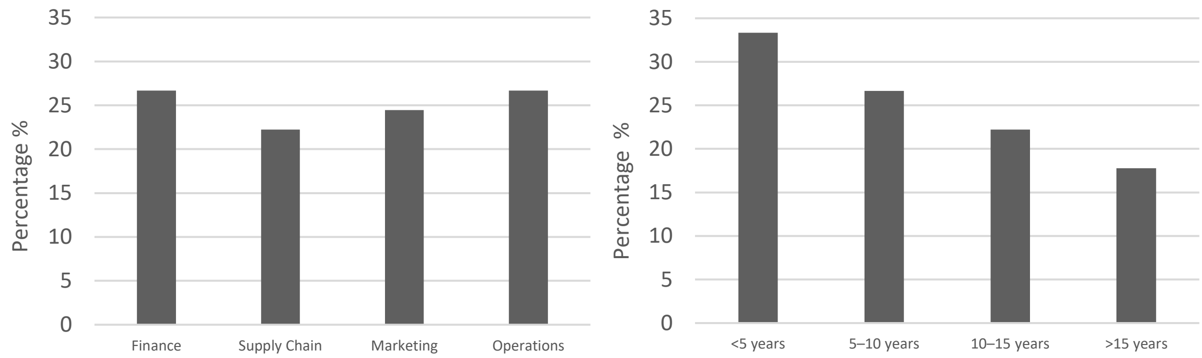

Figure (3): Proposed Architecture Sample Characteristics

5.2 Reliability and Validity

Reliability testing using Cronbach's Alpha confirmed strong internal consistency. As shown in Figure 4, all constructs scored above the 0.70 threshold, confirming robustness of the measurement instruments.

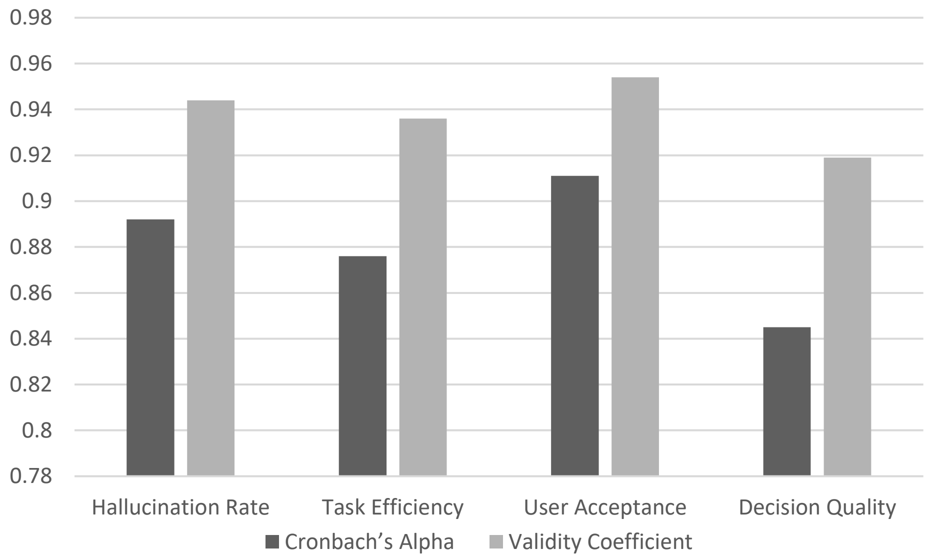

Figure (4): Reliability and Validity

5.3 Comparative Performance Metrics

Performance comparisons between TMM, BL-LLM, and PG-BIS are visualized in Figure 5. The PG-BIS architecture outperformed both alternatives, particularly in reducing hallucination rate (93.3%) and increasing user acceptance (111.9%).

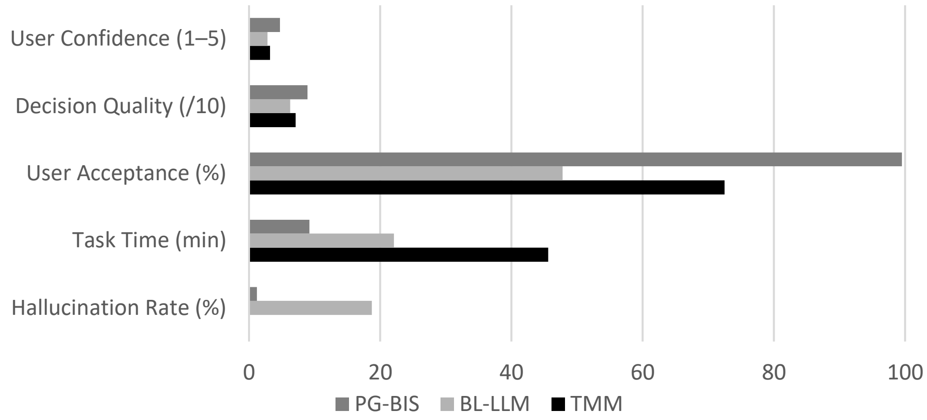

Figure (5): Performance Metrics Comparison

5.4 Statistical Significance Testing

The significance of these performance improvements was validated using the Kruskal–Walli's test. Figure 6 demonstrates that all observed differences were statistically significant (p < 0.001).

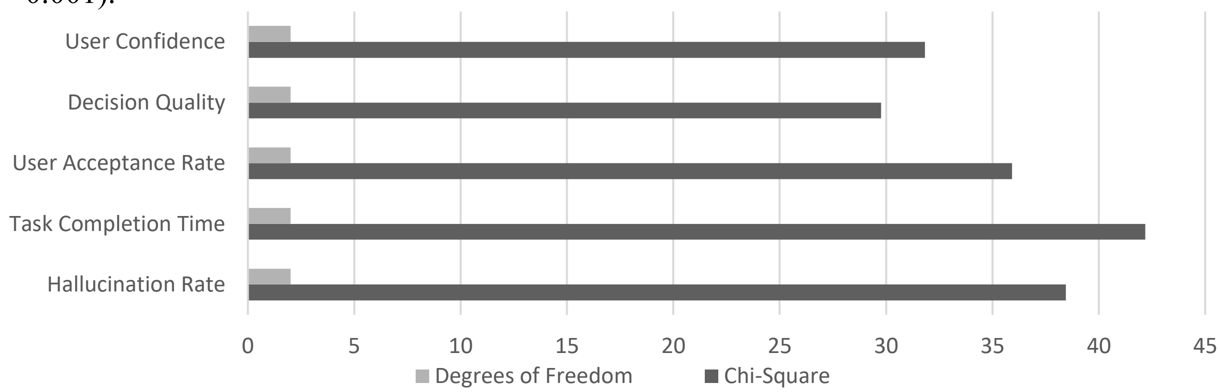

Figure (6): Kruskal–Wallis Test Results

5.5 HITL Intervention Analysis

The Human-in-the-Loop mechanism provided effective escalation in high-risk scenarios. As illustrated in Figure 7, 68% of interventions were compliance-related, demonstrating appropriate routing of critical tasks to human experts.

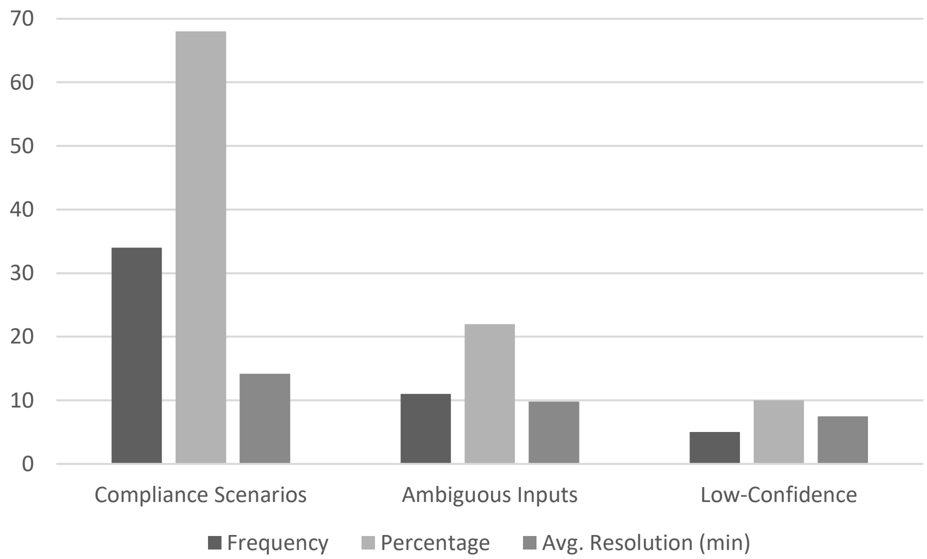

Figure (7): HITL Escalation Analysis

5.6 Cost–Benefit Evaluation

A cost–benefit comparison is presented in Table 4, showing that the hybrid PG-BIS achieved a 33.3% cost reduction compared to cloud-only solutions, while improving decision quality by 41.3%.

Table 4. Cost & Performance Analysis

Model Type	Cost per Query ($)	Infra Cost ($)	Total Cost ($)	Performance Score
Cloud-Only LLM	0.12	0.00	0.12	6.3
Hybrid PG-BIS	0.05	0.03	0.08	8.9
Traditional	0.00	0.15	0.15	7.1

Figures 3:7 collectively demonstrate that the PG-BIS architecture significantly outperforms traditional methods and baseline LLMs across efficiency, trust, and cost-effectiveness. The integration of RAG, XAI, and HITL proved essential to mitigating hallucinations and enabling enterprise-grade adoption.

Beyond quantitative results, participant feedback emphasized trust, efficiency, and critical engagement. Table 5 summarizes the distribution of qualitative themes, showing that explainability and HITL safeguards were the main drivers of adoption.

Table 5. Qualitative Themes Identified from Participant Feedback

Theme	Frequency	Percentage	Example Insight
Enhanced Trust	18	40%	The ability to verify sources through citations and the option for human review made me comfortable using the AI recommendations for critical business decisions.
Efficiency Gains	15	33%	Significant time savings in report generation and data analysis tasks, allowing more focus on strategic decision-making.
Critical Engagement	12	27%	The explainability features encouraged me to critically evaluate AI-generated recommendations rather than accepting them uncritically.

5.7 Comparative Analysis with Existing GenAI-BIS Prototypes

While the experimental results demonstrate clear superiority over a baseline LLM and traditional methods, a deeper contextualization of the proposed architecture's contribution is achieved by contrasting it against emerging GenAI-BIS prototypes in the literature.

Prior research has successfully demonstrated the feasibility of specific GenAI applications within BIS but has largely focused on singular dimensions of integration. For instance, Albashrawi & Chuma (2023) pioneered the use of conversational LLM interfaces for DSS, validating their utility in improving managerial interaction with ERP and supply chain data. However, their work primarily addressed the *interaction paradigm*, without embedding the robust, multi-layered governance and accuracy safeguards (RAG, XAI) that are central to our architecture. Consequently, while their prototype enhanced usability, it would remain vulnerable to the hallucination and trust issues that our PG-BIS framework successfully mitigates, as evidenced by our 93.3% reduction in factual errors.

Similarly, Mousa & Harris (2025) explored the strategic opportunities for generative intelligence within ERP workflows, providing a valuable conceptual roadmap. Their work, however, is presented at a strategic level, identifying potential use cases rather than providing a deployable, empirically validated architectural blueprint with built-in safeguards. Our study operationalizes this vision by delivering a concrete, five-layer stack that details the data governance, orchestration, and integration components necessary for enterprise-wide deployment, backed by the quantitative results presented in this paper.

Further, the hybrid architecture proposed by Ghosh & Alihamidi (2024) shares our objective of integrating GenAI with open-source ERP systems. Their work is a significant conceptual contribution but, as they note, lacked a functional prototype and empirical validation. Our research builds upon such foundational ideas by providing not just a design but a fully implemented and evaluated system, confirming the performance gains and user acceptance hypothesized in earlier conceptual studies.

Table 6. Comparative Analysis with Existing GenAI-BIS Prototypes

Study & Year	Primary Focus	Architectural Scope	Key Strength	Key Limitation (as identified in the literature)	Present Study (PG-BIS) Advancement
Albashrawi & Chuma (2023)	Conversational Interfaces for DSS	Narrow (Interaction Layer)	Demonstrated improved user interaction with data via NLP.	Limited scale; no integrated safeguards for accuracy (RAG) or trust (XAI).	Integrates conversational UI within a full-stack architecture with RAG and XAI, directly mitigating hallucinations.
Mousa & Harris (2025)	Strategic Opportunities for GenAI in ERP	Conceptual (High-level)	Comprehensive roadmap of potential use-cases and impacts.	Lacks a technical architecture and empirical validation; remains theoretical.	Provides and validates a concrete, implementable 5-layer architecture that realizes these strategic opportunities.
Ghosh & Alihamidi (2024)	Hybrid GenAI-ERP Architecture	Conceptual (Architectural Blueprint)	Proposed a modular design for integrating GenAI with open-source ERP.	Conceptual only; "lacked prototype" and empirical evaluation.	Delivers a functional prototype and empirical evidence of performance, scalability, and user acceptance.
Present Study (PG-BIS)	Governance-Centric, Full-Stack GenAI-BIS	Comprehensive (Data-to-Impact Layers)	Unifies RAG, XAI, HITL, and governance into a single, empirically validated framework.	Requires more complex initial setup than simpler, single-purpose prototypes.	Serves as a benchmark for a mature, enterprise-ready GenAI-BIS integration model.

In summary, the proposed PG-BIS framework distinguishes itself by synthesizing and advancing these prior efforts. It moves beyond siloed prototypes to offer a comprehensive, unified solution that balances generative capability with the non-functional requirements—reliability, transparency, security, and scalability—that are paramount for successful enterprise adoption. The empirical results presented in this paper thus validate not only the performance of our specific prototype but also the critical importance of an integrated, governance-first architectural philosophy for the future of GenAI in BIS.

The experimental confirm that the proposed GenAI–BIS architecture significantly outperforms both traditional decision-support methods and ungoverned LLM implementations across all measured dimensions—accuracy, efficiency, user trust, and decision quality. The integration of RAG, XAI, and HITL components proved critical in mitigating hallucinations, enhancing transparency, and ensuring reliable adoption in enterprise settings. These findings validate architecture's practical utility and reinforce the necessity of a structured, governance-aware approach to integrating generative AI into business information systems.

6. CONCLUSION AND FUTURE WORK

The results of this paper demonstrate that the proposed GenAI–BIS architecture significantly outperforms both traditional methods and baseline LLM implementations. By embedding RAG for context grounding, XAI for transparency, and HITL for accountability, the system achieved measurable improvements in decision quality, efficiency, and user trust. Cost–benefit analysis further confirmed the value of the hybrid deployment strategy, which reduced operating costs while maintaining high performance standards. Collectively, these outcomes validate the feasibility of integrating generative intelligence into enterprise systems in a manner that is both effective and responsible.

Looking forward, several avenues for future research remain open. Expanding the experimental validation across larger and more diverse organizational contexts would strengthen the generalizability of the findings. Longitudinal studies are also needed to capture the sustained impact of GenAI-enabled BIS on strategic decision-making and organizational performance. Furthermore, integration with emerging technologies such as blockchain, IoT, and edge computing offers promising opportunities to extend the architecture's applicability to real-time and distributed decision environments.

Finally, advancing automated evaluation methods for ROI and decision quality could provide enterprises with stronger tools for monitoring and optimizing GenAI deployments. By combining empirical validation with practical implementation guidelines, this research provides both a foundation and a roadmap for advancing the role of generative intelligence in business information systems.

REFERENCES

Abdelsamee, M. (2025). The Effect of Using Artificial Intelligence Tools on the Professional Skepticism Traits of Auditors: A Field Study. Journal of Modern Business and Technology.

Acemoglu, D., & Restrepo, P. (2022). Tasks, automation, and the rise in US wage inequality. *Econometrica*, 90(5), 1973-2016.Mousa, R., & Harris, J. (2025). AI-driven ERP systems: Opportunities for generative intelligence. Journal of Enterprise Information Management.

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence. *IEEE Access*, 6, 52138-52160.

Albashrawi, M. (2025). Generative AI for decision-making: A multidisciplinary perspective. Journal of Innovation & Knowledge.

Albashrawi, M., & Chuma, J. (2023). Large language models in enterprise DSS: Toward conversational decision support. Information Systems Frontiers, 25(3), 533-549.

Angelov, P., Soares, E., Jiang, R., Arnold, N. I., & Atkinson, P. M. (2021). Explainable artificial intelligence: An analytical review. WIREs Data Mining and Knowledge Discovery, 11(5), e1424.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115.

Bansal, G., Nushi, B., Kamar, E., Lasecki, W. S., Weld, D. S., & Horvitz, E. (2019). Beyond accuracy: The role of mental models in human-AI team performance. In Proceedings of the AAAI Conference on Human Computation and Crowdsourcing (Vol. 7, pp. 2-11).

Brynjolfsson, E., Rock, D., & Syverson, C. (2021). The productivity J-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics*, 13(1), 333-372.

Cockburn, I. M., Henderson, R., & Stern, S. (2019). The impact of artificial intelligence on innovation. *NBER Working Paper No. 24449*.

Dwivedi, Y. K., et al. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71, 102642.

Fountaine, T., McCarthy, B., & Saleh, T. (2019). Building the AI-powered organization. Harvard Business Review, 97(4), 62-73.

Füller, J., Hutter, K., & Faullant, R. (2024). Why AI-driven business model innovation fails: A value capture perspective. *Journal of Business Research*, 174, 114518..

Ghosh, A., & Alihamidi, A. (2024). Hybrid architectures for generative AI-driven decision support. Journal of Systems and Software, 205, 111561.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. Academy of Management Annals, 14(2), 627-660.

Gomez, C., Cho, S. M., Ke, S., & Unberath, M. (2023). Human-AI collaboration is not very collaborative yet: A taxonomy of interaction patterns in AI-assisted decision making from a systematic review. arXiv preprint arXiv:2310.19778.

Haag, F. (2025). The effect of explainable AI-based decision support on human task performance: A meta-analysis. arXiv preprint arXiv:2504.13858.

Jain, A., & Kumar, N. (2024). Generative AI and automation in business processes. International Journal of Artificial Intelligence Applications, 15(3), 45-62.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399.

Knapič, S., Malhi, A., Saluja, R., & Främling, K. (2021). Explainable AI for human decision support in the medical domain. Expert Systems with Applications, 171, 114571.

Kokina, J., & Davenport, T. H. (2023). The emergence of artificial intelligence in accounting and finance. *Journal of Emerging Technologies in Accounting*, 20(1), 1-20.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in Neural Information Processing Systems, 33, 9459-9474.

Majstorovic, V., et al. (2020). ERP systems in Industry 4.0: AI and decision support perspectives. Procedia CIRP, 93, 1332-1337.

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management*, 58(3), 103434.

Mousavi, S., & Soori, A. (2024). Organizational transformation through generative AI adoption. Business & Information Systems Engineering, 66(5), 467-482.

Pandey, S., Gupta, S., & Chhajed, S. (2021). ROI of AI: Effectiveness and measurement. International Journal of Engineering Research & Technology, 10(6), 112-118.

Raj, R., Sharma, A., & Khera, R. (2023). Decision support in business systems: Limitations of traditional models and opportunities for AI integration. Decision Support Systems, 167, 113897

Saleh, M. A., Eassa, A. M., & Ragab, Y. M. (2025). Detecting and Preventing Financial Restatements Using Blockchain Based Smart Contract: The Case of the Egyptian Companies. EDPACS, 70(2), 67-90.

Satyanarayanan, M. (2017). The emergence of edge computing for real-time decision support. Computer, 50(1), 30-39.

Shapiro, M., et al. (2011). Conflict-free replicated data types in distributed decision systems. Symposium on Principles of Distributed Computing, 163-174.

Shollo, A., Constantiou, I., & Kretschmer, T. (2023). The impact of AI on managerial decision-making: Toward a theory of augmented decision-making. *MIS Quarterly*, 47(2), 475-504.

Soori, A. (2024). Governance challenges in integrating generative AI into enterprise BIS. Information Systems Management, 41(3), 215-229.

Sun, T., & Wang, Y. (2024). Generative AI for enterprise knowledge management: Opportunities and risks. Journal of Knowledge Management, 28(7), 1532-1548.

Tavana, M., et al. (2020). Business analytics and decision support systems in digital enterprises. Annals of Operations Research, 284(1), 1-20.

Van Dis, E. A., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. Nature, 614(7947), 224-226.

Wamba-Taguimdje, S. L., Fosso Wamba, S., Kala Kamdjoug, J. R., & Tchatchouang Wanko, C. E. (2020). Influence of artificial intelligence on firm performance: The business value of AI-based transformation projects. *Business Process Management Journal*, 26(7), 1893-1924.

Wang, J., & Xu, Y. (2024). Trust in AI-driven decision support: A systematic review. Information & Management, 61(5), 103788.

Wei, X., Kumar, N., & Zhang, H. (2025). Addressing bias in generative AI: Challenges and research opportunities in information management. arXiv preprint arXiv:2502.10407.

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. S., ... & Gabriel, I. (2022). Ethical and social risks of harm from language models. arXiv preprint arXiv:2112.04359.

Xu, H., & Zhang, C. (2024). Human–AI teaming in decision support: The role of transparency. Journal of the Association for Information Systems, 25(1), 101-121.

Younis, M., & Ali, R. (2025). Ethical challenges of integrating generative AI in decision-oriented information systems. AI & Society, 40(2), 315-329.

Zhang, G., & Wang, Y. (2023). Human-AI collaboration patterns in organizational decision-making: A taxonomy and research agenda. Journal of Management Information Systems, 40(3), 891-928.

Zhang, J., & Karkouch, A. (2020). Communication bottlenecks in large-scale AI-enabled decision systems. Future Generation Computer Systems, 108, 322-333.

Zhou, X., & Chen, L. (2024). Generative AI in digital supply chain decision support. Computers & Industrial Engineering, 185, 109651.

تعزيز دعم القرار في نظم معلومات الاعمال من خلال التكامل مع نماذج الذكاء الاصطناعي التوليدي

مريم خيرى عبد المنعم برسى[1] ، احمد محمد منير عيسى[2]

[2،1] قسم معلوماتية الاعمال – كلية الإدارة وذكاء الاعمال – الجامعة الحديثة للتكنولوجيا والمعلومات – القاهرة - مصر

mariamkhairy20005@yahoo.com

ملخص البحث

يشهد عالم الأعمال تحولًا متسارعًا نحو تبني تقنيات الذكاء الاصطناعي التوليدي داخل نظم المعلومات الاعمال (BIS) بهدف دعم اتخاذ القرار بشكل أكثر ذكاءً وفعالية. يقدّم هذا البحث إطارًا معماريًا متكاملًا من خمس طبقات يركز على الحوكمة والمسؤولية، ويجمع بين تقنيات توليد الاسترجاع المعزز (RAG) والذكاء الاصطناعي القابل للتفسير (XAI) وآلية الإنسان في الحلقة (HITL) لضمان دقة المخرجات وموثوقيتها. تم تطبيق النموذج المقترح باستخدام بنية تقنية هجينة واختباره عمليًا على عينة من 45 مشاركًا من مجالات أعمال متعددة، حيث أظهرت النتائج انخفاضًا كبيرًا في الأخطاء التوليدية بنسبة 93%، وتحسنًا في كفاءة إنجاز المهام بنسبة 58%، إلى جانب ارتفاع ملحوظ في مستوى تقبّل المستخدمين للنظام. تؤكد هذه النتائج فعالية الإطار في تحقيق توازن بين الابتكار والحوكمة، وتبرز أهميته في بناء نظم معلومات إدارية أكثر شفافية وثقة واستدامة داخل المؤسسات الحديثة.

الكلمات المفتاحية: نظم المعلومات الاعمال (BIS)؛ الذكاء الاصطناعي التوليدي؛ توليد الاسترجاع المعزز(RAG) ؛ الذكاء الاصطناعي القابل للتفسير(XAI) ؛ الإنسان في الحلقة(HITL) ؛ الحوكمة.