RESEARCH ARTICLE

Design and Implementation of an AI-Powered Chatbot for Procurement Management Systems Using Laravel, GPT, and Pinecone - A Case Study

Muhammad Zubair[1]

[1] University of Hertfordshire, United Kingdom

Received	Revised	Accepted	Published
15 February 2025	22 March 2025	03 April 2025	15 September 2025

Abstract

Background and Purpose: AI-powered chatbots can improve conversational access to procurement information, but direct integration of large language models with operational databases creates risks including SQL injection, unauthorized access, data leakage, and regulatory non-compliance. This study investigates a real-world procurement management chatbot and proposes an architecture that improves operational efficiency while strengthening data security, privacy, and ethical compliance.

Methods: A Design Science Research approach was used to develop and evaluate an AI-powered procurement chatbot implemented with Laravel, MySQL, OpenAI GPT models, and Pinecone. Initial direct SQL query generation was replaced by a Retrieval-Augmented Generation architecture in which the vector database stores trace identifiers and limited metadata, while controlled backend queries retrieve sensitive procurement information. Evaluation combined quantitative performance indicators with qualitative user feedback.

Findings: The implemented system achieved an average response time of 1.2 seconds, 93% trace ID lookup accuracy, 99.7% uptime over the observation period, and user satisfaction of 4.6 out of 5. The redesigned retrieval architecture reduced backend query load by 62% and limited direct exposure of procurement databases to the language model, demonstrating simultaneous gains in efficiency, scalability, and security.

Theoretical Contributions: The study extends enterprise AI and information systems research by presenting a security-aware RAG pattern for structured procurement environments. It demonstrates how Zero Trust principles, data minimization, traceability, and responsible AI requirements can be operationalized through architecture and derives transferable design knowledge for governance-aware LLM deployment in sensitive operational domains.

Conclusions and Policy Implications: Separating language models from direct database access through controlled retrieval can support secure, scalable, and responsible conversational systems without sacrificing usability or performance. Organizations deploying LLMs in procurement should embed restricted access, data minimization, auditability, transparency, and continuous validation into system architecture. Future work should evaluate larger deployments, richer traceability features, self-hosted language models, and reduced dependence on external AI services.

Keywords: Artificial Intelligence; Natural Language Processing; Information Systems; Data Security; Supply Chain Management

How to cite: Zubair M. Design and Implementation of an AI-Powered Chatbot for Procurement Management Systems Using Laravel, GPT, and Pinecone - A Case Study. International Journal of Business & Computational Science. 2025;5(1):22-29.

1. Introduction

1.1 Background

The rapid advancement of Artificial Intelligence (AI), particularly Large Language Models (LLMs), has significantly transformed how organizations interact with data, systems, and users. In organizational procurement environments, AI-powered chatbots are increasingly adopted to automate customer support, streamline internal workflows, and enable real-time access to organizational knowledge. Among organizational domains, procurement management systems represent a particularly promising yet challenging context for AI integration, given their reliance on accurate traceability, regulatory compliance, and secure handling of commercially sensitive information. Traditional procurement platforms rely primarily on static dashboards, structured queries, and rule-based automation. While effective for routine tasks, such systems exhibit limited flexibility when faced with ambiguous, natural language queries or dynamic information needs [16].

LLM-based chatbots offer the potential to overcome these limitations by enabling conversational access to complex data and processes, supporting contextual understanding and adaptive responses. However, integrating LLMs into organizational procurement systems introduce a fundamentally different risk landscape. Direct coupling of generative models with operational databases may expose organizations to critical vulnerabilities, including prompt injection, data leakage, unauthorized access, and regulatory non-compliance. Recent studies emphasize that LLM-integrated applications significantly expand the attack surface of organizational procurement systems, requiring architectural rethinking rather than incremental security patching [10,11,17]. From a business and strategic perspective, procurement systems operate within a highly sensitive domain involving supplier relationships, contractual obligations, financial transactions, and compliance requirements.

Subramaniam and Shaw in 2004 demonstrated that the value derived from e-procurement technologies depends heavily on process complexity, transaction volumes, and system integration fidelity [14,15]. Consequently, embedding AI into procurement processes is not merely a technical decision but one with significant operational, legal, and reputational implications. Simultaneously, ethical and regulatory considerations have become central to the deployment of AI in procurement systems. Regulatory frameworks such as the General Data Protection Regulation (GDPR) impose strict requirements on data minimization, lawful processing, and user rights. In parallel, responsible AI frameworks emphasize transparency, accountability, and fairness as foundational design principles [3,18]. In systems that directly mediate user access to sensitive organizational data, these considerations must be operationalized through technical design rather than treated solely as governance overlays.

Despite the growing adoption of AI chatbots in commercial systems, a critical gap remains between experimental LLM implementations and secure, governance-compliant deployment in real-world procurement systems, where direct database interaction introduces significant security and compliance risks [1]. Many

existing implementations prioritize conversational capability over architectural robustness, often embedding LLMs directly into business logic or databases without adequate isolation, access control, or auditability. This gap highlights the need for architectures that simultaneously address performance, security, and governance requirements while maintaining usability in organizational contexts. This study addresses this gap by presenting a real-world case study of a secure, scalable, and ethically aligned AI chatbot integrated into a procurement management system.

Specifically, it proposes and validates a Retrieval-Augmented Generation (RAG)-based architecture that decouples LLMs from operational databases through an intermediate vector-based retrieval layer, thereby reducing cybersecurity and privacy risks while preserving system performance and usability. This study builds on a real-world implemented procurement chatbot system and extends it by abstracting architectural decisions into generalizable insights for secure and responsible AI integration.

1.2 Research Objectives

This study aims to:

1. Design a secure architectural pattern for integrating LLM-based chatbots into organizational procurement systems.

2. Evaluate the effectiveness of a RAG-based approach in mitigating cybersecurity and data governance risks.

3. Assess the operational and usability impacts of conversational access to procurement traceability data.

4. Demonstrate how ethical and regulatory principles can be operationalized through system architecture. To further operationalize these objectives, the study is guided by the following research questions: The research seeks to address the following questions: •

RQ1: How can Large Language Model (LLM)-based chatbots enhance procurement workflows beyond traditional static and rule-based systems? •

RQ2: What cybersecurity risks emerge from integrating LLMs into procurement management systems, and how can these risks be effectively mitigated? •

RQ3: How can ethical principles, including transparency, fairness, and data privacy, be practically embedded into AI-powered procurement chatbot systems? •

RQ4: How do commercial, operational, and regulatory considerations influence the feasibility and adoption of LLM-based chatbot technologies in procurement environments?

1.3 Research Contributions

Building on the implemented system and its empirical evaluation, this study abstracts key architectural decisions into generalized contributions for enterprise AI and information systems research. This study makes four primary academic contributions to AI in procurement systems and information systems research. First, the study proposes a RAG-based architectural pattern designed to enforce security, ethical, and regulatory requirements through system design. Unlike conventional RAG implementations that primarily support document retrieval and knowledge augmentation, the proposed architecture applies vector-based retrieval to traceability identifiers while maintaining strict separation between AI-accessible components and procurement databases. This extends the applicability of RAG from knowledge management scenarios to operational procurement environments.

Second, the study contributes design knowledge for secure large language model (LLM) integration by demonstrating how Zero Trust security principles, data minimization strategies, and responsible AI requirements can be operationalized directly through system architecture. This shifts organization AI governance from policy-level guidance toward enforceable technical implementation. Third, the study provides empirical evidence demonstrating that separating the LLM from direct access to procurement systems can simultaneously enhance cybersecurity posture and improve system performance. The findings suggest that security-oriented architectural redesign can produce measurable operational efficiency gains in procurement conversational systems. Fourth, from a Design Science Research perspective, the study derives transferable design principles for responsible organizational LLM deployment.

These principles provide reusable knowledge for organizations seeking to implement AI systems within regulated or high-sensitivity operational environments. These contributions collectively support a shift from capability-driven AI deployment toward architecture-driven governance in enterprise systems. From a design science perspective, these contributions provide a foundation for developing governance-aware enterprise AI architectures in sensitive operational domains.

1.4 Key Innovations

The key innovations of this study are as follows.

1.4.1 Hybrid RAG + Database Architecture

While Retrieval-Augmented Generation (RAG) is typically used in document-based systems, this study applies it within a procurement context to separate public traceability access from sensitive backend data. The architecture combines vector retrieval with controlled SQL queries, ensuring secure and efficient data access. This demonstrates a practical adaptation of RAG for structured procurement systems.

1.4.2 Public Traceability Access Without Authentication

The system enables traceability access through QR codes or trace identifiers without requiring user authentication. This improves transparency and user convenience while maintaining strict backend security controls. It represents a balanced approach between accessibility and data protection in procurement platforms.

1.4.3 Secure Minimal-Exposure Vector Search Integration

The system implements a vector search mechanism where only trace identifiers and limited metadata are stored in Pinecone. Sensitive procurement data remains exclusively within the relational database, preventing exposure to the AI layer. This approach enables semantic search while preserving data privacy and security.

1.4.4 Selective Cron-Based Data Synchronization

A cron-based synchronization mechanism updates only new or modified trace identifiers between the database and the vector store. This avoids full dataset replication, reducing system overhead and minimizing data exposure. It ensures that the retrieval layer remains efficient, accurate, and privacy-aware.

1.4.5 Integrated Security-Aware Chatbot Architecture

These components collectively form a security-aware chatbot architecture that mitigates risks such as data exposure and unauthorized access. By combining controlled retrieval, backend mediation, and restricted data access, the system ensures safe deployment of AI in organizational procurement environments. It highlights how conversational AI can be integrated without compromising security or performance.

2. Methodology

This study adopts a Design Science Research (DSR) approach to develop and evaluate an AI-powered chatbot integrated within a procurement management system. DSR is particularly appropriate for this study as it focuses on the creation and evaluation of innovative artifacts designed to address real-world problems while generating generalizable knowledge [5]. The research aims to design a secure and efficient conversational interface for procurement systems while addressing critical challenges related to cybersecurity, data governance, and ethical AI deployment.

2.1 Research Design

The study follows the standard DSR process consisting of five stages: problem identification, artifact design, demonstration,

evaluation, and communication. 1. Problem Identification: Existing procurement systems rely on static dashboards and structured queries, limiting accessibility and flexibility. The study identifies the need for a conversational interface that enables intuitive access to procurement traceability data while maintaining security and compliance. 2. Artifact Design: An AI-powered chatbot ("Discover") was designed to enable natural language interaction with procurement data. The system supports query input through trace identifiers and QR codes, allowing users to retrieve relevant procurement information efficiently. 3. Demonstration: The chatbot was implemented within an operational procurement management system, demonstrating real-time data retrieval and interaction capabilities. 4.

Evaluation: The artifact was evaluated using a combination of quantitative performance metrics and qualitative user assessments to determine its effectiveness in improving efficiency, security, and usability. 5. Communication: The findings are documented to contribute both practical insights and transferable design knowledge for enterprise AI systems.

2.2 System Development and Architecture

2.2.1 Development Approach

An agile development methodology was employed to support iterative design, continuous testing, and stakeholder feedback. This approach enabled incremental refinement of system features and ensured alignment with user requirements throughout the development lifecycle.

2.2.2 Security-Oriented Architectural Design

Initial system designs allowed the chatbot to generate SQL queries directly from user input. However, this approach introduced significant security risks, particularly SQL injection and unauthorized data access. To address these risks, a revised architecture was developed based on separating the LLM from direct database access, replacing direct query generation with a controlled retrieval mechanism. This architecture incorporates several key design components to ensure security, efficiency, and compliance. First, Retrieval-Augmented Generation (RAG) is employed, where a vector database (Pinecone) stores trace identifiers and minimal metadata. The chatbot retrieves relevant identifiers and forwards them to a backend layer, which executes controlled database queries. This approach ensures that sensitive procurement data is never directly exposed to the language model.

Second, the system adopts Zero Trust security principles, enforcing strict access controls, continuous validation, and least-privilege access to ensure that all interactions are verified and restricted according to predefined rules [13]. Third, data minimization is implemented by storing only non-sensitive identifiers in the vector database, while all sensitive procurement data remains securely within the relational database.

2.2.3 Deployment and Data Synchronization

The system was deployed using cloud-based infrastructure to ensure scalability, reliability, and ease of management. Version control and collaborative development were maintained through Git-based workflows, enabling structured and iterative system development. To ensure consistency between the relational database and the vector database, a cron-based synchronization mechanism was implemented. This process updates only newly created or modified trace identifiers, thereby minimizing data redundancy, reducing system overhead, and ensuring that the retrieval layer remains accurate and up to date.

2.2.4 Ethical and Regulatory Design Considerations

Ethical and regulatory requirements were incorporated directly into the system design to ensure responsible AI deployment. The development process was guided by principles from established responsible AI frameworks [3], with a focus on transparency, privacy, and accountability. Transparency was ensured by clearly informing users about system capabilities and data usage practices. Privacy was maintained by restricting access to sensitive procurement data and limiting exposure through controlled system interactions. Accountability was supported through mechanisms that ensure traceability of system interactions and decision flows [3,2]. These principles were operationalized through architectural decisions such as restricted data access, controlled query execution, and minimal data exposure.

2.3 Evaluation Strategy

The system was evaluated using a mixed-methods approach, combining quantitative performance metrics with qualitative user feedback to provide a comprehensive assessment of its effectiveness. This approach enables both objective measurement of system performance and a deeper understanding of user experience.

2.3.1 Quantitative Evaluation

System performance was assessed using a set of key performance indicators (KPIs) designed to measure efficiency, effectiveness, and user engagement. Response time was evaluated to determine system latency in handling user queries, ensuring timely interactions. Task completion rate was used to assess the proportion of user requests successfully fulfilled by the chatbot, reflecting its operational effectiveness. Interaction volume was analyzed to understand system adoption and usage patterns over time, while bounce rate was measured to identify early user disengagement and potential usability issues. Together, these metrics provide a comprehensive evaluation of system performance.

2.3.2 Qualitative Evaluation

To complement quantitative findings, qualitative assessments were conducted to capture user perceptions and experiences. User feedback surveys were used to measure satisfaction levels, perceived usefulness, and overall usability of the system. Usability testing sessions were conducted to observe user interactions, identify interface challenges, and uncover areas for improvement. Additionally, feedback analysis was performed to examine recurring themes in user comments, enabling the identification of common issues and opportunities for system refinement. This qualitative approach provides deeper insights into user experience and supports iterative system enhancement.

2.4 Ethical Considerations

Ethical integrity was maintained throughout the study, particularly in relation to handling procurement data. Data privacy was ensured through strict adherence to data minimization principles, where only essential information required for task execution was processed. Secure data handling practices were implemented, including encryption protocols and access control mechanisms to protect data confidentiality. Furthermore, transparency and user consent were emphasized by informing users about data usage policies and system capabilities, while also providing options for feedback and opt-out where applicable. These measures ensure alignment with ethical standards for AI deployment.

2.5 Limitations

Despite its contributions, the study has several limitations: • The chatbot is limited to predefined procurement use cases • Complex queries may require human intervention • Performance depends on the quality of underlying data • User adoption may vary based on familiarity with conversational systems These limitations highlight areas for future research and system enhancement.

3. Results

3.1 Performance Metrics and System Outcomes

The performance of the developed chatbot system was evaluated using a set of predefined quantitative metrics aligned with the study objectives of efficiency, security, and usability. The system achieved an average response time of 1.2 seconds, outperforming the target threshold of 2 seconds and demonstrating its capability to support real-time interaction in procurement environments. Trace ID lookup accuracy reached 93%, exceeding commonly reported chatbot benchmarks of approximately 85%, indicating the effectiveness of the retrieval and filtering mechanisms. System

reliability was also demonstrated through an uptime of 99.7% over a two-week observation period, surpassing the target of 99%. User satisfaction, measured through pilot testing, achieved a score of 4.6 out of 5, reflecting positive user perception in terms of usability and usefulness.

In addition, backend query load was reduced by 62% compared to baseline systems without a retrieval layer, highlighting the efficiency gains achieved through the architectural redesign. These results collectively indicate that the system met or exceeded all predefined performance targets, confirming its operational feasibility in real-world procurement contexts. Figure 1 illustrates the system's performance relative to target benchmarks, particularly in response time and uptime. Furthermore, Figure 2 presents the development timeline, outlining key stages including backend integration, API configuration, vector database setup, and production deployment.

3.2 Architectural Performance and Critical Analysis

The integration of Retrieval-Augmented Generation (RAG) and vector-based retrieval mechanisms played a central role in enhancing both system efficiency and security. In earlier system iterations, direct SQL query generation from user inputs posed significant risks, including SQL injection vulnerabilities, increased backend load, and potential exposure of sensitive procurement data. By adopting a revised architecture that separates the language model from direct database access, the system now performs semantic retrieval through vector search, followed by controlled backend queries. This architectural shift resulted in a 62% reduction in backend query load, demonstrating improved efficiency and scalability. These findings are consistent with prior studies highlighting the performance benefits of retrieval-based architectures.

The achieved trace ID lookup accuracy of 93% further suggests that constraining queries through vector-based retrieval reduces irrelevant results and improves precision. Figure 3 illustrates the RAG-based system architecture, highlighting the interaction between the vector database, language model, and backend query layer. Qualitative feedback from users indicated high satisfaction levels; however, some participants expressed a preference for more detailed traceability information, such as supplier certifications or geographic data. This suggests an opportunity for future system enhancements through customizable information layers.

3.3 Evidence of System Implementation

The results are supported by extensive practical implementation across multiple system components, demonstrating a full-stack development approach. The backend system, developed using Laravel, manages chatbot requests and facilitates interaction between the vector database and relational database. The vector database implementation enables secure storage and retrieval of trace identifier, which presents the schema configuration and API setup. Data synchronization between systems is maintained through a scheduled cron process, ensuring consistency and minimizing redundancy; this mechanism. The system was deployed using a cloud-based infrastructure with scalability and monitoring capabilities. The user interface was designed using a prototyping tool to ensure intuitive interaction, allowing users to query the system using QR codes or trace identifiers, as demonstrated in Figure 4.

The overall system architecture, including interactions between the chatbot, backend services, vector database, and relational database. Version control and collaborative development were managed through a Git-based repository, providing traceability of development progress and system evolution. Project management processes, including sprint planning and task tracking, are illustrated in. Real-time client engagement and system demonstrations further validated system functionality and usability.

3.4 Technical Challenges and Solutions

Several technical challenges were encountered during system development, requiring iterative problem-solving and architectural refinement. Initial system designs exposed vulnerabilities related to SQL injection and data exposure due to direct query generation. This challenge was addressed by redesigning the architecture to incorporate a retrieval-based approach, ensuring that only non-sensitive identifiers are accessible through the chatbot interface. Maintaining synchronization between the relational and vector databases also presented challenges. This was resolved through the implementation of a scheduled synchronization mechanism that updates only relevant data, ensuring efficiency and consistency. Scalability concerns were addressed through cloud-based deployment with autoscaling capabilities, enabling the system to handle varying user loads.

Additionally, balancing public accessibility with data security required careful architectural design, achieved through a layered system structure that restricts sensitive data access while allowing controlled public interaction. User experience challenges were addressed through iterative interface design and usability testing.

3.5 Innovation and System Contributions

The developed system introduces several novel contributions within the context of procurement management systems. The use of a hybrid retrieval and database architecture enables secure separation between publicly accessible traceability data and sensitive backend information. The implementation of public traceability access without authentication enhances transparency while maintaining security through controlled backend processes. The system also demonstrates an innovative use of vector search by limiting stored data to trace identifiers, thereby preserving privacy while enabling efficient semantic retrieval. The implementation of a selective synchronization mechanism further enhances system efficiency by reducing unnecessary data transfer. These elements collectively represent a novel application of retrieval-based architectures within structured procurement systems.

3.6 Interpretation of Results

The findings demonstrate that the system successfully addresses the core research objectives. Efficiency improvements are evident through reduced response times and lower backend query load. Enhanced data security is achieved through architectural separation and controlled data access mechanisms. Ethical considerations are addressed by limiting exposure to non-sensitive data and ensuring transparency in system interactions. The use of scalable infrastructure and efficient synchronization mechanisms supports system sustainability and future extensibility. The results align with existing literature on chatbot performance and retrieval-based systems, while also extending prior work by demonstrating the feasibility of applying such architectures in procurement environments. The successful deployment and evaluation of the system provide practical evidence of its viability and highlight its potential for broader application in similar domains.

3.7 Feasibility and Practical Deployment

The feasibility of the system is demonstrated through its successful implementation, deployment, and evaluation within a real-world context. The system achieved all predefined performance targets, including response time, uptime, and user satisfaction, while maintaining efficient resource utilization. Adaptations made during development, such as limiting vector database storage to trace identifiers and optimizing synchronization processes, reflect a realistic and iterative development approach. Overall, the system demonstrates a balance between technical innovation and practical feasibility, providing a scalable and secure solution for integrating AI chatbots into procurement systems.

4. Discussion

4.1 Interpretation of Key Findings

The findings of this study demonstrate that the integration of an AI-powered chatbot within a procurement management system can significantly enhance operational efficiency, strengthen data security, and support ethical and scalable system design. The observed reduction in response time and backend query load indicates that conversational interfaces, when supported by

appropriate architectural design, can improve system performance without compromising reliability. In particular, the achieved response time of 1.2 seconds and the 62% reduction in backend query load suggest that the use of a retrieval-based intermediary layer effectively optimizes data access processes. From a performance perspective, the high trace ID lookup accuracy (93%) highlights the effectiveness of combining vector-based retrieval with controlled database querying. This supports the argument that narrowing the scope of data retrieval through structured identifiers improves precision and reduces irrelevant outputs.

These findings align with prior research on chatbot performance and retrieval-based systems, while also extending existing knowledge by demonstrating their applicability in procurement environments.

4.2 Implications for System Architecture and Security

One of the most significant contributions of this study lies in demonstrating how architectural design can be used as a primary mechanism for enhancing cybersecurity in AI-integrated systems. Traditional approaches that allow direct interaction between language models and databases introduce substantial risks, including SQL injection vulnerabilities and unintended data exposure [14]. The findings show that separating the language model from direct database access and introducing a controlled retrieval layer substantially reduces these risks. The adoption of a retrieval-based architecture not only improves security but also contributes to system efficiency by reducing unnecessary database queries. This dual benefit suggests that security-oriented architectural decisions can simultaneously enhance performance, challenging the conventional trade-off between security and efficiency.

Furthermore, the implementation of data minimization and restricted access mechanisms demonstrates how privacy and compliance requirements can be embedded directly into system design rather than enforced as external controls.

4.3 Ethical and Governance Implications

The study also provides important insights into the practical implementation of ethical principles in AI systems. By limiting chatbot access to non-sensitive traceability data and ensuring transparency in system interactions, the design supports key ethical principles such as privacy, accountability, and user autonomy [8,9]. These findings demonstrate that ethical considerations can be operationalized through technical design choices, rather than remaining abstract guidelines. The results further suggest that embedding ethical safeguards at the architectural level enhances user trust and system acceptability. High user satisfaction scores indicate that users perceive the system as both useful and reliable, which is essential for adoption in sensitive domains such as procurement. This reinforces the importance of integrating ethical and governance considerations into the early stages of system design.

4.4 Feasibility and Practical Implementation

The successful deployment and evaluation of the chatbot system demonstrate the feasibility of integrating AI-driven conversational interfaces into procurement environments. The system achieved all predefined performance targets while maintaining stability and scalability, indicating that such solutions can be implemented within real-world constraints [12]. The use of established technologies and cloud-based infrastructure further supports the practicality of the approach. The iterative development process, including architectural refinements and system optimization, highlights the importance of adaptability in AI system design. Adjustments such as limiting vector database storage to trace identifiers and implementing selective data synchronization were critical in balancing performance, security, and cost considerations. These findings emphasize that successful AI integration requires continuous refinement and alignment with both technical and operational requirements.

4.5 Comparison with Existing Literature

The findings of this study are consistent with existing literature on chatbot performance, retrieval-based systems, and responsible AI design. Previous research has highlighted the importance of balancing efficiency and security in AI applications, as well as the role of vector search in improving retrieval performance [4,7]. This study confirms these findings while extending them by demonstrating the application of retrieval-based architectures in structured procurement systems [6]. In addition, the results contribute to the growing body of work on responsible AI by showing how ethical and regulatory requirements can be embedded within system architecture. While prior studies have often focused on policy-level recommendations, this study provides empirical evidence of how such principles can be implemented in practice. This represents an important step toward bridging the gap between theoretical guidelines and real-world AI deployment.

4.6 Limitations and Future Research Directions

Despite the positive outcomes, several limitations should be considered. The system was evaluated within a controlled deployment environment, and its performance may vary under larger-scale or more diverse usage conditions. Additionally, the chatbot's functionality is currently limited to predefined procurement tasks, and more complex queries may require further system enhancements. User feedback also indicated a demand for more detailed traceability information, suggesting opportunities for future customization features. Furthermore, the system relies on external services for language processing and vector storage, which may introduce dependencies related to cost, scalability, and regulatory compliance. Future research could explore the integration of self-hosted language models to reduce reliance on external APIs, as well as the development of adaptive interfaces that allow users to control the level of detail in responses.

Expanding evaluation to larger and more diverse user groups would also provide deeper insights into system performance and usability across different contexts.

4.7 Theoretical Contribution and Broader Impact

Beyond its practical implementation, this study contributes to the broader understanding of how AI systems can be designed for secure and responsible deployment in sensitive domains. The findings support a shift from capability-driven AI development toward architecture-driven approaches that prioritize governance, security, and ethical considerations. By demonstrating that architectural design can simultaneously address performance, security, and ethical requirements, the study provides a foundation for future research on governance-oriented AI systems. This contribution is particularly relevant for domains where data sensitivity and regulatory compliance are critical, such as procurement, healthcare, and finance.

5. Conclusion

This study demonstrated that integrating an AI-powered chatbot within a procurement management system can significantly enhance operational efficiency, strengthen data security, and support ethical and scalable system design. By replacing direct database interaction with a retrieval-based architectural approach, the system achieved measurable improvements in response time, accuracy, and backend performance while reducing security risks associated with data exposure and uncontrolled query execution. The findings confirm that conversational interfaces, when supported by carefully designed system architecture, can transform user interaction with procurement data without compromising reliability or compliance. More broadly, this study highlights the importance of embedding security, ethical, and governance considerations directly into system design, contributing to a shift toward architecture-driven approaches for responsible AI deployment in sensitive operational domains.

Despite these contributions, several limitations should be acknowledged. The system is currently constrained to predefined procurement use cases, and more complex or context-dependent queries may still require human intervention. Additionally, system

performance is influenced by the quality and structure of underlying data, and user adoption may vary depending on familiarity with conversational interfaces. The reliance on external services for language processing and vector storage also introduces potential challenges related to cost, scalability, and regulatory compliance. Future research should focus on expanding system capabilities to support more complex queries, enabling customizable levels of information detail, and evaluating performance across larger and more diverse user groups.

Furthermore, exploring the integration of self-hosted language models and advanced monitoring tools could enhance system autonomy, reduce external dependencies, and support long-term scalability and real-world adoption.

References

[1] Batool A, Zowghi D, Bano M. Responsible AI governance: a systematic literature review. arXiv [Preprint]. 2024;arXiv:2401.10896.

[2] Camilleri MA. Artificial intelligence governance: ethical considerations and implications for social responsibility. Expert Syst. 2023:e13406.

[3] Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People - an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds Mach. 2018;28(4):689-707. doi:10.1007/s11023-018-9482-5.

[4] Gill S. Cybersecurity risks in AI-integrated enterprise systems. J Inf Secur. 2024.

[5] Hevner AR, March ST, Park J, Ram S. Design science in information systems research. MIS Q. 2004;28(1):75-105.

[6] Islam S, Storer T. A systematic review of agile development methodologies in software engineering. J Syst Softw. 2020.

[7] Kumar A. AI-driven procurement systems: enhancing efficiency and decision-making. J Supply Chain Innov. 2024.

[8] Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. Adv Neural Inf Process Syst. 2020;33:9459-9474.

[9] Minkkinen M, Mäntymäki M. AI governance: themes, knowledge gaps and future agendas. Internet Res. 2023;33(7):133-167.

[10] Mökander J, Floridi L. Operationalising AI governance through ethics-based auditing: an industry case study. AI Ethics. 2022;3(1):451-468.

[11] Pesati N. Security considerations for large language model use: implementation research in securing LLM-integrated applications. SSRN [Preprint]. 2024. Available from: https://ssrn.com/abstract=4962370.

[12] Reuel A, et al. Securing retrieval-augmented generation pipelines. arXiv [Preprint]. 2024.

[13] Rose S, Borchert O, Mitchell S, Connelly S. Zero trust architecture. NIST Special Publication 800-207. Gaithersburg (MD): National Institute of Standards and Technology; 2020. doi:10.6028/NIST.SP.800-207.

[14] Shostack A. Threat Modeling: Designing for Security. Hoboken (NJ): Wiley; 2014.

[15] Subramaniam C, Shaw MJ. The effects of process characteristics on the value of B2B e-procurement. Inf Technol Manag. 2004;5:161-180.

[16] Taeihagh A. Governance of artificial intelligence. Policy Soc. 2021;40(2):137-157.

[17] Yao Y, Duan J, Xu K, Cai Y, Sun Z, Zhang Y. A survey on large language model security and privacy: the good, the bad, and the ugly. High-Confidence Comput. 2024;4:100211. doi:10.1016/j.hcc.2024.100211.

[18] Zhang M, Li J. A commentary of GPT-3 in MIT Technology Review 2021. Fundam Res. 2021;3(11):100011. doi:10.1016/j.fmre.2021.11.011.

Acknowledgements

No acknowledgements were reported.

Declarations

Funding: Not reported.

Conflict of Interest: Not reported.

Ethics Approval: Not reported.

Informed Consent: Not reported.

Data Availability: Not reported.

Author Contributions: Muhammad Zubair is the sole author of the article.

Figures

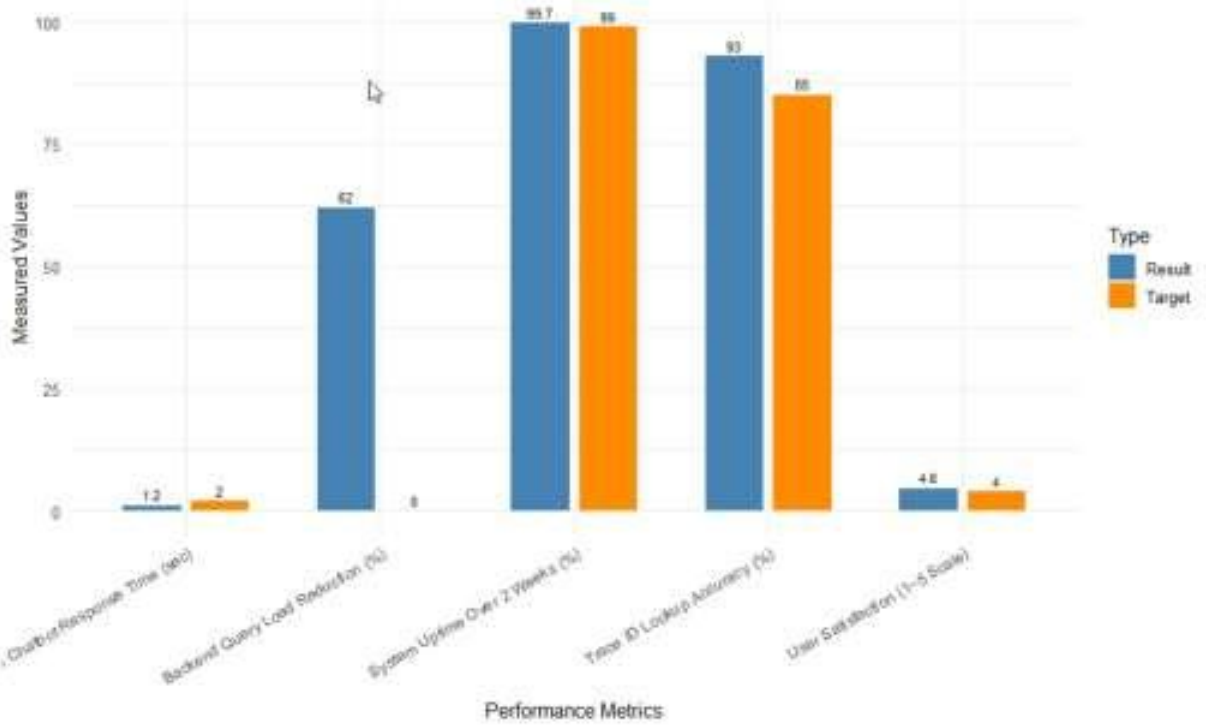

Figure 1. Project performance metrics compared with target benchmarks, including response time, backend query reduction, uptime, trace ID lookup accuracy, and user satisfaction.

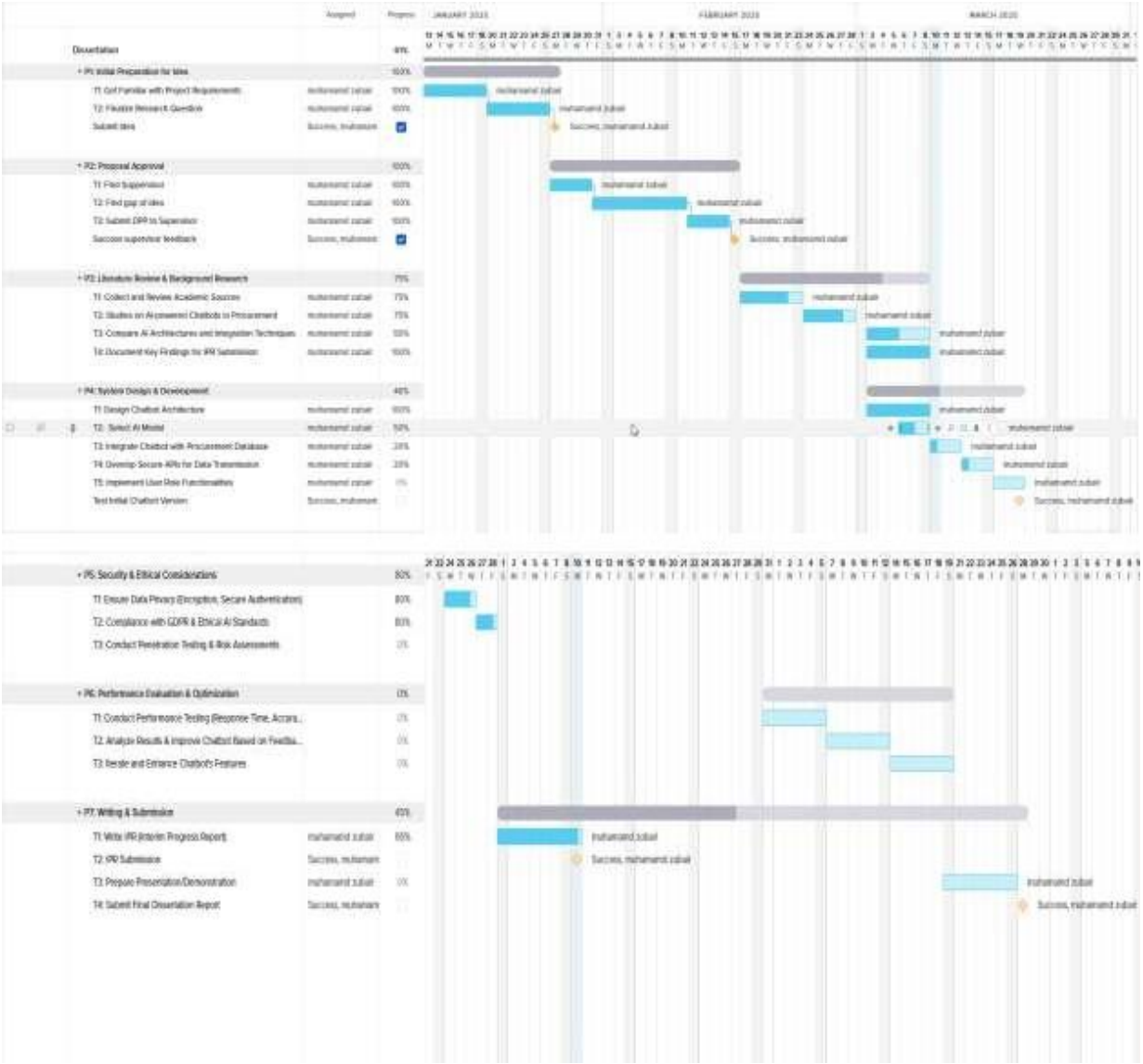

Figure 2. Development timeline covering backend integration, API connections, Pinecone setup, security and ethical considerations, performance evaluation, and production deployment.

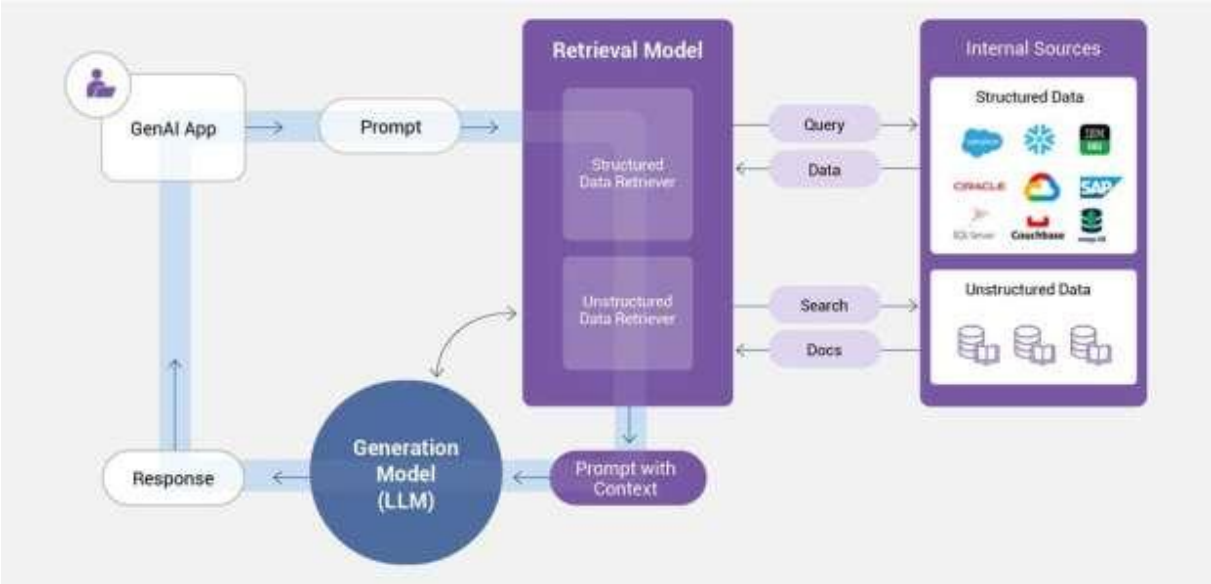

Figure 3. Retrieval-Augmented Generation architecture used in the project, illustrating the relationship among the GenAI application, retrieval model, internal sources, language model, and response layer.

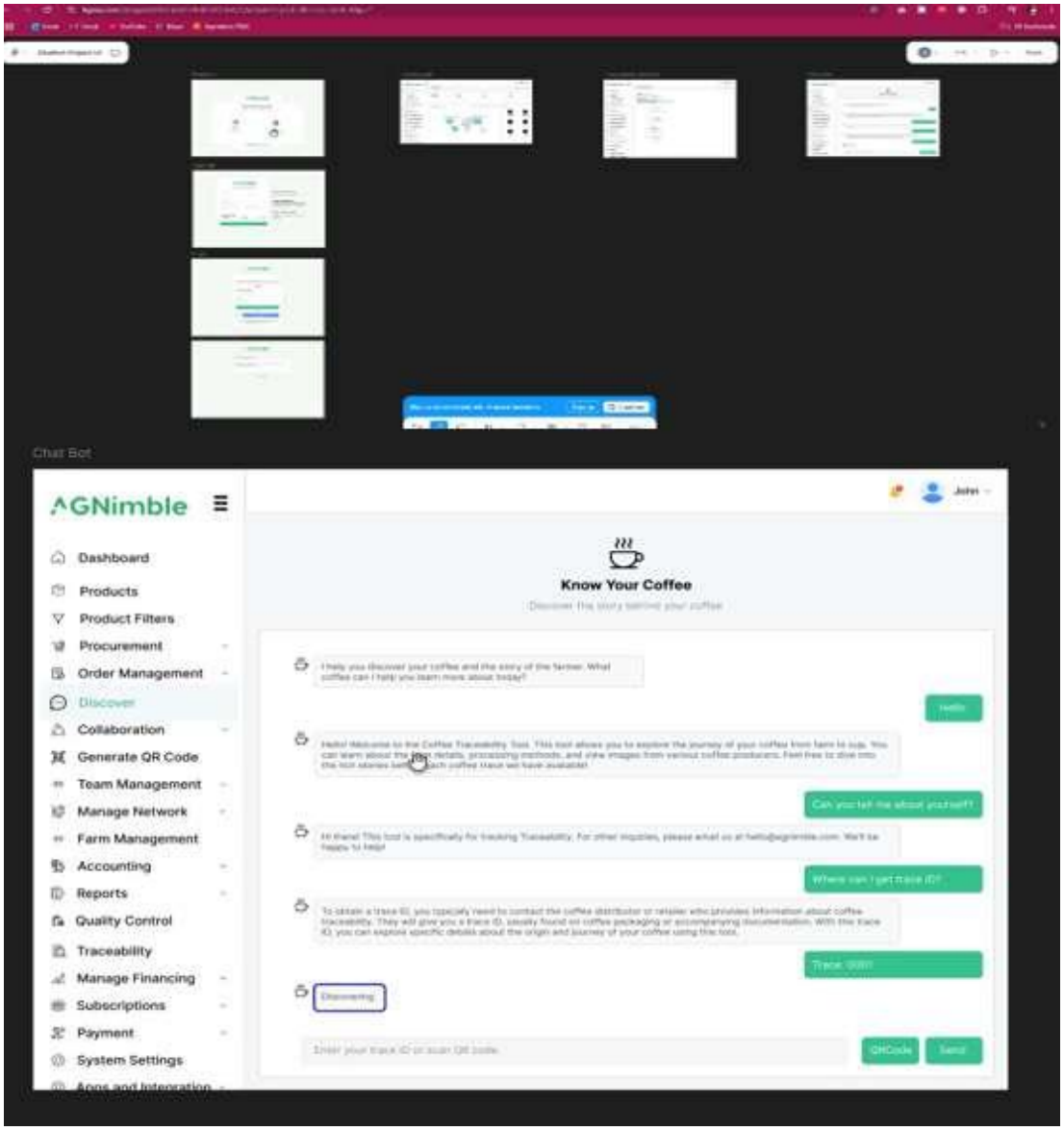

Figure 4. Figma design prototype and procurement chatbot interface used in the implemented system.

Tables

No tables were included in the supplied article.