Frontiers in Computer Science and Artificial Intelligence
ISSN: 2978-8048
DOI: 10.32996/fcsai
Journal Homepage: www.al-kindipublisher.com/index.php/fcsai

FCSAI
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Enterprise Readiness for Generative AI: The Critical Role of Data Engineering

YUVACHANDRA MARASANI
Director, Software Development, Data Science & Engineering, Healthcare Data & Analytics Platforms, Agentic AI
Corresponding Author: YUVACHANDRA MARASANI**, E-mail**: yuvachandramarasani@gmail.com

| ABSTRACT

Generative artificial intelligence (GenAI) is accelerating enterprise transformation by enabling natural language interfaces, automated knowledge work, and new modes of decision support. Yet many GenAI initiatives stall after pilot stages—not because models underperform, but because enterprises lack the data engineering maturity required to operationalize these systems at scale. This paper argues that enterprise readiness for GenAI is primarily a data engineering problem: GenAI outcomes depend on the ability to reliably ingest, integrate, govern, and serve high quality data across fragmented environments. Through a literature based conceptual analysis, the study synthesizes the data engineering capabilities most critical to GenAI adoption, including scalable data pipelines, modern storage and processing architectures, integrated enterprise data platforms, and robust governance mechanisms for security, privacy, lineage, and compliance. The paper proposes a conceptual framework that links enterprise readiness dimensions (digital infrastructure, leadership alignment, talent, and governance) to data engineering capability as a mediating layer that enables sustainable GenAI deployment. The findings position data engineering as a strategic prerequisite for trustworthy, cost effective, and scalable GenAI systems, and provide a practical lens for enterprises to assess and strengthen readiness before broad rollout.

| KEYWORDS

Generative Artificial Intelligence; Enterprise Readiness; Data Engineering; Data Infrastructure; Data Pipelines; Artificial Intelligence Adoption; Digital Transformation

| ARTICLE INFORMATION

ACCEPTED: 01 November 2024 **PUBLISHED:** 25 November 2024 **DOI:** 10.32996/fcsai.2024.3.2.9

1. Introduction

1.1 Background

Generative artificial intelligence (GenAI) has emerged as a major technological shift in recent years, driven by rapid advances in large language models (LLMs), diffusion models, and related generative methods capable of producing high-quality text, images, and code. As a result, enterprises across industries are increasingly exploring GenAI to improve productivity, automate complex workflows, and enhance decision support. Recent studies highlight accelerating adoption of generative AI among both large enterprises and small-to-medium organizations, where digital innovation is pursued as a source of competitiveness and operational advantage.

In parallel, GenAI adoption has been associated with organizational resilience and increased capacity for innovation, particularly in environments characterized by uncertainty and rapid change. As firms invest in GenAI to modernize business models and create new sources of digital value, the focus frequently centers on model capability and application scenarios. However, the extent to which these initiatives deliver sustainable impact depends on whether enterprises can integrate GenAI into existing operating environments and data ecosystems in a manner that is scalable, governed, and aligned with enterprise risk constraints.

1.2 Problem Statement

Despite growing interest and investment in GenAI, many organizations encounter significant barriers when moving from experimentation to scalable deployment. A primary challenge is the limited maturity of the data infrastructure required to support enterprise-grade GenAI. In many environments, data remains distributed across disconnected platforms, legacy systems, and inconsistent repositories, making it difficult to assemble reliable datasets for AI workflows. Additionally, organizations often lack robust and well-engineered pipelines for ingesting, integrating, and governing data across diverse sources.

Because GenAI systems depend on large volumes of high-quality data and reliable data flows, fragmented architectures and weak pipeline design directly constrain performance, trustworthiness, and operational scalability. Equally important, insufficient data governance - covering access control, lineage, auditability, and privacy enforcement-limits an organization's ability to deploy GenAI responsibly, particularly when models interact with sensitive or regulated enterprise information. Without governance mechanisms that define authoritative data sources, permissible usage, and accountability, GenAI outputs may be difficult to validate, reproduce, or defend in compliance and risk reviews—thereby constraining enterprise adoption beyond pilot environments.

1.3 Research Gap

While existing literature broadly documents the benefits, use cases, and adoption momentum of generative AI technologies, comparatively less attention has been paid to the data engineering readiness required for successful enterprise implementation. Many studies emphasize technological potential, organizational performance outcomes, or high-level readiness factors, but provide limited detail on the practical engineering capabilities that enable scalable GenAI delivery—such as robust pipeline architectures, integrated data platforms, and governance mechanisms that support secure and compliant AI operations.

In particular, the governance dimension is frequently addressed at a conceptual level, while the operational requirements—such as end-to-end lineage, policy-driven access, data quality controls, and audit trails for data movement and transformation - remain under-specified in discussions of GenAI readiness. This gap is significant because, in enterprise contexts, GenAI success depends not only on models and applications but also on the ability to operationalize consistent, governed, and high-quality data flows. A clearer articulation of data engineering readiness as a prerequisite for GenAI adoption is therefore needed to guide enterprise investment and implementation strategies.

1.4 Research Objective

To address this gap, the objective of this paper is to examine data engineering capabilities as a defining element of enterprise readiness for generative AI adoption. Specifically, the study analyzes how modern data infrastructure, scalable pipeline architectures, and governance mechanisms enable effective implementation, reliability, and operational scalability of GenAI systems in enterprise environments. Through a literature-based conceptual analysis, the paper positions data engineering not merely as a supporting function, but as a strategic enabler that mediates between organizational readiness and successful GenAI adoption outcomes. Additionally, the paper emphasizes governance as an operational capability—where enforceable controls for privacy, security, lineage, and accountability must be embedded into data pipelines and platform architectures to enable trustworthy enterprise deployment.

Contribution sentence:
This study contributes a readiness-oriented synthesis and conceptual framework that reframes enterprise GenAI adoption through measurable data engineering capability—particularly pipeline maturity and embedded governance controls—as the critical mediating mechanism between organizational readiness and sustainable, scalable deployment outcomes.

1.5 Structure of the Paper

The remainder of this paper is structured as follows. Section 3 reviews the literature on generative AI adoption, enterprise readiness, and the role of data engineering in enabling AI systems. Section 4 analyzes the data engineering capabilities required to support scalable and trustworthy GenAI deployments in enterprise settings. Section 5 discusses strategic and organizational considerations—including capability development and risk implications—for enterprises pursuing GenAI. Finally, Section 6 concludes by summarizing key findings and outlining directions for future research.

2. Literature Review

2.1 Generative AI Adoption in Enterprises

The rapid advancement of generative artificial intelligence (GenAI) has accelerated enterprise interest in deploying large language models and generative systems to improve operational efficiency, support knowledge work, and enhance innovation capacity. Across both large enterprises and small-to-medium organizations (SMEs), recent studies indicate increasing exploration and adoption of GenAI as a driver of productivity and competitive advantage in digital markets.

A recurring theme in the literature is that GenAI can create value across multiple enterprise functions—including content generation, customer interaction automation, software development assistance, and analytic augmentation—by enabling organizations to streamline workflows and expand decision support. Etemad (2024) highlights the transformational potential of GenAI for entrepreneurial enterprises, emphasizing its role in enabling new business models and operational agility. Similarly, Kochkina et al. (2024) discuss GenAI adoption as a strategic initiative linked to competitiveness and broader economic performance outcomes.

In addition to productivity and innovation benefits, GenAI adoption is increasingly associated with organizational resilience. Shore et al. (2024) demonstrate how SMEs can strengthen adaptability during disruption by leveraging AI-driven insights and automated decision support. Likewise, Kmiecik et al. (2024) describe the role of generative AI in improving resilience within complex operational environments such as supply chain and logistics contexts.

Despite these reported benefits, the literature also reflects uneven adoption maturity across organizations. While some firms have advanced toward broader integration of GenAI, many initiatives remain constrained by organizational preparedness factors—particularly technological capacity, data architecture maturity, and operational readiness. Collectively, these findings suggest that access to capable models alone does not ensure enterprise success; rather, sustained adoption depends on complementary enterprise capabilities that support implementation, governance, and scaling.

2.2 Enterprise Readiness for Artificial Intelligence

Enterprise readiness for artificial intelligence refers to an organization's capacity—across infrastructure, governance, and managerial systems—to implement, operate, and sustain AI solutions. The literature consistently frames AI adoption as not solely a technical implementation challenge, but a broader organizational change requiring strategic alignment and digital maturity. Felemban et al. (2024) emphasize that organizational readiness encompasses both technological preparedness and institutional capability to integrate AI into business operations responsibly and effectively.

Digital maturity is frequently identified as a foundational determinant of AI readiness. Organizations with mature digital platforms, standardized data practices, and scalable IT ecosystems tend to be better positioned to implement AI at enterprise scale. Lai and Su (2024) further argue that successful integration of emerging technologies depends on dynamic capabilities that enable organizations to combine knowledge, technological resources, and strategic decision-making processes—supporting adaptation to technological change and embedding AI within broader digital transformation initiatives.

Governance and leadership structures also emerge as central readiness dimensions. Effective AI adoption requires leadership commitment and governance mechanisms capable of aligning AI initiatives with business objectives while addressing risk, compliance, and ethical considerations. Sharma (2024) highlights the role of strategic governance in enabling responsible AI adoption, including accountability mechanisms that mitigate risks related to security, transparency, and organizational control.

Synthesizing this literature, AI readiness frameworks commonly include dimensions such as digital infrastructure, data management capability, talent availability, leadership alignment, and governance mechanisms. Together, these dimensions shape the extent to which enterprises can deploy and sustain AI initiatives. Importantly, a growing emphasis across readiness research is that data management and underlying data infrastructure are not peripheral considerations, but prerequisites—particularly for GenAI systems that depend on reliable, governed access to large-scale data.

Readiness Dimension	Description	Importance for Generative AI
Digital Infrastructure	Availability of scalable computing systems, cloud platforms, and enterprise data architectures that support AI deployment.	Enables large-scale model training, data storage, and processing required for generative AI systems.
Data Management Capability	Organizational ability to collect, integrate, clean, and manage data from multiple enterprise sources.	Ensures generative AI models are trained on high-quality, structured datasets.

Leadership and Strategic Alignment	Organizational leadership commitment and strategic vision guiding AI adoption and implementation.	Aligns generative AI initiatives with long-term business objectives and innovation strategies.
Skilled Talent and Technical Expertise	Availability of data engineers, AI specialists, and technical teams capable of developing and maintaining AI systems.	Supports model development, pipeline management, and continuous system optimization.
Governance and Ethical Frameworks	Policies and procedures for responsible AI usage, compliance, data privacy, and risk management.	Ensures generative AI systems operate securely, transparently, and within regulatory standards.

Table 1: Key Components of Enterprise AI Readiness

Source: Adapted from Felemban et al. (2024); Lai & Su (2024); Sharma (2024).

2.3 Data Engineering as the Foundation for AI Systems

While AI adoption discourse often prioritizes model architectures and algorithmic advances, the literature emphasizes that the quality, accessibility, and reliability of the underlying data environment are decisive determinants of AI system performance in enterprise contexts. Data engineering provides the architectural and operational foundation for AI by enabling organizations to collect, store, transform, and manage large volumes of data in usable and governed forms. Sridhar et al. (2024) describe data engineering as the structural basis for supporting AI and machine learning through effective data processing and integration capabilities.

A central contribution of data engineering is the development and operation of robust data pipelines that enable consistent movement of data from diverse sources into centralized or federated platforms where it can be processed for analytics and AI use. Without well-engineered pipelines, organizations may struggle to supply stable, high-quality data streams required for AI training, evaluation, and production deployment. Data transformation and integration are equally critical, as enterprises typically operate heterogeneous systems that generate data across multiple platforms and formats. Gudavalli (2024) positions data engineering as a key enabler of digital transformation, particularly through its role in unifying enterprise data to support better decision-making.

Scalability is another recurring requirement in the literature, driven by increasing data volume and growing complexity of AI workloads. Modern data engineering tools and frameworks are presented as essential to enabling high-performance processing across large and dynamic datasets. Mbata et al. (2024) survey pipeline tools that support contemporary data engineering needs, while Wu et al. (2024) discuss approaches for optimizing data pipelines to improve machine learning workflows. In parallel, Yan et al. (2024) illustrate how advanced data engineering workflows can support complex distributed environments and large-scale data processing requirements.

Taken together, these studies position data engineering not merely as a supporting technical function, but as a core enterprise capability that materially influences the feasibility of AI deployment. This observation is particularly relevant for GenAI systems, which require consistent access to large, curated datasets and dependable data flows. Where data engineering practices are weak—due to fragmented architectures, unreliable pipelines, or insufficient scalability—GenAI initiatives may fail to operationalize reliably or may produce inconsistent and low-trust outputs. Therefore, the literature supports treating data engineering maturity as a central component of enterprise preparedness for sustainable GenAI adoption.

3. The Role of Data Engineering in Enterprise Generative AI Adoption

The enterprise adoption of generative artificial intelligence (GenAI) extends well beyond deploying advanced model architectures or selecting high-value use cases. While model capability is necessary, it is rarely sufficient for sustained enterprise impact. In practice, GenAI systems require dependable access to large volumes of relevant data, consistent data movement across heterogeneous sources, and controls that ensure security, privacy, and accountability. These requirements place data engineering at the center of enterprise GenAI readiness. Data engineering enables organizations to operationalize data at scale by building pipelines, architectures, and governance mechanisms that transform raw, distributed data into curated, usable, and trusted inputs for AI workflows. As such, the ability to implement GenAI in a reliable and scalable manner is directly mediated by the maturity of the enterprise's data engineering foundations.

3.1 Data Infrastructure for Generative AI

The technological foundation for enterprise GenAI begins with data infrastructure: the integrated set of ingestion, storage, processing, and orchestration capabilities that enable data to be acquired, standardized, and served to downstream consumers. A robust data infrastructure supports the end-to-end lifecycle of GenAI workflows by ensuring that data is continuously captured, reliably stored, efficiently processed, and made accessible for modeling and inference tasks.

Data ingestion is a first-order requirement, as enterprises generate data across diverse channels and formats, including structured operational data and unstructured content. Effective ingestion mechanisms enable continuous capture and controlled onboarding of data, which is essential for maintaining completeness and timeliness of AI inputs. Without reliable ingestion, GenAI pipelines can become inconsistent, leading to incomplete training corpora or out-of-date context during inference.

Storage architecture is equally critical. Because GenAI systems often require access to large datasets, enterprises typically rely on scalable storage patterns such as data lakes, distributed storage, and cloud-based architectures that can retain raw data while enabling downstream processing. Such designs support both structured and unstructured data while preserving flexibility for evolving GenAI use cases.

Processing and transformation pipelines convert raw data into structured, high-quality datasets suitable for analytics and AI. These pipelines commonly include cleansing, standardization, integration, and enrichment steps, enabling consistency and improving downstream model reliability. In the enterprise setting, the ability to manage processing at scale—both in terms of data volume and operational complexity—is a defining characteristic of infrastructure readiness for GenAI.

Finally, enterprises increasingly require integrated orchestration of data and model workflows, where data pipelines align with training, validation, and deployment activities. This integration enables repeatable operational cycles and supports continuous improvement of GenAI systems as new data becomes available and as enterprise requirements evolve

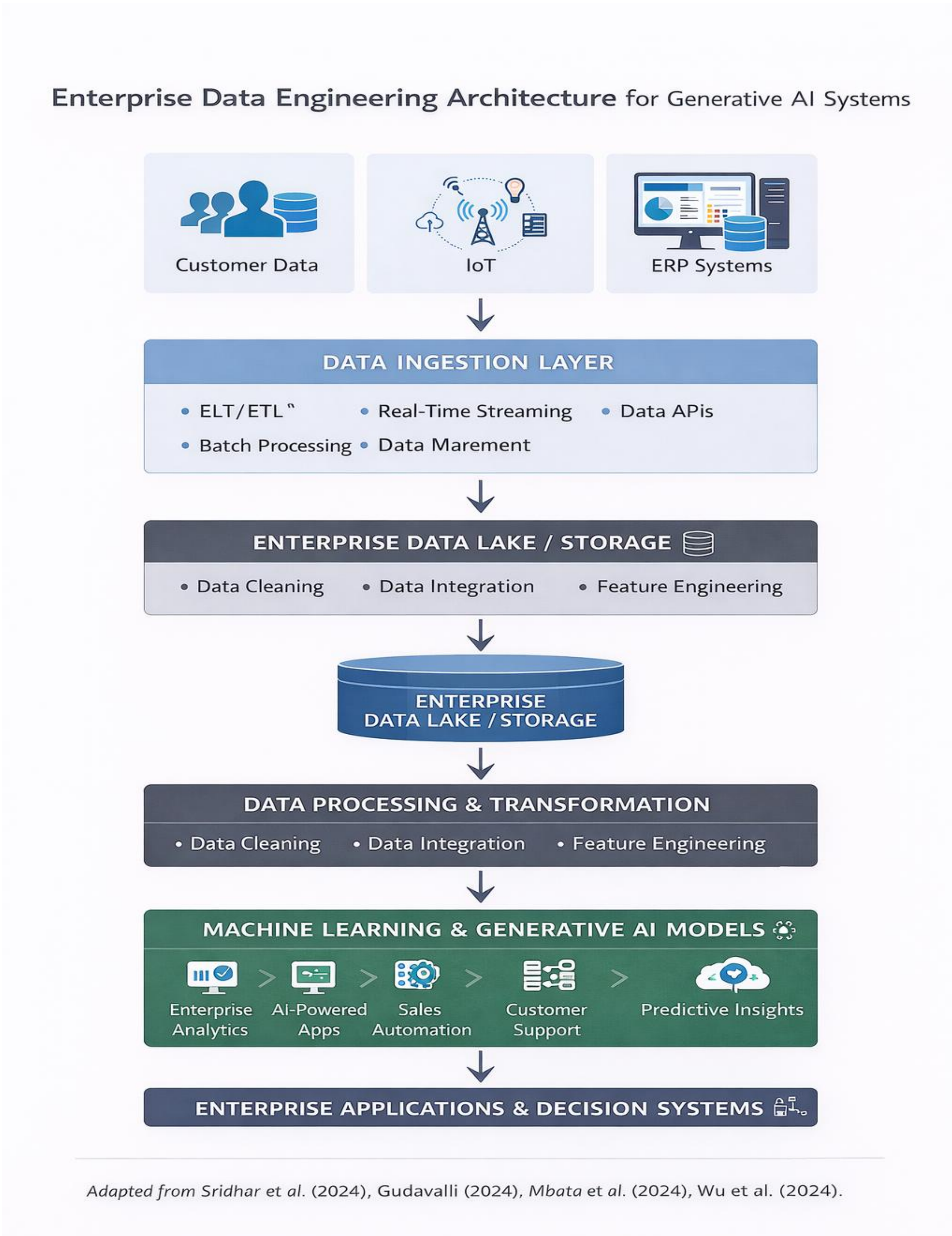


Figure 1. Enterprise data engineering architecture supporting generative AI systems in organizational environments.

Source: Adapted from Sridhar et al. (2024); Gudavalli (2024); Mbata et al. (2024); Wu et al. (2024).

3.2 Data Quality and Governance

Beyond infrastructure, **data quality and governance** represent decisive determinants of GenAI trustworthiness and enterprise viability. GenAI outputs are highly sensitive to the underlying data used for training, grounding, and operational decision support; therefore, poor data quality can directly lead to bias, inaccuracy, and inconsistent outputs that undermine user trust and business adoption. Data quality management in this context includes systematic data cleansing, standardization, validation, and curation practices to ensure that datasets are fit for purpose before they are used in GenAI pipelines.

Equally, enterprise readiness requires governance mechanisms that establish enforceable policies and procedures for data handling, access, and accountability. Governance frameworks support consistency and control by defining rules for data usage, strengthening security posture, and supporting compliance with regulatory requirements. In GenAI settings—where models may touch sensitive organizational or customer information—governance becomes operationally central rather than peripheral. Effective governance clarifies which data is authoritative, who can access it, how it can be used, and how decisions can be audited.

From an enterprise risk perspective, governance also supports transparency and accountability across the data lifecycle—spanning collection, transformation, storage, and consumption. Without these controls, organizations may face heightened risk of privacy violations, non-compliant usage of sensitive data, and limited ability to defend AI-assisted decisions in audit contexts. Consequently, data quality and governance should be treated as co-equal pillars of readiness alongside infrastructure and pipeline scalability.

3.3 Integration with Enterprise Systems

To create meaningful enterprise value, GenAI systems must be integrated with existing enterprise platforms and operational systems. Most enterprises operate complex ecosystems comprising enterprise resource planning platforms, operational databases, analytics systems, and specialized applications. GenAI initiatives that remain isolated from these systems often struggle to deliver sustained impact because they cannot reliably access the data required for context, nor can they deliver outputs into the workflows where decisions and operations occur. Data engineering enables this integration by establishing interoperable pipelines and architectures that connect data sources to GenAI workflows and downstream consumption layers.

Enterprise data lakes and centralized repositories are frequently described as key integration points, enabling organizations to consolidate data across domains and formats while supporting large-scale access for modeling. Integration also extends into analytics environments, where GenAI-augmented insights can be incorporated into reporting and decision support. Additionally, integration with machine learning pipelines and operational systems is essential to support the AI lifecycle—training, deployment, monitoring, and iterative improvement—while ensuring that data flows remain reliable and consistent across enterprise boundaries.

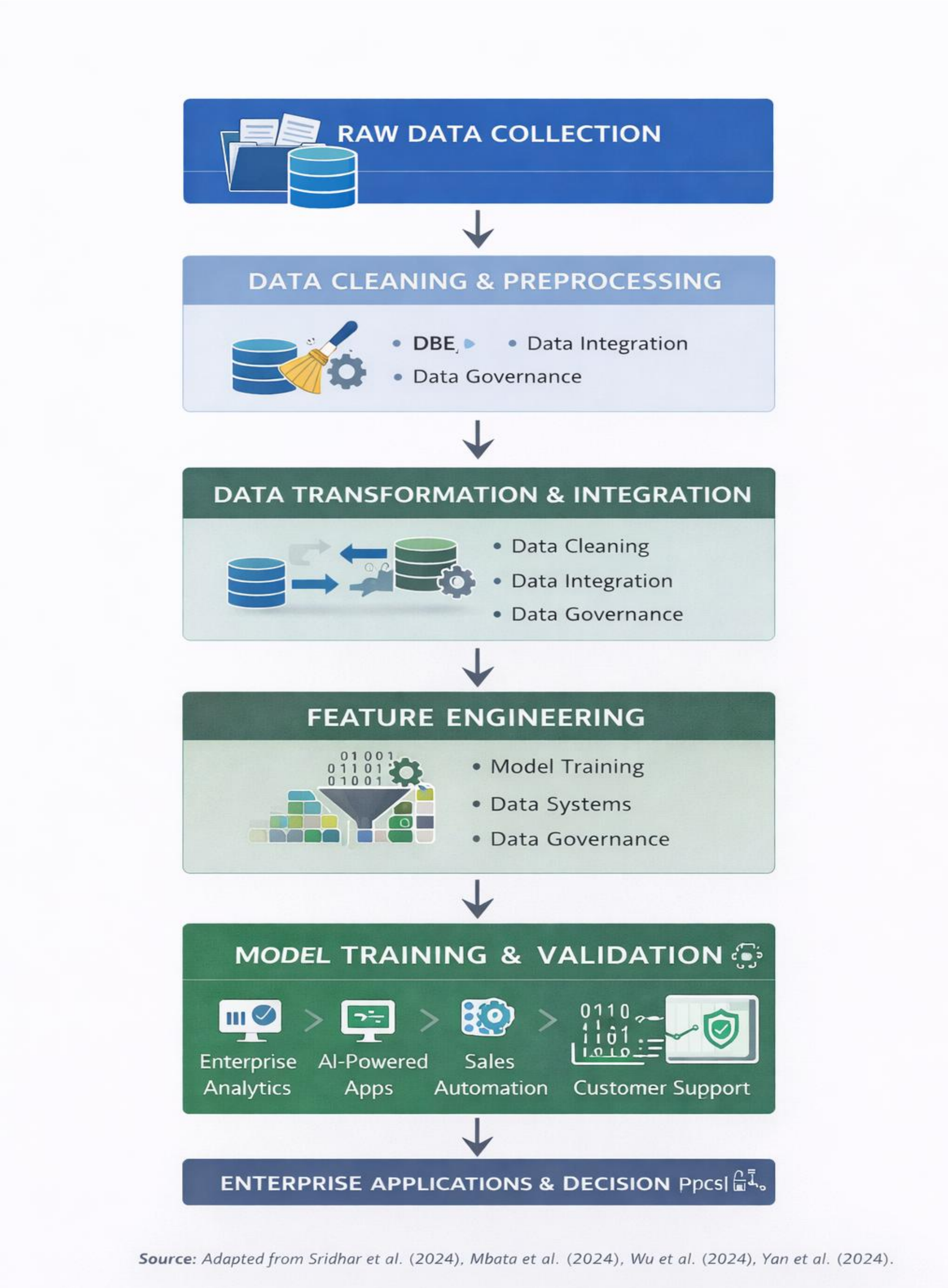


Figure 2. Data engineering pipeline supporting generative AI applications in enterprise environments.

Source: Adapted from Sridhar et al. (2024); Mbata et al. (2024); Wu et al. (2024); Yan et al. (2024).

3.4 Conceptual Framework

Building on the above analysis, this study conceptualizes enterprise GenAI adoption as a relationship mediated by data engineering capability. Enterprise readiness reflects the organization's overall capacity to adopt advanced technologies, including digital maturity, leadership alignment, governance structures, and enabling infrastructure. However, readiness alone does not ensure successful GenAI adoption unless the enterprise can translate high-level readiness into operational capability—specifically, the ability to deliver governed, scalable, and reliable data flows that GenAI systems require.

In this framework, data engineering capability acts as the mediating layer connecting enterprise readiness to GenAI adoption outcomes. Scalable architectures, robust pipelines, and governance frameworks provide the mechanisms through which raw and distributed enterprise data is transformed into structured, trustworthy inputs for GenAI systems. Enterprises with mature data engineering practices are therefore more likely to deploy GenAI systems reliably, scale them across business functions, and sustain them over time. Conversely, where data engineering capability is weak, GenAI initiatives may remain limited to pilots or experience failures due to unreliable data, operational instability, and governance risk.

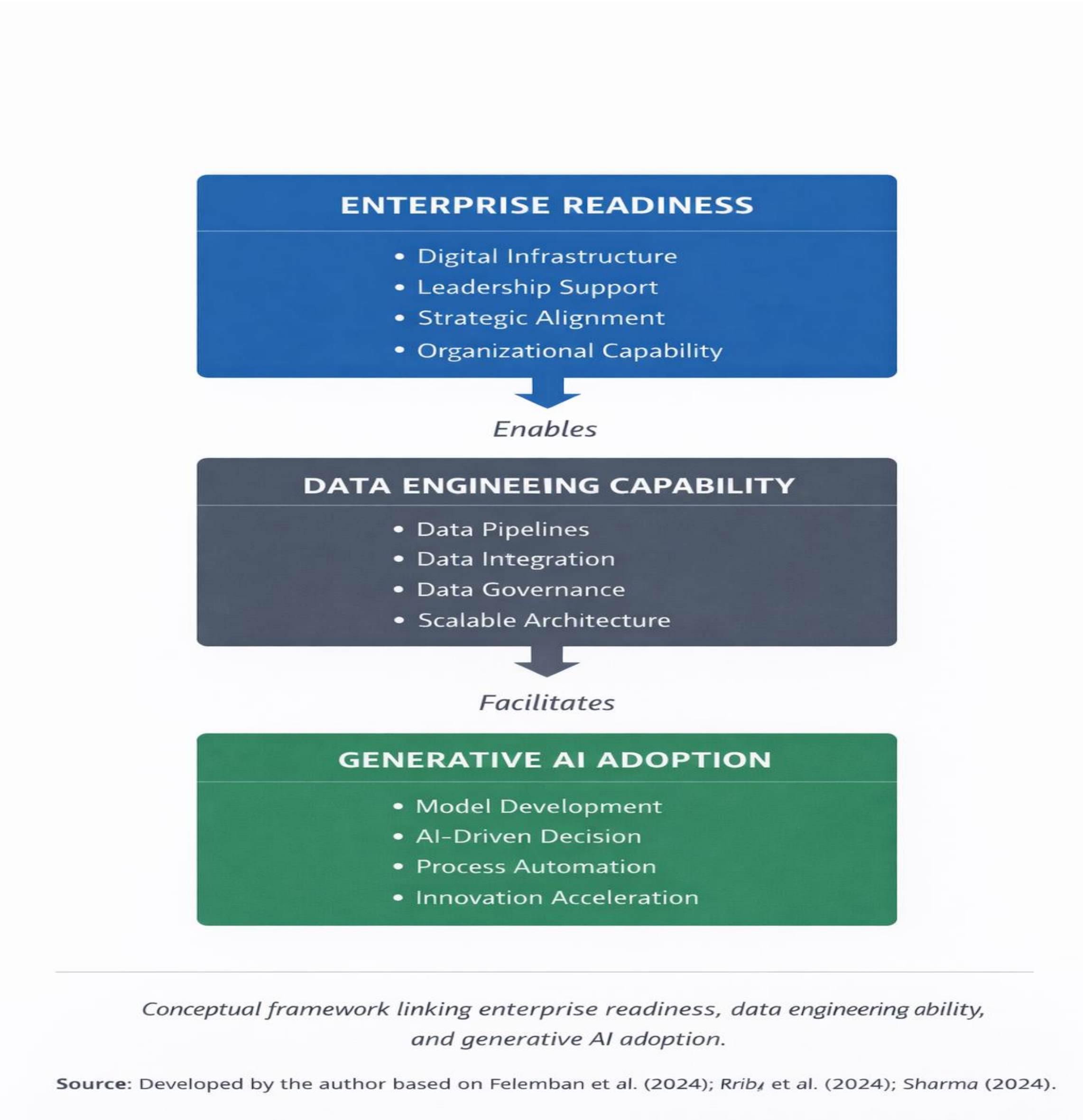
The conceptual relationship can be summarized as:

Enterprise Readiness → Data Engineering Capability → Generative AI Adoption

This framing reinforces data engineering as a strategic prerequisite for sustainable GenAI transformation in enterprise contexts and provides a foundation for readiness assessment and capability investment.

In summary, the literature and analysis in this section position data engineering as the operational "translation layer" that converts enterprise intent to adopt GenAI into scalable and trustworthy implementation capability. Robust data infrastructure and pipeline architectures determine whether enterprises can reliably ingest, curate, and serve the high-volume, high-variety datasets required by GenAI systems, while data quality and governance mechanisms provide the controls necessary for secure access, privacy protection, auditability, and regulatory alignment. Equally, effective integration with enterprise systems ensures GenAI solutions are not isolated experiments but embedded capabilities that can be operationalized within existing workflows and decision processes. Framing these elements as a mediating capability clarifies why model selection and use-case identification alone do not guarantee adoption outcomes: enterprise GenAI success depends on whether data engineering maturity can sustain reliability, scalability, and accountability under real-world constraints. Building on this foundation, the next section examines the strategic implications for enterprises—specifically, how organizations should prioritize investments, develop capabilities, and manage risks to move from pilot initiatives to sustainable GenAI deployment.

Figure 3. Conceptual framework linking enterprise readiness, data engineering capability, and generative AI adoption.



Source: Developed by the author based on Felemban et al. (2024); Sridhar et al. (2024); Sharma (2024).

4. Discussion: Strategic Implications, Capability Development, and Risk Considerations

This section discusses the strategic implications of the paper's central argument: while enterprises are increasingly motivated to adopt generative artificial intelligence (GenAI), sustainable success depends on whether the underlying data engineering foundations are sufficiently mature to support reliability, governance, and scale. In enterprise environments, GenAI adoption is constrained not only by model selection or proof-of-concept feasibility, but by operational realities—fragmented data landscapes, pipeline instability, and governance requirements—that determine whether GenAI solutions can be deployed responsibly and maintained over time. The discussion therefore focuses on (i) strategic implications for enterprise investment and platform priorities, (ii) organizational capability development required to operationalize GenAI, and (iii) risks associated with inadequate data infrastructure and weak governance controls.

4.1 Strategic Implications for Enterprises

A key implication of this study is that enterprises should treat data engineering maturity as a first-class strategic dependency for GenAI initiatives, rather than as a downstream implementation concern. GenAI systems require access to large-scale, high-quality datasets and dependable data flows that can support both experimentation and production workloads. As a result, investment priorities should emphasize scalable data infrastructure, robust pipeline architectures, and standardized integration patterns that reduce friction in sourcing and preparing data for GenAI use cases.

From a strategic perspective, the "pilot-to-production" transition is where many GenAI initiatives encounter structural limitations. Proof-of-concept systems can often operate with limited data scope and informal controls; however, enterprise deployment demands repeatability, throughput, and operational guarantees—such as timely refresh cycles, reliability under load, and consistent downstream serving. Modern pipeline architectures become critical in this transition because they enable continuous ingestion, transformation, and validation of data, ensuring that GenAI systems receive consistent and well-structured inputs across changing sources and evolving requirements.

Equally, enterprises must embed governance considerations into strategy rather than treating governance as a compliance afterthought. Strong governance frameworks are necessary to maintain data integrity, enforce appropriate access to sensitive data, support regulatory compliance, and ensure accountability for the data used by GenAI systems. These governance requirements are inseparable from data engineering execution because controls must be implemented operationally—through enforceable policies, auditable data movement, and consistent quality checks—across the data lifecycle

4.2 Organizational Capability Development

Beyond infrastructure investments, enterprise GenAI adoption requires purposeful capability development across people, process, and operating models. Organizations should establish or strengthen specialized data engineering teams responsible for designing scalable pipelines, managing enterprise data architectures, and enabling interoperability with machine learning and GenAI workflows. These teams play a foundational role in ensuring that data environments remain reliable, extensible, and aligned with production requirements, including performance, maintainability, and continuous delivery of pipeline changes.

In addition, enterprises must institute governance capabilities that support responsible GenAI operation. Effective governance includes policies for data access, privacy protection, lineage, and accountability, as well as operational processes for enforcing those policies consistently across platforms and teams. In practice, this implies establishing cross-functional mechanisms that connect engineering execution to oversight—so that GenAI systems can be monitored, audited, and managed as enterprise services rather than experimental tools. Such capability development also requires leadership commitment and strategic alignment so that data engineering priorities are funded, measured, and sustained as part of long-term digital transformation roadmaps.

More broadly, GenAI adoption should be integrated into enterprise digital maturity strategies. This includes strengthening foundational infrastructure, improving data management discipline, and ensuring the workforce has the skills required to operationalize and sustain AI-enabled systems. Enterprises that build these capabilities are more likely to achieve repeatable GenAI delivery, scale solutions across domains, and maintain system integrity as adoption expands.

4.3 Risks of Poor Data Infrastructure

Although GenAI promises significant organizational value, enterprises with insufficient data infrastructure and weak governance controls face substantial risks. One critical risk is degraded model output quality—such as bias, inaccuracy, or inconsistency—when training or inference depends on incomplete, stale, or poorly governed datasets. GenAI systems are sensitive to the quality and representativeness of the data they consume; therefore, weaknesses in data curation, integration, or validation can directly undermine trust in AI-generated outputs and limit business adoption.

A second risk involves operational reliability. When data pipelines are unstable or poorly integrated across enterprise systems, GenAI workflows can become brittle—leading to interruptions in AI services, inconsistent outputs, or failures during peak demand. Such instability is particularly problematic in enterprise settings where GenAI outputs may influence decision processes or customer-facing operations. In these cases, unreliable data flow can translate into unreliable AI behavior, amplifying operational risk.

Finally, weak governance introduces security, privacy, and compliance exposure. Where sensitive enterprise or customer information is involved, insufficient oversight of data usage and movement can increase the likelihood of data leakage, unauthorized access, or regulatory violations. Yigit et al. (2024) emphasize that responsible deployment of GenAI requires careful attention to governance and security considerations to protect critical systems and maintain trustworthiness.

Accordingly, enterprises should view robust data engineering and governance not only as enablers of GenAI value, but also as primary mechanisms for risk reduction. The risks commonly associated with inadequate data infrastructure—including data quality issues, integration failures, security and privacy risks, scalability limitations, and operational instability—underscore why

GenAI readiness must be evaluated through the maturity of data pipelines, platform integration, and governance controls (Table 2).

Risk Category	Description	Potential Impact
Data Quality Issues	Inconsistent or incomplete datasets used for AI training	Biased or inaccurate AI outputs
Data Integration Failures	Poor integration between enterprise systems and AI platforms	Disrupted AI workflows
Security and Privacy Risks	Weak governance over sensitive enterprise data	Data breaches and regulatory violations
Scalability Limitations	Inadequate infrastructure to handle large datasets	Reduced system performance
Operational Instability	Unreliable data pipelines and system failures	Interrupted AI services

Table 2. Key Risks of Inadequate Data Infrastructure in Generative AI Deployment

Source: Adapted from Yigit et al. (2024).

Collectively, the discussion emphasizes that enterprise GenAI adoption is best understood as a socio-technical transformation grounded in operational data realities. The strategic implications and capability requirements outlined above reinforce the paper's broader argument that data engineering maturity—spanning scalable infrastructure, dependable pipelines, integrated enterprise architectures, and embedded governance controls—determines whether organizations can move from isolated pilots to sustainable, enterprise-grade GenAI deployment. At the same time, the risk analysis illustrates that inadequate data foundations and weak governance do not merely slow adoption; they can materially undermine trust, reliability, and compliance, thereby limiting organizational willingness to institutionalize GenAI at scale. These observations motivate the concluding section, which synthesizes the paper's key findings and reiterates the central role of data engineering as a prerequisite for responsible and scalable GenAI readiness, while also outlining directions for future research and practical readiness assessment.

5. Conclusion

GenAI has created significant opportunities for enterprises to automate complex workflows, improve productivity, and augment decision-making. However, realizing these benefits in production environments requires more than selecting powerful models or identifying compelling use cases. This paper demonstrates that GenAI success is tightly coupled to enterprise data engineering readiness—the ability to operationalize consistent, governed, and scalable data flows that supply models and downstream applications with reliable information. Organizations that continue to operate with fragmented data landscapes, brittle pipelines, and weak governance frameworks face elevated risks of inaccurate outputs, poor reliability, compliance exposure, and an inability to scale beyond experimentation.

The analysis underscores data engineering as an enterprise-level capability that must be treated as a strategic investment, not a back-office technical function. Scalable ingestion and transformation pipelines, modern storage and compute architectures, and integrated platforms provide the performance and elasticity required for GenAI workloads, while data quality management and governance controls ensure trust, accountability, and regulatory alignment. The proposed conceptual framework reinforces that enterprise readiness is achieved when organizational enablers—digital infrastructure, leadership commitment, skilled talent, and governance—translate into measurable data engineering capability that can support continuous delivery and operational stability of GenAI systems.

As GenAI technologies evolve, enterprises should prioritize strengthening data foundations and governance as the fastest path to sustainable value creation. Future work can extend this study by validating the framework empirically across industries and by defining maturity metrics that quantify readiness in terms of data reliability, lineage coverage, security posture, and lifecycle operationalization. Ultimately, organizations that build robust data engineering ecosystems will be best positioned to deploy GenAI responsibly, cost-effectively, and at scale—turning experimentation into durable competitive advantage.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1]. Strzelecki, A., Pańkowska, M., & Rizun, M. (2024, December). Generative AI Adoption and Use by Micro-enterprises: Validation of the Measurement Instrument. In *International Conference on Marketing and Technologies* (pp. 77-91). Singapore: Springer Nature Singapore. doi.org/10.1007/978-981-96-3077-6_5

[2]. Etemad, H. Transformative potentials of generative artificial intelligence: Should international entrepreneurial enterprises adopt GEN.AI?. *J Int Entrep* **22**, 141–163 (2024). doi.org/10.1007/s10843-024-00363-8

[3]. Carayannis, E. G., Dumitrescu, R., Falkowski, T., & Zota, N. R. (2024). Empowering SMEs "harnessing the potential of gen AI for resilience and competitiveness". *IEEE Transactions on Engineering Management*, *71*, 14754-14774. doi: 10.1109/TEM.2024.3456820.

[4]. Shore, A., Tiwari, M., Tandon, P., & Foropon, C. (2024). Building entrepreneurial resilience during crisis using generative AI: An empirical study on SMEs. *Technovation*, *135*, 103063. doi.org/10.1016/j.technovation.2024.103063

[5]. Sridhar, R., Kanani, I. J., & Dhenia, R. N. K. (2024). The Role of Data Engineering in Enabling Machine Learning and Artificial Intelligence. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *5*(4), 214-226. doi.org/10.63282/3050-9262.IJAIDSML-V5I4P120

[6]. Gudavalli, Sunil, Role Of Data Engineering In Digital Transformation Initiative (October 20, 2024). Available at SSRN: https://ssrn.com/abstract=5068353

[7]. Mbata, A., Sripada, Y., & Zhong, M. (2024). A survey of pipeline tools for data engineering. *arXiv preprint arXiv:2406.08335*. doi.org/10.48550/arXiv.2406.08335

[8]. Yan, T., Li, S., Kraner, B., Zhang, L., & Tessone, C. J. (2024). Analyzing reward dynamics and decentralization in ethereum 2.0: An advanced data engineering workflow and comprehensive datasets for proof-of-stake incentives. *arXiv preprint arXiv:2402.11170* doi.org/10.48550/arXiv.2402.11170

[9]. Lai, I. C., & Su, H. N. (2024). Initiating technology convergence through knowledge integration: The critical role of dynamic capabilities. *IEEE Transactions on Engineering Management*, *71*, 11533-11550. doi: 10.1109/TEM.2024.3425656

[10]. Kochkina, N., Andriushchenko, I., & Gatto, G. (2024, October). Strategic AI adoption: economic impact, case studies from handy. AI, and industry readiness. In *2024 IEEE International Conference on Artificial Intelligence & Green Energy (ICAIGE)* (pp. 1-6). IEEE. doi: 10.1109/ICAIGE62696.2024.10776631.

[11]. Kmiecik, M., Maghroor, H. R., & Maj, M. (2024). GENERATIVE AI ROLE IN BUILDING 3PL COMPANIES'RESILIENCE. *Scientific Papers of Silesian University of Technology. Organization & Management/Zeszyty Naukowe Politechniki Slaskiej. Seria Organizacji i Zarzadzanie*, (212). doi: 10.29119/1641-3466.2024.212.14

[12]. Sharma, R. (2024). Architecting AI Solutions: A Blueprint for Generative AI. In *AI and the Boardroom: Insights into Governance, Strategy, and the Responsible Adoption of AI* (pp. 259-273). Berkeley, CA: Apress. doi.org/10.1007/979-8-8688-0796-1_21

[13]. Felemban, H., Sohail, M., & Ruikar, K. (2024). Exploring the readiness of organisations to adopt artificial intelligence. *Buildings*, *14*(8), 2460. doi.org/10.3390/buildings14082460

[14]. Wu, J., Wang, H., Ni, C., Zhang, C., & Lu, W. (2024). Data pipeline training: Integrating AutoML to optimize the data flow of machine learning models. *arXiv preprint arXiv:2402.12916*. doi.org/10.48550/arXiv.2402.12916

[15]. Yigit, Y., Ferrag, M. A., Sarker, I. H., Maglaras, L. A., Chrysoulas, C., Moradpoor, N., & Janicke, H. (2024). Critical infrastructure protection: Generative AI, challenges, and opportunities. *arXiv preprint arXiv:2405.04874*. doi.org/10.48550/arXiv.2405.04874