

AI-powered RFx Intelligence for Strategic Supplier Excellence

by

Johnny Gonzalo Paredes Delgado

Master of Business Administration, University of Maastricht, 2014

and

Nastasja Dinah Zaunick

Bachelor of Engineering in Industrial Engineering, University of Applied Sciences Düsseldorf,
2018

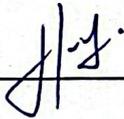
SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2025

© 2025 Johnny Gonzalo Paredes Delgado and Nastasja Dinah Zaunick. All rights reserved.

The authors hereby grant to MIT permission to reproduce and to distribute publicly paper and electronic copies of this capstone document in whole or in part in any medium now known or hereafter created.

Signature of Author: _____



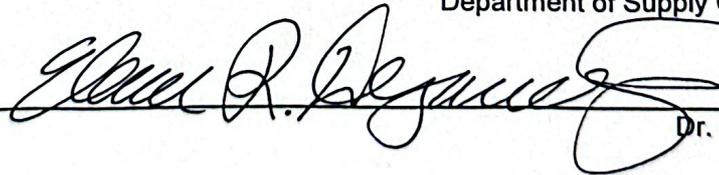
Johnny Gonzalo Paredes Delgado
Department of Supply Chain Management
May 9, 2025

Signature of Author: _____



Nastasja Dinah Zaunick
Department of Supply Chain Management
May 9, 2025

Certified by: _____



Dr. Elenna R. Dugundji
Research Scientist
Capstone Advisor

Accepted by: _____

Prof. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

AI-powered RFx Intelligence for Strategic Supplier Excellence

by

Johnny Gonzalo Paredes Delgado

and

Nastasja Dinah Zaunick

Submitted to the Program in Supply Chain Management
on May 9, 2025 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

Procurement teams often struggle to fully leverage the valuable supplier knowledge captured during RFx (Request for Information, Proposal, or Quotation) events, mainly due to the high volume of unstructured data and fragmented data storage, which limits access to past insights. This paper explores how generative AI can preserve, retrieve, and apply RFx knowledge to strengthen supplier discovery, streamline event execution, and support data-driven strategic sourcing decisions. To address this challenge, the project developed a custom-built generative AI chatbot named Raffa, designed to extract meaningful insights from a wide range of historical RFx documents, and evaluated its performance alongside the sponsor's internal GenAI Platform. Raffa was built on a Retrieval-Augmented Generation (RAG) framework integrating prompt engineering, structured metadata, and a vector database to enable context-rich retrieval. Both solutions were assessed against five criteria: truthfulness, answer accuracy, contextual relevance, handling of complex queries, and response structure. Results showed that metadata and prompt engineering significantly improved Raffa's contextual relevance and overall response quality. Raffa outperformed the GenAI Platform in three of the five evaluation areas, particularly excelling in delivering precise and context-aware answers, even for complex queries. Additionally, the project introduced an RFx Knowledge Intelligence Framework and Prompt Library to guide scalable generative AI adoption across the RFx lifecycle.

Capstone Advisor: Dr. Elenna R. Dugundji
Title: Research Scientist

ACKNOWLEDGMENTS

We would like to sincerely thank our advisor, Dr. Elenna Dugundji for her continuous support and thoughtful guidance throughout this capstone project. Her commitment to connecting us with thought leaders in the field and fostering an open, collaborative environment provided us with invaluable insights and inspiration. Special thanks to the individuals who engaged with us throughout the project and generously shared their perspectives. These conversations enriched our understanding of generative AI in procurement and helped shape the direction of this project. We are also deeply grateful to our project sponsor for their engagement, expert input, and the many hours of thoughtful discussion that contributed to the depth and relevance of this project. Finally, we would like to thank our writing advisor, Ms. Pamela Siska, for her thoughtful feedback and support in helping us to ensure that our final report reflects the significance of the work we set out to accomplish.

Johnny Paredes & Nastasja Zaunick

I would like to thank my capstone partner and friend, Nastasja Zaunick, whose dedication and pursuit of excellence were essential to the success of this work. I am deeply grateful to my wife, Luciana, for her unwavering support and for caring for our family during this demanding time. Along with our children, Jacobo and Aitana, she remains my greatest source of strength and motivation. I also thank my mother, whose values and resilience continue to guide me in pursuing my dreams. Finally, I extend my heartfelt appreciation to my siblings, extended family, colleagues, mentors, and coaches for their invaluable support throughout this extraordinary journey.

Johnny Paredes

I would like to thank my capstone partner and friend, Johnny Paredes, for his dedication and the spirit of collaboration he brought to every step of this journey. To my husband Louis: Thank you for your unwavering support and encouragement, especially in the moments when I doubted myself. I would not be where I am without you. Finally, I am deeply grateful to my family in Germany for their continuous belief in me and for always cheering me on from afar.

Nastasja Zaunick

TABLE OF CONTENTS

1	INTRODUCTION	5
1.1	Motivation	5
1.2	Problem Statement	6
1.3	Scope: Project Goals and Expected Outcomes	7
2	STATE OF PRACTICE	8
2.1	Generative AI in the Sourcing Process	9
2.2	Knowledge Management with Generative AI	11
2.3	Generative AI Optimization	12
2.3.1	Prompt Engineering	13
2.3.2	Retrieval-Augmented Generation (RAG).....	15
2.3.3	RAG Databases.....	16
2.3.4	Fine Tuning.....	16
2.3.5	Data Harmonization	17
2.4	Capabilities of GenAI Platform.....	18
3	METHODOLOGY.....	19
3.1	Data in Scope and Data Harmonization	23
3.2	Proof of Concept Creation	25
3.3	GenAI Platform Evaluation.....	27
4	RESULTS AND DISCUSSION.....	27
4.1	Comparative Analysis of Generative AI Solutions	28
4.1.1	Proof of Concept Assessment.....	28
4.1.2	GenAI Platform Assessment.....	32
4.1.3	Solution Performance Comparison	35
4.2	RFx Knowledge Intelligence Framework.....	36
4.2.1	Embedding Generative AI in the RFx Process	37
4.2.2	RFx Prompt Overview.....	39
5	CONCLUSION AND OUTLOOK	40
	REFERENCES	43
	APPENDICES.....	46

1 INTRODUCTION

“Knowledge management is the process of identifying, organizing, storing, and disseminating information within an organization” (IBM, 2022). When knowledge is not readily accessible, it is likely that the cost of reprocessing information increases, as time may be spent searching for relevant information instead of completing results-oriented tasks. Some of the main advantages of knowledge management include making informed decisions, preserving company knowledge, improving operational efficiency, enhancing communication, and ensuring data security. Knowledge management consists of three types: tacit knowledge (stemming from experience), implicit knowledge (know-how), and explicit knowledge which is captured in various kinds of documents such as manuals, reports, and guides including databases and technical documents (IBM, 2022). For the purposes of this project, the primary focus is on explicit knowledge.

Building on the foundation of explicit knowledge, Artificial Intelligence (AI) emerges as a powerful set of technologies designed to emulate human intelligence, enabling machines to process information and make autonomous decisions. By leveraging AI, organizations can analyze vast amounts of information from explicit sources to identify patterns and trends. This capability enhances various Sourcing and Procurement Objectives, such as Business Partner Engagement, Supplier Partnership Optimization, and the Management of Spend, Supply Risks, and the Procurement function itself (Gartner, 2023). In particular, the integration of AI in RFX processes showcases its transformative potential. By applying AI and machine learning algorithms, organizations can optimize operations through data structuring, highlighting the most relevant information, and suggesting best-fit suppliers based on historical RFX events. The targeted application of AI in RFX processes assists decision-makers in making more informed choices, thereby accelerating the process and improving supplier discovery and accuracy in the selection of suppliers.

1.1 Motivation

Our sponsor company is an American healthcare manufacturer. Its aim in this project is to optimize its knowledge management in the sourcing process, with a particular focus on RFX - Requests for Information (RFI), Requests for Proposal (RFP), and Requests for Quotes (RFQ), using generative AI. RFI is a research process in which information is collected from different suppliers before formally sourcing goods or services, to create a short list of suppliers to meet the company’s requirements. RFP is a competitive process that invites suppliers to find a

standardized solution for the company's needs. Finally, RFQ is a process designed to obtain the best commercial conditions from suppliers focused on prices.

The goal of RFX is to ensure reliable and resistant supply, drive innovation, and optimize cost. For example, RFX is used to support sudden demand, capacity, or location changes. In benchmarking RFX is used to achieve optimal cost to provide fair pricing. Our sponsor company uses RFX to engage with the external market and to support innovation following the pull approach, addressing customer needs or market demands, and the push approach starting with new developments created by a company and pushed into the market.

Our sponsor's procurement organization executes hundreds of RFX events per year. These events range in complexity from a simple quote request sent to three suppliers, to a complex event involving tens of suppliers and hundreds of requirements. During the RFX process, our sponsor company gains extensive knowledge about the suppliers including capability, capacity, reliability, cost structure, relationship management, and sustainability. Furthermore, the insights gained during the RFX process encompass the suppliers' product and service offerings, their potential for innovation, production scalability, past performance, pricing models and cost transparency, regulatory compliance, and ethical standards.

At present, the company uses two independent and disconnected tools for its RFX process. The choice of which tool to use depends on the complexity of the RFX event. Additionally, RFX events are conducted outside of these tools, and data is collected, in part, by e-mail, stored individually, and disconnected in various locations such as SharePoint.

The main challenge in the current RFX process is to leverage the knowledge from previous RFX events and data stored in various data sources in new bidding processes. Currently, buyers use two independent systems to conduct an RFX and have access to several templates that include questionnaires predefined by the tool owners and by users based on previous bidding processes. However, the buyers lack valuable information such as supplier capacity, performance indicators, and pricing strategies.

Our sponsor aims to develop an AI solution that can gather unstructured data from various formats and process the information to identify patterns based on purchasing categories and regions, thereby providing valuable insights to procurement professionals in future RFX.

1.2 Problem Statement

Our sponsor company conducts a high volume of RFX events per year, often inviting the same suppliers. Sourcing events take place across all categories and regions. Although the company has established processes and checklists for managing RFX events, the knowledge

gained from these events is not strategically applied to future projects because the company lacks a structured archive. The lack results in a failure to transfer knowledge between sourcing events. Once an event concludes, the collected data and insights are not leveraged effectively. Consequently, the knowledge captured during the RFX process remains siloed. Data from past RFX events is maintained in two independent tools used for executing RFX events.

Both tools offer the users the option to create customized questionnaires as part of the RFX process which are stored as structured data. Suppliers participating in the RFX events provide their answers and information in various formats, including structured tables, PDF documents (such as presentations of capabilities or technical specifications), Word documents, and PowerPoint presentations (for example company overviews) which is a challenge for data extraction. In this context, the key questions this project answers include:

- i. How can insights (such as supplier capabilities, supplier RFX performance and cost structure) from previous RFX responses, including cross-events, be generated for evaluation and data-driven decision-making?
- ii. How can the deep and broad knowledge gained throughout the RFX process be retained, and made accessible for the future, leading to more effective and efficient supplier discovery and RFX event execution processes?

1.3 Scope: Project Goals and Expected Outcomes

The project's overall goal is to deliver proof of concept that enables procurement professionals to seamlessly access targeted information about suppliers, their capabilities, capacities and performance, derived from past RFX events. By leveraging AI technologies, we organize and store data from diverse sources and formats in a centralized, structured manner and retrieve insights gained throughout the RFX process. This approach breaks down silos, making supplier information accessible across categories and regions and thus uncovering interdependencies, driving data-driven decision-making, and retaining institutional knowledge.

By developing a user-friendly chatbot interface we enable procurement professionals, experienced and inexperienced, to ask targeted questions and receive actionable insights on suppliers, facilitating supplier discovery, fostering greater collaboration and efficiency in procurement processes to expedite future RFX events, and optimizing their outcomes. Insights gained from the AI solution can, for instance, support decisions on which suppliers to invite to an RFX event by leveraging results and knowledge from historical events. Additionally, the AI solution can provide repetitive supplier information, for suppliers frequently invited to RFX events and support with the RFX evaluation by synthesizing large volumes of information into concise

summaries across multiple suppliers allowing for a more comprehensive and consistent comparison. We expect the AI solution to save procurement professionals time, for example, by suggesting key questions based on user inputs to include in bid documents, helping users to identify the most relevant information and questions based on data from previous events related to similar categories and regions. We further anticipate the outcomes described will encourage procurement professionals to use the RFX tools instead of conducting RFX events outside of the tools. Ensuring that the RFX events are conducted through the available tools promotes process adherence and maintains data availability and knowledge retention which presents an additional benefit for the company.

AI technologies in scope include Retrieval Augmented Generation (RAG) to expand the knowledge of pre-trained Large Language Model (LLM) using data retrieved from previous RFX events. To ensure that the users are providing inputs that will result in accurate responses, we will further develop an RFX Knowledge Intelligence Framework and a Prompt Library of optimized, targeted prompts, leveraging techniques in prompt engineering aimed at simplifying and enhancing user interaction with the LLM and RFX data.

The Capstone Project “Integrating Generative AI to Drive Efficiency and Spend Intelligence and Negotiation Strategy”, conducted by MIT students in 2024, with our sponsor company, was a source of information for this project. It offered useful information about the limitations and challenges of current enterprise solutions. The project deliverables include the following:

- i. Provide Supplier Intelligence
- ii. Support RFX Creation and Execution and Data-Driven Decision Making
- iii. Apply knowledge from previous RFX to future RFX

Building on these goals, our project demonstrates how Generative AI and Retrieval-Augmented Generation (RAG) can transform tacit procurement knowledge into structured, accessible insights, enabling data-driven decision-making and strategic supplier selection.

2 STATE OF PRACTICE

The objective of this capstone is to explore how our sponsor company can leverage historical data from past RFX events with the help of generative AI to optimize future RFX events, enhance decision-making, and enable more efficient supplier discovery and selection. To address this problem, the four key areas of focus are:

- i. **Generative AI in the Sourcing Process:** Explore the procurement and sourcing community's perspective on the state of generative AI in sourcing and procurement, including benefits and potential applications. Examine the specific use cases where these solutions are implemented and identify the associated risks that must be considered in our methodology and final solution. Additionally, it provides an outlook on the future of generative AI, exploring upcoming advancements and their potential impact on the sourcing and procurement landscape.
- ii. **Knowledge Management with generative AI:** Evaluate how knowledge is currently transferred and investigate which technical solutions can effectively support knowledge management in the context of generative AI.
- iii. **Generative AI Optimization:** Generative AI solutions are commonly applied to support knowledge management, data-driven decision-making, and process optimization. To ensure that the solution resulting from this project provides the most accurate output, it is essential to determine which optimization techniques are the most suitable, how the data must be cleaned and harmonized, and what types of databases are available to support the project's aim of providing procurement professionals with relevant and accurate RFx information.
- iv. **Capabilities available in our sponsor company:** The GenAI Platform was developed by our sponsor company. The research assesses the solution's suitability for achieving the project objectives and the constraints.

2.1 Generative AI in the Sourcing Process

Generative AI transforms the sourcing and procurement landscape by driving efficiency and providing procurement professionals with tools that offer advanced capabilities in data analysis and predictive insights, enabling more strategic decision-making (McMillan, 2024). According to Gartner's Hype Cycle (Sommers et al., 2024), generative AI for procurement is at the peak of inflated expectations, a stage that highlights both the potential and challenges of this transformative technology. The Hype Cycle, which tracks innovations from their introduction to real-world application (Gartner, 2018), underscores the urgency of adoption, as "procurement organizations that do not embrace AI technologies will find themselves at a cost and agility deficit compared to their competitors." (Keck et al., 2024). This emphasizes the critical role generative AI plays in maintaining a competitive edge in the evolving procurement landscape.

The procurement and sourcing community views generative AI as a transformative technology ready to enhance efficiency, decision-making, and the overall procurement process.

It offers new capabilities, including creating content, analyzing unstructured data, and enabling faster and more informed decisions (Ivalua, 2024). Generative AI stands apart from traditional AI in its ability to create new content rather than relying solely on predictive models derived from historical data (Ivalua, 2024).

The integration of generative AI into procurement offers several benefits, such as streamlining workflows, automating processes, and freeing up procurement professionals' time for strategic tasks. Specifically, generative AI can support procurement professionals by aiding in supplier discovery, category intelligence, and market research (Ivalua, 2024, p.4). Its application in e-sourcing allows non-professional sourcing staff to execute tasks such as RFX creation and supplier identification, simplifying traditionally complex workflows (Keck et al., 2024). Moreover, the technology is being used for content creation such as RFX questionnaires, summarization of documents and RFX responses, and analysis of supplier information, enhancing efficiency and enabling faster insights (Gartner, 2023). By integrating structured and unstructured data, procurement professionals can make data-driven decisions, increasing the overall effectiveness of sourcing processes.

Additionally, generative AI enables procurement teams to maintain up-to-date knowledge bases, integrating domain-specific information, regulatory requirements, and best practices into a single source of truth (CPOstrategy, 2024, p.24). This improves both internal operations and supplier engagement by ensuring data utilization is central to decision-making processes (CPOstrategy, 2024, p.37). For example, organizations can improve the supplier discovery process by leveraging knowledge databases, AI and other technologies. Through the creation of an AI driven knowledge management solution for RFX and supplier related knowledge, the supplier discovery process becomes more efficient (Sommers et al., 2024). Thereby enabling organizations to mitigate risk and build supply chain resilience and ensuring that the organizations can quickly adapt and ensure that sources are available.

Despite its benefits, generative AI introduces several risks that must be addressed. One critical concern is the accuracy and veracity of its outputs, as hallucinations and incorrect information generation remain prevalent issues (Ivalua, 2024; Chandrasekaran & Ramos, 2024). This highlights the need for robust quality control mechanisms and verification processes within any generative AI-enabled solution. Ensuring outputs are verified against reliable data sources and aligned with organizational objectives will be essential in mitigating these risks.

Another significant risk is data quality. The effectiveness of Generative AI relies on clean, accurate, and complete data, which can be challenging given that procurement data is often dispersed across multiple systems (Ivalua, 2024). Poor data quality not only impacts the reliability

of insights but also increases the likelihood of compliance risks, especially when handling confidential or regulatory-sensitive information (CPOstrategy, 2024).

Security concerns also demand attention. Enterprises must establish stringent governance frameworks to ensure that sensitive procurement data is handled responsibly (Chandrasekaran & Ramos, 2024). Additionally, ensuring user trust through transparency and robust prompting libraries can mitigate risks associated with misuse or misunderstanding of Generative AI tools (CPOstrategy, 2024).

Research indicates that generative AI is already being utilized across various use cases in sourcing and procurement. However, its rapid evolution promises to introduce new functionalities and significantly impact both sourcing and procurement, in the near future (Polk et al., 2024). For instance, autonomous sourcing is expected to streamline routine tasks, enhance efficiency, and ensure compliance, allowing procurement professionals to focus on more strategic, value-driven activities (Sommers et al., 2024). Emerging functionalities such as agentic reasoning, multimodality, and AI agents are expected to reshape procurement further. Agentic reasoning will enable autonomous decision-making, multimodality will enhance accessibility and user experience by making AI more intuitive, and AI agents will redefine procurement roles by automating critical activities.

To successfully implement these advanced use cases, procurement leaders should prioritize privacy standards and ensure that high quality data from structured and unstructured sources is available to train, fine-tune or apply Retrieval-Augmented Generation (RAG) on Large Language Models (Polk et al., 2024). Additionally, applying advanced prompt engineering techniques, such as the Chain of Thought methodology, can help break down complex procurement challenges into manageable steps to derive effective solutions. Section 2.3 Generative AI Optimization will expand on optimization techniques such as fine-tuning, RAG and prompt engineering.

2.2 Knowledge Management with Generative AI

Companies manage knowledge to drive continuous innovation, enhance productivity, foster collaboration, and maintain a competitive advantage (O'Dell & Grayson, 1998). The creation of knowledge is an important pillar within knowledge management; for this reason, Nonaka and Takeuchi (1995) introduce the SECI model, which describes the knowledge creation process in four stages: socialization, externalization, combination, and internalization. Through this process, tacit knowledge (personal knowledge, difficult to express in words) is converted into

explicit knowledge (formal and shareable) and vice versa, creating a dynamic cycle of learning and improvement.

In the current digital era, artificial intelligence, especially through machine learning algorithms and natural language processing, has further empowered organizations by automating the capture and analysis of knowledge. AI systems can detect patterns, extract meaningful insights from vast amounts of data, and provide personalized recommendations, thereby enhancing real-time access to information and decision-making (Dalkir, 2011) which complements and accelerates the knowledge creation process described in the SECI model.

Furthermore, it is important to emphasize that knowledge management is not only a matter of technology; it also requires a culture that values learning, collaboration, and transparency. Leadership is key to the success of knowledge management, as leaders must foster an environment where knowledge sharing is encouraged and information silos are broken down (Dalkir, 2011).

For this project, we will make use of an existing knowledge management platform (GenAI Platform) that allows for the structured storage, organization, and sharing of information within the procurement organization of our sponsor. The platform can be applied primarily in the knowledge creation stages of Combination for a procurement area, as it facilitates the creation, organization, and access to structured knowledge.

In the Combination stage, which involves organizing, classifying, and merging existing knowledge to generate new knowledge applicable across the entire procurement organization, the GenAI Platform can act as a repository where documents, reports, and analyses on suppliers, market conditions, and RFX processes are gathered and combined. Users can access this information in a centralized manner, facilitating the creation of aggregated reports or the comparison of data between suppliers. This allows knowledge to be structured and accessed in a way that supports strategic decision-making and the optimization of procurement processes. A thorough explanation of the GenAI Platform can be found in section 2.4 Capabilities of GenAI Platform.

2.3 Generative AI Optimization

General-purpose Large Language Models (LLM) such as GPT, BERT or LLaMA are versatile tools, commonly applied for a wide variety of use cases. These include conversing with a user in natural language, answering general questions, and summarizing content (Dhamani & Engler, 2024). LLMs are trained on large data sets, based on publicly available data, including content from the internet, social media, and other public domains. LLMs are considered

generalistic as they rely on their pre-trained knowledge to generate output and have considerable limitations, including “a knowledge cut-off, the point at which their training was last updated” (Belcic, 2024).

To adapt an LLM to domain-specific tasks or use case and to improve the LLM’s output organizations can choose between various methods which differ in complexity and applicability. The following will discuss and compare three commonly used techniques to customize and enhance generative AI solutions: Prompt Engineering, Retrieval-Augmented Generation and Fine tuning. The goal is to identify which technique is the most beneficial for the purpose of this Capstone project.

2.3.1 Prompt Engineering

Prompt engineering is the practice of creative input prompts that guide Large Language Models to the desired outputs, utilizing the full capabilities of generative AI models (Lorica, 2023). It involves designing prompts that guide LLMs to produce accurate results by optimizing prompt length, structure, and style (Boonstra, 2024). For success, prompt engineering tools must be flexible, transparent, and user-friendly. Such tools should handle various data types, be robust against noisy data, and be usable across platforms. Transparency enables users to understand model functionality and make informed decisions when using AI tools (Lorica, 2023). To achieve these objectives, several techniques of prompt engineering, along with their advantages and disadvantages, are outlined in Table 1.

Table 1. Techniques of Prompt Engineering (Boonstra, 2024; Wei et al., 2022)

Prompting Technique	Description	Advantages	Disadvantages
Zero Shot Prompting	Relies on task descriptions without providing examples. The model uses its pre-trained knowledge to respond.	<ul style="list-style-type: none"> - Easy and fast to implement. - Suitable for tasks where instructions are straightforward. - No additional data preparation needed. 	<ul style="list-style-type: none"> - Limited reliability, especially for complex or nuanced tasks. - May generate incorrect or generic responses due to lack of guidance.
Few Shots Prompting	Provides 1 to 5 examples within the prompt to demonstrate the desired response pattern. Helps the model infer the expected behavior based on these examples.	<ul style="list-style-type: none"> - Improves accuracy and relevance of responses. - Helps the model handle new tasks effectively with minimal data. - Allows for greater control over output style and format. 	<ul style="list-style-type: none"> - Requires carefully curated examples that are diverse, high-quality, and relevant. - Limited scalability for very large or varied datasets.
Chain of Thought Prompting	Encourages the model to break down complex reasoning tasks into intermediate steps, like how humans solve multi-step problems (e.g., math word problems). The process involves decomposing the problem, solving each step, and then combining results into the final answer.	<ul style="list-style-type: none"> - Highly effective for tasks requiring logical reasoning. - Enhances explainability of the model's output. - Useful for debugging and understanding model performance in reasoning tasks. 	<ul style="list-style-type: none"> - Generates longer responses, increasing computational cost. - Slower processing time. - May require tuning for tasks to avoid excessive verbosity.
Other Prompting Techniques	Includes various specialized techniques: <ul style="list-style-type: none"> - System Prompting: Defines the big picture or high-level task for the model. - Contextual Prompting: Adds background details or task-specific nuances. - Role Prompting: Assigns a persona or voice for the model. - Step Back Prompting: Splits the task into smaller prompts sequentially. - Tree-of-Thought Prompting: Explores multiple reasoning paths simultaneously. 	<ul style="list-style-type: none"> - Highly flexible and adaptable. - Can address specific task requirements or challenges. - Role and contextual prompting enhance the model's alignment with task needs. - Supports exploratory tasks. 	<ul style="list-style-type: none"> - Requires expert knowledge to implement effectively. - More complex prompts may result in harder-to-predict outputs. - Scalability depends on the task and context.
Automatic Prompt Engineering	Utilizes models to generate and evaluate their own prompts, optimizing performance across multiple tasks. Includes steps such as generating candidate prompts, scoring outputs (e.g., BLEU, ROUGE), and selecting the best-performing prompts automatically.	<ul style="list-style-type: none"> - Scalable across multiple tasks. - Automates prompt creation, saving time and effort. - Produces more consistent results by leveraging systematic evaluation criteria. 	<ul style="list-style-type: none"> - High setup cost and technical complexity. - Initial design and evaluation process require expert knowledge. - May not handle highly creative or subjective tasks effectively.

To improve the accuracy of prompt engineering, it is essential to first align the desired outcomes with the specific goals of the procurement organization. Prompts should be crafted with clarity and precision (DeepLearning.AI, 2024), leveraging past examples through techniques such as few-shot learning to guide the model, and including relevant and sufficient context. For more complex tasks, adopting a Chain of Thought (CoT) method can enhance logical reasoning and coherence in responses.

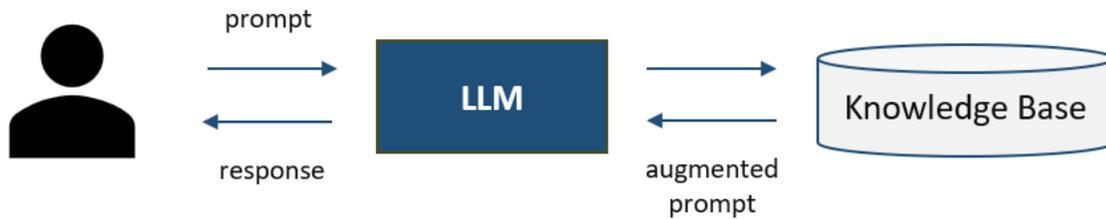
On the other hand, the structure and complexity of prompts should be adjusted to fit the task, balancing detail with conciseness. Regularly reviewing and refining prompts based on observed outcomes, alongside parameters including temperature, control the degree of randomness in the selection of the most prescribed token (Google, 2024) and nucleus sampling (Top-P). How many possible words to consider (Google, 2024), can further enhance precision and consistency. Finally, achieving success depends on iterative improvement and feedback to ensure alignment between the prompts and the procurement organization's objectives.

2.3.2 Retrieval-Augmented Generation (RAG)

Another method to improve the LLMs output is Retrieval-Augmented Generation (RAG). RAG enables LLMs to retrieve knowledge from additional knowledge bases on top of the pre-trained knowledge it possesses. Knowledge bases include data relevant to specific use cases and can be updated frequently with new information. By integrating the specialized, real-time knowledge in the output generation process, LLMs can answer domain-specific questions related to internal company data more accurately (Zhou, 2024), and the need of model retraining and the likelihood of hallucinations is reduced. However, it is important that the knowledge base contains comprehensive, high-quality data to ensure that high quality responses are generated (IBM Technology, 2023).

RAG technology can be summarized in two steps: the ingestion and the retrieval (McKinsey & Company, 2024). In the ingestion phase, the data is loaded, broken down into smaller chunks and stored in a knowledge base (LangChain, 2024). In the retrieval phase, the user input triggers a retriever to retrieve the relevant information from the knowledge base. Relevant content from the retrieved data is used to augment the user prompt. The augmented prompt is given to the LLM which generates enhanced output. (Belcic, 2024). Figure 1 shows a simplified model of RAG technology.

Figure 1. RAG Technology – Simplified Model (Adapted from IBM Technology, 2023)



2.3.3 RAG Databases

RAG technology retrieves knowledge from a database and combined with the original user query, generates the output or response. The choice of the type of database is critical for the performance of the system in terms of speed and response accuracy (Xinghan, 2025) and depends on the requirements for final solution and the characteristics of the data (Behairy, 2023).

The research shows that Knowledge Graph databases and Vector databases are commonly used in the context of RAG. Behairy (2023) and Xinhghan (2025) agree that Knowledge Graph databases excel in use cases with structured data which require the understanding of complex relationships between data points. While Knowledge Graph databases are strong in transparent reasoning, they require a well-defined structure (Behairy, 2023). Vector databases are a good choice for large, unstructured datasets (Xinghan, 2025). Their ability to understand context and interpret underlying meaning (Behairy, 2023) is specifically useful for chatbots where the structure and clarity of user queries differ between users. With their ability to incorporate a variety of data types and retrieve data quickly based on vector similarity, Vector databases offer high scalability (Behairy, 2023).

2.3.4 Fine Tuning

Fine tuning is a commonly used technique to customize and optimize LLMs for specific use cases. Fine Tuning describes the method of training a pre-trained LLM with targeted domain-specific data, tailoring the LLM for specific tasks and a narrow scope (Zhou, 2023). While RAG supplements the LLM with the use of a knowledge base; fine tuning instills context and intuition into the model. This improves the LLMs performance on a specific task, influences how the model behaves, and is beneficial for instance to keep the tone or style of a brand (IBM Technology, 2024). Fine-tuning also improves the model's ability to consistently produce reliable outputs, especially for tasks requiring complex prompt adherence or handling diverse edge cases.

Furthermore, it enables the model to learn new skills or tasks that may not be easily specified through standard prompting (OpenAI, n.d.).

Fine-tuning and Retrieval-Augmented Generation offer both advantages and limitations for enhancing LLMs. Fine-tuned models are optimized for specific tasks, requiring smaller prompts and providing faster inference, making them suitable for scenarios where speed is essential. However, they come with a fixed cut-off date for knowledge, require substantial computing power for training, and risk overtraining, which can limit their flexibility and adaptability. In contrast, RAG is better suited for dynamic data sources where continuous updates are needed. It does not modify or enhance the base model but instead retrieves information from an external knowledge base. This allows RAG to provide sources for its output, ensuring transparency. However, maintaining RAG's knowledge base involves manual updates and a verification process is recommended to review documents before adding them. While fine-tuning excels in efficiency for static datasets, RAG is ideal for evolving data environments where dynamic updates and source traceability are critical (IBM Technology, 2024).

2.3.5 Data Harmonization

The International Data Corporation forecasts that the global datasphere will expand from 149 zettabytes in 2024 to 394 zettabytes by 2028 (IDC et al., 2024). This significant increase in data presents an opportunity for companies to gain insights that enhance decision-making across various processes. However, this data is often dispersed, stored in silos, and exists in different languages, formats, and structures. Therefore, it is essential to standardize databases, which will serve as raw material, before implementing a Generative AI (Gen AI) solution to improve the accuracy of the responses that can be obtained. This standardization process is known as data harmonization, defined as the process of converting raw data from various sources and formats into a coherent, standardized, and comprehensive structure for analysis (EPAM SolutionsHub, 2022).

Given the rapid growth of data, companies must be ready to harmonize their data efficiently and effectively. Table 2 illustrates the potential impact of employing techniques such as machine learning and AI to enhance the data harmonization process within organizations. The integration of these technologies can ensure higher accuracy in the responses generated using Generative AI. Furthermore, it offers benefits such as reduced data processing time and improved scalability (Sedinkina, 2023).

Table 2. Advantages of Machine Learning and AI Usage in Data Harmonization (Sedinkina, 2023)

Key Reason	As Is	To Be
Scalability and Efficiency	<ul style="list-style-type: none"> - Manual harmonization - Dealing with many labels - High labor and time-consuming 	<ul style="list-style-type: none"> - Automate the harmonization process - Lower time and labor
Adaptability to Changing Data	<ul style="list-style-type: none"> - Global expansion - Diversity of data sources, languages, and high variation in labeling 	<ul style="list-style-type: none"> - Adapt models and learn patterns from data - Handle different labeling conventions
Improved Accuracy and Consistency	<ul style="list-style-type: none"> - Human errors and inconsistencies 	<ul style="list-style-type: none"> - Models learn from existing labeled data and apply statistical techniques - Improve accuracy, consistency, and reliability
Language and Contextual Understanding	<ul style="list-style-type: none"> - Labels in databases written in different languages - Similar concepts represented differently 	<ul style="list-style-type: none"> - Natural Language Processing - Understand the semantics and context of labels - Enhance Efficiency in the Harmonization Process

2.4 Capabilities of GenAI Platform

The GenAI Platform, developed by our sponsor company, is a promising tool for enhancing knowledge management. The GenAI Platform consists of two core capabilities: the Content Editor and the Content Chat User.

The Content Editor enables users to build custom knowledge bases by uploading unstructured documents from previous RFX processes in formats such as DOC, PDF, PPTX, and TXT. Leveraging a built-in RAG functionality, the editor retrieves information from the curated knowledge bases to augment the prompts provided to the LLM, which is part of our sponsor's intelligent stack architecture (Sponsor Company, 2024). The Content Chat User enables users to interact with a chatbot by following suggested prompts or ask specific questions. The platform then delivers responses or generates relevant insights based on the information available in the knowledge base (Sponsor Company, 2024).

As a result, the platform enables procurement professionals to accelerate the process of accessing relevant information, reducing the time spent on repetitive tasks and helping procurement professionals derive insights to establish strategies for their upcoming RFX projects. However, a challenge remains in the platform's current inability to support structured data formats, such as Excel, which are often critical in RFX events. Furthermore, fostering the adoption of this platform as an integral part of procurement professionals' daily operations and ensuring its scalability across various organizational departments are key considerations for maximizing its effectiveness.

After reviewing the four key areas, we have determined that our proof of concept will focus on building a custom chatbot based on RAG technology, utilizing the existing GenAI Platform to upload unstructured documents received from suppliers during RFX processes and compare the custom solution created for the procurement use case with the general solution available in our sponsor company. The built-in RAG technology will allow users to retrieve relevant information and generate text-based responses to assist procurement professionals. Furthermore, by applying prompt engineering techniques in the custom solution, we aim to enable users to interact seamlessly with the platform and extract valuable insights for future RFX processes. Given the current limitation of the GenAI Platform in handling structured data, such as Excel files, and the focus of this project on unstructured data, we will test the model's response when structured information is transformed into text and integrated into the platform.

3 METHODOLOGY

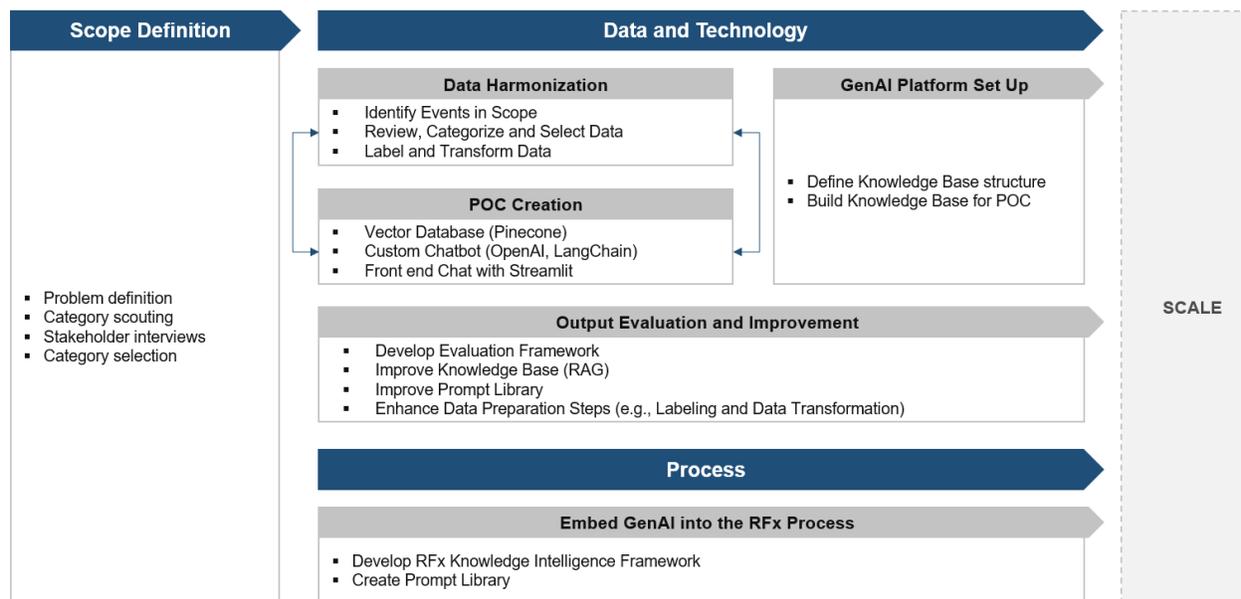
This chapter outlines the methodology applied in this capstone project, covering the approach for data collection and harmonization, proof of concept development, GenAI Platform setup and the methodology for evaluation and improvement of the response quality of both the proof of concept and the GenAI Platform.

The literature review demonstrates that the objective of this capstone project, to make historical RFX data accessible and extract actionable insights for future RFX events by retrieving explicit knowledge from a large amount of unstructured data, and to compare large amounts of unstructured data collected in ongoing RFX, is an excellent use case for the application of generative AI. Multiple optimization techniques have the potential to improve the model's output and provide procurement professionals with valuable information. Among these, fine-tuning, a method of training a pre-trained LLM with domain-specific data, presents limitations for this use case. It depends on the fixed cut-off date for knowledge and requires substantial computing power and cost for training, making it less suitable for this use case where RFX events are conducted frequently, leading to the need for continuous updates of the knowledge base. Retrieval-Augmented Generation (RAG), which augments user prompts by leveraging a use case specific knowledge base, is a more suitable approach for our project. RAG is effective in answering complex questions that involve specialized knowledge such as the information obtained in past RFX events. By leveraging RFX specific data, RAG enhances the capabilities of general Large Language Models, thereby creating more accurate and relevant responses and reducing hallucinations (Zhou, 2023). Additionally, RAG can handle dynamic data, such as RFX data, as

the knowledge base can be updated continuously with minimal effort which is ideal for this use case. Finally, RAG aligns with our sponsor company’s existing GenAI Platform capabilities.

Our approach addresses data, technology and process challenges to ensure scalability and alignment with business needs. We developed a proof of concept in the form of a customized chatbot named “Raffa” that leverages a Retrieval-Augmented Generation (RAG) approach and a vector database to store and retrieve knowledge collected in the RFX process. In addition, we assessed how to embed generative AI in the RFX process to guide procurement professionals in effectively applying generative AI throughout the RFX lifecycle, maximizing user adoption, value and insight generation. Figure 2 summarizes the project methodology, outlining the key phases and activities involved in achieving the project objectives.

Figure 2. Methodology Overview



The methodology for this project follows a structured process starting with Scope Definition. The problem was identified, relevant categories were scouted, and insights were gathered through interviews with our sponsor company’s procurement teams. The scope definition was followed by identifying the data in scope, data collection and harmonization which involved identifying the events in scope in collaboration with the subject matter experts, reviewing and categorizing relevant data, downloading necessary datasets, and performing labeling and transformation to ensure usability.

Next, we created a proof of concept in form of a RAG-powered chatbot. We integrated technologies such as the LangChain Framework, and OpenAI’s Large Language Model and

Embedding Model. After reviewing the vector databases Chroma (open source) and Pinecone, we decided to proceed with Pinecone (managed database) as our vector database of choice due to its scalability, automatic indexing, and Python support allowing for easy deployment (Woyera, 2023). Pinecone stores the vectors created in the chunking and embedding process. Finally, the proof of concept was complemented by a front-end user interface using Streamlit. To ensure data privacy, the proof-of-concept was developed and tested with synthetic RFX data only.

In the subsequent phase, we set up the knowledge base in the GenAI Platform by uploading and structuring data, while leveraging our learnings regarding formats and content structure from the proof-of-concept phase.

During the output evaluation phase, we documented prompts and responses to analyze the response quality. Following the framework proposed by Es et al. (2023) we identified three key metrics: truthfulness, answer relevance and context relevance, to evaluate the model's robustness. In addition to these foundational metrics, we introduced two supplementary metrics, Handling of complex queries and Response structure, to tailor the evaluation to the practical needs of our use case. Table 3 shows the final response evaluation framework, including the five key metrics, tailored to the specific requirements of enterprise and procurement knowledge intelligence, ensuring both technical robustness and practical applicability.

Table 3. Response Evaluation Framework (Adapted from Es et al., 2023)

Evaluation Metric	Definition	High	Medium	Weak
Truthfulness	Does the model's output align with the facts presented in the knowledge base, or do we observe hallucinations?	All points in the answer are supported by the provided context.	Most points are well-supported, but there are slight factual gaps.	Includes points that are either unverified or invented by the model.
Answer Accuracy	Does the answer address the user's prompt?	Covers the question in full, staying clear and to the point.	Covers part of the topic, though parts feel repetitive or off-point.	The response doesn't cover the key aspects of the question.
Contextual Accuracy	Is the output focused on the intended topic and excludes irrelevant information?	Information is clear, relevant, and free of unnecessary detail.	Contains useful and irrelevant information.	Too much irrelevant information makes the context hard to follow.
Handling Complex Queries	How well does the model respond to multi-layered questions?	Effectively handles questions with multiple parts or deeper meaning.	Basic understanding but overlooks key parts of the question.	Covers the question inaccurately or ignores its detailed parts.
Response Structure	Is the model's output organized and easy to understand?	The response flows logically with a clear and structured format.	The content is fine, but the presentation is not organized.	The response is scattered and not easy to follow.

Truthfulness is a foundational metric which measures whether the chatbot's responses are factually correct and aligned with the source material (Es et al. 2023). Maintaining truthfulness is essential in the RFX process, where incorrect information can result in expensive mistakes during decision-making. Answer accuracy assesses how precisely the chatbot addresses the specific question asked. Evaluating answer accuracy ensures that the chatbot not only accesses the right information but also understands the intent behind the question. Contextual accuracy evaluates whether the chatbot can maintain the correct context between the question, the data provided, and the answer delivered. This criterion is important because even accurate facts, when placed outside their intended context, can potentially lead the procurement professional to make the wrong decisions. Handling complex queries focuses on the chatbot's ability to manage nuanced prompts that require deeper reasoning and synthesis of multiple pieces of information. In strategic sourcing and supplier assessment, queries are rarely simple; thus, this capability

determines how practical the model is performing the tasks it is intended to do. Finally, response structure focuses on how well the answers are organized, coherent, and easy to follow. Organizing a response allows procurement professionals to quickly understand and use the insights for future RFX events.

Finally, the methodology addresses output improvements, by refining the knowledge base, customizing and improving system prompts and the prompt library, and revisiting data preparation steps (e.g. metadata and data transformation). This optimization enhanced the output quality of both the proof of concept and the GenAI Platform, while supporting the scalability of the solution and ensuring alignment with the overall project objectives.

To address the challenge of embedding generative AI effectively in the RFX process, we developed an RFX Knowledge Intelligence Framework. This framework aligns knowledge upload and knowledge retrieval with the five core phases of the RFX process: Initiation, Exploration, Execution, Evaluation, and Awarding. It enables the structured capture and reuse of relevant RFX data in each stage of the process and supports the project objective of leveraging historical RFX data in future RFX, drive informed decision making and compare ongoing RFX data in real time.

3.1 Data in Scope and Data Harmonization

Data in scope

Based on the methodology described above, two data sets were included in the scope of this project: a synthetic dataset and a real-world dataset provided by the sponsor company.

To support the development and evaluation of the proof of concept, we created a synthetic dataset using generative AI. This approach was necessary due to the sponsor company's strict AI policy framework and the use of Pinecone as the vector database, which is currently not approved within our sponsor company's IT environment. Appendix A outlines the initial prompt used to generate the synthetic data. The dataset consists of thirty documents spanning two procurement categories and eight fictional RFP's including RFX event details, questionnaires, supplier responses, evaluation forms, and supplementary material such as annual reports. For each RFP, the responses and supporting materials from three fictional suppliers were generated. The format and structure of these synthetic documents were modeled after the actual RFX documents from our sponsor company to ensure alignment of the synthetic data with real-world formats, compatibility and seamless adoption of the solution.

To evaluate the GenAI Platform using real-world data, the dataset of an RFX event from our sponsor company was utilized. This dataset stems from an extensive ongoing RFP within a selected procurement category that met all requirements and involved over 40 participating

suppliers. The real-world dataset includes internal records from the RFX event set up and evaluation data, and external content from supplier submissions. These materials are stored within our sponsor company's inquiry management tool and span multiple formats, including PowerPoint Presentations (pptx), PDF Documents (pdf), and Excel Spreadsheets (xlsx). The inquiry management tool organizes and stores event details, questionnaires, supplier responses, and evaluations in a structured, uniform format exportable to Excel. For the purpose of this project, all data was exported and stored offline, preserving the original file formats and folder structure prior to preprocessing and data harmonization.

Data Harmonization

The review of the data in scope from our sponsor company revealed that information related to RFX events was stored in multiple formats, including structured Excel spreadsheets. To facilitate ingestion of relevant data into a RAG model and enable more effective knowledge retrieval we transformed RFX event data from Excel spreadsheets into unstructured formats in PDF and Word Documents. For this proof-of-concept we developed two python-based preprocessing scripts to transform two key components of an RFX event: General Event Information and Supplier Questionnaire Responses.

These scripts automated the parsing of the Excel spreadsheets, extracting relevant information such as event ID and title, timelines, procurement category and stakeholder information, supplier names, questions and their corresponding answers.

To ensure accuracy, we hardcoded the field mapping, based on the use of standardized templates across RFX events, which ensured a consistent, uniform structure of the input data. To teach the model about missing data and avoid null interpretation, we included missing value handling, filling blanks with placeholders such as "missing" or "no answer". The resulting output was organized following a predefined structure, with consistent formatting elements including page breaks and labeled headers to provide the model with additional information to better differentiate between suppliers.

By supplying the model with clean, context-rich and consistently structure data the quality of the embeddings and retrieval improved. The preprocessing step improved the performance of the RAG-based solution by enabling more precise search and content generation while reducing hallucinations and mismatches. Section 4.1 Comparative Analysis of Generative AI Solutions further elaborate on the advantages of this data transformation and labeling approach.

3.2 Proof of Concept Creation

For the proof of concept, we developed an AI-powered chatbot, Raffa, designed to facilitate the retrieval of RFX related insights using a Retrieval-Augmented Generation (RAG) approach. The solution was developed using Python, with the LangChain framework handling backend processing and Streamlit serving as the front-end interface.

As discussed in section 2.3.3 RAG Databases, the type of database chosen for the knowledge base is crucial for the performance of the final solution and depends on the characteristics of the data and the use case. Given that a large volume of unstructured data collected during the RFX process is in scope of the project, we chose to use a vector database, Pinecone, for its scalability and ease of integration, for the proof of concept.

To effectively retrieve relevant procurement insights, the chatbot utilizes OpenAI's text-embedding-ada-002 model to convert unstructured RFX in formats such as PDF documents and text files into vector representations. These are stored in the Pinecone vector database for similarity-based retrieval.

When a user submits a query, the chatbot converts the query into an embedding using the same OpenAI embedding model text-embedding-ada-002. It then performs a similarity search in the vector database to retrieve the most relevant document chunks. The content of the retrieved chunks, in combination with the metadata (such as RFX name, supplier, category and document type) is then assembled into a structured context string, which is passed to the Large Language Model.

The response generation is conducted by OpenAI's GPT-4o model. The model receives a customized system prompt (see Appendix B). This prompt includes instruction-based formatting to define how the model should structure its outputs, to ensure consistency, clarity and accuracy. The system prompt instructs the chatbot to consider the following when generating the responses: use of markdown formatting and bullet points where applicable, maintain a concise and professional tone, follow a step-by-step reasoning approach, and a set of few-shot examples showing how to respond to common RFX related questions. Additionally, the system prompt enforces the use of the retrieved content only and instructs the chatbot to respond "I don't know based on the current information" if no relevant data is found. This approach reduces hallucinations and improves response truthfulness and context accuracy.

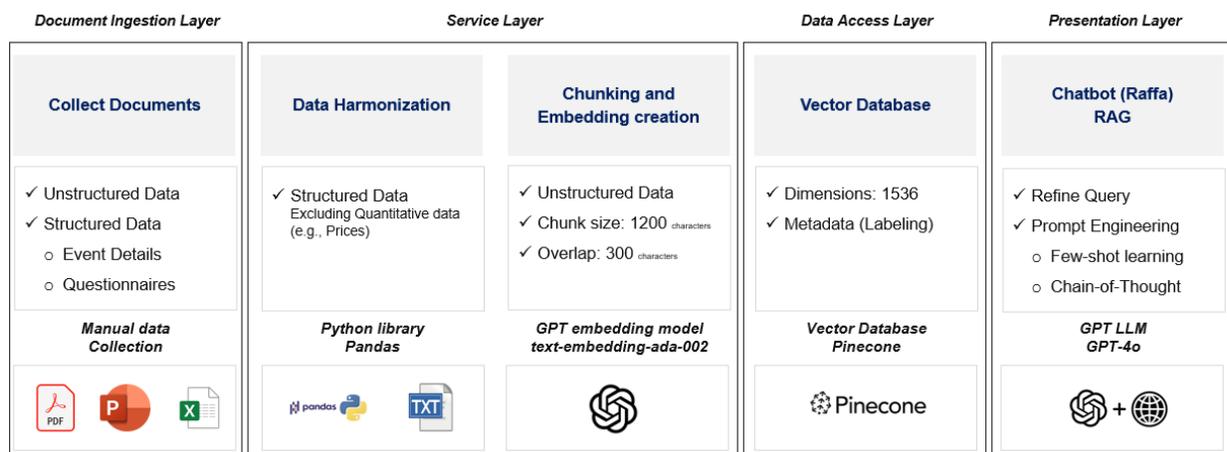
To further improve the retrieval precision and response quality, the chatbot performs automated query refinement for every user input. For the query refinement the GPT-4o model is used to rephrase the user's query, incorporating the latest conversation history and optimizing it for retrieval of the relevant content from the vector database. Appendix C shows examples of

query refinements based on prompts related to the RFX use case. If the initial similarity search returns less than two results with a similarity score above 0.75, the chatbot uses fallback refinement. The query is rephrased again, using even broader simplification and focus on keywords, to improve the likelihood of receiving relevant results. The refining process allows the chatbot to handle both vague and overly specific queries more effectively and improves the chatbots ability to retrieve optimal results for high-quality responses.

After the chatbot retrieves RFX information from the vector database and passes it as contextual input to the Large Language Model, the model formulates a natural language response based on the retrieved information. This end-to-end workflow enables the chatbot to provide contextually relevant and accurate responses, supporting more informed procurement decision-making by incorporating historical insights.

Figure 3 illustrates the proof-of concept architecture, highlighting the four core layers: Document Ingestion, Service Layer, Data Access Layer and Presentation Layer.

Figure 3. Proof-of-Concept Architecture



The chatbot performance was evaluated based on synthetic data, because our prototype was not operating behind our sponsor company’s firewall. Additional information about the dataset is provided in section 3.1 Data in Scope and Data Harmonization. The accuracy of the output was evaluated through a set of standardized questions aligned to the RFX stages defined in the RFX knowledge intelligence framework (see section 4.2.1 Embedding Generative AI in the RFX Process) and benchmarked against five key metrics defined in the evaluation framework: truthfulness, answer accuracy, contextual accuracy, handling of complex queries, and response structure.

Throughout the iterative development phase, several enhancements were made to improve the performance of the model. Initially, the chatbot operated on unstructured RFX data alone, without leveraging structured metadata or prompt engineering. While this provided a functional baseline, we observed limitations in retrieving precise and contextually relevant information. To address these limitations, we assessed the impact of adding structured metadata containing RFX event attributes, category information and supplier details to the embeddings as well as prompt engineering techniques. The impact of these enhancements on the response quality was assessed using the five evaluation metrics outline in the evaluation framework.

3.3 GenAI Platform Evaluation

The GenAI Platform which has been developed by our sponsor company is evaluated in a two-step approach.

1. Output Evaluation: Analyzing the GenAI platform's performance using real data.
2. Performance Comparison: Assessing the GenAI platform against the chatbot Raffa using synthetic data.

The evaluation follows an approach similar to that used to evaluate Raffa: leveraging the response evaluation framework. The assessment is intended to provide a comprehensive understanding of the platform's capabilities and limitations in both real-world and controlled scenarios.

4 RESULTS AND DISCUSSION

This chapter presents the results of the evaluation of Raffa, with a focus on the retrieval performance and quality of the responses. The analysis explores how well Raffa delivers accurate, contextually relevant and structured answers to a range of RFX-related questions. Here we also compare the performance of the GenAI Platform provided by the sponsor company, and Raffa's performance against the GenAI Platform, using a standardized set of prompts and the evaluation metrics defined in the evaluation framework. Key differences in system flexibility, retrieval precision, user friendliness and deployment readiness are discussed.

To support the integration and organizational adoption of generative AI in the RFX process, we introduce an RFX Knowledge Intelligence framework which maps how knowledge is created, captured and leveraged across the RFX lifecycle. The framework highlights key objectives, data input and knowledge flow at each stage of the RFX process and serves as a guide for embedding a generative AI solution into the RFX process.

4.1 Comparative Analysis of Generative AI Solutions

To compare our chatbot results, we established a structured evaluation framework based on five evaluation criteria as described in Table 3. Each criterion was assessed concerning the responses generated by the chatbot Raffa and the GenAI Platform as well as the information previously loaded into the respective knowledge bases. The selected criteria are: truthfulness, answer accuracy, contextual accuracy, handling of complex queries, and response structure.

The evaluation prompts were designed to assess each solution's ability to support activities across the RFX lifecycle including:

- Extraction of details from previous RFX to inform the creation of new requisitions.
- Identification of suppliers with specific capabilities.
- Generating supplier overviews.
- Summarizing and comparing supplier responses.
- Benchmarking evaluation outcomes and highlight strengths and weaknesses of suppliers.

Section 4.1.1 Proof of Concept Assessment and section 4.1.2 GenAI Platform Assessment contain a selection of prompts and generated responses. Appendix D contains the full list of prompts considered for the output evaluation.

4.1.1 Proof of Concept Assessment

The evaluation of the proof-of-concept chatbot, Raffa, was conducted using synthetic RFX data to evaluate the chatbot's retrieval and response generation capabilities. The system initially processed unstructured RFX documents divided into chunks of 1,200 characters and an overlap of 300 characters, without metadata labeling. This baseline approach demonstrated some limitations in retrieving precise information, as the absence of structured references made it difficult for the system to differentiate between similar RFX events. To enhance retrieval accuracy, manual metadata labeling was introduced. The following metadata was added to the Pinecone embedding: RFX ID, RFX Name, RFX Type (RFI, RFP, RFQ), Year, Category ID, Category Name, Supplier Name, Document Type. Adding structured metadata improved searchability and retrieval relevance, allowing Raffa to return more precise responses based on procurement professionals' queries. The comparison, however, also uncovers some limitations of the solution. While we have not encountered any incorrect information provided by the chatbot, the chatbot answered "*I don't know based on the current information.*", in response to some queries for which the data was available in the database. This can be explained through the system prompt: "If the context does not contain the answer, say: "I don't know based on the current information."" (see Appendix B).

The chatbot's response quality was evaluated by testing a set of procurement-related queries connected to the five stages of the RFX process, before and after implementing metadata and prompt engineering techniques. Table 4 illustrates that the introduction of structured metadata, such as RFX name, RFX type, RFX year, supplier name and document type in combination with prompt engineering significantly improved the accuracy of retrieved responses. In the proof of concept, the metadata was added manually, however metadata could be introduced through machine learning to make this improvement scalable as discussed in section 2.3.5 Data Harmonization.

In the first question, corresponding to the Initiation stage, the chatbot retrieved the six requested questionnaire categories from both the baseline (BL) and the advanced solution (AL). In both cases the answers are precise and display a clear structure aligned with the information uploaded in the questionnaires. The answer accuracy for AL is higher than for BL; AL offered more detailed information for each category, while BL only displays a selection of questionnaire categories.

In the second question, which relates to the exploration stage, both the BL and AL responses are accurate and truthful. However, the AL response is better structured and handles complex queries more effectively, as it includes details such as the name of the RFP in which suppliers participated. This is particularly valuable for procurement professionals, as it provides additional context to validate the accuracy of the answer.

The third question seeks common responses describing R&D capability among suppliers. The BL response provides a generalized synthesis of typical R&D responses based on data from the knowledge base and demonstrates the tool's capability to summarize patterns. However, the response is pretty broad and could apply to almost any domain. In contrast, the AL response provides examples of supplier specific responses and includes quantitative data, which when validated against the information in the documents, proves to be accurate and truthful. Both responses are well structured and easy to read. This comparison demonstrates that the labeling process and the fine-tuning of prompts improves the tool's ability to handle complex queries and retrieve more specific answers.

Table 4. Output Assessment of Proof of Concept

Prompt	Results before Labeling	Results incl. Metadata and Prompt Engineering
<p>Initiation Stage</p> <p>List the categories of the Questionnaire used in RFPs for Augmented Reality and Cybersecurity Solutions for healthcare.</p>	<p>Correct answer: “1. Business Philosophy, 2. Software Compliance & Cybersecurity Readiness, 3. ESG (Environmental, Social, and Governance), 4. Company Background, 5. Technical Capabilities, 6. Deployment & Support Services”</p>	<p>The categories of the Questionnaire used in Requests for Proposals (RFPs) for Augmented Reality and cybersecurity solutions in the healthcare industry are:</p> <ul style="list-style-type: none"> • Business Philosophy • Software Compliance & Cybersecurity Readiness (for Augmented Reality RFP) • Software Certification & Regulatory Compliance (for Cybersecurity RFP) • Environmental, Social, and Governance (ESG) • Company Background • Technical Capabilities • Deployment & Support Services (for Augmented Reality RFP) • Capabilities (for Cybersecurity RFP) • SLA & Support Commitments (for Cybersecurity RFP)
<p>Exploration Stage</p> <p>Which suppliers have experience with conducting work related to Augmented Reality Solutions?</p>	<p>The suppliers listed in the context with experience related to Augmented Reality Solutions are Company A, Company B, and Company C.</p>	<p>The suppliers that have experience and have been evaluated for their work related to Augmented Reality solutions in the healthcare sector in the context of RFP 2024_003 are:</p> <ul style="list-style-type: none"> • Company A • Company B • Company C <p>These suppliers participated in the RFP titled "Augmented Reality Solutions for Surgical Training" and their responses were evaluated.</p>

Table 4 (Continued). Output Assessment of Proof of Concept

Prompt	Results before Labeling	Results incl. Metadata and Prompt Engineering
<p>Execution Stage</p> <p>What are common responses to the question: 'Describe your R&D capacity'?</p>	<p>Common responses to the RFP question 'Describe your R&D capacity' include:</p> <ol style="list-style-type: none"> 1. The scope of R&D, which may encompass multiple global centers that integrate various fields such as medical physics, embedded systems, and neural networks to enhance the product pipeline. 2. The composition of the R&D team, which might include a mix of engineers and data scientists specializing in areas like embedded systems, UX design, and machine learning. These answers highlight the breadth and depth of R&D capabilities, showcasing the integration of cutting-edge technology and expertise to drive innovation. 	<p>Based on the current information, typical responses from suppliers regarding their research and development (R&D) capacity for Augmented Reality solutions in healthcare include:</p> <ul style="list-style-type: none"> • Company A reports that 40% of their workforce is dedicated to R&D across AR development, simulation design, and data analytics. • Company B invests 20% of revenue into R&D and maintains AR research partnerships with five leading universities. • Company C states that 30% of their team works in R&D, supported by funding from two NIH grants and a DARPA research pilot.
<p>Evaluation Stage</p> <p>Provide an overview of the number of employees reported by all suppliers participating in the RFP. Use data from their submitted company profiles or supporting documents.</p>	<p>"I don't know"</p>	<p>Based on the current information:</p> <ul style="list-style-type: none"> • Company A has 180 employees. • Company B has 300 employees. • Company C has 95 employees.

In the fourth question related to the Evaluation stage, we tested the chatbots ability to extract and aggregate numeric values across multiple suppliers. We found that the advanced solution outperformed the baseline. BL responded "I don't know". While the model does not hallucinate when no retrievable data is found, the solution shows low performance across all metrics, especially answer and contextual accruacy, as it is failing to retrieve available information. On the other hand, the AL response is accurate in addressing the question, presenting a three-point structure with a clear and precise answer about the number of employees for each supplier. This demonstrates that the use of metadata and prompt engineering in our

model enables the handling of complex queries and delivers truthful answers with high contextual and answer accuracy.

Finally, the results show that adding metadata and introducing prompt engineering techniques, leads to more precise responses, higher answer and contextual accuracy and a clearer structure for simple and complex queries. Compared to the BL version, the AL responses demonstrated higher truthfulness, better handling of complex queries, improved contextual accuracy, and stronger answer structures across all stages.

Despite some limitations that were revealed during the evaluation phase, the proof of concept demonstrates the strong potential of integrating embedding models, vector databases, and large language models within a Retrieval-Augmented Generation (RAG) framework to enhance RFX knowledge retrieval. While the current solution only supports unstructured data, posing a challenge especially for Requests for Quotation (RFQ), that primarily rely on prices and structured data, it effectively addressed key challenges such as limited access to historical knowledge. By leveraging these AI-driven techniques, the solution effectively addressed key procurement challenges related to historical data accessibility, supplier evaluation, and knowledge retention.

4.1.2 GenAI Platform Assessment

Our sponsor company's generative AI platform consists of two parts: a knowledge database, comparable to the vector database used in the proof of concept, and a GenAI Chatbot that allows users to retrieve information that is stored in the knowledge database.

While setting up the knowledge database, it became apparent that it faces several usability and reliability issues that make it difficult to use effectively. Uploading documents is often a tedious process, files frequently fail to upload, and there's no option to implement a customized folder structure to organize content. When trying to upload more than a few documents at once, the tool either crashes or displays vague and sometimes incorrect error messages, making it hard to tell what has been uploaded. Occasionally, uploaded files don't appear at all, further reducing confidence in the system. The page navigation is inconsistent, and page changes often don't work properly, with files from the previous page still being shown. Useful features include the tool's ability to identify duplicate content even when file names differ, helping avoid redundant uploads as well as the reference to source documents which lead to increased user trust. Overall our qualitative assessment shows that the tool is not user-friendly, lacks key organizational features, and is not scalable in its current state. These persistent issues significantly impact user trust and make it challenging to rely on the solution for managing content efficiently.

As discussed in section 3.3 GenAI Platform Evaluation, the overall solution performance evaluation follows a two-step approach. In this section, we are discussing the output evaluation following the evaluation framework to analyze the GenAI platform's performance using real data.

We started our evaluation using real data from previous RFX proposals involving more than 40 suppliers. We assessed the responses based on the five metrics defined in Table 3 Evaluation Framework. The following findings were observed:

First, we prompted for lists, for example: "Provide me the list of X suppliers" or "Give me the complete list of suppliers.", the solution was not capable of listing the complete list of suppliers, despite them being stored in one document. This shows that the GenAI platform cannot retrieve the full context, leaving the response incomplete. Next, we asked the tool to create a list of suppliers that answered "X" to a specific question. The tool was not capable of doing it, which means the GenAI platform cannot handle complex queries. Finally, we prompted specific yes/no questions, such as: "Did the majority of suppliers answer YES or NO to question Y?" In both cases, the system responded incorrectly. It answered "No" when asked what most suppliers responded to a Yes/No question. It also answered "No" when the majority responded "Yes" (yes is true / no is wrong). These findings show that the responses lack truthfulness and do not meet the required level of accuracy.

For this reason, we decided to assess the solution based on a synthetic dataset (see section 3.1 Data in Scope and Data Harmonization) that we can preprocess, thereby improving the data quality to understand the effect of data harmonization, and to refine our prompts before proceeding with the second approach of the evaluation.

To test the GenAI platform, we applied the same logic used to evaluate the results of the chatbot Raffa. As shown in Table 5, we selected specific questions corresponding to each stage of the RFX process, illustrated in Figure 4. It is important to highlight that for these tests, we used the same synthetic data and the same prompts as for Raffa. However, unlike Raffa, which has access to the metadata of each vector in Pinecone, in this case, we had to perform labeling for each document. This was necessary because we could not modify the structure or access the programming layer of the GenAI tool provided by the sponsor company.

Table 5. Output Assessment of GenAI Platform

Prompt	Result
<p>Initiation Stage List the categories of the Questionnaire used in RFPs for Augmented Reality and Cybersecurity Solutions for healthcare.</p>	<p>The categories of the Questionnaire used in RFPs for Augmented Reality and cybersecurity solutions for healthcare are as follows:</p> <ul style="list-style-type: none"> - Section 1: Business Philosophy - Section 2: Software Compliance & Cybersecurity Readiness - Section 3: Environmental, Social, and Governance (ESG) - Section 4: Company Background - Section 5: Technical Capabilities - Section 6: Deployment & Support Services
<p>Exploration Stage Which suppliers have experience with conducting work related to Augmented Reality Solutions?</p>	<ul style="list-style-type: none"> - Company A has experience in Augmented Reality Solutions, specifically focusing on surgical training and developing co-designed AR solutions with surgeons. - Company B also has experience in Augmented Reality Solutions for surgical training, offering an adaptive AR surgery trainer that changes complexity based on user performance.
<p>Execution Stage What did Company B respond to the question: 'Describe your R&D capacity'?</p>	<p>Company B invests 20% of its revenue into research and development.</p> <ul style="list-style-type: none"> - The company maintains AR research partnerships with five leading universities to enhance its R&D capacity.
<p>Evaluation Stage Provide an overview of the number of employees reported by all suppliers participating in the RFP. Use data from their submitted company profiles or supporting documents.</p>	<p>I am sorry, I don't have the information to address your question.</p>

In the first query related to the Initiation stage, the response was truthful. The response structure was clear, displaying the six categories and even indicating the number of each section. Additionally, the response provided the source from which the information was retrieved. However, the answer is missing expected components and does not provide categories of the Cybersecurity Solutions Questionnaire, therefore the answer accuracy is rated medium.

In the subsequent query related to the Exploration stage, the data contains three suppliers with experience in activities related to Augmented Reality Solutions. However, the solution only retrieved two of the three suppliers. While truthful, the answer lacked accuracy. The structure of the response was again satisfactory and the context accurate.

The third prompt related to the Execution stage was more specific and intended to test the ability to answer targeted questions from the questionnaire for a particular supplier. The system

generated truthful and accurate answer in terms of context and answer, validated against the source.

For the fourth inquiry related to the Evaluation stage, the task was to retrieve information not just from one supplier, but from the documents of each supplier. As shown in Table 5, the system responded with *“I am sorry, I don’t have the information to address your question.”* demonstrating that the GenAI platform was not able to handle complex queries effectively to deliver a clear answer. However, it managed to avoid hallucinations by not fabricating any information.

In conclusion, we can see that the GenAI platform performs well when prompted for general questions and summarization. When we tested specific questions from different documents, the tool showed a limitation in handling complex queries.

4.1.3 Solution Performance Comparison

Table 6 presents the result of the performance comparison of both generative AI solutions, Raffa and the GenAI Platform. To ensure a fair comparison, both solutions were trained on the same synthetic dataset (see section 3.1 Data in Scope and Data Harmonization).

Table 6. Result comparison for Synthetic Data - POC Chatbot and GenAI Platform

Category	“Raffa” Chatbot (POC)	GenAI Platform
Truthfulness	High	High
Answer Accuracy	Medium	Medium
Contextual Accuracy	High	Medium
Handling Complex Queries	Medium	Weak
Response Structure	High	Medium

Both solutions demonstrated a high degree of truthfulness and performed well in identifying suppliers that met specific requirements (e.g., capabilities and past RFX participation) and summarizing general supplier information. However, notable differences were observed across other key differences. Raffa achieved higher contextual accuracy which is grounded in the metadata-rich content retrieved from Pinecone. Despite our labeling attempts within key documents, the GenAI Platform lacks important metadata which leads to decreased contextual accuracy.

We further observed both solutions initially struggling with prompts that required filtering by a specific timeframe. The chatbot Raffa correctly identified RFX events within the requested year and was programmed to respond with “I don’t know” when no relevant information was

available. The GenAI platform underperformed in time-specific queries, producing incorrect answers four times. Notably, adding a document summarizing key RFP details, such as RFP type, RFP ID, and RFP year, improved its accuracy, allowing it to return the correct results when prompted about an RFP from a specific year.

The handling of complex queries was a challenge for both solutions. By rephrasing the prompts for complex queries, we observed that the responses of Raffa improved. Raffa was benefitting from prompt engineering including step-by-step reasoning, few-shot examples and query refinement logic.

Finally, we observed that the response structure from Raffa was superior compared to the GenAI Platform, using markdowns bullet points and logical flow. During the conversation, the chatbot furthermore incorporated key information such as RFX title and category in the answers which helps with connecting responses and following up on specific categories or RFX events.

These findings suggest that a procurement adapted and customized RAG solution, which allows customization related to metadata management, prompt engineering and parameter adjustments, offers advantages in contextual relevance, structure and adaptability.

In addition to the performance evaluation, a qualitative comparison was conducted to assess the potential for customization, user experience and deployment readiness of the two generative AI solutions. The GenAI Platform, while already embedded in our sponsor company's IT environment, was not deployed for RFX intelligence use cases. The platform offers limited flexibility and makes it challenging to adapt to procurement-specific applications and the data upload process remains challenging. In contrast, the custom-built chatbot Raffa, which was designed to offer greater flexibility, supports an interchangeable vector database and is highly adaptable, allowing for prompt engineering, parameter control and query refinement techniques.

4.2 RFX Knowledge Intelligence Framework

One of the challenges to successful generative AI implementation which we have identified in the course of the project is the lack of a clear operationalization path. To address this, we propose an RFX Knowledge Intelligence Framework, designed to address this lack and provide a structured approach for the application of generative AI in the inquiry management process. The framework serves as a guide for procurement teams in various industries to manage knowledge effectively across categories, regions and RFX events to maximize the value of historical data and the benefits of generative AI in RFX Knowledge Intelligence.

This framework focuses on two key use cases:

- Use Case 1: Leveraging historical data to inform and enhance the design and execution of future RfX events.
- Use Case 2: Evaluating data collected during ongoing RfX to enable faster, more consistent, and insight-driven RfX assessments.

4.2.1 Embedding Generative AI in the RfX Process

The RfX process can be outlined in five stages: Initiation, Exploration, Execution, Evaluation, and Awarding. Each of these stages presents distinct opportunities for generative AI to generate insights and enhance the RfX process. Figure 4 illustrates when and how knowledge can be both gained and uploaded during each stage to create a searchable, structured foundation that enables procurement professionals to leverage existing knowledge more effectively.

Figure 4. Generative AI Application in the RfX Process

	Initiation	Exploration	Execution	Evaluation	Awarding
Objective	- Create Requisition	- Supplier Shortlist - Know Capabilities	- Comparable Proposals	- Preferred Supplier List - SLA/Contractual terms	- Supplier is awarded
Gain Knowledge	Requirements for similar requests from previous RfX Type of work required for similar requirements	Potential suppliers Bidding history Winning history Market Capabilities	RFI Status RfX Availability RfX Structure RfX Questionnaire Common Questions	Supplier Experience Supplier Scores RfX Benchmark	
Upload Knowledge		RfX Questionnaire RfX Data received	RfX Questionnaire RfX Data received	Evaluation Result	Awarded Supplier

The first stage, Initiation, requires the participation of internal users and procurement professionals in identifying business requirements and defining the scope of work. Sometimes in this part of the process, the internal users have an idea of needs but do not have all the information necessary to structure a requirement. For this reason, they can pull insights from previous proposals or requirements through the GenAI tool. Nevertheless, procurement professionals evaluate the maturity of requirements and decide what RfX process to follow.

In the second stage, Exploration, procurement professionals decide to launch a Request for information (RFI) to get a supplier discovery and market research based on the company's needs. The objective in this stage is to get supplier shortlist capabilities, the purchaser not only gains the shortlist from the market but also pulls insights from previous RFIs or market strategies for procurement categories through the Generative AI tool. Furthermore, purchasers can upload

information with flags to have warnings about the procurement process. For instance, it was incomplete or the RFI considered only suppliers from a specific country.

The third stage is Evaluation, in which procurement professionals decide to launch a Request for proposal (RFP). It is important to have standard specifications to receive comparable proposals from the market. Also, purchasers can set criteria to evaluate proposals such as quality, cost, delivery, and flexibility. At this point, they can compare current proposals and previous proposals based on the criteria for uploading new proposals to the Generative AI platform and using prefix prompts to get the right insights. In addition, they can upload the results of their evaluations with flags as explained above.

The fourth stage is Execution in which procurement professionals and internal users participate. Purchasers make a Request for quotations (RFQ) from the shortlisted supplier with homologated requirements by internal users. After the negotiation process, the goal is to allocate the goods and services to the preferred supplier following the strategy of each category and the best commercial conditions for the company. Generative AI tool supports procurement professionals with insights such as last prices and payment terms. Also, procurement professionals can upload information on the list of preferred vendors and their capabilities.

Finally, in the fifth stage, Awarding, the selected supplier is awarded as a result of the evaluation. For the purpose of RFX Knowledge Intelligence, the awarding process foresees knowledge retention by storing the collected data in a knowledge database. Retaining targeted, accurate, and clean knowledge in a structured way is the most important stage to improve knowledge management in the organization.

Use Case 1 applies primarily to the Initiation, Exploration and Execution stages. At these stages, generative AI helps to identify relevant suppliers, retrieve bidding and award history, and provide visibility into prior RFX structures and RFI statuses as well as frequently asked supplier questions, enabling more informed event design and supplier engagement. Use Case 2 is most impactful in the Evaluation stage, where generative AI synthesizes information from questionnaires and supporting documentation. This allows procurement professionals to access relevant insights faster and conduct more comprehensive and consistent supplier comparisons.

Though not in the scope of this project, we would like to point out that generative AI can be leveraged beyond the RFX process. Some use cases are verifying whether the selected supplier is an existing vendor, assessing past supplier performance, and determining whether contractual agreements are in place. By leveraging supplier data beyond the data collected in RFX informed decision-making can be ensured before awarding the contract.

4.2.2 RFx Prompt Overview

Section 2 highlights the significance of prompts and prompt engineering. Maximizing the benefits of generative AI relies on formulating the right questions. To achieve high quality outputs, we mapped the RFx stages with targeted prompts, as illustrated in Figure 5, to inspire the user and guide them through the process. These questions are designed to be applicable across industries but primarily focus on the RFI and RFP process for indirect materials and services. As a starting point, the prompt library presented in Figure 5 was developed with insights from both business and academic viewpoints, including the support from a Category Director of our sponsor company and is intended to assist users in formulating effective prompts to navigate the RFx process efficiently.

Figure 5. RFx Prompt Library and Allocation

Initiation	Exploration	Execution	Evaluation	Awarding
When have we bid out this type of work/project/service before?	What suppliers have bid on / were awarded this type of work before?	What type of RFx was executed previously (standard bid, reverse auction, etc.)	What suppliers have high experience with requirement 1 in markets A/B/C?	Is this an existing supplier?*
What RFx events have been previously conducted for this type of work?	What RFxs has Supplier X bid on / won in the last Y timeframe?	What questions are typically asked as part of an RFx for this type of work? What was the scoring criteria for similar RFx?	Which suppliers have experience in all X regions in all requirements?	What is the Supplier Performance?*
What questions are typically asked as part of an RFx for this type of work?	Which suppliers have experience conducting work X in in Region A?	What was the total duration of a similar RFx?	Which are the 3 suppliers with the most favorable commercial conditions?	Does a frame agreement / contract exist?*
What are the terms and conditions of RFx for this type of work?	Of suppliers with experience conducting work X, which ones have experience with work Y?	What are the most common questions from suppliers for this type of RFx?	Which suppliers scored 9 on the question X?	What is the expiration of existing agreements (if any)?*
What is the latest questionnaire or template of this type of RFx event?	Are there existing/expired contracts for this type of work?*	What are the most common answers for question X?	How did suppliers score in previous RFx?	What are current Terms and Conditions?*

*Out of scope for the purpose of this project

The Initiation stage focuses on understanding the business requirements. By leveraging historical insights procurement professionals, in collaboration with their internal customers, establish a strong foundation for an RFx event. Reviewing past bidding activities and identifying previous RFx events conducted for similar work help narrow down the scope. By analyzing commonly asked questions, examining the standard terms and conditions, and accessing the latest RFx questionnaire, additional insights can be derived to formulate the requirements and prepare the requisition. By consolidating historical information, our sponsor company can ensure a well-informed and structured approach to initiating the RFx process.

The Exploration stage focuses on analyzing past supplier performance and supplier capabilities to create a supplier shortlist. The prompts related to this process step support user in identifying suppliers that have previously bid on or been awarded similar types of work, tracking RFX participation history for suppliers in scope, and assessing supplier experience for specific types of work and/or regions. Additionally, GenAI can help to evaluate which suppliers with relevant experience in one area also have expertise in related work. This stage helps refine supplier selection and inform strategic sourcing decisions.

The Execution stage focuses on understanding past RFX events and refining the approach for the current event. In this stage prompts target the identification of RFX types previously executed (e.g., standard bid, reverse auction) and analyzing commonly asked questions in the RFX, scoring criteria, and total duration of similar RFX events. Additionally, the most frequent supplier inquiries can be identified to further refine the RFX. These insights can help with the RFX planning and timeline, speed up the overall RFX process and improve the clarity leading to an improved supplier experience. In the Execution Stage, Generative AI capabilities can furthermore be used to assess the data provided by suppliers in an ongoing RFX, providing insights into most common answers or summarizing supplier capabilities and other requested information.

The Evaluation stage focuses on assessing supplier performance. Generative AI can speed up the evaluation process by supporting procurement professionals in identifying suppliers with high experience in specific markets, those meeting all regional and requirement criteria, and the top suppliers with the most favorable commercial terms. Additionally, after the evaluation was finalized, Generative AI can support with examining supplier scores on key questions and compare scores from past RFX event to support decision-making.

The Awarding stage, which includes contract creation and supplier onboarding, is out of scope for this use case and is therefore not addressed in this report.

5 CONCLUSION AND OUTLOOK

The goal of this capstone project was to create an AI-driven proof-of-concept that addresses knowledge silos in the inquiry management process and delivers value to our sponsor company's global procurement organization. In a company that manages thousands of RFX events with global suppliers across multiple business sectors, category teams and regions, the need for structured knowledge management is evident. With our project, we delivered a solution that enables structured and efficient knowledge storage and retrieval, turning tacit knowledge into explicit, shareable insights.

Our work demonstrates the potential of Retrieval-Augmented Generation (RAG) and generative AI to address the challenge procurement functions are facing, to retain and leverage organizational knowledge for informed-decision making and strategic supplier selection. By developing our chatbot, Raffa, we delivered a proof-of concept that bridges disconnected teams and RFX events, offering a user-friendly interface and supporting procurement professionals with answering inquiry related questions based on historical RFX data, supplier responses and supporting information collected during the inquiry process.

Throughout the project, we assessed how the responses provided to the procurement professionals could be improved, with a focus on truthfulness, answer accuracy, contextual accuracy, handling of complex queries, and response structure. We identified critical success factors including data preprocessing, the use of metadata and prompt engineering as well as limitations of large language models related to knowledge management in large organizations. During the project, we considered the limitations our sponsor company is facing, especially in terms of volume, format and storage location of the data in scope. We developed a systematic approach and preprocessing techniques to broaden the scope of usable data. In particular, we transformed structured data into unstructured formats where it was suitable, to enable the ingestions by our RAG model. This approach enhances the transferability and scalability of our solution as it widens the usable data and is not limited to a specific data source.

While choosing the right technology was a key focus, it is equally important to understand the business context and the way procurement teams operate. Generative AI adoption depends on aligning the tool's capabilities with the user demands and real business needs requiring a thorough understanding of existing processes and a methodological approach to determining where and how AI can accelerate results. Provided that AI literacy varies across teams, it is essential to provide structured guidance to ensure users can apply the tool efficiently and evaluate its outputs critically. To support adoption, we developed an RFX Knowledge Intelligence Framework that provides a structured way to capture and access knowledge throughout the RFX process to deliver value in the daily work of procurement professionals, and making its integration into existing workflows both practical and impactful.

Generative AI continues to evolve at an incredible speed and our project lays the foundation for a scalable tool which can be adopted across categories and regions. During the course of our work, rapid advancements in generative AI opened up a lot of new opportunities for application within this use case. Table 7 highlights the AI technologies that align closely with the objectives of this capstone project.

Table 7. Future Use Cases of Generative AI in Procurement

Impacts	Use Case	Description
Agentic reasoning	Procurement Assistants	Intelligent assistant, providing strategic insights, historical data analysis and decision-making support based on real-time enterprise data to procurement staff in a dynamic and context-aware manner, going beyond the current models, which offer static responses and require user input for each query. (Polk et al., 2024)
Multimodality	Category Intelligence	Analyze data, summarize and draw insights from various formats such as text, images, and video. (Polk et al., 2024)
Hybrid RAG	Category Intelligence	Combines the power of vector databases and knowledge graphs in Retrieval Augmented Generation (RAG), enabling the extraction and analysis of both structured and unstructured data, and providing procurement professionals with hard metrics alongside contextualized, qualitative insights. (Sarmah et. al., 2024)
AI Agents	Semi-Autonomous Procurement	Analyze a wide range of data (e.g., business needs, potential suppliers, supplier performance history, financial stability and market reputation) to recommend the best suppliers and initiate downstream activities, such as drafting a contract, starting the negotiation process and populating ERP systems. (Polk et al., 2024)

Among these technologies, HybridRAG has significant potential for knowledge management in the RFX Process. By expanding the data scope and integrating, structured and unstructured data, it enables improved quantitative comparison, for example related to pricing, supports trend tracking and improved analytics while contextualizing and offering qualitative insight. While structured data integration was not within scope of this project, we strongly recommend that our sponsor company explores the integration of structured data in the RAG solution to further strengthen the solution and its benefits and provide procurement professionals with a powerful tool for procurement related decision-making capabilities.

Our capstone project demonstrates the great potential of generative AI in procurement. By building a scalable, AI-powered solution that provides procurement professionals easy access to the knowledge available across RFX events in the organization, our project provided value to our sponsor company's procurement organization. It sets the stage for a time where knowledge is no longer siloed but continuously captured, shared and used at scale rather than being isolated. In addition to providing a technical solution, our work enables procurement professionals to make more informed and faster decisions, which ultimately improves the operational and strategic sourcing efficiency.

REFERENCES

- Behairy, A. (2023). *RAG: Vector Databases vs Knowledge Graphs?*. Medium. Retrieved 03-04-2025, from <https://medium.com/@ahmedbehairy/rag-vector-databases-vs-knowledge-graphs-f22697b1a940>
- Belcic, I. (2024). What is RAG (retrieval augmented generation)? Retrieved 11-19-2024, from <https://www.ibm.com/think/topics/retrieval-augmented-generation>
- Boonstra, L. (2024). *Prompt engineering: Best practices for enterprise AI*. Kaggle. Retrieved 11-19-2024, from <https://www.kaggle.com/whitepaper-prompt-engineering>
- Chandrasekaran, A., Ramos, L. (2024). *Hype Cycle for Generative AI, 2024*. Gartner.
- CPOstrategy. (2024). *AI in procurement playbook*. CPOstrategy Media. Retrieved 11-25-2024, from <https://cpostrategy.media/ai-in-procurement-playbook/>
- Dalkir, K. (2011). *Knowledge management in theory and practice* (2nd ed.). MIT Press.
- DeepLearning.AI. (n.d.). *ChatGPT prompt engineering for developers*. DeepLearning.AI. Retrieved 11-19-2024, from <https://www.deeplearning.ai/short-courses/chatgpt-prompt-engineering-for-developers/>
- Dhamani, N., & Engler, M. (2024). *Introduction to Generative AI*. Simon and Schuster
- EPAM SolutionsHub. (2022, October 21). *What is data harmonization?* EPAM SolutionsHub. Retrieved 4-3-2025, from <https://solutionshub.epam.com/blog/post/what-is-data-harmonization>
- Es, S., James, J., Espinose-Anke, L., Schockaert, S. (2023). *RAGAS: Automated Evaluation of Retrieval Augmented Generation*. Retrieved 12-05-2024, from <https://arxiv.org/abs/2309.15217>
- Gartner. (2018). *Understanding Gartner's Hype Cycles*. Gartner Research. Retrieved 11-26-2024, from <https://www.gartner.com/en/documents/3887767>
- Gartner. (2023). *Generative AI Use Cases in Sourcing and Procurement*
- Google. (2024). *GEN AI course - Lesson 1 [Video]*. YouTube. Retrieved 11-29-2024 from <https://www.youtube.com/watch?v=kpRyiJUUFxY&list=PLqFaTlg4myu-b1PlxitQdY0UYIbys-2es&index=1>
- IBM. (2022). *Knowledge management*. IBM. Retrieved 11-29-2024, from <https://www.ibm.com/topics/knowledge-management>
- IBM Technology. (2023). *What is Retrieval-Augmented Generation (RAG)?*- Youtube. Retrieved 11-22-2024, from <https://www.youtube.com/watch?v=T-D1OfcDW1M&t=3s>
- IBM Technology. (2024). *RAG vs. Fine Tuning*- Youtube. Retrieved 11-22-2024, from <https://www.youtube.com/watch?v=00Q0G84kq3M&t=1s>

- IDC, Statista, & Various sources. (2024, May 31). *Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2023, with forecasts from 2024 to 2028 (in zettabytes)* [Graph]. Statista. Retrieved 04-04-2025, from <https://www.statista.com/statistics/871513/worldwide-data-created/>
- Ivalua. (2024). *Generative AI in Procurement – A Practical Guide*. Retrieved 12-05-2024, from <https://info.ivalua.com/reports/generative-ai-procurement-roadmap>
- Keck, M., Sommers, K. Mahendra, N., Polk, R., O’Doherty, M., Phelan, L. Mehrzai, F. (2024). *Predicts 2024: CPOs Adjust to Technology’s Impact on Procurement*. Gartner.
- LangChain. (2024). *Build a Retrieval Augmented Generation (RAG) App*. Retrieved 11-19-2024, from <https://python.langchain.com/docs/tutorials/rag/>
- Lorica, B. (2023, September 19). *The future of prompt engineering: Getting the most out of LLMs*. Gradient Flow. Retrieved 11-19-2024, from <https://gradientflow.com/the-future-of-prompt-engineering-getting-the-most-out-of-llms/>
- McKinsey & Company. (2024). *What is retrieval-augmented generation (RAG)?* Retrieved 11-19-2024, from <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-retrieval-augmented-generation-rag>
- McMillan, A. (2024). *How Gen AI is Shaping Sourcing and Procurement*. Procurement Mag. Retrieved 11-19-2024, from <https://procurementmag.com/technology-and-ai/genai-shaping-sourcing-and-procurement>
- Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford University Press.
- O’Dell, C., & Grayson, C. J. (1998). *If only we knew what we know: The transfer of internal knowledge and best practice*. Free Press.
- OpenAI. (n.d.). *Fine-tuning*. Retrieved 11-19-2024, from <https://platform.openai.com/docs/guides/fine-tuning>
- Polk R., Keck, M., Sommers, K., Curtin, C., O’Doherty, M. (2024). *How Generative AI Progress Will Shape the Future of Procurement*. Gartner.
- Sarmah, B., Hall, B., Rao, R., Patel, S., Pasquali, S., Mehta, D. (2024, August 9). *HybridRAG: Integrating Knowledge Graphs and Vector Retrieval Augmented Generation for Efficient Information Extraction*. arXiv. Retrieved 03-30-2025, from <https://arxiv.org/abs/2408.04948v1>
- Sedinkina, M. (2023, July 21). *Data harmonization using machine learning and language models*. Medium. Retrieved 4-3-2025, from <https://medium.com/datareply/data-harmonization-using-machine-learning-and-language-models-4e215da19e3f>
- Sponsor Company (2024). *What is [GenAI Platform]?* Confluence.
- Sommers, K., Keck, M., Curtin, C., Phelan, L., Mahendra, N., Paradarami, C. (2024). *Hype Cycle for Procurement and Sourcing Solutions, 2024*. Gartner.

- Wei, J., Wang, X., Schuurmans, D., et al. (2022). *Chain-of-thought prompting elicits reasoning in large language models*. arXiv. <https://arxiv.org/abs/2201.11903>
- Woyera. (2023). *Pinecone vs. Chroma: The Pros and Cons*. Medium. Retrieved 12-05-2024, from <https://medium.com/@woyera/Pinecone-vs-chroma-the-pros-and-cons-2b0b7628f48f>
- Xinghan, P. (2025). *Comprehensive Guide to Choosing the Right Database for RAG Implementation: Leveraging Elasticsearch, Vector Databases, and Knowledge Graphs*. Medium. Retrieved 03-04-2025, from [Comprehensive Guide to Choosing the Right Database for RAG Implementation: Leveraging Elasticsearch, Vector Databases, and Knowledge Graphs](https://medium.com/@panxinghan/comprehensive-guide-to-choosing-the-right-database-for-rag-implementation-leveraging-elasticsearch-vector-databases-and-knowledge-graphs) | by Pan Xinghan | Medium
- Zhou, Y. (2023). *Optimizing GenAI: Comparing Model Training, Fine-Tuning, RAG, and Prompt Engineering*. Medium. Retrieved 11-19-2024, from <https://medium.com/generative-ai-revolution-ai-native-transformation/optimizing-genai-comparing-model-training-fine-tuning-rag-and-prompt-engineering-7a7c6c65e0f0>
- Zhou, Y. (2024). *A New Paradigm in Generative AI Applications: Navigating the Power of LLM-RAG*. Medium. Retrieved 11-19-2024, from <https://medium.com/generative-ai-revolution-ai-native-transformation/a-new-paradigm-in-generative-ai-applications-navigating-the-power-of-llm-rag-f3186200302e>

APPENDICES

Appendix A

Prompt entered in ChatGPT for synthetic data creation

"I need support creating Unstructured Synthetic Data for a Database.
The use case is a database that includes RFP documents.

I want to include the following type of documents:

- RFP general information (similar to Event Details template)
- Questionnaire and Questionnaire response by supplier (similar to Questionnaire template)
- RFP Evaluation (score by category and overall, similar to Evaluation Template)
- Supporting documents by supplier (e.g. Annual Report, Certificates)

I need the following number of documents:

- At least 30 documents
- Covering at least two procurement categories
- Covering at least two RFP's per procurement category
- At least 3 participating suppliers per RFP
- 1 RFP General Information Sheet per supplier (similar to Event Details template attached)
- 1 Questionnaire response per supplier (similar to Event Details template attached)
- 1 RFP Evaluation per RFP event
- At least one supporting document per supplier (e.g. Annual Report or Certificate)

The categories and RFP's in scope should be related to an indirect material category in healthcare and technology. Make up company names and company details.

Please follow the structure of the attached templates."

Appendix B

Chatbot System Prompt

system_template = ""

You are Raffa, a highly knowledgeable assistant specialized in procurement, RFx processes, and supplier intelligence.

Your job is to answer user questions as accurately and truthfully as possible using only the provided context.

Instructions:

- ONLY use information from the provided context to answer.
- If the context does not contain the answer, say: "I don't know based on the current information."
- Be concise, clear, and professional in your tone.
- If applicable, use bullet points or a markdown table to organize information.
- Never make up supplier names, RFx IDs, or scores.
- Assume the user is familiar with procurement concepts, but may need clarification on specific details.
- If only partial data is available, answer with what's present and state what's missing.

Few-shot examples:

Question: What was the compliance score of Company B in the AR training RFP?

Answer:

- Company B scored 9.5 out of 10 for compliance.

Question: List the categories of the Questionnaire used in RFPs for Augmented Reality solutions for healthcare?

Answer:

- Section 1: Business Philosophy
- Section 2: Software Compliance & Cybersecurity Readiness
- Section 3: Environmental, Social, and Governance (ESG)
- Section 4: Company Background
- Section 5: Technical Capabilities
- Section 6: Deployment & Support Services

Question: Provide an overview of the number of employees reported by all suppliers participating in the RFP "2023-003".

Answer:

The suppliers participating in RFP "2023-003" are:

- Company A Inc: 180 employees
- Company B: 300 employees
- Company C: 95 employees

Reasoning instructions:

Answer the question based on the context below.

Follow this structured reasoning approach before answering:

1. Identify the applicable **category** based on the context.
2. List the applicable **RFP IDs and titles**.
3. Identify the **suppliers in scope** based on the RFP(s) mentioned.

After completing the steps above, provide a final concise answer in 2–3 sentences.

<BEGIN CONTEXT>
{context}
<END CONTEXT>

<BEGIN HISTORY>
{history}
<END HISTORY>

Now answer the following question based on the context above: Question: {query}""

Appendix C

Comparison table of original queries entered to “Raffa” and the refined query after application of the query refinement mechanism.

Process Stage	Original Prompt	Refined Prompt
Initiation	List the categories of the Questionnaire used in RFPs for Augmented Reality and cybersecurity solutions for healthcare?	What are the categories included in the questionnaire for evaluating Augmented Reality and cybersecurity solutions in the healthcare sector RFPs?
Initiation	What are the typical questions included in a questionnaire to evaluate Augmented Reality solutions in the healthcare sector?	What are the common questions used in questionnaires to evaluate Augmented Reality solutions specifically for the healthcare sector?
Exploration	Which suppliers have experience with conducting work related to Augmented Reality Solutions?	"Which suppliers participating in the Augmented Reality Solutions for Surgical Training RFP (2024_003) have experience in conducting work related to Augmented Reality Solutions?"
Exploration	Summarize Company C's performance based on their RFP submission, evaluation scores, and annual report. Highlight strengths, weaknesses, and overall competitiveness.	Summarize Company C's performance in the Augmented Reality Solutions for Surgical Training RFP (2024_003) using their RFP submission, evaluation scores, and annual report, focusing on strengths, weaknesses, and overall competitiveness.
Execution	What are common responses to the question: 'Describe your R&D capacity'?	"What did Company B and other suppliers say in response to the question: 'Describe your R&D capacity' in their RFP submissions?"
Execution	What did Company B respond to the question: 'Describe your R&D capacity'?	"What did Company B report regarding their R&D capacity in their RFP submission documents?"
Evaluation	What is the average supplier performance based on evaluation scores for the RFP on Augmented Reality solutions in healthcare?	What is the average evaluation score for suppliers in the Augmented Reality solutions RFP for healthcare?

Comparison table continued

Process Stage	Original Prompt	Refined Prompt
Evaluation	Provide an overview of the number of employees reported by all suppliers participating in the RFP.	Provide a summary of the number of employees reported by each supplier participating in the Augmented Reality Solutions for Surgical Training RFP (2024_003), based on their submitted company profiles or supporting documents.
Evaluation	How many employees do Company A, Company B, and Company C have?	What is the total number of employees for Company A, Company B, and Company C based on all documents submitted for the Augmented Reality Solutions for Surgical Training (RFP 2024_003)?

Appendix D

Comprehensive table of prompts and responses generated by the POC chatbot Raffa and the GenAI Platform

Prompt	Chatbot Raffa: Result before Labeling	Chatbot Raffa: Result including Metadata and Prompt Engineering	Gen AI Platform
<p>Initiation Stage</p> <p>List the categories of the Questionnaire used in RFPs for Augmented Reality and Cybersecurity Solutions for healthcare.</p>	<p>1. Business Philosophy, 2. Software Compliance & Cybersecurity Readiness, 3. ESG (Environmental, Social, and Governance), 4. Company Background, 5. Technical Capabilities, 6. Deployment & Support Services</p>	<p>The categories of the Questionnaire used in Requests for Proposals (RFPs) for Augmented Reality and cybersecurity solutions in the healthcare industry are:</p> <ul style="list-style-type: none"> • Business Philosophy • Software Compliance & Cybersecurity Readiness (for Augmented Reality RFP) • Software Certification & Regulatory Compliance (for Cybersecurity RFP) • Environmental, Social, and Governance (ESG) • Company Background • Technical Capabilities • Deployment & Support Services (for Augmented Reality RFP) • Capabilities (for Cybersecurity RFP) • SLA & Support Commitments (for Cybersecurity RFP) 	<p>The categories of the Questionnaire used in RFPs for Augmented Reality and cybersecurity solutions for healthcare are as follows:</p> <ul style="list-style-type: none"> - **Section 1: Business Philosophy** - **Section 2: Software Compliance & Cybersecurity Readiness** - **Section 3: Environmental, Social, and Governance (ESG)** - **Section 4: Company Background** - **Section 5: Technical Capabilities** - **Section 6: Deployment & Support Services**
<p>Initiation Stage</p> <p>What are the typical questions included in a questionnaire to evaluate Augmented Reality solutions in the healthcare sector?</p>	<p>I don't know.</p>	<p>Based on the current information, the questions asked to suppliers about Augmented Reality solutions in the healthcare sector include:</p> <ul style="list-style-type: none"> - Employee count and core business areas. - Recent milestones achieved in the past 3 years. 	<p>Typical questions included in a questionnaire to evaluate Augmented Reality solutions in the healthcare sector focus on various aspects of the offerings. These questions may include:</p> <ul style="list-style-type: none"> - **Business Philosophy and Alignment**: - Do you agree to follow the provided business philosophy?

		<ul style="list-style-type: none"> - Differentiation of the AR solution from competitors. - Customization options available for the AR platform. - Research and development capacity. - Average deployment timeline after contract signing. - Agreement to follow a specified business philosophy and demonstration of alignment with it. - Innovations delivered that support the specified business philosophy. <p>Updated prompt: Provide a summary of questions that have been asked to suppliers about Augmented Reality solutions.</p>	<ul style="list-style-type: none"> - How does your company demonstrate alignment with this business philosophy in your operations? - Describe one innovation your company has delivered that supports this business philosophy. - **Software Compliance & Cybersecurity Readiness**: - What compliance standards does your software follow? - How do you ensure data security for your AR system? - Do you conduct regular cybersecurity audits or penetration testing? - **Environmental, Social, and Governance (ESG)**: - What environmental initiatives has your company implemented in the past 2 years? - Describe your approach to social responsibility and community impact. - Have you received any ESG ratings or certifications? - **Company Background**: - When was your company founded, and where is it headquartered? - How many employees do you currently have, and what are your core business areas? - Describe one milestone from the past 3 years.
--	--	--	--

Comprehensive table of prompts and responses generated by the POC chatbot Raffa and the GenAI Platform continued

Prompt	Chatbot Raffa: Result before Labeling	Chatbot Raffa: Result including Metadata and Prompt Engineering	Gen AI Platform
<p>Exploration Stage</p> <p>Which suppliers have experience with conducting work related to Augmented Reality Solutions?</p>	<p>The suppliers listed in the context with experience related to Augmented Reality Solutions are Company A, Company B, and Company C.</p>	<p>The suppliers that have experience and have been evaluated for their work related to Augmented Reality solutions in the healthcare sector in the context of RFP 2024_003 are:</p> <ul style="list-style-type: none"> • Company A • Company B • Company C <p>These suppliers participated in the RFP titled "Augmented Reality Solutions for Surgical Training" and their responses were evaluated.</p>	<ul style="list-style-type: none"> - Company A has experience in Augmented Reality Solutions, specifically focusing on surgical training and developing co-designed AR solutions with surgeons. - Company B also has experience in Augmented Reality Solutions for surgical training, offering an adaptive AR surgery trainer that changes complexity based on user performance.
<p>Exploration Stage</p> <p>Summarize Company C's performance based on their RFP submission, evaluation scores, and annual report. Highlight strengths, weaknesses, and overall competitiveness.</p>	<p>Company C's performance based on their RFP submission, evaluation scores, and annual report shows both strengths and weaknesses. In terms of evaluation scores, they received a total of 30.0 out of 40, indicating that while they have strong pricing, their technical fit and compliance were not as strong as their competitors. Their annual report highlights that they are one of the fastest-growing innovators in medical augmented reality, with expansions in their simulation platform and multi-user collaboration environments. They reported a revenue of \$22M with a year-over-year growth of 14% and significant R&D investment, which shows their commitment to innovation. Their annual report also emphasizes their</p>	<p>Based on the current information, Company C participated in the RFP titled "Augmented Reality Solutions for Surgical Training" (RFx ID: 2024_003). Their evaluation scores were as follows: Technical Fit: 7.0, Pricing: 9.0, Compliance: 6.5, and Support: 7.5, resulting in a total score of 30.0 out of 40.</p> <p>Strengths:</p> <ul style="list-style-type: none"> - Strong pricing score (9.0), indicating competitive pricing strategies. - Commitment to environmental initiatives, such as transitioning to 100% recycled materials and carbon-neutral cloud services. <p>Weaknesses:</p> <ul style="list-style-type: none"> - Lower compliance score (6.5), suggesting potential areas for 	<ul style="list-style-type: none"> - Company C is recognized as a rapidly growing innovator in the medical augmented reality sector, achieving significant milestones in 2023, such as expanding its simulation platform and increasing support for nonprofit training. - Financially, Company C reported revenues of \$22 million with a year-over-year growth of 14%, and an operating profit of \$4.3 million. The company employs 95 people, with 30% of the workforce dedicated to research and development, reflecting a strong commitment to innovation. - In terms of strengths: - The company has developed gesture-enabled, multi-user AR surgical training

	<p>achievements, such as the adoption of their platform by a leading UK medical school and securing significant grants, which speaks to their credibility and potential. However, their lower compliance score in the evaluation suggests a potential area of improvement. Overall, Company C is competitive in pricing and innovation but may need to strengthen their technical fit and compliance aspects to improve their competitiveness in future RFPs.</p>	<p>improvement in meeting RFP requirements.</p> <p>Overall Competitiveness:</p> <ul style="list-style-type: none"> - Company C demonstrates a solid presence in the AR market with a focus on innovation and sustainability, but there is room for improvement in compliance to enhance their competitiveness in RFP evaluations. 	<p>platforms that enhance skill-building and collaboration.</p> <ul style="list-style-type: none"> - They have secured important grants, including two from NIH and one from DARPA, indicating trust and investment in their technology. - Company C has made strides in sustainability by ensuring all packaging is 100% recyclable and transitioning to carbon-neutral hosting. - Weaknesses are evident in their RFP evaluation: <ul style="list-style-type: none"> - Company C received a total score of 30 out of 40, indicating some areas for improvement, particularly in compliance (6.5) and technical fit (7.0). - Competitors like Company B achieved higher scores in critical categories such as technical fit and compliance, highlighting a competitive gap. - Overall, Company C demonstrates competitiveness through its innovative solutions and strong financial performance, but must address weaknesses in compliance and technical fit to enhance its standing in the market.
--	---	--	---

Comprehensive table of prompts and responses generated by the POC chatbot Raffa and the GenAI Platform continued

Prompt	Chatbot Raffa: Result before Labeling	Chatbot Raffa: Result including Metadata and Prompt Engineering	Gen AI Platform
<p>Execution Stage</p> <p>What are common responses to the question: 'Describe your R&D capacity'?</p>	<p>Common responses to the RFP question 'Describe your R&D capacity' include:</p> <ol style="list-style-type: none"> 1. The scope of R&D, which may encompass multiple global centers that integrate various fields such as medical physics, embedded systems, and neural networks to enhance the product pipeline. 2. The composition of the R&D team, which might include a mix of engineers and data scientists specializing in areas like embedded systems, UX design, and machine learning. These answers highlight the breadth and depth of R&D capabilities, showcasing the integration of cutting-edge technology and expertise to drive innovation. 	<p>Based on the current information, typical responses from suppliers regarding their research and development (R&D) capacity for Augmented Reality solutions in healthcare include:</p> <ul style="list-style-type: none"> • Company A reports that 40% of their workforce is dedicated to R&D across AR development, simulation design, and data analytics. • Company B invests 20% of revenue into R&D and maintains AR research partnerships with five leading universities. • Company C states that 30% of their team works in R&D, supported by funding from two NIH grants and a DARPA research pilot. 	<ul style="list-style-type: none"> - Common responses to the question regarding R&D capacity include the percentage of the workforce dedicated to R&D and the sources of funding for research initiatives. - Company A dedicates 40% of its workforce to R&D across various areas including AR development and simulation design. - Company B invests 20% of its revenue into R&D and collaborates with five leading universities for research partnerships. - Company C has 30% of its team working in R&D, supported by two NIH grants and one DARPA research pilot.
<p>Execution Stage</p> <p>What did Company B respond to the question: 'Describe your R&D capacity'?</p>	<p>Company B responded to the question "Describe your R&D capacity" with the following: "R&D spans three global centers, integrating medical physics, embedded systems, and neural networks into our product pipeline."</p>	<p>Yes, Company B has R&D capacity. They invest 20% of their revenue into research and development and maintain research partnerships with five leading universities.</p>	<p>Company B invests 20% of its revenue into research and development.</p> <ul style="list-style-type: none"> - The company maintains AR research partnerships with five leading universities to enhance its R&D capacity.
<p>Evaluation Stage</p> <p>What is the average supplier performance based on evaluation scores for the RFP on Augmented Reality solutions in healthcare?</p>	<p>To calculate the average performance of the suppliers based on their evaluation scores, we add up the total scores for all suppliers and divide by the number of suppliers. The total scores are as follows:</p> <p>Company A: 31.0 Company B: 36.5 Company C: 30.0</p> <p>The sum of these scores is $31.0 + 36.5 + 30.0 = 97.5$. There are 3 suppliers. The</p>	<p>Category: Augmented Reality & Cybersecurity RFP IDs and Titles: 2024_003 - Augmented Reality Solutions for Surgical Training Suppliers in Scope: Company A, Company B, Company C The average performance of the suppliers based on evaluation scores for the RFP 2024_003 is as follows:</p> <ul style="list-style-type: none"> - Company A: $31.0 / 40$ 	<ul style="list-style-type: none"> - The average performance of suppliers based on evaluation scores for the RFP on Augmented Reality solutions in healthcare can be calculated by taking the total scores of each supplier and dividing by the number of suppliers. - The total scores for each supplier are as follows: - Company A: $31.0 / 40$ - Company B: $36.5 / 40$ - Company C: $30.0 / 40$

	average performance score is $97.5 / 3 = 32.5$. Therefore, the average performance of suppliers based on their evaluation scores is 32.5 out of 40.	<p>- Company B: 36.5 / 40 - Company C: 30.0 / 40</p> <p>The average score across these suppliers is 32.5 out of 40. Company B had the highest score, indicating strong overall performance compared to the other suppliers.</p>	<p>- To calculate the average: - Sum of scores: $31.0 + 36.5 + 30.0 = 97.5$ - Number of suppliers: 3 - Average score = $97.5 / 3 = 32.5 / 40$ - Therefore, the average performance score of the suppliers is 32.5 out of 40.</p>
<p>Evaluation Stage</p> <p>Provide an overview of the number of employees reported by all suppliers participating in the RFP. Use data from their submitted company profiles or supporting documents.</p>	I don't know	<p>Based on the current information:</p> <ul style="list-style-type: none"> • Company A has 180 employees. • Company B has 300 employees. • Company C has 95 employees. 	I am sorry, I don't have information to address your question.