AI-Driven Decision Models for Subcontractor Evaluation: Integrating ChatGPT 4.0 for Interactive Performance Analysis

Stylianos Chatzakis, 6068553

MSc in Construction Management and Engineering

March 18th, 2025

Graduation committee

Dr. Tong Wang

Ir. J.P.G. Hans Ramler

Dr. Ir. Ranjith Kuttantharappel Soman

Company supervisor

Richard Bouman

Table of Contents

Preface

This thesis marks the culmination of my journey through the MSc program in Construction Management and Engineering at TU Delft. This journey would not have been possible without the support, encouragement, and guidance of many individuals and institutions to whom I owe my deepest gratitude.

First and foremost, I would like to thank HOCHTIEF Nederland for providing me with the opportunity to conduct this research within their organization. Their openness, as well as the friendly and collaborative environment they created, was vital in shaping the direction of this work. Special thanks go to my company supervisor, Richard Bouman, for his constructive feedback, constant encouragement, and the trust he placed in me throughout this project.

I would also like to extend my special thanks to Patrick van Hofwegen for his keen interest in my project, his constant support, and his eagerness to help me achieve the best possible results and become a better professional.

To my other colleagues at HOCHTIEF, thank you for generously sharing your insights, experience, and time during interviews, discussions, and evaluations. Your perspectives brought real-world depth to this work and made the research process engaging and meaningful—special thanks to Sam Huckle, Vasco Ribeiro, and Lohen Campbell for their mentorship.

I am also profoundly grateful to my academic supervisors, Dr. Tong Wang, Hans Ramler, and Dr. Ir. Ranjith Kuttantharappel Soman, for their insightful guidance, critical comments, and unwavering support. Their expertise helped me stay focused, think critically, and refine the core contributions of this research.

I also want to thank my friends who made this academic journey not only intellectually rewarding but also socially enriching. Their constant support and motivation made this journey possible. I have some great memories and stories to tell, and I am certain I have made friends for life over the past two years.

Last but certainly not least, I owe heartfelt thanks to my parents and family. Their unwavering belief in me, their patience, and their endless support, both emotional and practical, carried me through the most demanding moments of this journey.

To all of you: thank you for helping me reach this milestone.

Stylianos Chatzakis
Delft, August 2025

Abstract

Subcontractor evaluations are a critical component of post-project reviews in the construction sector, significantly influencing future procurement decisions, subcontractor selection, relationship management, and risk mitigation strategies. Despite their strategic value, existing subcontractor evaluation practices often suffer from several persistent shortcomings: high subjectivity, lack of standardized performance metrics, limited documentation, exclusion of subcontractor perspectives, and minimal use of structured or historical evaluation data. In response to these challenges, this thesis examines whether and how integrating ChatGPT 4.0 with a Multi-Criteria Decision Analysis (MCDA) framework can enhance the consistency, objectivity, and strategic value of subcontractor evaluations within HOCHTIEF Nederland.

Building on a comprehensive literature review and exploratory interviews with key procurement and project staff, the study identifies gaps in traditional evaluation processes, particularly the overreliance on static rating templates and the absence of dual-perspective analysis. To address this, a novel AI-enhanced evaluation framework is developed using a design science methodology. The framework enables structured performance assessments based on both internal HOCHTIEF feedback and subcontractor self-evaluations. ChatGPT 4.0, embedded in HOCHTIEF's internal AI assistant "NextChat" supports this process by interpreting qualitative data, prompting for clarification when needed, generating justifications for ratings, and summarizing insights into actionable reports.

The framework was implemented in a pilot case study involving a subcontractor working on a real HOCHTIEF data centre project. Evaluation inputs from both parties were processed using the AI assistant and benchmarked against HOCHTIEF's existing manual evaluation methods.

Validation of the framework was multi-faceted. Process validation demonstrated that AI-generated reports aligned closely with manual evaluations, with a Mean Absolute Percentage Deviation (MAPD) of less than 10%, indicating high accuracy. Stakeholder validation was conducted through structured surveys with HOCHTIEF personnel, assessing insightfulness, transparency, clarity of follow-up queries, added value, and perceived limitations. The results were consistently positive.

While the results indicate strong potential for improving subcontractor evaluations through AI integration, the study also highlights critical limitations and risks. These include the need for high-quality, context-rich input data, the need for human oversight and verification of the results, data availability challenges, and the necessity of embedding the tool within existing procurement databases to ensure organizational consistency. Furthermore, successful implementation depends on training and change management strategies, particularly in organizations with limited digital procurement maturity.

The study contributes theoretically by empirically validating the integration of AI and MCDA in construction procurement and practically by providing a scalable tool that enhances the structure, transparency, and usability of post-project subcontractor evaluations. It is supported by established adoption frameworks, including the Technology–Organization–Environment (TOE) model and the Unified Theory of Acceptance and Use of Technology (UTAUT). Overall, this research offers a data-informed, dual-perspective, and AI-supported approach to subcontractor assessment, positioning it as a robust enhancement to current construction management practices.

Chapter 1: Introduction

1.1 Background

Procurement is the cornerstone of successful construction projects, directly impacting firm profitability by determining the quality, efficiency, and reliability of subcontractors. Subcontractor evaluation and selection are crucial components of procurement and supply chain management, significantly impacting operational efficiency, project costs, and risk mitigation. Traditional evaluation methods focus heavily on quantitative metrics such as cost, delivery performance, and compliance with specifications (Jiang et al., 2013; Moretto et al., 2017). While these metrics are essential, they often overlook qualitative factors like communication effectiveness, relationship management, and long-term reliability, which are equally crucial for comprehensive assessments. Effective communication can lead to better coordination and fewer project delays, while strong relationship management fosters trust and collaboration, which are essential for resolving unforeseen challenges during project execution.

The advent of Artificial Intelligence (AI) and Natural Language Processing (NLP) has introduced new opportunities to improve subcontractor evaluation processes. AI models have demonstrated significant potential in automating supplier assessments, reducing human biases, and accelerating decision-making processes (Guida et al., 2023; Wilson, 2024). Despite these advancements, the successful implementation of AI in procurement faces several challenges, including data quality, bias mitigation, and model interpretability (Abdulla & Baryannis, 2024).

1.2 Significance of the Study

Construction firms, like HOCHTIEF, invest a significant portion of their revenues in procurement activities, making procurement strategies directly impactful on firm profitability (Guida et al., 2023). The integration of AI into procurement processes, particularly in post-project evaluations, has the potential to make a significant impact on how subcontractor performance is assessed by combining both quantitative and qualitative data. However, the current literature lacks robust frameworks and empirical studies that demonstrate the successful implementation and measurable benefits of AI-driven subcontractor evaluations (Guida et al., 2023). This study aims to fill this gap by developing and validating an AI-driven subcontractor evaluation framework that integrates ChatGPT 4.0 with a Multi-Criteria Decision Analysis (MCDA) approach.

By leveraging ChatGPT 4.0's interactive question-generation capabilities, the proposed approach aims to improve the comprehensiveness and accuracy of subcontractor assessments. This integration facilitates data clarifications by addressing ambiguities and ensuring that subcontractor evaluations are based on complete and reliable information. The research not only contributes to academic knowledge by providing a theoretical framework that combines AI with established MCDA methodologies but also offers practical insights through a detailed case study within HOCHTIEF's ongoing construction projects.

Overall, this study aims to bridge the existing knowledge gap by offering an innovative, AI-enhanced framework that integrates both quantitative and qualitative assessments in post-project subcontractor evaluations. This contribution advances both academic research and industry practices, promoting more informed, accurate, and transparent procurement decisions in the construction sector.

1.3 Structure of Research

This thesis is organized into ten primary chapters, each addressing a distinct aspect of the research process. The first chapter is a small introduction to the research. The second chapter delves into existing subcontractor evaluation frameworks, AI applications in procurement, advancements in Machine Learning (ML) and NLP for supplier evaluation, and the integration of MCDAs with Large Language Models (LLMs). This comprehensive review synthesizes insights from academic journals, industry reports, and case studies, establishing a robust theoretical foundation for the study. The third chapter identifies the knowledge gap in the current literature, articulates the specific problem that this research addresses, and formulates the primary and secondary research questions guiding the study. In the fourth chapter, the research design is presented, including qualitative and quantitative data collection methods, the validation method, AI and MCDA integration processes, and project phases. The fifth chapter presents the evaluation process currently employed by HOCHTIEF, identifies the involved departments, the evaluation criteria currently being used, and challenges and potential improvements to the current evaluation process. All the insights presented in Chapter 5 are derived from interviews with company employees.

Chapter 6 presents the subcontractor evaluation criteria derived from the literature review, details the integration of these criteria into the AI-driven framework, and explains how they have been validated and assigned weights based on their importance through interviews with company employees. Following that, Chapter 7 explains the development of the proposed framework, which includes the creation of a new evaluation document and a new AI-based evaluation assistant. A case study, where the proposed approach is tested, will be presented in Chapter 8. The case study involves implementing an AI-driven evaluation framework to assess the performance of two subcontractors in an ongoing data center project led by HOCHTIEF. The results will be validated through a survey of company employees and compared to manual evaluations conducted using the company's current process. Next, Chapter 9 provides a discussion of the findings, analyzing their implications for the research questions, how they compare to the literature findings, and the limitations of the research. Finally, Chapter 10 offers conclusions, practical recommendations for industry adoption, and a reflection on the methodology and the researcher's learning throughout the thesis process.

Chapter 2: Literature Review and Background

This literature review aims to provide a comprehensive understanding of both traditional subcontractor evaluation methods and recent advancements in AI and NLP technologies, which promise to change and improve procurement processes in modern construction management. It will also explore how AI, specifically LLMs, have been implemented in the construction sector, examining the potential benefits of their application and the issues that hinder or prevent their implementation.

As part of this literature review, focus has been given on Multi-Criteria Decision Method (MCDMs). More specifically, the review investigates how MCDMs have been used in construction and combined with AI to create evaluation frameworks. It examines existing evaluation frameworks that predominantly rely on quantitative metrics such as cost efficiency, timeliness, and technical compliance, as well as studies that highlight the importance of qualitative factors like communication effectiveness, relationship management, and risk management (Jiang et al., 2013; Moretto et al., 2017; Silva et al., 2022). It further investigates how integrating AI tools, including LLMs and NLP techniques, can address current gaps by incorporating detailed qualitative insights. Lastly, through this research, the aim is to identify evaluation criteria for subcontractor assessments that can then be used in the farmwork proposed by this research. The criteria derived from this literature will be presented and explained in Chapter 6.

To ensure a comprehensive search, the following keywords were employed across academic databases like Google Scholar and ScienceDirect:

- “Evaluat* AND LLM AND Suppliers”
- “Assess* AND Criteria AND Subcontractors”
- “Evaluat* AND Criteria AND Suppliers”
- “Evaluat* AND LLM AND Subcontractors”
- “Assess* AND LLM AND Suppliers”
- “Assess* AND LLM AND Subcontractors”
- “AI AND Construction AND Benefits”
- “AI AND Construction AND Disadvantages”
- “AI AND Construction AND Advantages”
- “AI AND Construction AND Negatives”

This literature review incorporates studies published since 2000 to capture recent innovations and address emerging research challenges.

The remainder of this chapter is structured as follows: Subchapter 2.2 examines subcontractor evaluation frameworks by contrasting traditional quantitative methods with emerging qualitative and hybrid approaches. These approaches combine MCDA techniques, such as the Analytic Hierarchy Process (AHP) and the Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE), to enhance decision-making. Subchapter 2.3 examines the role of Artificial Intelligence and Natural Language Processing in procurement, highlighting their benefits, including improved efficiency and transparency, as well as challenges such as data quality, bias, and ethical considerations. Additionally, it discusses the integration of LLMs in supplier evaluations. Subchapter 2.4 reviews theoretical frameworks and conceptual models, including the Technology-Organization-Environment (TOE) framework and the Unified Theory of Acceptance and Use of Technology (UTAUT), to provide a conceptual basis for AI adoption in procurement. Subchapter 2.5 presents empirical evidence and case studies that validate the integrated evaluation frameworks in real-world

scenarios, synthesizing findings and illuminating future research directions. Finally, subchapter 2.6 summarizes the findings.

2.1 Subcontractor Evaluation Frameworks

2.1.1 Traditional and Qualitative Evaluation Methods

Traditional subcontractor evaluation methods in the construction industry predominantly focus on objective, quantitative metrics like cost efficiency, adherence to project timelines, and compliance with industry standards (Jiang et al., 2013; Moretto et al., 2017). While these provide measurable bases for comparison, they often fail to capture nuanced aspects such as innovation capabilities, problem-solving skills, and adaptability to project changes (Silva et al., 2022). Yin et al. (2017) note that numerical metrics, though essential, often overlook softer performance aspects such as communication effectiveness, relationship management, and adaptability. Furthermore, Upadhyaya et al. (2021) suggest that qualitative parameters, like stakeholder feedback and on-site performance narratives, are critical for a nuanced evaluation but are frequently absent from traditional frameworks. Qualitative data are crucial and should be included as subcontractors with strong interpersonal skills and relationship management contribute to smoother project execution and higher client satisfaction (Sundquist et al., 2018). Cannavacciuolo et al. (2015) assert that subjective evaluations derived from unstructured data can significantly enhance decision-making by capturing insights that raw numbers cannot. Sustainability practices and innovation are also critical evaluation criteria as the industry shifts toward more environmentally responsible methods (Silva et al., 2022; Keshavarz-Ghorabaee et al., 2020). Thus, it is important to not only focus on quantitative factors when evaluating a subcontractor's performance, as several qualitative factors can change the outcomes of the assessment if considered.

2.1.2 The Emergence of Qualitative and Hybrid Methods

Integrated evaluation frameworks provide a comprehensive method for assessing subcontractor performance by combining both quantitative and qualitative criteria (Abdulla & Baryannis, 2024). Using extensive data sets and advanced analytical techniques, these frameworks offer a complete view of performance, supporting more informed and balanced decision-making. Basu et al. (2017) propose a multi-dimensional framework that identifies nineteen attributes for evaluating subcontractors. These attributes are grouped into five main factors: Technical Experience, Financial Competency, Resource Adequacy, Job Quality and Safety, and Local and Other Factors. The framework was tested in a real-world scenario at a housing project using monolithic concrete construction with aluminum formwork. The implementation demonstrated better decision-making regarding subcontractor evaluation.

A growing body of research advocates for employing MCDA to integrate these varied data types. Akmaludin et al. (2023) state that combining the AHP with the PROMETHEE helps practitioners assign appropriate weights to both qualitative and quantitative criteria. Johnson et al. (2023) further show that real-time feedback integrated into MCDA frameworks can enhance responsiveness and flexibility in contractor selection.

A key feature of integrated evaluation frameworks is their ability to combine different types of data into a single assessment model. For example, Polat (2021) introduced a framework that combines the AHP with the PROMETHEE. AHP structures the evaluation criteria into a hierarchy, enabling evaluators to compare and assign relative weights based on importance. PROMETHEE then uses these weighted

criteria to evaluate and rank subcontractors by assessing their performance against each criterion using predefined preference functions. This method ensures a balance between objective metrics and subjective judgments, leading to more reliable subcontractor rankings (Polat, 2021; Abdulla & Baryannis, 2024).

Furthermore, integrated evaluation frameworks use MCDA techniques to manage the complexity of assessing multiple, often interconnected, criteria. MCDA offers a structured decision-making process that considers various evaluation factors. Emmanouil-Kalos (2024) notes that adopting MCDA improves decision-making efficiency and transparency while allowing for the nuanced handling of competing priorities. This is especially important in integrated frameworks where stakeholders may have different goals and values, requiring a structured way to evaluate trade-offs among options (Stratil et al., 2020; Huang et al., 2023). In such cases, MCDA helps identify and prioritize stakeholder perspectives, which is crucial for achieving consensus when objectives differ. Using MCDA alongside techniques like scenario planning has proven effective for strategic decisions, enabling decision-makers to compare strategies against predefined criteria and potential future scenarios (Montibeller et al., 2006).

Another important feature of integrated evaluation frameworks is their flexibility to adapt to different project settings and organizational goals (Keshavarz-Ghorabaee et al., 2020). These frameworks can be tailored to meet specific project needs and strategic targets. For instance, a project focused on sustainability might weigh environmental compliance and green practices more heavily. Conversely, a project prioritizing timely completion might emphasize schedules and deadlines. This flexibility helps maintain the relevance and effectiveness of these frameworks across various industries and project types (Silva et al., 2022).

Implementing integrated evaluation frameworks also encourages greater stakeholder participation and consensus (Upadhyaya et al., 2021). By incorporating input from departments such as procurement, project management, and quality assurance, these frameworks ensure multiple perspectives are included in the subcontractor assessment. Engaging stakeholders in setting criteria and weights leads to more representative evaluations that consider different interests and foster stakeholder support (Bartolini & Viaggi, 2010). This collaborative process not only improves assessment accuracy but also promotes buy-in, resulting in smoother procurement processes and stronger subcontractor relationships (Sundquist et al., 2018; Thompson, 2023).

Finally, integrated evaluation frameworks support ongoing improvement through iterative feedback and data-driven updates (Abdulla & Baryannis, 2024). As subcontractor performance data accumulates over time, organizations can adjust their evaluation criteria and weights to better match evolving project needs and industry standards. This continuous refinement ensures the evaluation process remains relevant and practical, adapting to changing project landscapes and organizational objectives.

2.2 Artificial Intelligence and NLP in Procurement

2.2.1 The Role and Impact of AI in Procurement

The rapid digitization of procurement functions is reshaping supplier evaluation processes. Guida et al. (2023) observe that AI can digitize vast amounts of procurement data, enabling enhanced visibility and precise risk mitigation. AI techniques automate routine tasks, streamline demand forecasting, and optimize resource allocation. Lee (2023) describes how machine learning algorithms applied in AI systems have reduced procurement cycle times by up to 25%, demonstrating improved

operational efficiency. By employing ML and data analytics, organizations can streamline procurement activities, analyze market trends, and enhance supplier relationships, ultimately contributing to a more resilient supply chain (Althabatah et al., 2023; Obinna & Kess-Momoh, 2024). The integration of AI technologies, such as predictive analytics and natural language processing, allows businesses to identify procurement opportunities and risks more effectively, thus optimizing sourcing strategies (Althabatah et al., 2023; Obinna & Kess-Momoh, 2024).

In addition to predictive analytics, AI-powered dashboards facilitate real-time monitoring of supplier performance, allowing for prompt identification of issues and swift remedial actions (Ogundipe et al., 2024). Hu and Ren (2023) emphasize that such efficiencies not only reduce costs but also improve overall performance quality by integrating objective data with contextual insights.

2.2.2 Benefits and Challenges of AI and NLP Integration

Adopting AI in procurement processes offers many benefits that can significantly impact operational effectiveness, risk management, and strategic decision-making. Elmousalami et al. (2025) note improvements in decision-making accuracy and operational efficiency, while Tai et al. (2023) emphasize enhanced supplier risk assessment through the integration of LLMs and life cycle assessment techniques. Improved real-time data processing from AI systems facilitates better negotiation, higher supplier performance measurement, and overall cost reduction.

AI-driven analytics automate routine tasks, reduce human errors, and provide actionable insights that inform strategic procurement decisions (Handfield et al., 2019). They provide procurement teams with the ability to focus on strategic decision-making rather than mundane administrative tasks, leading to improved negotiations and decision-making (Morgan, 2021; Tatini, 2025). Furthermore, AI improves transparency and fairness by standardizing evaluation criteria and mitigating subjective biases (Scott et al., 2015). AI facilitates better inventory management through predictive analytics, which helps organizations anticipate demand fluctuations and adjust procurement strategies accordingly.

Moreover, AI can enhance fraud detection in procurement systems by analyzing historical data for anomalies or fraudulent behavior. This function helps organizations mitigate risks associated with procurement corruption, ultimately leading to cost savings and improved operational integrity (Adelakun et al., 2024; Ezeji, 2024). With AI systems continuously adapting through learning algorithms, they can identify shifts in procurement patterns, alerting teams to potential issues before they escalate (Adelakun et al., 2024).

Through an extensive literature review of 85 papers, Guida et al. (2023) indicate 27 benefits stemming from the adoption of AI throughout the procurement process, which are extensively mentioned throughout the literature. An overview of these benefits is presented in Table 1.

***Table 1**. Benefits stemming from the adoption of AI throughout the procurement process (Guida et al., 2023)*

Benefit Number	Benefit Description
1	Higher visibility and control
2	Risk reduction along the supply chain
3	Higher accuracy in the planning process
4	Real-time adaptation to external requests
5	Improved negotiation and supplier selection by analyzing historical data and market trends
6	Enhanced communication channels with suppliers
7	Reduction in time spent on non-value-added tasks
8	Identification of cost-saving opportunities
9	Monitoring of internal purchasing performance
10	Alignment of internal systems with external partners' systems
11	Supplier performance measurement
12	Reduced cycle time from request-for-quote to order issuance
13	Improvement in finished product quality
14	Prediction of commodity price volatility
15	Reduction of costs associated with downtime and errors
16	Reduced time required for reviewing and approving contracts
17	Increased participation of the purchasing department in innovation partnerships
18	Establishment of closer relationships with key suppliers
19	Improvement of supply chain relationships
20	Better integration of the purchasing department with other organizational functions
21	Enhanced spend classification
22	Greater alignment of the purchasing department with corporate strategy
23	Rationalization of suppliers
24	Optimal specification definition
25	Improvement in the usability of information systems
26	Identification of suspicious expenses
27	Alignment of the company to market standards

By integrating AI with MCDA, the evaluation of complex datasets is automated, and the accuracy of criterion weighting and alternative ranking is improved (Abdulla & Baryannis, 2024). AI-driven approaches analyze vast procurement data to refine MCDA processes (Abdulla et al., 2023). Studies demonstrate the practical benefits of integrating AI with MCDA in procurement, such as improved supplier selection accuracy and operational efficiency in the oil and gas industry (Abdulla et al., 2023). Comparative studies also show that AI-MCDA models enhance accuracy and consistency in supplier rankings compared to traditional methods (Abdulla & Baryannis, 2023; Wang et al., 2025).

AI adoption in procurement offers numerous advantages. However, it also poses significant challenges. A critical issue is data availability and systematization (Wilson, 2024). Procurement operations often deal with fragmented datasets that hinder the integration into a cohesive analytical system, which is essential for the effective functioning of an AI model. Concurrently, data quality and the presence of

large, reliable datasets are crucial (Guida et al., 2023), demanding robust data governance practices (Pandit, 2025).

Additionally, the reliance on AI can result in a lack of accountability, as decisions made by AI systems may not be easily traceable or explainable, leading to ethical dilemmas and potential legal issues (Nagbøl et al., 2021; Alzubaidi et al., 2023). In the procurement process, where transparency, accountability, and trust are essential, algorithm aversion can lead procurement professionals to reject AI-generated recommendations, reducing the effectiveness of AI-driven decision-making. Procurement involves high-stakes financial transactions, regulatory compliance, and supplier relationships. If an AI system makes an erroneous or biased decision, tracing responsibility becomes difficult. This can lead to legal disputes, reputational damage, and operational inefficiencies. Another disadvantage is the risk of discrimination and bias inherent in AI algorithms. If the data used to train AI systems is biased, the decisions made by these systems can perpetuate existing inequalities (Nagbøl et al., 2021).

Ensuring ethical deployment involves addressing bias and fairness to achieve equal treatment for all suppliers (Akintayo et al., 2024; Obinna & Kess-Momoh, 2024). Alongside these ethical considerations, risks of data privacy and security arise from reliance on shared information between organizations and suppliers (Alshehhi et al., 2023).

Integration and interoperability challenges are exacerbated by existing procurement systems' difficulty accommodating advanced AI technologies (Cui et al., 2022). Due to this issue, there is a need for continuous investment in infrastructure and personnel training (Matsuno et al., 2014), something particularly straining for smaller organizations with limited budgets (McBride et al., 2021).

Furthermore, there is an ongoing deficiency in the internal analytical skills required to interpret AI-generated insights (McBride et al., 2021). Over-reliance on AI models without adequate human oversight could lead to poor decision-making if AI-derived insights are not thoroughly critiqued (Spreitzenbarth, 2021; Cooper, 2024). Lastly, job displacement due to automation remains a concern, contributing to employee resistance against AI adoption (He, 2023). Organizations can mitigate this by engaging in transparent communication and implementing training initiatives.

According to the literature review carried out by Guida et al. (2023), the adoption of AI in procurement has been associated with 17 distinct challenges. A complete overview of these challenges can be found in Table 2.

***Table 2**. Challenges to the adoption of AI throughout the procurement process (Guida et al., 2023)*

Disadvantage Number	Disadvantage Description
1	Extensive initial investments required in AI-based solutions
2	Ethics and governance challenges
3	Data protection and integrity issues
4	Lack of standardization in AI systems
5	Lack of trust in AI outcomes
6	Lack of skilled workforce to operate AI systems
7	Exploitation by hackers, cybercrimes, and privacy intrusions
8	Reluctance to adopt new technologies
9	Risk and cost associated with construction projects
10	Requirement of explainable AI (XAI)
11	Uncertain processing and functions of AI algorithms
12	Unclear profits and advantages from AI adoption
13	Erroneous AI algorithms
14	Expensive and continuous maintenance requirements
15	Fragmented and project-based nature of the industry
16	Legal and contractual issues
17	Frequent interruptions in power and internet connectivity

2.2.3 The Integration of LLMs in Procurement

LLMs are transforming procurement by enhancing automation, streamlining processes, and improving decision-making. By analyzing vast amounts of data, LLMs can assist procurement teams in tasks such as contract review, supplier selection, market analysis, and risk management. These models can quickly identify patterns, generate insights, and optimize procurement strategies, saving time and reducing human error. LLMs can not only address the limitations of traditional MCDM methods in handling complex multidimensional tasks but also highlight their great potential and advantages in decision analysis (Wang, 2025).

One of the primary advantages of employing LLMs in procurement is the potential for improved transparency and efficiency. As outlined by Mishra et al. (2021), the integration of an improved software ecosystem utilizing data sharing frameworks such as Data Mesh and Service Mesh architecture can enhance the visibility of procurement processes, leading to productivity gains and cost savings across sectors. This integrated approach allows organizations to process vast amounts of procurement data, enabling them to identify inefficiencies and opportunities that may not be apparent through traditional analysis methods. Moreover, LLMs facilitate negotiations in procurement, a challenging yet critical aspect of the purchasing process. As discussed in the work of Li (2024) on AI-powered negotiations, integrating LLMs can simplify the negotiation process by automating the generation of negotiation strategies and ensuring that all relevant data is considered during discussions. LLMs can analyze previous agreements, comprehend negotiation contexts, and provide real-time advice, thus enhancing negotiation outcomes by maximizing benefits. However, the success of this integration hinges on the negotiator's ability to leverage these technological tools effectively, which can vary based on individual skills and external factors.

Another significant application of LLMs in procurement is maintaining regulatory compliance. As Li (2025) notes, industries must adhere to a variety of regulatory requirements, and the compliance landscape is often complex and fragmented. LLMs can assist organizations by conducting continuous compliance checks, assessing documentation against evolving regulations, and synthesizing regulatory updates into actionable directives (Li, 2025). However, this application raises concerns regarding privacy and data security, necessitating robust frameworks to safeguard sensitive information while ensuring compliance.

Despite these potential benefits, several challenges arise when integrating LLMs into procurement systems. A fundamental challenge is the reliance on high-quality data for the practical training of these models. As highlighted by Earley and Mehta (2024), effective personalization and user connection in applications like procurement rely on rich and diverse datasets. However, the fragmented nature of procurement data can lead to inconsistencies that diminish the effectiveness of LLMs. Organizations need comprehensive strategies to manage data quality and integrity within these models to prevent biases and inaccuracies in procurement processes. Furthermore, as organizations move towards integrating LLMs, they face challenges in managing the transition from traditional procurement methods to technology-driven approaches. The implementation of LLMs requires significant infrastructure upgrades and changes in organizational culture. As indicated in Armstrong et al.'s (2024), facilitating this transition often falls to IT departments, which may lack procurement-specific expertise necessary for optimal deployment of LLM solutions. This misalignment can result in suboptimal integration, where the full benefits of LLM capabilities are not realized.

2.3 Theoretical Frameworks and Conceptual Models

The integration of AI into procurement processes requires a structured understanding of the factors that influence its adoption and implementation. To guide the effective adoption of AI technologies within procurement systems, various theoretical frameworks have been developed. These frameworks provide insights into the technological, organizational, and environmental factors that shape the adoption and use of AI in procurement.

2.3.1 Technology-Organization-Environment (TOE) Framework

One of the foundational frameworks for understanding AI adoption is the Technology-Organization-Environment (TOE) framework. The TOE framework, as outlined by Tornatzky and Fleischer (1990), offers a robust model for understanding technology adoption in procurement, which suggests that successful technology adoption depends on three critical dimensions:

- **Technological Readiness:** The availability and maturity of AI systems, data infrastructures, and LLMs.
- **Organizational Capacity:** The readiness of organizations to implement new technologies, including skilled human resources and effective change management protocols.
- **External Environmental Pressures:** Factors such as regulatory demands, competitive intensity, and market dynamics that drive technology adoption.

In the context of procurement, the TOE framework helps explain how an organization's internal technological infrastructure and external pressures can impact the decision to adopt AI. For instance, in procurement, technological readiness might include the availability of AI-driven tools and the organization’s ability to integrate them effectively into existing procurement workflows. The organizational dimension considers factors such as the expertise of procurement teams and their

ability to manage AI tools. In contrast, environmental factors assess how the external market, competitive pressures, or regulatory changes may influence the adoption of AI in procurement activities. Research by Uddin et al. (2021) confirms that firms with high organizational capacity and strong external support are more likely to adopt AI-enhanced procurement strategies successfully.

2.3.2 Unified Theory of Acceptance and Use of Technology (UTAUT)

Another important model is the Unified Theory of Acceptance and Use of Technology (UTAUT), which provides an understanding of the factors that influence the acceptance and use of technology. The UTAUT model identifies key determinants such as performance expectancy, effort expectancy, social influence, and facilitating conditions that shape how individuals and organizations perceive and use technology (Venkatesh et al., 2003). In procurement, performance expectancy refers to the perceived effectiveness of AI tools in enhancing procurement efficiency and decision-making. Effort expectancy reflects how easy or difficult stakeholders find it to use AI tools. At the same time, social influence involves the degree to which peer organizations or industry leaders influence the decision to adopt AI. Facilitating conditions refer to the resources and infrastructure available to support AI adoption, including the availability of training and support systems for procurement teams.

2.4 Empirical Evidence and Case Studies of the Application of AI in Procurement

Several case studies illustrate the practical application of AI in procurement across various sectors. Advanced frameworks that integrate ML with MCDA, such as the ML-MARCOS model, offer substantial advantages in managing multi-criteria decision problems (Abdulla et al., 2023). These frameworks leverage ML algorithms to optimize decision-making by analyzing large datasets, identifying patterns, and providing real-time insights into supplier performance (Wang et al., 2025; Abdulla et al., 2023). This integration enhances the procurement process by providing more profound, more nuanced insights into the performance of potential suppliers, ensuring that decisions are based on a broader range of factors. To ensure that these advancements are practical and effective, academia and industry need to collaborate closely in developing and testing these models, ensuring their applicability in real-world procurement scenarios.

Akbar and Şimşek (2024) have demonstrated that the integration of LLMs, particularly GPT and BERT, into MCDA frameworks can enhance the processing of qualitative information. Hadi et al. (2023) report that LLMs are effective in extracting sentiments and critical signals from supplier communications, customer reviews, and contractual texts. Furthermore, fine-tuning these LLMs for domain-specific applications has been shown to improve their relevance and accuracy in specialized contexts (Pandit, 2025). Guida et al. (2023) reported that companies implementing AI in procurement achieved significant improvements in cycle times and cost reductions. Similarly, studies by Tai et al. (2023) in green supplier risk assessment have shown that integrating NLP with life cycle assessment enhances decision accuracy and supplier risk prediction.

The Qatar Foundation explored the potential of AI-powered procurement systems to enhance operational efficiency and decision-making (Bahameish et al., 2023). In South Korea, AI's impact on procurement processes has led to significant improvements in purchasing efficiency, demonstrating its broad applicability in different contexts (Lee, 2023). In another relevant case study, organizations utilized AI technologies for demand forecasting and vendor selection, which led to streamlined procurement workflows and improved project outcomes (Ogundipe et al., 2024). Lumanauw et al.

(2023) document improved subcontractor performance evaluation using a weighted sum method combined with KPI measurement, indicating that the integration of AI frameworks can lead to more reliable supplier performance ratings.

Comparative analyses in the literature underscore the superiority of these hybrid methods over traditional evaluation techniques. Mwangata and Chrine (2024) found that evaluating supplier performance using such hybrid models resulted in a more comprehensive understanding of both quantitative and qualitative aspects.

2.5 Synthesis and Conclusions

Chapter 2 provides a comprehensive review of the literature surrounding subcontractor evaluation in the construction industry, with a particular focus on the evolving role of AI and ML technologies. This literature review has traced the evolution of subcontractor evaluation and procurement practices from traditional quantitative methods to sophisticated, integrated frameworks that leverage AI and NLP. Traditional evaluation methods, while robust, overlook important qualitative dimensions that can be captured through stakeholder feedback and unstructured data analysis (Yin et al., 2017; Upadhyaya et al., 2021).

AI is introduced as a transformative tool in procurement, capable of automating routine tasks, improving decision-making, and enhancing operational efficiency. By integrating AI technologies such as predictive analytics and natural language processing, organizations can optimize procurement processes, analyze market trends, and improve supplier relationships. The chapter further discusses how AI can enhance MCDA, improving supplier selection accuracy and operational efficiency.

The application of AI and NLP opens new avenues for analyzing large volumes of unstructured supplier data, which enhances risk management and decision-making accuracy (Akbar & Şimşek, 2024; Hadi et al., 2023). Empirical studies and case analyses underscore both the benefits and challenges of AI integration, with significant improvements noted in cycle time reduction, cost efficiency, and predictive accuracy, balanced against issues like data integration and ethical concerns (Guida et al., 2023; Wilson, 2024). Apart from the benefits, the chapter also outlines challenges associated with AI adoption, including ethical concerns, the need for specialized knowledge, data privacy risks, and resistance to automation.

This literature review has demonstrated that while traditional subcontractor evaluation methods offer essential quantitative metrics, they cannot capture the nuanced qualitative insights needed for comprehensive assessments. By synthesizing findings from various studies, this review highlights a significant research gap: the need for an integrated, AI-enhanced evaluation framework that combines quantitative rigor with qualitative depth.

Chapter 3: Research Problem, Objectives, and Questions

3.1 Knowledge Gap

While significant progress has been made in incorporating AI into procurement and subcontractor evaluations, several critical gaps remain. First, the existing literature mainly emphasizes quantitative metrics, which do not fully capture the scope of subcontractor performance (Jiang et al., 2013; Mutai, 2016). Qualitative factors are often neglected, limiting the thoroughness of supplier assessments. Second, there is a notable lack of empirical studies comparing AI-driven subcontractor evaluations with traditional manual assessments. Most current research focuses on theoretical frameworks and model development without enough real-world validation, making it hard to understand the practical impact and effectiveness of AI-enhanced evaluation methods. Additionally, most studies focus on the prequalification phase of a construction project, leaving a gap in empirical research on evaluations conducted after project completion. The potential of NLP in analyzing unstructured qualitative data from reports, feedback, and communications is also underused. This gap is crucial because qualitative insights can significantly improve risk assessment and decision-making.

3.2 Problem Statement

As noted in Chapter 2.2.1, traditional subcontractor evaluation methods rely heavily on static quantitative metrics, which fail to capture the complex nature of subcontractor performance assessments fully. This dependence can lead to incomplete and biased evaluations because qualitative factors are often overlooked. As a result, procurement decisions may not accurately reflect a subcontractor's actual performance, potentially leading to misjudging capable subcontractors, choosing underperforming firms for future projects, straining relationships with qualified specialists, causing delays, or compromising the quality of upcoming projects. Additionally, it may result in financial losses, as poor performance goes unaddressed, negatively impacting overall project outcomes and the company's reputation. By harnessing the processing power of LLMs and integrating them with established decision-making frameworks, the aim is to enhance the accuracy, transparency, and informativeness of subcontractor evaluations. This approach enables construction companies like HOCHTIEF to achieve a more detailed and comprehensive assessment of subcontractor performance, aligning procurement decisions more closely with strategic project goals and operational needs.

3.3 Research Objectives

The main goal of this research is to develop and validate an AI-based subcontractor evaluation framework that combines ChatGPT 4.0 with an MCDA approach. Besides this primary objective, several secondary goals are also set. First, the study aims to identify and analyze key criteria for evaluating subcontractors based on existing literature and industry standards. Recognizing and weighting these criteria is essential for creating a framework that accurately captures the most important performance factors for HOCHTIEF. Second, the study will explore how to integrate LLMs with MCDA to enhance the company's subcontractor evaluation process. By using LLMs, the research aims to make evaluations more comprehensive and precise.

A comparison will be made between AI-generated ratings and traditional manual evaluations to determine if AI-powered assessments can provide greater accuracy and depth than conventional methods. Achieving these objectives will help improve procurement practices at HOCHTIEF, ensuring

subcontractor assessments are thorough and aligned with the company's strategic goals. The final output will be an advisory report including ratings on subcontractor performance, a summary of their collaboration with HOCHTIEF, lessons learned, and recommendations for future projects. This report will serve as a valuable resource for the company when planning future engagements with subcontractors.

3.4 Research Questions

Primary Research Question:

- How can ChatGPT 4.0, integrated with an MCDA framework, lead to more informed subcontractor evaluations in HOCHTIEF's construction projects?

To answer and support the main research question, the following question will be addressed:

Secondary Research Questions:

- Which performance criteria are most critical for HOCHTIEF's construction projects, and how can they be effectively structured for AI analysis?
- How can ChatGPT 4.0 be used to handle discrepancies or missing details within subcontractor data?
- How can ChatGPT 4.0 be fine-tuned to generate performance-related and clarifying questions to gather more information and provide context for subcontractor evaluations?
- In what ways does integrating AI with an MCDA framework affect subcontractor ratings compared to manual evaluations?
- What lessons learned and best practices emerge from the integration process, and how can they be utilized by HOCHTIEF for future projects?

Chapter 4: Methodology and Project Planning

4.1 Research Design

This study adopts a mixed-methods research design, combining qualitative and quantitative approaches to develop and validate an AI-driven subcontractor evaluation framework tailored to HOCHTIEF's procurement needs. The design involves creating a subcontractor evaluation assistant using HOCHTIEF's AI tool, NextChat, based on OpenAI's GPT-4.0 language model. This assistant will be fine-tuned to enhance the subcontractor assessment process by integrating natural language understanding with decision-support capabilities. The research begins with semi-structured interviews conducted with HOCHTIEF employees. These interviews aim to gain insights into current subcontractor evaluation processes, challenges, and priorities. They will also be used to identify key evaluation criteria and opportunities for AI integration. The qualitative data from these conversations will guide the framework's design to ensure alignment with HOCHTIEF's operational context and strategic goals. Following the interviews, historical procurement data from a previous project will be used to fine-tune the GPT-4.0-powered NextChat assistant. This process will help the AI understand HOCHTIEF's specific language, evaluation logic, and performance standards. Additionally, NextChat's question-generation capabilities will be improved to create relevant, context-aware queries that address inconsistencies, clarify vague responses, and uncover more profound insights into subcontractor performance. The assistant will also be trained to generate actionable, evidence-based recommendations, avoiding generic or unfounded suggestions.

To quantitatively assess subcontractor performance, the study applies the WSM. This model is chosen for its simplicity and usefulness in fast-paced construction environments. The WSM will combine scores across evaluation criteria to produce a unified, comparative score for each subcontractor. These criteria will be identified through a comprehensive literature review (see Chapter 5) and validated and weighted, as mentioned above, through stakeholder interviews. This structured approach ensures that the importance of each criterion aligns with HOCHTIEF's strategic goals. The combined framework will be tested in a real-world case study involving a subcontractor working on a data centre project executed by HOCHTIEF Nederland. This subcontractor is responsible for steel construction. The case study will use the AI evaluation assistant and WSM scoring model to assess the subcontractor in an operational setting. Evaluation data will come from two sources: formal assessment forms filled out by HOCHTIEF employees and the subcontractor, and responses to AI-driven follow-up questions during face-to-face meetings. These follow-ups will be guided by earlier responses and the AI assistant's ability to identify gaps or issues. After completing the evaluation, NextChat will produce a detailed advisory report. The report will include overall performance scores, summaries of key collaboration insights, and data-driven suggestions for future improvements. It will also highlight strengths and concerns, serving as a valuable decision-making tool for ongoing and future subcontractor partnerships. The entire AI-powered evaluation process is illustrated in a flowchart (see Figure 1), showing each step from data input to report generation.

4.2 Data Collection Methods and Ethics Considerations

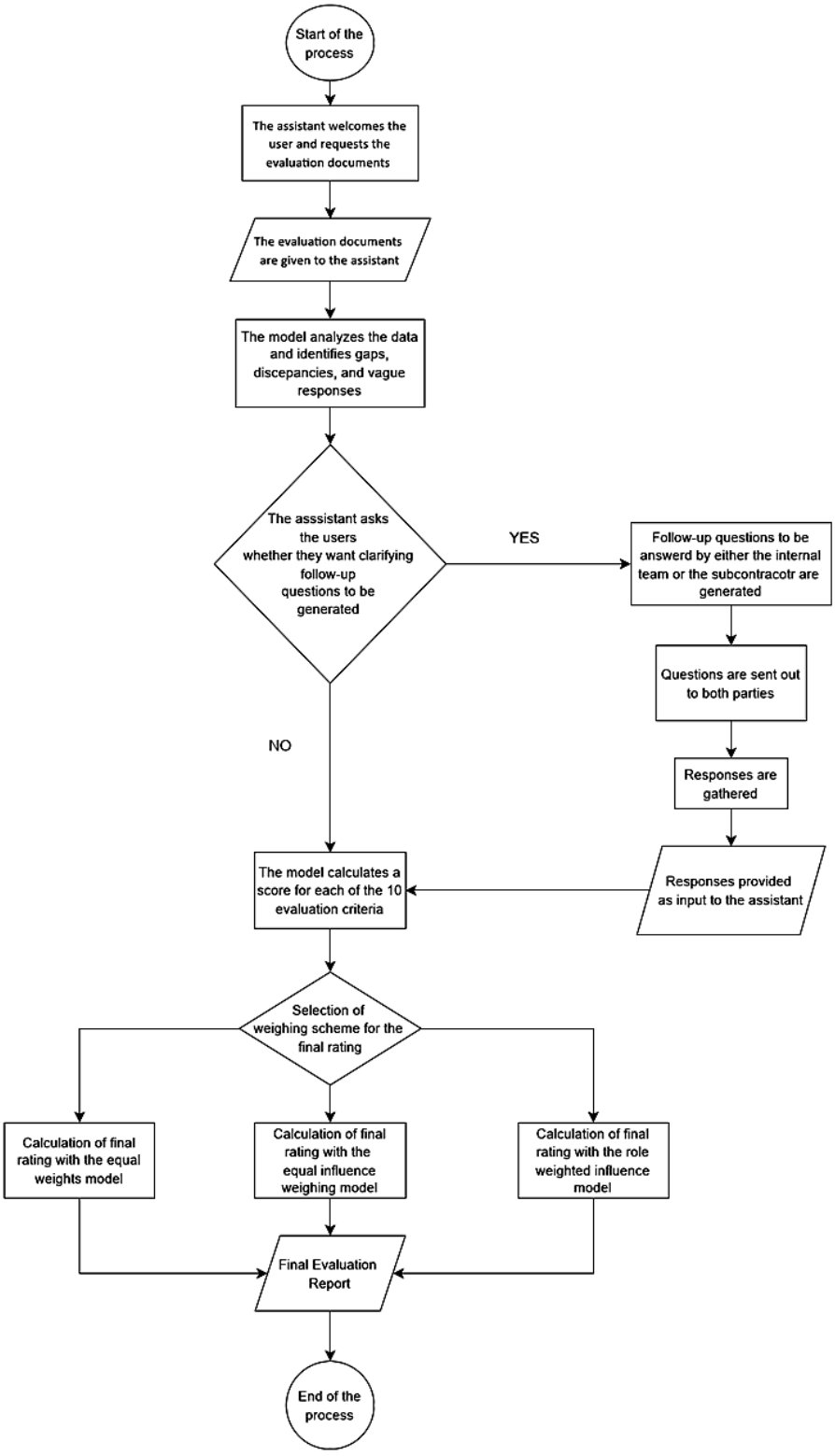
The literature review, presented in Chapter 2, serves as the foundational step in this research by providing a broad understanding of current AI applications in procurement and AI-driven evaluations. In Chapter 5, the evaluation criteria for assessing subcontractors will be introduced. These criteria will be based on the same literature review, ensuring that the proposed approach is

rooted in established knowledge and best practices. As mentioned in Chapter 4.1, semi-structured interviews will be conducted with key stakeholders at HOCHTIEF to assign appropriate weights to each criterion within the MCDA framework. Afterward, a case study will be carried out.

The data sources for this case study include responses from two evaluation documents as well as responses to AI-generated follow-up questions asked during face-to-face meetings. These follow-ups are based on insights from the initial evaluation responses. The case study aims to provide empirical evidence of the framework's impact, offering insights into its practical application and effectiveness within HOCHTIEF's procurement processes.

Additionally, it is important to note that ethical considerations have been carefully addressed in the planning and execution of this research. Approval for both the semi-structured interviews and the case study was officially granted by TU Delft's Human Research Ethics Committee. This ensures that all procedures have been reviewed to meet recognized standards for ethical research practice.

***Figure 1**: Illustration of AI-Powered Evaluation Framework: Process Flow for Subcontractor Assessment, from Document Analysis to Final Rating Calculation and Reporting*



4.3 Validation

The validation process aims to determine whether the tool provides measurable improvements over the current evaluation method by assessing the quality, transparency, and usefulness of its outputs. As mentioned earlier, the final output is a report (in PDF format) that includes the calculated scores for each criterion, the overall performance score, summaries from both perspectives, lessons learned, and recommendations. A structured validation framework has been developed to evaluate the proposed approach using six criteria, each aligned with the main research objective and the research questions outlined in Chapter 3.4.

The validation of the proposed framework will be carried out using two complementary strategies. To verify the numerical consistency of the AI framework with HOCHTIEF's established manual process, a quantitative alignment procedure will be implemented. This involves calculating the mean absolute percentage deviation (MAPD) between the final subcontractor performance scores produced by the AI system and those assessed manually by HOCHTIEF's procurement teams. If the MAPD is less than 10%, the results will be considered accurate. Although there is no universal benchmark for Mean Absolute Percentage Deviation (MAPD) or Mean Absolute Percentage Error (MAPE) specific to the construction industry, foundational forecasting literature provides broadly accepted standards. Lewis (1982) classifies an MAPE under 10% as highly accurate, while values between 10% and 20% are deemed good. Wilson and Keating (2009) similarly consider errors below 10% as indicating excellent predictive capability. Bowerman et al. (2005) reinforce this rule, describing MAPE values under 10% as demonstrating strong model performance. Given the consistency of these guidelines across reputable sources, this study adopts a MAPD threshold of less than 10% to signify strong agreement between the AI-based evaluation system and manual assessments.

Besides quantitative benchmarking, a structured validation survey will be administered to HOCHTIEF professionals. Each participant will evaluate the AI-generated report based on six validation criteria. Where relevant, outputs from the traditional evaluation process will serve as a baseline for comparison. To ensure uniformity, all participants will receive the same report materials and brief instructions on how to interpret AI-generated content. The survey will be completed online, with responses anonymized to encourage honest and critical feedback.

The first criterion, *"Insightfulness",* measures how well the AI-generated report offers a more comprehensive and meaningful analysis of subcontractor performance compared to current methods. The second criterion, *"Clarity and Relevance of Follow-up Questions",* assesses the assistant's ability to generate meaningful, context-aware follow-up questions in response to inconsistencies or missing information in the evaluation data. Participants will review both HOCHTIEF's and the subcontractor's responses for one evaluation criterion and then evaluate the follow-up questions generated by the assistant for that criterion, rating their clarity, relevance, and usefulness in addressing data gaps or contradictions.

The third criterion, *"Justification and Transparency of Ratings*", examines whether each rating is clearly explained and backed by evidence from the evaluation data. Participants will review excerpts from the AI-generated report containing individual ratings and justifications. The fourth criterion, "Quality of Lessons Learned and Recommendations," evaluates the practical value and specificity of the AI-generated lessons and suggestions. To assess this, participants will review the "Lessons Learned" and "Recommendations" sections in the report.

The second-to-last criterion, *"Perceived Added Value",* aims to capture overall user satisfaction and the perceived usefulness of the AI-enhanced evaluation tool. The final criterion, *"Perceived Limitations*

and Risk Awareness", gauges users' recognition of potential shortcomings of the assistant, such as oversimplification, lack of context, or overreliance on written input, and their judgment on whether the outputs are reliable enough for real-world application.

For each criterion, participants will review specific sections of the AI-generated report and respond to statements on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree). A threshold of 3.5 has been set for all six criteria. For the first five criteria, an average rating above 3.5 indicates success. For the last criterion, which contains negatively worded statements, an average below 3.5 suggests the tool's outputs are not perceived as problematic. The complete survey is available in Appendix G.

Table 3 outlines each criterion, its purpose, and how it will be measured.

***Table 3**: Validation Criteria for Assessing the Effectiveness of the AI-Enhanced Subcontractor Evaluation Framework*

Criterion	What It Evaluates	Measurement Method	Success Threshold
1. Insightfulness	The extent to which the AI-generated evaluation report provides more detailed, meaningful insights than existing reports	Likert-scale statements (1-5)	Mean Likert score of **≥ 3.5**
2. Clarity and Relevance of Follow-up Questions	The assistant's ability to generate helpful and context-aware follow-up questions to resolve inconsistencies	Likert-scale statements (1-5)	Mean Likert score of **≥ 3.5**
3. Justification and Transparency of Ratings	Whether ratings are justified with logical reasoning and supporting data	Likert-scale statements (1-5)	Mean Likert score of **≥ 3.5**
4. Quality of Lessons Learned and Recommendations	The specificity and practical relevance of the AI-generated insights for future projects	Likert-scale statements (1-5)	Mean Likert score of **≥ 3.5**
5. Perceived Added Value	Overall user satisfaction and perceived utility of the AI-enhanced tool compared to the existing process	Likert-scale statements (1-5)	Mean Likert score of **≥ 3.5**
6. Perceived Limitations and Risk Awareness	Users' recognition of the assistant's potential shortcomings	Likert-scale statements (1-5)	Mean Likert score of **< 3.5**
7. Accuracy of Scores	The consistency between AI-generated and human-generated scores	Numerical comparison (MAPD across scores)	Mean deviation **≤ 10%**

4.4 Project Phases and Timeline

The project will be conducted in six phases spanning from March 18th to July 28th, 2025.

- **Phase 1: Overview of Evaluation Process and Defining Key Evaluation Criteria for Subcontractor Performance (18th of March – 8th of April 2025)**

In this initial phase, interviews will be conducted with HOCHTIEF's employees to gain a comprehensive understanding of the current subcontractor evaluation process. The objective is to map out the steps followed by HOCHTIEF when evaluating subcontractors, identify the main issues and bottlenecks, and understand the criteria currently used. Concurrently, a literature review will be conducted to identify evaluation criteria for subcontractors' performance. The literature review will explore various frameworks and metrics used in the construction industry. During the interviews, the criteria will be ranked, and weights will be assigned to them. The weights will be used in later stages to conduct the WSM.

- **Phase 2: Designing the Evaluation Template and Weighting Scheme (9th of April – 30th of Apr 2025)**

Phase 2 focuses on collecting the necessary performance data related to the identified evaluation criteria. Utilizing this data, an evaluation template will be created, standardizing how data is collected from subcontractors. It will include questions for each of the criteria identified in the literature review. The answers to these questions will be qualitative and quantitative, for example, a rating of 1 through 5, open-text answers, or data like agreed contract price, actual expenses, or safety incidents reported. It ensures consistency and comprehensiveness in subcontractor assessments across different projects. Two versions will be created. Company employees will answer one version, and the subcontractor will answer the other. Simultaneously, the weighting scheme for each evaluation criterion will be finalized.

Once the draft templates are created, they will be validated by a small group of stakeholders to assess their clarity and effectiveness. Feedback from the pilot group will be used to refine and improve the evaluation templates before wider deployment. Once the documents have been finalized, they will be distributed to the subcontractor and the company employees involved in the evaluation process.

- **Phase 3: Gathering Responses and Fine-Tuning NextChat assistant (1st of May –24th of May)**

While awaiting questionnaire responses, a virtual assistant will be developed using NextChat. That assistant will act as an evaluator and will be used to assess the performance of the subcontractor. This assistant will be fine-tuned using past evaluations from HOCHTIEF's Amaliahaven project. The aim is to improve NextChat's ability to detect ambiguities and generate relevant follow-up questions.

Once the answered evaluation documents are collected, they will be given to NextChat to generate follow-up questions. These questions will then be asked to company employees and the subcontractor either through mail or during a face-to-face meeting.

- **Phase 4: AI-Generated Rating & Comparative Benchmarking (26th of May – 13th of June)**

In Phase 4, the answers to NextChat's follow-up questions will be integrated with the responses to the evaluation document to generate an AI-driven subcontractor score for each criterion. Using the Weighted Sum Model (WSM), NextChat will calculate an overall score for the subcontractor's performance. The final deliverable of the analysis will be a PDF report, which will include the final AI-

generated rating, a summary of the collaboration between the subcontractor and HOCHTIEF, the lessons learned, and recommendations for future projects.

- **Phase 5: Results Analysis, Recommendations, and Final Refinements (16th of June – 28th of July 2025)**

Phase 5 focuses on analyzing and validating the AI-supported subcontractor evaluation approach developed in this study. The validation will be done as described in Chapter 4.4. The results will be extensively analyzed and discussed alongside the limitations of this research. Based on the results from the validation and the insights captured in the interviews with HOCHTIEF employees, a set of recommendations will be developed to improve HOCHTIEF's procurement practices. Lastly, the implications for the construction industry will be discussed, and recommendations for future research will be given.

Chapter 5: Analysis of HOCHTIEF's Current Subcontractor Evaluation Process

This chapter examines the evaluation procedure from project completion, highlighting key steps, criteria, departmental involvement, challenges, and potential improvements. The contents of this chapter have been collected from interviews with HOCHTIEF employees with different roles in the company, including project managers, process managers, work preparators, procurement managers, quantity surveyors, and HSE managers. In total, seven interviews were conducted. The questionnaire used can be found in Appendix A.

5.1 Evaluation Process Steps

The company's evaluation process consists of eight steps:

Step I: Create an Evaluation Document Specific to the Project

The subcontractor evaluation begins with the creation of a project-specific evaluation document. This document includes specific questions tailored to the project requirements, client, and partners. These questions are designed to capture nuances of the project's execution and subcontractor performance and are specific to the criteria used by the company. An example of the questions included in an evaluation document for a past project can be found in Appendix H in Table H1.

Step II: Distribute Among Evaluators

Once the evaluation document is crafted, it is distributed to project-specific evaluators. These evaluators include employees who have had firsthand experience with subcontractors involved. Their insights are crucial in creating a comprehensive evaluation that reflects actual performance.

Step III: Discuss Results During Meetings

Initially, in the prequalification phase, it is explained to the subcontractors what HOCHTIEF expects from them and what they need to do. According to the interviews, during the first meeting, the evaluators are guided on interpreting results and what action needs to be pursued. Throughout the project's duration, escalation meetings are held as needed, based on the results or when organizational changes necessitate additional scrutiny. The goal of these meetings is to act as a proactive measure to avoid unpleasant situations or unwanted results and to improve the existing process and quality of work.

Step IV: Guide Approval of the Evaluation Document

The evaluated document goes through an approval flow, where results are summarized and verified. The document must be signed off by project managers, ensuring the evaluation reflects accurate project experiences and feedback.

Step V: Conduct Follow-Up Meeting with Supplier

If necessary, a follow-up meeting with the subcontractor is organized based on internal discussions. This meeting aims to address areas needing improvement and align subcontractor expectations with project standards.

Step VI: Summarize Scores to Add to Procurement Plan

Scores from the evaluations are compiled and reflected in the procurement plan, integrating subcontractor performance into future project decision-making processes.

Step VII: Archive Evaluation Documents on SharePoint

The evaluation documents are archived on SharePoint, ensuring accessibility and record-keeping for reference in future project planning and procurement processes.

Step VIII: Add Data to Vendor Database

The summarized data is added to the vendor database, contributing to a comprehensive record that aids in prequalification and selection processes in subsequent projects.

5.2 Departments Involved

In the evaluation process, employees with different roles participate. The project manager is responsible for assessing the quality of the work delivered, adherence to the schedule, progress made, and the subcontractor's collaboration with HOCHTIEF. The Health, Safety, Security, and Environment (HSSE) manager evaluates safety awareness, subcontractors' certifications, and their approach to sustainability and environmentally friendly practices. Finally, the quantity surveyor or the quality assurance manager oversees the quality of the documentation and ensures that the delivered work meets the contractual specifications.

Through interviews, it was concluded that subcontractors are typically informed of areas needing improvement during project execution. Final evaluations are shared with key subcontractors to close any loose ends, though the feedback varies. Extensive feedback is provided when subcontractors receive poor evaluations, prompting discussions about potential improvements throughout the project lifecycle and in the final meeting.

The only feedback HOCHTIEF receives from subcontractors and suppliers they collaborate with is that given during progress or escalation meetings, if necessary. There is no official document or template for HOCHTIEF’s partners to provide feedback or share their perspective on collaboration.

5.3 Current Criteria for Subcontractor Evaluation

Subcontractors are evaluated based on several criteria, which are the same for all types of work and projects. These criteria were determined by the head office of HOCHTIEF in Essen and are the same for all departments of HOCHTIEF. These criteria include

- Quality: Engineering and product quality.
- HSE Compliance: Adherence to health, safety, and environmental standards.
- CSR Practices: CO2 reduction strategies, waste management, and compliance with sustainability standards.
- Planning & Timeliness: Adherence to timelines and deliveries.
- Cooperation & Documentation: Compliance with agreements and thorough documentation practices.

For each criterion, specific questions are asked in the current evaluation document that HOCHTIEF has created. From the table H1, it is evident that the answer to all of the questions in the current document is a rating from 1 to 5:

- Very bad (Grade 1),
- Insufficient (Grade 2),
- Satisfactory (Grade 3),
- Well sufficient (Grade 4) and
- Very good (Grade 5)

The evaluators are also given the field titled “Comments” for each question, where they can add an explanation for their rating and provide any additional information they believe is important. The final question of the evaluation document asks the evaluators if they would be willing to work with that particular subcontractor or supplier in the future. The average of the ratings provided is then the final rating of the subcontractor’s performance.

According to the interviews, the results of the evaluations can lead to a subcontractor's exclusion from future projects if their overall score falls below 2.5. This exclusion is part of the prequalification process. So, in the prequalification process for the next project, a subcontractor with a rating of 2.5 in a previous one may be excluded from consideration. Severe safety or behavior issues may result in permanent blockage in the vendor database or expulsion from tender or project considerations.

However, that decision is not always simple. According to one employee, “I have seen low scores that should have led to elimination, but it never happened. I think it happens because the score depends on the people you have worked with in the company. You do not want to judge the whole company based on a few employees. It is relationship management”. Another employee mentioned, “Even though a supplier or subcontractor has a bad evaluation, we might still need them because they might be the only ones who can carry out a certain job. Maybe the team was bad, not the overall company. For example, a bad result might come because a company does a project for the first time in the Netherlands, the reason is not always obvious”. From these comments, it is evident that the results of the evaluation are important and play a role in a future project. However, sometimes due to unavoidable circumstances, the final decision may override the evaluation results.

5.4 Challenges and Potential Improvements

During the interviews, employees were asked to identify challenges they face with the current process. One of the employees noted that evaluations are difficult to find in the company’s database. More specifically, the employee commented: “Evaluations are not easily accessible, limiting learning and insight-sharing across projects”. Another employee added to that challenge and mentioned, “I have never received anything. From my perspective, at least there is no standardized template. The evaluations are not shared at all levels, so you cannot manage your expectations. Even though we have information about the subcontractor from past collaborations or previous evaluations, they are not shared, so it is not known to you “. Based on these answers, it is evident that even though there is a standardized template and all the evaluations are available on the company’s database, employees are not aware of their existence if they are not the ones doing them, or often have difficulties finding them.

Another issue mentioned is that staff turnover leads to the loss of evaluation data, which in turn affects informed future decisions. Lessons learned are not captured effectively, and people taking over a role in the project are often left in the dark. When they eventually conduct their evaluations, they have

limited information because data from previous phases or stages of the project is unavailable to them. One employee mentioned, “Lessons learned are not carried forward. When a person is leaving a project, there is usually a formal process of data transfer. That handover is not very effective. It happens, but sometimes not, so information is lost, and you take over a job without knowing what to expect”. It was also mentioned that with the current template, there is a lack of context in the evaluations. Interviewees mentioned that “When you look at the evaluation document, you just see some questions and a rating. However, that rating is often not accompanied by an explanation, so you do not get much out of it. You see someone's opinion. That person might not even be part of the company anymore, so you cannot even ask them for more information. So important information might be lost”.

Furthermore, one employee noted a lack of transparency in the evaluation process, while another observed that evaluators' emotions often influence their assessments, making them less objective. Lastly, because specific projects are relatively new to the Dutch market, companies capable of executing the necessary work are in high demand. Consequently, evaluators are often reluctant to evaluate them negatively, as this might impact HOCHTIEF’s future collaboration options. More specifically, an employee commented, “When it comes to new projects, sometimes the options are limited, and you do not have many options to work with. So even when a subcontractor's or a supplier’s performance is low, you do not want to give negative feedback because there are not a lot of other companies who can do that job for you in a similar project in the future”.

As part of the interview process, employees were also asked to provide possible improvement measures that could be implemented. One of the few suggestions was to implement a company-wide database containing standardized evaluation data to ensure accessible information for guiding future decisions and to avoid repeating mistakes. The same employee emphasized the importance of completing an evaluation of the project's progress up to their departure. This evaluation should be handed over to the employees taking over the project and included in the project archives. The employee mentioned, “It should be company policy that before somebody leaves, they should fill in evaluations, and these should be part of the handover and become part of the project archives to be shared internally and used in other projects. We need to create a common database with evaluations not to repeat mistakes and to make informed decisions when selecting subcontractors”.

Moreover, two employees suggested that more evaluations should take place throughout the project. This way, the company can provide regular feedback to the subcontractors, be more aware of the project's status and the performance of everyone involved, and be more prepared to act if the situation does not improve. These evaluations should be part of the monthly reports and not only include ratings as in the current template, but also include key performance indicators (KPIs) to be even more informative. More specifically the employees commented: “Do more evaluations throughout the project so then depending on the results we can inform them, take action and if nothing happens or if the change is not enough, we make a change” and “Evaluations should be part of the Monthly Report to have continuous monitoring of the project and should also include KPIs”.

In addition to these suggestions, interviewees emphasized the importance of clearly communicating expectations from the outset and informing subcontractors at the beginning of the project about the evaluation criteria they will be assessed against. One of the interviewees mentioned that “The solution would be to make sure that we are very clear about our expectations both to the client and to the subcontractors. That starts with the contracts. Sometimes, certain tasks are not included in the scope of some requirements. From the tender phase, we need to be very clear about our expectations”.

Another employee commented: “Subcontractors should be aware from the start of the project, based on which criteria they will be evaluated. The questions you ask them and the things you look at to determine if they are a good fit during the prequalification phase should be similar to the criteria you use in the end to evaluate their performance”.

5.5 Integration of AI in HOCHTIEF’s procurement process

The interviewees were also asked to provide their opinions on integrating AI into the company's procurement process. They were asked about the potential benefits of using AI and about specific issues that might arise if that integration does happen. Most of them were not familiar with AI or had never used it for procurement purposes, including subcontractor evaluations.

From the seven interviews, it was concluded that by adopting AI, they expect quicker and more efficient analysis of data, a less biased evaluation, and a more informative final report of the subcontractors’ performance. However, they expressed concerns regarding data quality and bias in data, while also stressing that before AI can be used more extensively, careful validation is necessary to ensure effectiveness.

5.6 Conclusion

The current evaluation process at HOCHTIEF includes eight clear steps, from the creation of a project-specific evaluation document to archiving and integrating evaluation results into future procurement decisions. The use of standardized criteria such as quality, HSE compliance, planning, cooperation, and CSR practices provides a solid framework for ensuring consistent assessments across projects. However, interviews with staff across different roles have shown that the current process's practical implementation faces several limitations, particularly in consistency, accessibility, and feedback mechanisms. The evaluations are primarily top-down and one-sided, with minimal feedback from subcontractors, indicating an area where communication and mutual assessment could be significantly improved.

One of the most pressing issues is the inaccessibility of evaluations within the company’s database, leading to missed opportunities for shared learning and informed decision-making. Moreover, the evaluation process suffers from subjectivity, emotional bias, and inconsistencies. Staff turnover further exacerbates this by interrupting the continuity of information, which is critical for long-term subcontractor relationship management. The strategic dilemma of retaining poor-performing subcontractors due to their unique capabilities or limited market alternatives further complicates and hinders the evaluations.

To address these shortcomings, interviewees suggested several recommendations, which will be discussed more extensively in Chapter 10. Lastly, Interviewees were asked about integrating AI into the company's procurement process. Most had little to no experience with AI in this context. While they saw potential benefits, such as faster data analysis, reduced bias, and more insightful subcontractor reports, they also raised concerns about data quality and the risk of bias within the data itself. Overall, they emphasized the need for thorough validation before AI can be used more broadly.

Chapter 6: Criteria Identification and Weighing

This chapter begins by presenting the key criteria identified from the literature, which will be used in the proposed framework. Then it explains how appropriate weights will be assigned to them. This chapter conducts a comparative analysis of two weighting scenarios. These are Scenario A, which applies equal weighting, and Scenario B, which uses role-weighted influence. This examination aims to highlight the effectiveness of each approach and its impact on the ranking and weights of selected evaluation criteria.

6.1 Criteria Identified by Literature

The literature review conducted in Chapter 2 made it apparent that practical subcontractor evaluation in construction projects requires a multidimensional framework combining both quantitative and qualitative criteria. Several key criteria for subcontractor evaluation were identified. These criteria provide a comprehensive framework for assessing subcontractor performance by ensuring that many different aspects are considered.

6.1.1 Quality of Workmanship

This criterion evaluates the subcontractor's ability to produce high-standard interior work that meets design specifications and industry quality benchmarks. Consistent high quality across various project phases is crucial for maintaining a firm's reputation and client satisfaction (Liu et al., 2018; Polat, 2015). High-quality workmanship is pivotal because it directly impacts project longevity and client satisfaction by reducing maintenance costs and ensuring safety. El-Khalek et al. (2018) emphasize that superior workmanship is paramount because it not only reduces future maintenance costs but also enhances the overall reputation of the project, thereby contributing directly to project success. Basu et al. (2017) further advocate that consistent job quality is critical to meeting both safety and performance goals. According to Polat (2015), factors such as defect rates, precision in execution, and adherence to verified quality standards are critical for distinguishing high-performing subcontractors.

6.1.2 Timeliness and Schedule Adherence

This criterion assesses a subcontractor's ability to deliver materials and complete tasks according to the predetermined project timelines. Timeliness is critical to prevent cascading delays and cost overruns. Effective schedule management mitigates the risk of project overruns and enhances operational efficiency (Maestrini et al., 2018). Consistent on-time performance is a key indicator of operational efficiency and is supported by findings from Mutai (2016), who noted the critical role of punctuality in enhancing project outcomes. Research by Olanrewaju et al. (2021) underlines that on-time delivery of materials and the efficient management of critical activities are among the most significant predictors of a subcontractor's overall performance. Additionally, El-Khalek et al. (2018) note that delays rooted in schedule slippages can jeopardize project success, thus underscoring the necessity of proactive monitoring and management of timelines to ensure uninterrupted project progress.

6.1.3 Cost Management

Cost Management focuses on the subcontractor's capability to control expenses and adhere to budgets throughout the project lifecycle. It involves real-time cost tracking, variance analysis, and an overall financial oversight to detect inefficiencies early. Accurate financial oversight is essential for maintaining project profitability and achieving cost efficiency (Hu & Ren, 2023; Murekatete & Dushimimana, 2023). According to Polat (2015), traditional decision-making models have long incorporated price and cost-risk factors as principal evaluation attributes. Effective cost management not only affects profitability but also minimizes financial risks for the prime contractor. Practices such as real-time cost tracking and variance analysis ensure that cost overruns are detected promptly, allowing for immediate corrective measures. Sadeghpour and Isaac (2015) provide insights into how robust cost management practices contribute to the overall viability of projects. Furthermore, Olanrewaju et al. (2021) highlight that robust financial documentation is essential to avoid cost overruns. Sudarsanam et al. (2022) further demonstrate that AI-based approaches, such as Fuzzy AHP, contribute to a more systematic interpretation of cost data, ensuring that economic decisions are both transparent and resilient to fluctuations.

6.1.4 Safety and Compliance

This criterion measures how well subcontractors adhere to established safety protocols and industry regulations. Safety and compliance are non-negotiable in construction projects. Protocols include documented safety records, provisions of personal protective equipment, and certifications that ensure a safe working environment. Silva et al. (2022) and Honhon et al. (2012) state that strict safety measures not only reduce the risk of on-site accidents but also contribute to a more reliable and compliant execution of projects, thereby protecting both human resources and material investments. Olanrewaju et al. (2021) further recommend that rigorous safety practices should be non-negotiable when pre-qualifying subcontractors for involvement in complex construction projects. This claim is also supported by Polat (2015), who emphasizes that consistent implementation of safety standards and adherence to health regulations are crucial determinants of reliability, with numerous models highlighting the importance of safety indices in prequalifying subcontractors.

6.1.5 Communication and Collaboration

This criterion will assess the subcontractor's ability to maintain clear and effective communication with all stakeholders and to foster collaborative teamwork, criteria that are crucial for conflict resolution and efficient operations, as highlighted by Yin et al. (2017). According to Polat (2015), interpersonal exchanges and cooperative behavior are important qualitative indicators. Structured communication protocols—such as regular inter-disciplinary meetings, digital collaboration platforms, and stakeholder surveys—greatly enhance coordination among project teams (Yin et al., 2017). These mechanisms ensure that problems are promptly addressed and that there is a continuous flow of information between subcontractors and the main contractor.

6.1.6 Reliability

Reliability measures the consistency and dependability of a subcontractor's performance over multiple projects. Sundquist et al. (2018) note that dependable performance is fundamental to sustaining successful collaborations. According to the authors, a strong track record of reliability is confirmed not only through past performance data but also by the presence of robust contingency measures to

handle unexpected disruptions. Furthermore, they emphasize that a reliable subcontractor is a key indicator of underlying operational stability and is integral to maintaining long-term partnerships.

6.1.7 Technical Ability and Qualifications

This criterion examines both the practical skills and the formal qualifications of a subcontractor. Ongoing professional development, certification renewals, and exposure to emerging technologies are highlighted as key drivers that support sustained technical excellence (Polat, 2015). The focus on continuous technical improvement ensures that subcontractors remain competitive and capable of addressing complex construction challenges. According to Ngowi and Pienaar (2005), subcontractors with strong technical foundations are more likely to succeed in meeting the multifaceted demands of modern construction projects.

6.1.8 Financial Stability

Ensuring that subcontractors possess robust financial health is critical to the sustainability of any project because stable financial health minimizes the risk of project delays and cost overruns due to cash flow issues (Polat, 2015). This criterion evaluates whether a subcontractor possesses the fiscal strength to manage their operations without risking insolvency, ensuring that they can sustain project investments. Sadeghpour and Isaac (2015) stress the need for a secure financial footing and propose a multi-layered approach to evaluating financial stability, which includes analyzing financial statements, credit ratings, and historical payment performance. Such comprehensive financial evaluations help guarantee that subcontractors can support long-term project demands without jeopardizing timely delivery due to financial constraints.

6.1.9 Sustainability Practices

Sustainability practices have become increasingly significant as projects aim to minimize their environmental impact and enhance energy efficiency. Sustainability practices assess a subcontractor's commitment to environmentally friendly methods and responsible resource utilization, focusing on activities such as waste management and energy efficiency; Silva et al. (2022) and Osiro et al. (2021) provide data-backed insights that underscore the increasing importance of sustainable construction practices in the evaluation process.

6.1.10 Innovation and Problem-Solving

Innovation and Problem-Solving measure a subcontractor's ability to adopt new technologies and address unexpected challenges creatively, which is essential for continual improvement and adaptability in dynamic project environments. According to Wang et al. (2025), innovative approaches can lead to improved predictive accuracy and more effective decision-making. Subcontractors who invest in innovative technologies and continuous improvement programs achieve higher project performance. These subcontractors often demonstrate an aptitude for quickly adapting to changes in project scope or overcoming unexpected obstacles (Keshavarz-Ghorabaee et al., 2018).

6.2. Assigning Weights to the Evaluation Criteria

All identified criteria from the literature are incorporated into the evaluation template to ensure a comprehensive assessment of the subcontractor's performance. Each criterion is assigned a specific weight based on its importance to HOCHTIEF's strategic objectives, as determined through stakeholder interviews. During these interviews, stakeholders were asked to rank the 10 criteria based on their importance. As seven employees participated in the interviews, seven different rankings were created (see Table 4).

***Table 4**: Individual Rankings of Subcontractor Evaluation Criteria by Role*
This table presents the ranking (1 = most important, 10 = least important) of key subcontractor evaluation criteria as assessed by various project roles, where SWP is Senior Work Preparator. ProjM is Project Management, ProcM is Procurement Manager, SM is Safety Manager, QAM is Quality Assurance Manager, and QS is Quantity Surveyor. A dash implies that the interviewee could not provide a rank for that criterion.

Criterion	SWP 1	SWP 2	ProjM	ProcM	SM	QAM	QS
Quality of Workmanship	2	2	4	3	2	2	2
Timeliness and Schedule Adherence	3	3	3	2	3	3	5
Cost Management	6	5	6	4	8	4	3
Safety and Compliance	1	1	2	1	1	1	1
Communication and Collaboration	4	4	5	6	5	6	4
Reliability and Repeat Engagement	5	6	7	5	7	6	5
Technical Ability and Qualifications	3	3	1	7	4	5	5
Financial Stability	7	8	8	8	10	4	4
Sustainability Practices	8	9	9	10	9	8	-
Innovation and Problem-Solving	9	7	10	9	6	7	-

These rankings were used to create an average ranking for each criterion. The average rank was calculated by using the following equations:

$$Average\ Rank = \frac{Sum\ of\ Ranks}{Number\ of\ Raters} \qquad (5.1)$$

The lower the average rank, the higher the importance of the criterion.

Initially, the interviewees were asked to assign a percentage weight based on importance. However, some of them found it challenging to assign percentages that would ultimately equal 100%. So instead, the average rank was converted to a weight for each criterion. The following process was used:

The average rank was inverted:

$$Inverse\ Rank = \frac{1}{Average\ Rank} \qquad (5.2)$$

1. Then normalized into a 0–100% scale:

$$Weight\ (\%) = \frac{Inverse\ Rank}{\sum All\ Inverse\ Ranks} \times 100 \quad (5.3)$$

This ensures that all weights add up to 100%, with the criteria ranked as most important (lower rank) receiving a higher percentage.

However, given the diversity of roles and the differing levels of involvement in the subcontractor evaluation process, a key methodological consideration was how to aggregate these multiple perspectives into a single set of criterion weights that reflects both inclusiveness and decision-making realities.

The initial approach averaged the weights provided by all seven interviewees equally, assuming each person's input had an identical influence. While this method is objective and straightforward, it may not reflect the actual decision-making structure within the organization, where specific roles hold more responsibility or authority over subcontractor evaluations.

To address this, two weighting scenarios were developed and compared:

- **Scenario A: Equal Influence Model** — All interviewees' weights were averaged with equal importance.
- **Scenario B: Role-Weighted Influence Model** — Interviewee weights were aggregated using influence weights assigned to each role, reflecting their actual involvement and authority in subcontractor evaluation.

6.2.1 Scenario A: Equal Influence Model

In this scenario, each of the seven interviewees was treated as an equally weighted contributor to the evaluation process. Instead of providing direct percentage weights for each criterion, each interviewee was asked to rank the criteria in order of importance. The rankings were converted into percentage weights using the method described above. The calculated weight for each criterion can be found in Table 5.

Table 5: *Final ranking, average ranking, and weight for each criterion in Scenario A*

Rank	Criterion	Avg Rank	Weight (%)
1	Safety and Compliance	**1.14**	**31.23%**
2	Quality of Workmanship	**2.43**	**14.64%**
3	Timeliness and Schedule Adherence	**3.14**	**11.34%**
4	Technical Ability & Qualifications	**4.00**	**8.90%**
5	Cost Management	**4.86**	**7.32%**
6	Reliability and Repeat Engagement	**5.14**	**6.92%**
7	Financial Stability	**5.86**	**6.07%**
8	Communication and Collaboration	**7.00**	**5.09%**
9	Innovation and Problem-Solving	**8.00**	**4.45%**
10	Sustainability Practices	**8.83**	**4.03%**

This scenario is based on the idea that everyone interviewed has the same level of knowledge and influence. It uses clear and simple math to be fair and straightforward. By ignoring the usual hierarchy in an organization, it avoids judging based on someone's position or power, thereby reducing potential bias from emphasizing certain positions or individuals excessively.

In terms of advantages, this method is simple and transparent, making it easy to understand and implement. Treating all perspectives equally allows it to circumvent subjective bias that might arise from assigning varying levels of influence. Furthermore, this approach reflects a collaborative and inclusive attitude, ensuring all voices are given uniform attention and consideration in the evaluation process.

However, there are some downsides. It fails to acknowledge the realistic organizational disparity where roles carry differing degrees of influence in subcontractor evaluations. Different people have different levels of influence when evaluating subcontractors, depending on their role. Consequently, individuals who are not directly involved in the evaluation process may have an impact equal to that of the final decision-makers, which might not accurately reflect actual practice. This method may oversimplify the weighting process and potentially lead to inaccurate prioritization of criteria. Additionally, it risks diminishing the input from those with the most significant expertise or decision-making power, which could affect the overall effectiveness of the evaluation.

6.2.2 Scenario B: Role-Weighted Influence Model

In Scenario B, weights were assigned to each interviewee based on their role within the organization. This scenario reflects the reality that in construction firms, different professionals have different levels of authority, involvement, and expertise in evaluating subcontractors. In this scenario, each interviewee's input is weighted based on their role in the company, reflecting their actual influence on subcontractor-related decisions.

The project manager (25%) and procurement manager (25%) were assigned the highest weights due to their central roles in subcontractor selection and overall project execution. The project manager holds primary responsibility for coordinating subcontractor performance on-site, while the procurement manager oversees the selection, contracting, and negotiation processes. The quality assurance manager (15%) and safety manager (15%) were weighed based on their critical roles in ensuring subcontractor compliance with quality standards and safety regulations, respectively. The senior work preparators (5% each) were assigned lower weights, as they do not formally participate in subcontractor evaluation decisions. Based on the interviews conducted, their role is limited to providing technical input and advisory opinions during meetings, rather than being involved in final evaluations. In contrast, the quantity surveyor (10%) was assigned a higher weight because, despite primarily handling cost tracking and budget management, they are one of the final evaluators involved in assessing subcontractor performance. Table 6 shows the adjusted weights assigned to each interviewee role for Scenario B, based on their actual involvement in the subcontractor evaluation process. These weights reflect the relative influence of each role, as determined through interviews conducted with company personnel.

***Table 6**: Adjusted role-based weights and justifications used in Scenario B*

Role	Assigned Weight	Justification
Project Manager	0.25	Has final responsibility for project execution and subcontractor performance; typically, a key decision-maker.
Procurement Manager	0.25	Directly responsible for selecting and contracting subcontractors; central to commercial and operational decision-making.
Safety Manager	0.20	Ensures regulatory and on-site safety compliance; critical evaluator of subcontractor behavior and compliance.
Quality Assurance Manager	0.10	Oversees compliance with specifications and quality standards; plays an active role in evaluating subcontractor output.
Quantity Surveyor	0.10	Manages budgeting and cost tracking but has less direct influence on performance-based evaluation decisions.
Senior Work Preparator 1	0.05	Provides technical input but does not formally participate in evaluations.
Senior Work Preparator 2	0.05	Same as above; consulted for technical opinions but not a formal evaluator.
Total	**1.00**	—

In Scenario A, each interviewee's ranking contributed equally. In Scenario B, the rankings are adjusted by applying a weight to each interviewee's input, based on their role.

The formula for the weighted average rank of each criterion is:

$$Weighted\ Average\ Rank_i = \frac{\sum_{j=1}^{N} R_{ij} \times W_j}{\sum_{j=1}^{N} W_j} \quad (5.4)$$

Where:

i is the index for each criterion
j is the index for each respondent/role
N is the total number of respondents who provided a rank for that criterion.
R_{ij} is the rank that respondent *j* gave to criterion *i*
W_j is the weight assigned to the opinion of respondent *j* based on their role

The following steps are similar to scenario A. Since a lower rank indicates higher importance, ranks need to be converted by inverting them:

$$Inverted\ Score_i = \frac{1}{Weighted\ Average\ Rank_i} \quad (5.5)$$

The weights must add up to 100%, so they must be normalized. So:

$$Final\ Weight_i = \frac{Inverted\ Score}{\sum_{i=1}^{10} Inverted\ Score_i} \times 100 \quad (5.6)$$

Where:

i is the index for each criterion

Table 7 includes the calculated weighted average rank for each evaluation criterion based on the input from different roles and their assigned role weights. The final rank and normalized weight (as a percentage) for each criterion are also included.

***Table 7**: Final ranking, average ranking, and weight for each criterion in Scenario B*

Final Rank	Criterion	Weighted Avg Rank	Final Weight (%)
1	Safety and Compliance	**1.25**	**30.12%**
2	Quality of Workmanship	**2.75**	**13.69%**
3	Timeliness and Schedule Adherence	**2.95**	**12.76%**
4	Technical Ability and Qualifications	**4.00**	**9.18%**
5	Communication and Collaboration	**5.15**	**7.31%**
6	Cost Management	**5.35**	**7.04%**
7	Reliability and Repeat Engagement	**6.05**	**6.22%**
8	Financial Stability	**7.55**	**4.99%**
9	Innovation and Problem-Solving	**8.28**	**4.55%**
10	Sustainability Practices	**9.11**	**4.03%**

This approach of incorporating hierarchical influence and relevance of expertise into subcontractor evaluations is more aligned with the decision-making dynamics observed in real-world settings. By factoring in the importance and weight of each role within HOCHTIEF, this method acknowledges that decision-making is often shaped by the varying levels of influence and specialized knowledge inherent in different positions. This allows for a more realistic depiction of the evaluation processes within a company, offering insights that are both practical and reflective of operational realities.

However, despite its strengths, this method introduces an element of subjectivity. The task of assigning weights based on the importance of roles requires interpretation, which inevitably brings in subjective judgment.

6.2.3 Comparative Analysis of Scenario A and Scenario B

Table 8 provides a structural context for understanding how the two scenarios differ in approach and implication. Scenario A treats each evaluator's input with equal weight, representing a consensus-driven model. In contrast, Scenario B assigns different weights according to professional roles and their relevance to overseeing subcontractors, thereby aligning the evaluation more closely with the realities of project execution.

***Table 8**: Methodological Overview of Scenario A and Scenario B.*

Aspect	Scenario A	Scenario B
Weighting Method	Equal weighting of all roles	Weighted roles based on their relevance/authority in subcontractor evaluation
Assumption	All individuals have an equal impact on the final criteria rankings	Certain roles (e.g., Project Manager, Procurement, Safety) have more influence on the result
Result Interpretation	Reflects a democratic average of expert input	Reflects a more hierarchical, responsibility-weighted assessment

Table 9 presents the average rank, the final rank, and the final weight for each criterion from Scenario A and Scenario B. The side-by-side rank columns allow for an immediate visual comparison of how criteria are prioritized and weighed depending on the evaluation method.

***Table 9**: Comparison of Outcomes in Scenario A and Scenario B.*

Criterion	Avg Rank (A)	Final Rank (A)	Final Weight (A)	Avg Rank (B)	Final Rank (B)	Final Weight (B)
Safety and Compliance	1.14	1	31.23%	1.25	1	30.12%
Quality of Workmanship	2.43	2	14.64%	2.75	2	13.69%
Timeliness and Schedule Adherence	3.14	3	11.34%	2.95	3	12.76%
Technical Ability and Qualifications	4.00	4	8.90%	4.00	4	9.18%
Communication and Collaboration	4.86	5	7.32%	5.15	5	7.31%
Cost Management	5.14	6	6.92%	5.35	6	7.04%
Reliability and Repeat Engagement	5.86	7	6.07%	6.05	7	6.22%
Financial Stability	7.00	8	5.09%	7.55	8	4.99%
Innovation and Problem-Solving	8.00	9	4.45%	8.28	9	4.55%
Sustainability Practices	8.83	10	4.03%	9.11	10	4.03%

From Table 9 it is evident that Safety and Compliance remains the top priority in both scenarios. However, its weight slightly decreases in Scenario B due to the influence of role-specific perspectives. This shift illustrates how assigning different levels of influence to evaluator roles can moderate the dominance of any single criterion. Quality of Workmanship also maintains consistent importance, showing that regardless of weighting models, quality is universally valued across all organizational levels and project phases.

Timeliness and Schedule Adherence gain prominence in Scenario B. At the same time, Technical Ability, Communication and Collaboration, and Cost Management show only minor shifts, with the latter becoming slightly more emphasized when budget-focused roles carry more influence. Reliability and Repeat Engagement and Financial Stability maintain stable but mid-tier importance, reinforcing their role in long-term partnerships and organizational resilience. Meanwhile, Innovation

and Problem-Solving and Sustainability Practices consistently rank lowest, suggesting a tendency to prioritize compliance and immediate deliverables over long-term or transformative strategies.

6.3 Conclusion

Chapter 6 starts by presenting ten evaluation criteria that will be used in the proposed AI-driven MCDA framework. These criteria were derived from an extensive literature review. This multidimensional framework ensures that both technical execution and strategic alignment are factored into performance assessments.

The core of the chapter lies in assigning weights to the criteria by comparing two weighting models: the *Equal Influence Model* (Scenario A) and the *Role-Weighted Influence Model* (Scenario B). Scenario A promotes egalitarian input but risks oversimplifying complex decision dynamics. In contrast, Scenario B aligns with organizational hierarchies by giving more weight to roles such as project managers and procurement leads, thereby better reflecting real-world decision-making priorities. The ranking of the criteria remained the same in both cases. However, the weights showed fluctuation, highlighting how considering the importance of certain roles' opinions affects the focus and emphasis of the assessment process.

By capturing both the inclusive and role-specific perspectives, the chapter offers insights into how weightings influence subcontractor rankings and decision-making processes. These insights form a critical bridge to the development of the AI-enhanced evaluation assistant in Chapter 7, which will use the identified criteria and their assigned weights to evaluate subcontractors' performance.

Chapter 7: Development of the Proposed Hybrid Approach

This chapter presents the development, structure, and functionality of the AI-powered subcontractor evaluation assistant, a tool designed to enhance the evaluation process at HOCHTIEF. This assistant functions as a semi-autonomous evaluation analyst, capable of reviewing structured and unstructured content from multiple PDF evaluation reports and generating standardized, data-driven assessments. This section explains the logic behind the assistant's development, the structure of its workflow, the reasoning behind its decision-making, and how it addresses limitations in the existing evaluation process at HOCHTIEF. The tool's features were not developed in isolation but were directly informed by challenges observed in the current subcontractor evaluation process presented in Chapter 5.

7.1 Understanding the Evaluation Document

This chapter describes the new evaluation document that was created to act as input for the developed AI assistant. It begins by outlining the document's structure and then presents the reasoning behind it, explaining the challenges of the current process identified in Chapter 5 that it aims to address. It concludes by describing how the document's final format was created and validated by HOCHTIEF employees.

7.1.1 Structure of the Document

The foundation of the AI-powered subcontractor evaluation framework lies in the structured design of the evaluation documents used as input. These Excel-based forms (later converted into PDF and processed by the AI assistant) are the central data collection instruments that capture both quantitative and qualitative performance feedback from two perspectives:

1. HOCHTIEF's employees, and
2. The subcontractor

Each version begins with an introductory text that sets the context and thanks the respondents for participating. The introductory text also provides instructions on how to fill out the evaluation, ensuring it is completed correctly. This is followed by a section on "Respondent Information", where the respondent is asked to fill out their name, the company they work for, their position in that company, their role in the project for which the evaluation is taking place, and finally, the date they filled out the document and their contact details. Both versions can be found in Appendix B.

The core of the form is structured around the evaluation criteria identified in Chapter 5, which are consistently applied across both versions. Each criterion is assessed through multiple questions. For example, under Safety and Compliance, a respondent might be asked: *"Did the subcontractor have all of the required safety certifications and were they valid during the whole project duration?"* or *"How effective is HOCHTIEF in maintaining safety standards and ensuring regulatory compliance?* Importantly, each section is completed by the person within the organization who holds the relevant area of responsibility. For example, responses related to Safety and Compliance come from the HSSE Manager, ensuring that subject-matter experts provide the input with direct oversight of that aspect.

In addition to the detailed, criterion-based sections, both versions of the evaluation document include an overall evaluation box and an overall comment box, designed to capture high-level impressions and closing feedback. Other prompts invite reflection on key weaknesses, project risks introduced by the subcontractor, and whether the team would want to work with this subcontractor again. An “Overall Comments” section concludes the document, providing an open space for any remaining insights or observations.

Similarly, the subcontractor’s version includes an overall evaluation box that asks for their overall impression of working with HOCHTIEF, the key strengths displayed by HOCHTIEF during the project, and areas where the company could improve. Subcontractors are also asked to rate their overall satisfaction with HOCHTIEF on a 1-to-5 scale and to provide an explanation for their rating. The form concludes with questions about willingness to work with the company again in the future and an “Overall Comments” section, where subcontractors can share any further suggestions or feedback regarding their experience.

Responses can vary in format, depending on the type of data needed. They may include:

- **A 1–5 Likert scale**, used for scoring responses quantitatively.
- **Open-text answers**, allowing for detailed explanations or subjective input.
- **Yes/No/Maybe selections**, providing straightforward responses without additional context.
- **Quantitative data**, such as contract price, project schedule (in months), number of safety violations, and other measurable figures.

Each response is accompanied by a comment box labeled “Comments”, where respondents can elaborate on their answers. Whether a respondent is rating on a 1–5 Likert scale, selecting Yes/No/Maybe, or entering quantitative data, they are prompted to explain their answer, with a comment box alongside every question, regardless of the response type. This design ensures every answer is accompanied by a rationale, reducing ambiguity and giving future reviewers insight into the reasoning behind the scores. This feature directly addresses the concern voiced by employees that ratings are often submitted without explanation, making them difficult to interpret, especially if the evaluator is no longer available for follow-up.

The document also improves transparency and objectivity by standardizing evaluation criteria across both internal and subcontractor versions of the document. These criteria, established earlier in the research, guide feedback toward clearly defined aspects of performance such as safety, communication, and compliance. As a result, evaluations are less likely to be driven by personal opinions or emotions, since each response must relate to a specific topic. By anchoring feedback to structured prompts and requiring supporting comments, the form mitigates the risk of emotional or biased judgments that employees identified as a problem in the current process.

Moreover, the evaluation tool includes a variety of question formats that allow respondents to provide both objective data and subjective insights. This balance ensures that evaluations are not one-dimensional, capturing a fuller picture of the subcontractor's performance. The inclusion of dual perspectives adds another layer of transparency, which did not exist before. As described in Chapter 5.2, subcontractors are not provided with a document or a questionnaire to provide their feedback and perspective. With the creation of two versions of the new evaluation document and since both sides participate in the evaluation, their feedback can be compared to identify alignment or discrepancies. This cross-verification helps expose bias or emotional overtones in either account, strengthening the fairness and reliability of the overall evaluation.

Finally, by including these summary-level questions in both forms, the evaluation process ensures that both quantitative scores and qualitative impressions are considered, capturing a well-rounded perspective of the collaboration from both parties. The responses gathered in these overall evaluation and comment sections are particularly valuable for informing the project summary, as they distill the most significant strengths, weaknesses, and experiences of each side. By encouraging both internal teams and subcontractors to offer overarching reflections and additional comments, the process draws out themes and patterns that might otherwise be missed in the other sections. These high-level inputs also serve as a rich source of insights for the “Lessons Learned” and “Recommendations” sections of the final report. Specific comments about what differentiated the subcontractor, or areas where HOCHTIEF’s processes could be improved, help identify best practices and recurring challenges across projects. Similarly, open-ended feedback regarding opportunities, risks, and future collaboration preferences provides context and concrete examples that can be referenced directly when crafting actionable recommendations.

7.1.2 Validation of the Evaluation Document

The questions included in the evaluation document have undergone a validation process to ensure their relevance, clarity, and effectiveness across different roles within the company. To achieve this, draft versions of the evaluation forms were distributed to employees occupying a range of positions. In total, seven people provided feedback on the evaluation document, including a procurement manager, a project manager, a construction manager, a financial controller, a quantity surveyor, a safety manager, and a quality assurance manager. These individuals were selected to reflect the diversity of perspectives and experiences involved in subcontractor oversight.

Each participant was asked to review the questions (more relevant to their roles) critically, considering whether they were clear, appropriately worded, and aligned with the realities of day-to-day project work. Employees provided feedback on elements such as the phrasing of questions, the usefulness of the response options, and the overall structure of the form. Based on this feedback, several revisions have been made. These included rewording ambiguous questions, refining the evaluation criteria to better reflect actual performance indicators, and adjusting response formats to suit the requested information type.

This validation process helped ensure that the questions were not only theoretically sound but also practical and grounded in the real-world experience of those who would use them. Since individuals with varying areas of expertise reviewed the forms, many of whom actively participate in filing out evaluation documents, the resulting questions are relevant, straightforward, and quick to answer. Most importantly, they are grounded in information that is readily available and routinely tracked within the company, ensuring that responses can be provided accurately and without unnecessary effort.

7.2 Development and Design of Virtual Assistant

This chapter analyzes the underlying design philosophy of the virtual assistant that will be used by the proposed hybrid approach to evaluate subcontractors working with HOCHTIEF. It begins by examining the evolution of the personality box from its first version to the second and final version. A personality box is a structured set of instructions that defines how a virtual assistant behaves, interprets input, and generates output. It acts as the assistant’s "operating manual" or behavioral blueprint, guiding its tone, reasoning style, interaction flow, and decision-making logic. In essence, the personality box

transforms a general-purpose AI into a task-specific collaborator. For HOCHTIEF's subcontractor evaluation framework, the personality box was critical in shaping the assistant into a transparent, fair, and structured evaluator that reflects the company's internal standards and addresses specific operational challenges. The chapter continues by exploring how each version was structured to address specific challenges in HOCHTIEF's subcontractor evaluation process. It highlights how Version 1 established a foundational, logic-driven workflow focused on traceability and procedural fairness, while Version 2 introduced enhancements aimed at improving usability, interaction quality, and the clarity of outputs.

7.2.1 Design Philosophy Behind the Virtual Assistant

The design of the virtual assistant is based on two versions of a "personality box", which are sets of precise instructions that tell the assistant how to behave, understand input, and create output. Each version was improved step by step to help the assistant better handle unclear information, follow HOCHTIEF's internal rules, and produce valid and accurate results.

At the core of the assistant is OpenAI's ChatGPT 4.0. By employing NLP techniques, the assistant facilitates analysis of both qualitative and quantitative data, thus providing a comprehensive evaluation that accounts for multifaceted performance metrics. The assistant then employs the WSM, which offers a methodical approach to aggregating diverse criteria into a single evaluative score. This model emphasizes the importance of balanced assessment by assigning weights to each criterion, reflective of HOCHTIEF's strategic priorities.

The virtual assistant was created to improve and organize how subcontractors are evaluated in HOCHTIEF's construction projects. The tool's features were not developed in isolation but were directly informed by challenges observed in the current subcontractor evaluation process. Version 1 created a basic, logic-based workflow designed to ensure transparency and fairness in the evaluation process. In contrast, Version 2 added improvements to make the assistant easier to use, more effective in its interactions, and more transparent in the way it presents results. In both versions, the evaluation process is described in 7 simple steps, which include clear instructions for the virtual assistant. Both versions can be found in Appendix C.

7.2.3 Personality Box Version 1: Functional and Procedural Foundation

The first version was developed to formalize the assistant's behavior into a logical, seven-step evaluation process. This version reflects a strong emphasis on data traceability, fairness, and factual reasoning.

The **first** step begins by welcoming the user and explaining the steps that will follow. The assistant extracts structured data from the two evaluation documents (explained in chapter 7.1). By merging the two reports into a shared knowledge base, the assistant lays the groundwork for a multi-perspective evaluation. As both sides have provided their views and opinions about how the collaboration went, a more objective, informed, and unbiased evaluation can be achieved.

The **second** step involves sorting the information under the 10 standard evaluation criteria. (see chapter 6.1). In the personality box, the assistant is instructed to locate all questions and answers related to each criterion across both documents. To ensure that all the answers, regardless of their format, will be included and to ensure that they are interpreted correctly, the assistant is instructed to

use the full context of responses, including numbers, explanations, and yes/no inputs. As it is important for the evaluation to be objective and data-based, the assistant combines all available information to form a complete picture of the subcontractor's performance on that topic, but without relying only on sentiment analysis. Instead, it is instructed to assess performance based on the facts and data in the answers.

In the **third** step, the assistant is instructed to detect discrepancies, gaps, or vague input and asks whether follow-up questions should be generated. This mimics how a trained human evaluator would handle inconsistencies but does so faster and more systematically. The user has the option to decline the generation of follow-up questions. This option is included to maintain flexibility and human oversight within the evaluation process. While the assistant is designed to detect discrepancies, missing details, or vague responses, not every situation requires additional clarification. For example, if the collaboration with the subcontractor went smoothly and either party reported no issues, there may be no need for further questions. In such cases, the information provided is already sufficient for a fair and accurate evaluation.

Additionally, follow-up questions are intended to be discussed in meetings with the subcontractor. However, not all projects warrant this level of interaction, especially when the subcontractor was involved in a small-scope task or had a low-value contract. In these instances, organizing a follow-up meeting may not be necessary or practical. By giving users the ability to decline follow-up question generation, the assistant respects the context of the project and avoids placing an unnecessary burden on teams, ensuring that the evaluation process remains efficient, relevant, and proportionate to the scale of the work completed by each subcontractor.

As the questions will be asked in a follow-up meeting with the subcontractor, the assistant is instructed to develop accurate and well-informed questions based on a comparison between the subcontractor's responses and those provided by the internal team. Finally, a distinction is made between the questions intended for the subcontractor and the ones intended for Hochtief's employees.

The **fourth** step introduces weight selection, allowing the user to choose between three options. The first option is equal weights, or the two predefined weighting schemes based on stakeholder rankings explained in chapter 6.2: the *Equal Influence Model* and the *Role-Weighted Influence Model.* This feature is important because it introduces flexibility into how the final evaluation score is calculated, allowing users to select the scoring method. By offering these options, the assistant adds analytical depth without requiring manual effort. Users can choose the model based on their preferences.

Next is the **fifth** step. At this stage of the process, the assistant calculates a score for each of the 10 criteria. Clear instructions are given, as the assistant is asked to analyze the evaluation documents while taking into consideration numerical inputs, free-form justifications, yes/no responses, and comments left for every answer. If provided, the assistant will also use the answers to the follow-up questions for further clarification and information. For every score assigned, the assistant is instructed to provide a clear explanation, citing relevant data from the answers.

The **sixth** step is the calculation of the final performance score. The final rating is calculated using the WSM. The assistant multiplies each criterion score by its assigned weight, which was selected in the fourth step. The final score is calculated by summing the weighted scores.

The **seventh** and final step is where the assistant goes beyond scoring to add strategic value. It generates:

- An extensive "*Summary*" of the overall performance, describing how the collaboration between Hochtief and the subcontractor went, explaining in detail both perspectives and providing an objective assessment of the situation
- A "*Lessons Learned*" section, which highlights practical insights based on the behavior and performance observed. These are tied directly to specific events or scores, making them evidence-based and valuable for future planning.
- A set of "*Recommendations*", separately directed to HOCHTIEF and the subcontractor. These concrete next steps provide insights into how HOCHTIEF and the subcontractor can improve in general, as well as advice on enhancing a potential future collaboration based on the latest experience.

The assistant then generates a standardized PDF report that brings all the evaluation data together into one coherent document.

7.2.4 Personality Box Version 2: Human-Centered and Output-Focused Refinement

Version 2 builds on the logic of the first version but introduces several enhancements based on usability feedback and observed bottlenecks in early trial runs. Past evaluation documents were used to transition from Version 1 to Version 2 of the virtual assistant, serving as real-world training material for fine-tuning its behavior, logic, and output quality. These documents, which included completed evaluations from HOCHTIEF's internal teams, were used systematically to improve several critical features of the assistant, like the generation of follow-up questions, the formulation of lessons learned, and the structure of actionable recommendations. As it stands (see chapter 5.2), HOCHTIEF does not directly involve subcontractors in providing feedback with an official document, so a questionnaire like the one that was created for this proposed framework does not exist. Thus, only the internal team's evaluations were used to fine-tune the AI-powered assistant and to create Version 2.

In total, seven evaluation documents were used. These documents were taken from the archives of the Amaliahaven project, where HOCHTIEF, together with its partners Ballast Nedam and Van Oord, completed the construction of approximately 2.5 kilometers of quays and retaining walls on behalf of the Port of Rotterdam Authority. For this project, HOCHTIEF collaborated with a total of 316 different companies. To select the final sample for the fine-tuning of the assistant, an initial group of 14 companies was chosen based on performance scores: the eight highest-scoring companies, the three lowest-scoring ones, and 3 with average scores. From this group of 14, only those classified as subcontractors were retained, resulting in a final selection of 7 companies. The remaining 7, consisting of hired labor, material suppliers, rental equipment providers, and service providers, were excluded from the analysis, as they did not fall under the subcontractor category.

Most of the significant additions were made at the start of the personality box, where some strict data grounding policies were defined. To ensure accurate outputs, the assistant was instructed to generate information strictly based on explicit, factual content present in the submitted evaluation documents or direct user follow-up answers. The instructions are very strict and specific. The following points are some of the rules that were added at the start of the second version of the personality box:

- Never invent, interpolate, generalize, extrapolate, fill narrative gaps, or use hypothetical scenarios unless they are direct quotations or paraphrases from the supplied material.

- If a field, justification, or explanation is blank or absent, state: “No explanation/data provided in the submitted reports”.
- Always present actual extracted answers, ratings, and comments for each criterion from each report in the order they appear, before any interpretation, scoring, or summary.
- All explanations, summaries, and follow-up questions must be directly based on explicit, factual evidence from the uploaded files or direct user follow-up. No extrapolation, generalization, or assumptions are permitted.

In Version 1, the assistant was built with a general logical structure and could generate outputs such as follow-up questions and summaries based on basic patterns and rule-based reasoning. However, early trial runs using the evaluation documents mentioned above revealed that the assistant often struggled to identify subtle inconsistencies or lacked the specificity needed to produce high-quality, context-relevant outputs. For example, it might generate a vague follow-up question like “Can you clarify the schedule performance?”. For that reason, in the second version, examples were added in the third step, along with specific examples. These examples were produced by examining the past evaluation documents mentioned above. They were provided to the assistant to familiarize them with the type of questions they will have to ask. More specifically, these examples were:

- “You mentioned 3 schedule revisions, which were initiated by Hochtief and which by the subcontractor?”
- “The internal team rated collaboration as 2/5 but gave no detail, please explain.”
- “You rated technical quality as 3/5, but provided no comments, could you explain the issues behind this rating?”
- "You mentioned that 95% of the work required rework or corrections but rated the quality of the subcontractors work as 5. Could you please explain?"
- “You rated your company’s safety compliance as a 4 but our manager rated it as 2 and mentioned 24 reported safety incidents-how do you explain that?
- “The subcontractor claims all certifications were valid, but you flagged issues with documentation — what specifically was missing?”

The last two examples were added as possible questions that can be asked in case of discrepancies in the answers to the evaluation questionnaire.

Similarly, in the **fifth step** of the second version, specific examples were given for the justifications the assistant should provide for every score it assigns to a criterion. These examples were created based on patterns found in real project evaluations and are used to help the assistant understand what a strong, fact-based justification looks like. This change was necessary because, in the first version, the assistant sometimes produced vague or overly abstract justifications, which made it difficult for users to assess the reasoning behind the scores. By including clear and realistic examples, the assistant can now model its behavior more consistently and produce explanations that are concise, objective, and directly tied to observable performance. This improves transparency and trust in the evaluation process, especially in cases where multiple stakeholders are involved or where scores may later need to be defended or reviewed.

Some of the examples that were added are the following:

- “Subcontractor missed 2 milestones, submitted 3 late updates, and caused 60 days of delay, score: 2”

- “All safety certifications were valid, and no incidents were reported during the project, score: 5”
- “Subcontractor met all quality standards but had two minor documentation errors, score: 4”
- “Subcontractor’s communication was inconsistent; emails were often unanswered for days, score: 2”

Furthermore, significant improvements were made in the formulation of the “Summary”, the “*Lessons Learned*”, and the “Recommendations” that are included in the PDF report generated in **step seven**. Previously, the generated evaluations often included broad statements or unsupported conclusions, which resulted in reports that lacked credibility and transparency. To correct this, new instructions were implemented that require the assistant only to include facts, numerical data, and explanations that are directly present in the supplied information. The assistant is now explicitly barred from adding assumptions or using “typical” reporting language, even for the sake of narrative flow. This change ensures that every point made in the report can be traced back to a real, verifiable source. Explanations, lessons learned, and recommendations must all reference data that was recorded or provided by follow-up answers, which helps make the reasoning behind each rating or observation transparent and accountable.

As the virtual assistant primarily focused on questions related to the evaluation criteria, the final parts of the evaluation, specifically “Overall Evaluation” and “Overall Comments,” were omitted in the first version and were not considered. This was a significant issue as these parts provide very valuable insights that could be used for the summary, lessons learned, and recommendations. To ensure their inclusion, the assistant was instructed to extract and fully consider the responses from the “Overall Evaluation”, “Overall Comments”, or equivalent final reflection/conclusion sections in both the Hochtief internal and subcontractor self-evaluation reports. These sections must then be used as essential sources when drafting the project summary, the lessons learned, and both sets of recommendations.

Apart from the general instructions described above, some more specific ones were added for every section of the report. Initially, the summary produced with the first version was often generic, short, and did not provide any helpful information about the project. To address this, specific instructions were added in the second version. The assistant was instructed to create a 250-to-300-word summary that references specific events, patterns, and decisions. The assistant was also instructed to include specific facts and data points (e.g., “60-day delay due to material delivery bottlenecks”, “3 formal schedule revisions”, “communication breakdowns in Q2”, “3 schedule revisions due to delayed workforce ramp-up”) and to reflect both perspectives. The end report must clearly describe how the subcontractor performed.

To add to this, in the outputs of the first version, the assistant would often generate lessons that were too generic without tying them back to specific project events or behaviors. By referencing the real-world evaluation forms from the Amaliahaven project, which included valuable insights, the assistant was given specific examples and instructions about generating lessons grounded in the specifics of project performance and supported by clear examples. Each lesson should:

- Identify why it happened (e.g., “lack of formal update protocol”)
- Suggest what should be done differently next time

The assistant was instructed to generate concise, high-impact lessons that are directly supported by the evaluation content.

The "*Recommendations"* section benefited similarly. Initial outputs tended to be broad or overly polite, such as "Consider improving documentation". Through targeted use of past evaluations, the assistant learned to generate role-specific and actionable recommendations. For instance, when subcontractors failed to submit safety documentation on time, a refined recommendation might be: "Subcontractor to implement a checklist-based submission process to ensure timely delivery of safety certifications in future projects". This made the assistant's recommendations far more helpful to both internal and external stakeholders. Other examples that were added:

- "Include a contractual clause requiring the subcontractor to submit updated schedules every two weeks with variance explanations for any deviation beyond 3 days."
- "Request digital submission of all required safety certifications two weeks before site access is granted and schedule a pre-mobilization audit."
- "Require that the first completed unit of work (e.g., the first installed ceiling section) undergo a detailed quality walkthrough with both parties to establish a shared standard for the rest of the project."

The last change that was made in the second version was the addition of "Behavioral Notes". This set of instructions was added to ensure the tool produces high-quality, professional, and actionable evaluation reports that meet the practical needs of HOCHTIEF's subcontractor assessment processes. The instructions are used to ensure that the assistant is specific, impartial, and objective, and that they produce data-driven and specific responses.

7.2.5 Mapping Features to Challenges at HOCHTIEF

As mentioned at the start of the chapter, each of the features of the assistant was developed to address some of the challenges in the current evaluation process at HOCHTIEF, as identified through interviews with the company employees. One of the key weaknesses is the lack of justification accompanying numeric scores. Evaluators often select ratings (e.g., 1–5) without providing supporting explanations, which makes it difficult for future users of the data to interpret or trust the results. The assistant addresses this issue by requiring a brief, factual justification for each score. This ensures that ratings are based on and supported by actual data provided by both the internal team and the subcontractor, for instance.

Another critical gap in the current process is the failure to address inconsistencies or vague responses systematically. The assistant tackles this by automatically detecting contradictions between the two versions and suggesting context-specific follow-up questions. Evaluations are also currently scattered across documents and formats, making them difficult to access or share. The assistant solves this by generating a centralized, well-structured PDF report. It brings together all key elements into a coherent and professional format. Moreover, subjectivity and emotional bias have also been identified as problems in previous evaluations, especially when scores are given based on personal impressions rather than standards. To counter this, the assistant incorporates feedback from both the internal HOCHTIEF team and the subcontractor by processing their separate evaluation forms and integrating both perspectives into the scoring logic. This two-sided approach encourages fairness, improves buy-in, and enables a more balanced assessment.

Finally, internal teams often fail to capture insights that could inform future projects, while actionable steps are rarely included. This means that the process fails to feed into continuous improvement. To tackle this, the assistant includes a dedicated "*Lessons Learned*" section in the final report, linking key takeaways to specific project behaviors or outcomes. The assistant also provides tailored recommendations for both HOCHTIEF and the subcontractor. These recommendations are practical, role-specific, and based on documented performance.

7.3 Conclusion

In this chapter, the development of the proposed hybrid framework for evaluating subcontractor performance through an AI-driven approach was explained in detail. By integrating NextChat with the WSM, the proposed methodology represents a significant advancement in addressing the challenges associated with the current process of subcontractor evaluations at HOCHTIEF.

The newly structured evaluation documents serve as a foundational element for this framework, designed to capture both quantitative and qualitative insights while incorporating dual perspectives, enabling a more comprehensive and balanced assessment. Validation efforts ensured that the evaluation documents were refined based on diverse feedback from HOCHTIEF employees, ensuring practical applicability and relevance to day-to-day operations.

The development of the virtual assistant is central to the framework. Through structured instructions provided in the personality box, the assistant processes inputs to deliver data-driven assessments and strategic recommendations. The initial version was extensively improved, leading to a more advanced second version incorporating feedback for improved usability, interaction quality, and output specificity. Overall, Version 2 of the assistant brought important refinements that made its outputs more transparent, more actionable, and easier to use, while ensuring that all evaluations are grounded in real data and aligned with HOCHTIEF's project realities. By mapping its features to HOCHTIEF's practical needs, the assistant effectively counters some of the issues identified in the current process.

Chapter 8: Evaluating Subcontractor Performance – A Case Study of Framework Implementation

This chapter presents the empirical case study conducted to evaluate the AI-driven subcontractor evaluation framework. The study was implemented within a HOCHTIEF data center construction project focusing on subcontractors responsible for concrete and steel construction tasks. The chapter presents the case study setup, the application of the AI-assisted evaluation process, a comparative analysis with traditional manual assessments, and the validation of the results through a survey with company employees. A more in-depth analysis of the results will be carried out in Chapter 9.

8.1 Case Study Background

The purpose of this case study was to assess whether integrating ChatGPT 4.0 with a Multi-Criteria Decision Analysis (MCDA) framework could deliver more transparent, accurate, and comprehensive subcontractor evaluations compared to existing manual methods. In response to the limitations of traditional evaluation techniques (see chapter 2.2.1), the study explored an innovative hybrid approach that incorporates interactive clarifications and data-driven follow-up questions.

The selected case study was part of an ongoing HOCHTIEF data center project. The project involves the repurposing of an existing building in Amsterdam into a data center. The project is now finishing its 2nd phase and will move into the 3rd and final phase. For that reason, the subcontractors involved in the 2nd phase will have to be evaluated. One subcontractor working with HOCHTIEF has been selected for this case study, subcontractor X. That particular subcontractor is responsible for the steel constructions inside the data center and will be evaluated using both the traditional manual assessment performed by the procurement team and the AI-enhanced framework. The manual evaluations of these subcontractors have already been completed and can be found in Table H2 in Appendix H.

According to the manual evaluation, subcontractor X exhibits strong performance in both quality and HSE. They score high in all HSE-related criteria, with ratings ranging from 4 to 5, reflecting strong safety compliance, use of certified equipment, and a culture of safety awareness. Their engineering work and service quality are consistently rated at 4, showing technical competence and reliable delivery. The cooperation and documentation categories also show solid performance, with scores of 3, indicating dependable communication, agreement adherence, and proper handling of administrative requirements. In the CSR category, multiple criteria are marked as “Not Applicable” (N/A). This absence of data may suggest either that the subcontractor has not provided sufficient environmental documentation or that these metrics do not apply to the nature of their work. Where scores are given, such as for material recycling, they score a 3, suggesting some room for improvement in sustainable practices. Planning is rated as average with consistent scores of 3, showing that while timelines are met, there may be opportunities for greater efficiency or communication. Finally, subcontractor X received a final score of **3.6**, reflecting above-average overall performance.

In a future prequalification, this evaluation would not provide handy information to the decision maker due to its lack of depth and supporting detail. Decision makers rely on comprehensive, evidence-based assessments to determine whether a subcontractor is capable of meeting project

requirements. However, this evaluation primarily consists of numerical scores with no accompanying narrative or justification, making it difficult to understand the reasons behind high or low ratings. Without context, these scores could be interpreted inconsistently or fail to reflect the subcontractor's actual performance.

8.2 Implementation of the AI-Driven Evaluation Framework

The case study started with the distribution of the two evaluation documents (described in Chapter 7). The first document was filled out by the internal team, which included a Senior Project Manager, an HSSE Manager, a Procurement Manager, a Quantity Surveyor, and a Sustainability & Environment Coordinator. The subcontractors themselves filled out the second document. Both documents were sent out by email to the participants of this case study. The responses to both documents can be found in Appendix D.

The evaluation documents were then given to the virtual assistant to generate follow-up questions based on discrepancies and gaps. They were divided into two groups, one to be answered by HOCHTIEF's internal team and the other by the subcontractor. The AI-generated questions can be found in Table H3 in Appendix H.

The follow-up questions for HOCHTIEF's employees were distributed via email. This approach was chosen for efficiency, as organizing face-to-face meetings would have required additional time and resources. Additionally, the number of questions for each employee was limited, making email communication a more practical method. In contrast, the AI-generated follow-up questions for the subcontractor were addressed during an in-person meeting. Subcontractor X was invited to HOCHTIEF's offices, where all follow-up questions were discussed and answered during a one-hour session. The full responses from both the internal team and the subcontractor are included in Appendix E.

After the follow-up questions were answered, they were given to the virtual assistant. The next step was the selection of the weighing method for the evaluation criteria. For this case study, all three weighing methods (see chapter 6.2) are used. Thus, in the final report, subcontractor X was given three different ratings. Each of these scores will be compared to the score from the manual evaluation.

Once the weighting method was selected, the virtual assistant scored each of the ten evaluation criteria while justifying each one. In the same chat with the virtual assistant, the ratings for each criterion were calculated three times to check if the scores would change. Table 10 below shows the three simulations and the scores for each criterion.

***Table 10**: Evaluation Scores of Subcontractor X Based on Three Model Runs*

Criterion	Ratings from 1st calculation	Ratings from 2nd calculation	Ratings from 3rd calculation
Safety and Compliance	4	4	4
Quality of Workmanship	4	4	4
Timeliness and Schedule Adherence	3	3	3
Technical Ability and Qualifications	5	5	5
Cost Management	3	3	4
Reliability and Repeat Engagement	4	4	4
Financial Stability	4	4	4
Communication and Collaboration	3	3	3
Innovation and Problem Solving	4	4	4
Sustainability Practices	3	4	4

From the table above, it is evident that the scores remained consistent across all three processing runs for the majority of evaluation criteria. The only exceptions were Cost Management and Sustainability Practices, which exhibited slight variations. Specifically, Cost Management was rated 3 in the first two runs and 4 in the third, while Sustainability Practices received a 3 in the first run, increasing to 4 in subsequent runs. Table H4 in Appendix H presents the justifications for the ratings of the two criteria (Cost Management and Sustainability Practices) that did not receive identical scores across all three processing runs. It highlights how the assistant's explanations evolved in parallel with these minor score variations.

After scoring each criterion, the virtual assistant continues with the calculation of the final rating. As mentioned above, all three weighing methods were used (see Table 11).

***Table 11:** Overall Ratings of Subcontractor X Across Three Weighting Approaches*

Weighing Method	Ratings from 1st calculation	Ratings from 2nd calculation	Ratings from 3rd calculation
Equal Weights	3.70	3.8	3.90
Equal Influence	3.81	3.85	3.92
Role Weighted Influence	3.78	3.82	3.89

Once the final score was calculated, the virtual assistant generated its final output, which was a PDF report including the final ratings and the score breakdown from the third calculation (see Tables 10 and 11), justification for each score, a summary of the collaboration, the lessons learned, and the recommendations to both Hochtief and the subcontractor. The summary, the lessons learned, and the recommendations are presented below. The entire report can be found in Appendix F.

Summary

The evaluation of Subcontractor X demonstrates strong overall performance across all primary subcontractor criteria. The team maintained all required safety certifications and operated incident-free, though minor general site-wide PPE lapses were noted. Quality of work was high, confirmed by minimal rework (1%) and positive ratings from Hochtief; minor workflow disruption and a single miscommunication

on staircase specifications were promptly addressed. Both parties cited challenges in timeliness and scheduling, chiefly due to unstructured, lengthy meetings, lack of lean/efficient scheduling, and frequent relocation by project management rather than the subcontractor. All technical qualifications were satisfied, with both parties rating technical skill and problem-solving abilities highly.

Cost management was fundamentally sound: disputes regarding scope and financial settlements stemmed from the project commencing before a contract or payment schedule was agreed, and were not due to subcontractor error. Reliability was strong, with Subcontractor X adapting to shifting site conditions and fulfilling all commitments, despite negative impacts from overall project organization. While communication and collaboration processes (notably meeting length and structure) could be improved, both parties successfully collaborated using multiple channels. Innovation and sustainability standards were met, as evidenced by certifications, implemented solutions, and adaptive support on technical and sustainable goals.

Lessons learned highlight the importance of advance contractual agreement, structured site management, and explicit goal communications at project outset.

Lessons Learned

- *Starting work without a signed contract or clear scope caused disputes and financial processing delays. All contractual and financial structures must be in place before site mobilization.*
- *Unstructured, overly large scheduling meetings led to inefficiency. Leaner, smaller, and more focused coordination sessions are recommended.*
- *Frequent area reassignments and lack of a unified building model disrupted workflow for multiple trades. Early measurement, accurate site modeling, and disciplined project scheduling are critical to reliable site progress.*
- *General project communication and sustainability requirements should be shared explicitly and early to ensure full alignment.*

Recommendations for Hochtief

- *Finalize contracts, scope, payment plans, and invoicing structure before authorizing site access or commencement of work.*
- *Implement structured, lean scheduling meetings limited to relevant stakeholders. Introduce building-wide project models (e.g., BIM) for coordinated sequencing.*
- *Provide site-wide communications and clear expectations for all subcontractors regarding safety, quality documentation, and sustainability processes at project launch.*
- *Ensure subcontractors are fully briefed on expected documentation, reporting formats, and sustainability goals via pre-mobilization workshops.*

Recommendations for the Subcontractor

- *Continue the existing good practice of adaptability and technical problem-solving.*
- *Document and promptly communicate disruptions or changes driven by client or site conditions to maintain transparency in progress and claims.*
- *Pursue early, proactive alignment on contract terms and scope before mobilization on future projects.*
- *Request early involvement in scheduling/model-coordination reviews to highlight sequencing issues and avoid site inefficiencies.*

Chapter 9: Results, Implications, and Limitations

This chapter provides a critical analysis of the results obtained from the case study presented in the previous chapter. The chapter starts by providing an overview of the results and continues by presenting the validation process. The validation was carried out through a survey with 10 HOCHTIEF employees with different roles in the company, all involved in the evaluation of subcontractors. The performance of the AI-based subcontractor evaluation framework for all the criteria evaluated in the survey is examined and explained. Then the final rating for the subcontractor's performance is compared with the one from HOCHTIEF's existing manual evaluation process. After the validation results are analyzed, they are re-linked to the literature review carried out in Chapter 2. The chapter concludes by addressing the research questions that this study aimed to answer.

9.1 Overview of the Results

The AI-driven framework was applied to evaluate Subcontractor X, who performed steel construction work on a HOCHTIEF data center project. The evaluation followed three runs in the same "Chat" using structured data from dual-perspective evaluation documents, supplemented with AI-generated follow-up questions. These questions addressed gaps and discrepancies in stakeholder feedback and were answered by both the internal HOCHTIEF team and the subcontractor.

Ratings across 10 performance criteria showed high consistency between the three AI-generated runs, with only minor differences in "Cost Management" (3, 3, 4) and "Sustainability Practices" (3, 4, 4). The rest of the criteria produced identical scores across all runs, indicating strong internal consistency within the assistant's reasoning model. Upon application of the three MCDA weighting strategies, the AI produced three final scores per run, resulting in a total of nine final performance scores for Subcontractor X.

The model's final output was a detailed PDF evaluation report for Subcontractor X. This report presented the three final performance scores generated during the third AI run (each corresponding to a distinct weighting method) and the score subcontractor received for each of the ten criteria, along with a justification for each score. In addition to the quantitative ratings, the report featured a performance summary, key takeaways from HOCHTIEF's experience working with the subcontractor, and tailored recommendations aimed at improving future collaboration. The comprehensive format ensured that both HOCHTIEF and the subcontractor received actionable insights grounded in data provided either in the evaluation documents or the answers to the follow-up questions.

9.2 Validation and Analysis of the Results

As described in Chapter 4.4, the validation of the proposed AI-driven subcontractor evaluation framework was structured around two pillars:

1. a detailed user validation survey assessing multiple dimensions of the tool's performance and perceived value, and
2. A quantitative alignment analysis comparing the AI-generated subcontractor scores with the manual assessments conducted by HOCHTIEF's procurement team, employing the Mean Absolute Percentage Deviation (MAPD) as an accuracy metric.

A total of 10 professionals from HOCHTIEF participated in the validation survey. While this may appear to be a modest sample size, it is entirely appropriate given the organizational context and the study's targeted objectives. As previously discussed in Chapter 5.2, subcontractor evaluations at the company are performed by a small, specialized group of stakeholders directly involved in project delivery and supplier management. These individuals hold critical roles, such as project management, procurement, quality assurance, safety oversight, and quantity surveying, whose perspectives collectively define the organization's subcontractor performance standards. By engaging precisely these key decision-makers, the study ensured that the validation results are deeply grounded in the insights of those who drive subcontractor assessments and selections within HOCHTIEF. This targeted sampling approach thus prioritizes depth, relevance, and practical applicability over broad generalization, aligning with best practices for evaluative research in professional settings where expertise is concentrated among a focused team.

The aim was to determine whether the framework could deliver measurable improvements over the existing evaluation process, particularly in providing more insightful, transparent, and actionable evaluations, while remaining numerically consistent with established practice. Table 12 presents the consolidated results of the validation process, comparing the performance of the proposed AI-driven subcontractor evaluation framework against the predefined success thresholds for each evaluation.

Overall, the results demonstrate that the framework met or exceeded the success thresholds for all dimensions/ In the case of *Perceived Limitations and Risk Awareness*, a lower score signifies stronger performance, as it reflects that users identified fewer shortcomings and expressed less concern about relying on the tool. This indicates they generally found the outputs credible and were not overly worried about issues such as oversimplification or lack of contextual grounding. The average score for every statement can be found in Appendix G.

***Table 12**: Validation outcomes for the AI-enhanced subcontractor evaluation framework compared against predefined success thresholds.*

Criterion	Success Threshold	Results
1. Insightfulness	Mean Likert score of **≥ 3.5**	**3,84**
2. Clarity and Relevance of Follow-up Questions	Mean Likert score of **≥ 3.5**	**4,16**
3. Justification and Transparency of Ratings	Mean Likert score of **≥ 3.5**	**4,18**
4. Quality of Lessons Learned and Recommendations	Mean Likert score of **≥ 3.5**	**4,25**
5. Perceived Added Value	Mean Likert score of **≥ 3.5**	**4,24**
6. Perceived Limitations and Risk Awareness	Mean Likert score of **< 3.5**	**3,08**

9.2.1 Insightfulness

Highest individual statement score under this criterion:

- *The report enables clearer identification of performance strengths and weaknesses.* (4.00)

Lowest individual statement score under this criterion:

- *The evaluation report includes sufficient context to understand the subcontractor's role in the project.* (3.10)

The first validation criterion assessed whether the AI-driven framework provided a deeper, more meaningful understanding of subcontractor performance compared to traditional manual assessments. With an overall average of 3.84, it exceeded the predefined success threshold of 3.5, indicating that participants generally viewed the tool as improving the insightfulness of the evaluation process.

The highest scores were tied to the framework's ability to highlight key performance dynamics. Participants rated *"The report enables clearer identification of performance strengths and weaknesses"* at 4.00, while *"The AI report provides more detailed and insightful content than the traditional report*" also scored 4.00. Additionally, the framework showed strength in revealing less obvious performance aspects, with statements like *"The evaluation report presents new or non-obvious information"* and *"The report enables individuals without prior interaction with the subcontractor to make informed decisions",* each earning a respectable 3.80. This indicates the AI framework successfully moved beyond static numerical scores to extract deeper operational insights, supporting more informed decision-making.

However, this criterion also revealed one of the weaker areas found in the validation study. The lowest-scoring statement, mentioned above, suggests that while the AI framework was effective at extracting insights from structured evaluation data, it was somewhat less effective at capturing and providing information about the broader project context in which the subcontractor operated. This observation aligns with challenges described in the literature review (Chapter 2), which pointed to AI's typical reliance on explicitly provided inputs and its limited ability to infer external situational nuances independently.

9.2.2 Clarity and Relevance of Follow-up Questions

Highest individual statement score under this criterion:

- *The tone and wording of the follow-up questions were appropriate for professional use.* (4.60)

Lowest individual statement score under this criterion:

- *The assistant's follow-up questions demonstrated an understanding of the evaluation context.* (3.80)

This criterion examined whether the AI assistant was able to generate meaningful, context-sensitive follow-up questions that effectively resolved inconsistencies, gaps, or ambiguities in the original evaluation data. The overall mean score was 4.06, well above the success threshold of 3.5, indicating strong participant endorsement of this component of the framework.

The results show that participants generally found the follow-up questions to be highly targeted, clear, and professionally phrased. Several statements achieved exceptionally high averages. For

instance, respondents strongly agreed that *"The tone and wording of the follow-up questions were appropriate for professional use"* (4.60), that the questions were *"concise and easy to understand"* (4.50), and that they would *"use these follow-up questions in a real subcontractor evaluation meeting"* (4.50). These consistently strong scores underscore that the assistant not only identified areas needing clarification but did so in a manner that upheld professional standards and facilitated productive dialogue.

Slightly more moderate scores were found for statements like *"The assistant's follow-up questions demonstrated an understanding of the evaluation context"* (3.80) and *"The follow-up questions helped uncover details about the subcontractor's performance that were not initially evident"* (3.90). These slightly lower, though still strong, averages highlight that while the questions were generally well-received, there remains room to enhance their contextual precision and ability to elicit deeper, previously hidden information.

9.2.3 Justification and Transparency of Ratings

Highest individual statement scores under this criterion:

- *The rationale for each rating is clear and easy to follow.* (4.30)
- *The justifications highlighted both strengths and areas for improvement.* (4.30)

Lowest individual statement score under this criterion:

- *The justifications demonstrated fairness and objectivity in the evaluation process.* (4.00)

This validation category evaluated whether the AI-generated justifications for each rating were clear, logically grounded, and sufficiently detailed to support transparency, which were areas where manual evaluations frequently fell short according to both company interviews and the literature review. The AI-enhanced framework achieved an overall average score of 4.08, comfortably exceeding the 3.5 benchmark.

Participants consistently rated statements in this category highly, indicating that the assistant successfully addressed a major shortcoming of existing practices by making the rationale behind scores explicit and data-linked. For instance, strong agreement was expressed with *"The rationale for each rating is clear and easy to follow"* (4.30) and *"The justifications highlighted both strengths and areas for improvement in the performance"* (4.30). Particularly notable was the agreement that compared to the manual method, *"the AI report offers a better explanation of scores"* (4.00), directly validating one of the core hypotheses of this research.

9.2.4 Quality of Lessons Learned and Recommendations

Highest individual statement score under this criterion:

- *I would consider applying these insights in upcoming selections.* (4.70)

Lowest individual statement score under this criterion:

- *The recommendations align with HOCHTIEF's internal goals and evaluation standards.* (3.60)

One of the main goals of deploying this AI-enhanced framework was to move beyond static performance scores and produce actionable lessons learned and specific, data-backed

recommendations. For this criterion, the tool achieved the highest overall average score of 4.25 across all primary validation criteria.

Respondents especially valued how the lessons and recommendations aligned with HOCHTIEF's strategic needs and could meaningfully influence future subcontractor selections. The standout item was *"I would consider applying these insights in upcoming selections"*, which scored an impressive 4.70, underscoring direct perceived operational value. Similarly, the scores for the statements *"The lessons and recommendations help prevent similar issues in future projects"* (4.50) and *"The integration of lessons learned and recommendations adds significant value to the evaluation"* (4.60) further emphasize the framework's practical impact on continuous improvement processes.

The only relatively moderate score in this cluster was for *"The recommendations align with HOCHTIEF's internal goals and evaluation standards"* (3.60), suggesting that while participants generally found the outputs strategically helpful, there may still be room to better tailor or adjust recommendations to the company's unique procedural standards or evolving policy objectives.

9.2.5 Perceived Added Value

Highest individual statement score under this criterion:

- *The integration of lessons learned and recommendations adds significant value to the evaluation.* (4.60)

Lowest individual statement score under this criterion:

- *The structured format of the report made it easier to understand and use.* (3.80)

This criterion was designed to capture overall satisfaction and the perceived utility of the AI-enhanced tool compared to the traditional process. With an overall average of 4.24, it confirmed that participants viewed the tool as substantially enriching the evaluation process.

The highest endorsements were found in statements highlighting the reduction of ambiguity (4.00), the addition of objectivity (4.40), and the significant value added by the integration of lessons learned and recommendations (4.60). The relatively lower scores for statements like *"The structured format of the report made it easier to understand and use"* (3.80) may point to occasional complexity or length in the generated outputs, as also flagged under perceived limitations.

9.2.6 Perceived Limitations and Risk Awareness

Lowest individual statement score under this criterion:

- *The assistant sometimes asked questions that felt too general or not tailored to the situation.* (2.50)

Highest individual statement score under this criterion:

- *The assistant's report should always be checked by a person before it is used officially.* (4.00)

Unlike the other validation categories, this criterion was designed so that lower scores signal a stronger outcome, reflecting fewer perceived risks, shortcomings, or reasons for concern about the AI-assisted evaluation process. The overall average for this criterion was 3.08, indicating that while participants recognized some limitations, they generally did not strongly agree with the more critical statements presented to them.

Notably, the lowest-scoring statements under this criterion, ranging from 2.50 to 2.90, represent reassuring findings, as they indicate that participants generally disagreed with or did not see these concerns as serious problems. The relatively low scores show that users generally found the AI outputs to be contextually relevant, sufficiently detailed, and cautious with incomplete information.

Slightly higher scores in the 3.30–3.60 range suggest moderate, healthy caution rather than serious red flags. The statement "Some of the scores in the report felt higher than the situation deserved", scored 3.30, indicating that participants perceived occasional optimism or overrating, but not to a problematic extent. Also, the statement "The scores were sometimes based more on the explanations than on actual performance" scored 3.60, suggesting some awareness that AI-generated narratives can weigh heavily on justification text, potentially smoothing over subtle performance issues.

Perhaps most importantly, the highest individual score in this category was 4.00 for the statement "A person should always check the assistant's report before it is used officially." Rather than pointing to a flaw in the tool, this strong agreement demonstrates that participants fundamentally understood the appropriate role of AI in such evaluations as a decision support mechanism, not a standalone authority. This view perfectly aligns with principles from the literature reviewed in Chapter 2, which stressed that integrating AI into project environments requires maintaining human interpretive oversight to ensure alignment with on-site realities and organizational standards.

9.2.7 Quantitative Alignment: MAPD Analysis

Complementing the qualitative validation, the study also assessed how closely the AI-generated scores aligned with HOCHTIEF's traditional manual evaluation. This was achieved by calculating the MAPD across three different weighting scenarios: Equal Weights, Equal Influence, and Role-Weighted Influence. In all cases, the MAPD values were well below the predefined threshold of 10%, indicating strong numerical alignment with the existing manual process.

Table 13 presents a comparative analysis of the final overall subcontractor scores generated by three different AI weighting scenarios against the manual evaluation score. Each scenario was run three times to assess repeat consistency and to measure deviation from the traditional manual score.

Table 13: *Comparison of AI-Enhanced and Manual Evaluation Scores with Deviation Analysis. This table presents the final subcontractor scores under three AI weighting scenarios compared to the manual evaluation score, along with percentage deviations and Mean Absolute Percentage Deviation (MAPD) as an indicator of average divergence.*

Weighing Method	AI Score 1	AI Score 2	AI Score 3	Manual Score	Deviation 1	Deviation 2	Deviation 3	MAPD
Equal Weights	3.70	3.8	3.90	3.60	2.78%	5.56%	8.33%	**5.56%**
Equal Influence	3.81	3.85	3.92	3.60	5.83%	6.94%	8.89%	**7.22%**
Role-Weighted Influence	3.78	3.82	3.89	3.60	5.00%	6.11%	8.06%	**6.39%**

The Equal Weights model produced AI scores of 3.70, 3.80, and 3.90, with deviations of 2.78%, 5.56%, and 8.33%, respectively. This approach assumes that all evaluation criteria are equally important, regardless of their actual impact on project outcomes. Although it achieved the lowest MAPD (5. 56%), indicating strong numerical alignment, it may oversimplify the evaluation by ignoring the relative importance of different criteria.

The Equal Influence model assigns equal decision-making weight to all evaluators, regardless of their role or seniority. This method resulted in slightly higher AI scores (3.81, 3.85, 3.92) and a MAPD of 7.22%. These results demonstrate moderate alignment with the manual benchmark, with deviations ranging from 5.83% to 8.89%. The consistent scores suggest that the Equal Influence model captures a broad range of perspectives, which can help smooth out individual biases. However, it might also dilute the impact of domain-specific expertise. Conversely, some might argue that it balances numerical accuracy with organizational fairness.

The Role-Weighted Influence model, on the other hand, adjusts each evaluator's input based on their position within the project hierarchy. This method yielded scores of 3.78, 3.82, and 3.89, with a slightly lower MAPD of 6.39%. The marginal improvement in alignment suggests that weighing inputs by role may better reflect decision-making in practice, especially in hierarchical project settings. However, the similarities in scores between this approach and Equal Influence indicate that the added complexity of role-based weighting might not produce significantly different outcomes where evaluator perspectives are already aligned. While this method aligns more closely with organizational decision-making structures, it may also introduce subjectivity and bias by favoring specific roles over others. It could also discourage open feedback from less senior team members.

These findings are significant in two ways. First, they confirm that, although the AI-driven framework generally provided slightly more favorable assessments of subcontractors, the deviations were modest, supporting the framework's reliability and accuracy. Second, the increasing MAPD values from the Equal Weights to Role-Weighted Influence suggest that giving more weight to high-responsibility roles slightly raises the final ratings, implying these roles may emphasize positive performance aspects more.

Overall, the low MAPD across all scenarios highlights that AI evaluations align well with traditional assessments while offering greater depth, transparency, and justification, as demonstrated in the broader validation.

9.3 Linking Findings to the Literature on AI in Construction Evaluation

This section examines how the empirical results of the validation study connect to key themes from the literature on AI adoption in the construction industry. Through detailed analysis, it becomes evident that the proposed AI-driven framework both delivers on many of the anticipated strengths highlighted in prior research and simultaneously surfaces enduring concerns that underscore the necessity of careful governance and hybrid human–AI approaches.

9.3.1 Empirical support for established benefits of AI adoption

The validation results strongly corroborate the literature’s emphasis on combining qualitative dimensions with traditional quantitative metrics. Silva et al. (2022) stress that innovation, sustainability, and adaptability must accompany core criteria like cost and compliance to achieve

meaningful assessments, while Upadhyaya et al. (2021) emphasize the critical role of stakeholder narratives and behavioral feedback. These conclusions are directly operationalized in this study's high participant agreement that the AI framework identified non-obvious subcontractor strengths and weaknesses (4.00 under *Insightfulness*). Such outcomes also closely align with benefits like higher visibility and control, higher accuracy in planning processes, and improved supply chain relationships mentioned by Guida et al. (2023), by making nuanced contractor capabilities more transparent and strategically visible.

Participants strongly agreed that rationales were clear and superior to traditional manual explanations. This directly validates the benefits mentioned by Guida et al. (2023), like improved supplier performance measurement and rationalization of suppliers, which are inherently supported by structured, data-grounded explanations that allow more precise differentiation of subcontractor capabilities. Furthermore**,** the high scores under *Clarity and Relevance of Follow-up Questions*, with specific agreement on resolving discrepancies and ambiguities, strongly support Abdulla & Baryannis (2024) and Polat (2021)'s arguments on MCDA frameworks' ability to integrate diverse data types and systematically flag unclear areas. Participants rated *Perceived Added Value* highly, particularly agreeing that the tool added objectivity and reduced ambiguity, closely supporting Handfield et al. (2019) and Morgan (2021), who found that AI-supported evaluations minimize intuition-driven variability.

9.3.2 Empirical Support for Multi-Perspective Evaluation in Subcontractor Assessment

A distinct contribution of the validated framework lies in its structured integration of dual stakeholder inputs, thereby operationalizing a multi-perspective evaluation approach. This design directly addresses concerns highlighted by Upadhyaya et al. (2021), who noted that traditional models in subcontractor evaluation often neglect the subcontractor's perspective, leading to a partial understanding of performance dynamics. By facilitating the submission of parallel assessments and using AI-generated clarifying questions to reconcile divergences, the framework introduces procedural symmetry and accountability into post-project evaluations.

The results suggest that the framework not only enhanced the contextual richness of the evaluations but also improved the traceability and fairness of the final ratings. This aligns empirically with Guida et al.'s (2023) documented benefit of better supplier performance measurement**.** The empirical implementation also reflects Upadhyaya et al.'s (2021) argument that stakeholder engagement improves both evaluation quality and stakeholder buy-in. By structuring contractor-subcontractor input symmetrically and enhancing transparency through textual justifications, the framework operationalizes collaborative governance principles.

9.3.3 Empirical Validation of AI-Enhanced MCDA Integration

This study offers empirical confirmation of literature propositions concerning the integration of AI with MCDA methods. Abdulla and Baryannis (2024) and Polat (2021) have previously argued that AI can automate the processing of multi-criteria datasets, enhancing consistency and supporting strategic adaptability. In line with this, the implemented framework incorporated three distinct weighting models. Across these configurations, the MAPD between AI-generated and manual scores remained consistently below 7.3%, thereby satisfying the predefined threshold of 10% for acceptable predictive alignment. This confirms the framework's robustness in accommodating different prioritization strategies without compromising accuracy.

9.6.4 Reflection of Challenges Documented in Literature

Despite the overall effectiveness of the AI-enhanced evaluation framework, multiple limitations observed in the empirical validation align closely with challenges documented in the literature. One of the most prominent challenges confirmed by the validation is the AI's limited ability to autonomously infer or reconstruct a broader project context without explicit supporting information. In the *Insightfulness* criterion, the statement "The evaluation report includes sufficient context to understand the subcontractor's role in the project" received an average score of 3.10. This result directly supports the literature review's conclusion that, while AI-enhanced frameworks can synthesize and clarify structured data, their performance is fundamentally bounded by the input provided. This confirms that AI tools are typically dependent on explicitly supplied inputs and have limited capacity to infer external situational nuances autonomously (Silva et al., 2022; Wilson, 2024). Thus, the AI model is effective at drawing deeper insights only as far as the underlying evaluation documents are themselves rich in context and detail.

Survey respondents also strongly reinforced the literature's warnings regarding the necessity of human oversight. Under the *Perceived Limitations and Risk Awareness* criterion, the statement *"The assistant's report should always be checked by a person before it is used officially"* received a notably high average score of 4.00, indicating broad agreement that expert judgment remains indispensable. This closely echoes the emphasis by Guida et al. (2023), who underscore that while AI can bring efficiency and transparency, it simultaneously raises governance challenges (ethical deployment, bias, data privacy), necessitating vigilant human involvement. Meanwhile, Scott et al. (2015) emphasize how AI can standardize and improve fairness but implicitly require human validation to ensure decisions are contextually appropriate and free from over-automation risks.

Furthermore, the literature review of Chapter 2 documented the challenges of integrating AI systems into existing organizational practices, citing the importance of robust data quality, change management, and digital infrastructure (Guida et al., 2023; Wilson, 2024). The need for structured, well-maintained datasets is particularly critical, as AI frameworks depend on this foundation to produce reliable evaluations. This concern was also expressed by a survey participant who noted that "training could be a threshold", affirming literature insights that organizations adopting AI must invest in not only technical tools but also staff upskilling and data governance reforms.

Finally, while the AI-enhanced framework was widely credited with improving transparency and justification, with strong scores like 4.30 for *"The rationale for each rating is clear and easy to follow"*, the validation also revealed residual concerns that scores might sometimes be driven more by narrative reasoning than by hard performance outcomes. This finding is consistent with the literature's emphasis on transparency and accountability challenges (Nagbøl et al., 2021; Alzubaidi et al., 2023) and the risks noted by Spreitzenbarth (2021) and Cooper (2024) that, without thorough human critique, organizations might inadvertently rely too heavily on seemingly persuasive but potentially superficial AI rationales.

9.4. Interpretation of Results about Research Objectives

Following the linkage of the results to the literature, this section synthesizes the empirical findings of the study and directly addresses the primary and secondary research questions set out in Chapter 3 by explicitly connecting the outcomes of the framework implementation and validation to each research question. First, the sub-questions will be addressed, leading up to the final resolution of the main research question.

9.4.1 Critical Performance Criteria

Which performance criteria are most critical for HOCHTIEF's construction projects, and how can they be effectively structured for AI analysis?

The identification of critical subcontractor performance criteria for HOCHTIEF's construction projects emerged through a two-tiered methodological process that combined a systematic literature foundation with targeted organizational customization. Initially, an extensive literature review consolidated insights from diverse studies on subcontractor evaluation and procurement decision-making. Thus, the final set of criteria comprised:

1. Safety and Compliance
2. Quality of Workmanship
3. Timeliness and Schedule Adherence
4. Technical Ability & Qualifications
5. Cost Management
6. Reliability and Repeat Engagement
7. Financial Stability
8. Communication and Collaboration
9. Innovation and Problem-Solving
10. Sustainability Practices

Building on this theoretical foundation, the study then engaged HOCHTIEF's internal experts in semi-structured interviews. In total, seven people were interviewed. Participants were asked to rank and discuss the relative importance of the evaluation criteria identified from the literature, specifically within HOCHTIEF's risk profiles and strategic project objectives. This approach produced two weighting perspectives:

- the Equal Influence Model,
- the Role-Weighted Influence Model.

Empirical results established that the criterion "*Safety and Compliance"* dominated both models, underscoring its criticality across all functional perspectives. This was followed by "Quality of Workmanship" and "Timeliness and Schedule Adherence", directly aligning with established industry imperatives that prioritize the mitigation of safety risks and the assurance of quality and schedule fidelity. In contrast, *"Sustainability Practices"* and *"Innovation and Problem-Solving"* received the lowest weights, indicating they were regarded as strategically supportive but not primary drivers of subcontractor selection and evaluation decisions.

To structure these criteria in an AI-ready format, dual-perspective evaluation documents were designed. Each criterion was addressed with a plethora of questions. Each section is integrated:

- Closed, factual questions, such as "What type of contract was signed?" or "Were there additions to the scope?"
- Quantitative data fields, e.g., contract sums, actual costs, percentage deviations.
- Open narrative prompts, inviting stakeholders to explain reasons behind budget variations, site changes, or technical adaptations.
- Likert-scale performance ratings, where respondents scored aspects like substantiation of charges or overall cost management on a scale from 1 to 5.

This dual perspective approach created the essential precondition for ChatGPT 4.0's analytical engine: structured, multi-source inputs that allowed the assistant to detect inconsistencies, generate clarifying follow-up questions, and build well-justified, transparent performance assessments.

9.4.2 Addressing Discrepancies and Missing Details

How can ChatGPT 4.0 be used to handle discrepancies or missing details within subcontractor data?

The second supporting question examined how the integrated AI could address typical data quality problems, notably discrepancies or gaps between subcontractor self-assessments and HOCHTIEF's internal evaluations. A central feature of the framework was its ability to identify these inconsistencies automatically. By parsing the structured dual evaluations, the assistant flagged areas where data was lacking (e.g., *You mention a 'disagreement' over final financial aspects and rated HOCHTIEF's cost management/support as 3. Could you elaborate on the nature of this disagreement and what specific issues arose?")* or was contradicting (e.g., "*You answered 'No' to 'Does the subcontractor qualify in the necessary technical skills for the project?' but rated technical ability as 5/5. Please reconcile these two statements—is the subcontractor fully qualified, and if not, what justified a perfect score?"*). It then generated context-sensitive follow-up questions to probe these gaps.

Empirical validation results strongly supported this functionality. The high scores under the criterion *Clarity and Relevance of Follow-up Questions* (see chapter 9.2.2) indicate that the AI assistant effectively generated follow-up questions that were clear, professionally phrased, and well-suited for use in formal evaluation settings. Participants responded positively overall, suggesting that the questions were easy to understand and appropriate for their intended context. This reflects strong performance in crafting responses that align with professional communication standards and support productive dialogue. The questions helped clarify the evaluation documents and provided even more information that would have been missed otherwise.

9.4.3 Fine-tuning ChatGPT 4.0

How can ChatGPT 4.0 be fine-tuned to generate performance-related and clarifying questions to gather more information and provide context for subcontractor evaluations?

The third secondary research question explored the fine-tuning of ChatGPT 4.0's mechanisms for generating clarifying questions that could elicit further detail and context. In this study, the assistant's prompts were tightly constrained to data present in the structured evaluation forms through a rigorous design. Extensive pilot testing and multi-round case simulations ensured that the questions were not hypothetical or excessively generic but directly anchored in identified ambiguities or explicit rating divergences. Past evaluation documents from the Amaliahaven project were used to train the assistant's logic behind the generated questions. Once the questions met the required standards, they were used as examples in the assistant's personality box. The assistant was explicitly instructed that these questions are just examples and should be adjusted to the data available in the evaluation documents or the answers to the follow-up questions.

By embedding explicit ground rules that only detected discrepancies or incomplete narrative triggers would initiate follow-ups, the framework avoided generating overly broad, boilerplate questions. Instead, it focused on extracting targeted clarifications that concretely enriched the factual basis of the evaluations. This approach proved highly effective, as shown in the validation results.

9.4.4 Addressing Secondary Research Question 4

In what ways does integrating AI with an MCDA framework affect subcontractor ratings compared to manual evaluations?

A critical dimension of the study was quantifying how this AI-enhanced approach influenced final ratings relative to HOCHTIEF's established manual processes. This was evaluated using the MAPD across three weighting scenarios. Across all calculation scenarios, the MAPD ranged from 5.56% to 7.22%, remaining well below the 10% threshold for acceptable predictive consistency.

9.4.5 Lessons Learned and Best Practices

What lessons learned and best practices emerge from the integration process, and how can they be utilized by HOCHTIEF for future projects?

One of the most significant lessons learned was the benefit of incorporating dual-perspective evaluations, where both HOCHTIEF staff and subcontractors contribute feedback. This approach not only enhanced the fairness and depth of evaluations but also facilitated mutual understanding and accountability. It addressed a common gap in traditional assessments, which often failed to capture the subcontractor's perspective or explain discrepancies in perceived performance.

Another important takeaway was the value of structured and AI-generated follow-up questions, which helped clarify vague input, address inconsistencies, and extract more meaningful insights from evaluation data. By prompting evaluators to provide specific justifications, the AI assistant improved the transparency and justification of performance scores, which are frequently missing in manual reports. This clarity was particularly beneficial in improving the documentation of lessons learned and providing actionable recommendations, both of which are often neglected or inconsistently captured in traditional processes.

The project also highlighted the need for human oversight and high-quality data input. While the AI assistant performed very well, its effectiveness heavily depended on the quality and completeness of the input data it received. Without thoughtful and accurate evaluator input, the system's outputs could be misleading. Moreover, the team recognized the risk of overreliance on AI, particularly in high-stakes procurement decisions. As such, AI should be viewed as a support tool that augments human expertise rather than replaces it. Finally, several best practices emerged from this experience, which will be extensively analysed in Chapter 10.

9.4.6 Synthesizing the Primary Research Question

How can ChatGPT 4.0, integrated with an MCDA framework, lead to more informed subcontractor evaluations in HOCHTIEF's construction projects?

Drawing these findings together, the study demonstrates that integrating ChatGPT 4.0 within a structured MCDA framework can fundamentally elevate subcontractor evaluations at HOCHTIEF by achieving a synthesis of greater analytical depth, improved data transparency, and stable quantitative alignment. The AI's ability to systematically parse dual-perspective inputs, identify discrepancies, and generate targeted follow-ups resolved many traditional gaps in contractor assessments. Simultaneously, the framework maintained close numerical fidelity to established manual processes, with MAPD consistently below 7.3%, ensuring that enhanced interpretability and strategic learning did not come at the cost of decision stability.

Critically, however, the study also reaffirmed the literature's guidance that AI should serve as a decision support tool rather than an autonomous evaluator. The consistently high endorsement for expert review of AI reports illustrates the framework's appropriate positioning within HOCHTIEF's governance structure: as a rigorous, transparent augmentation of human judgment rather than a replacement. In this configuration, the integration of ChatGPT 4.0 with MCDA demonstrably leads to more informed subcontractor evaluations by enriching the factual and contextual basis on which decisions rest, thereby better supporting HOCHTIEF's strategic objectives in project delivery and risk management.

Chapter 10: Conclusions, Recommendations, and Reflection

Building on the results of the previous chapter, Chapter 10 synthesizes the overall contributions, limitations, and strategic implications of the study. It provides a critical interpretation of the validation outcomes, considering the research questions, outlines actionable recommendations for HOCHTIEF, and reflects on broader applications within the construction sector. The chapter concludes with forward-looking pathways for research and personal reflections on the skills and lessons gained during the thesis journey.

10.1 Discussion

10.1.1 Key Takeaways and Conclusions

This study aimed to explore whether integrating ChatGPT-4.0 into a structured Multi-Criteria Decision Analysis (MCDA) framework could lead to more informed, balanced, and transparent subcontractor evaluations within HOCHTIEF's construction operations. The research used a rigorous multi-stage validation process to establish empirical credibility.

From a quantitative perspective, the framework showed strong alignment with traditional evaluation methods. Across three MCDA weighting scenarios, the MAPD from manual evaluations ranged from 5.56% to 7.22%, all comfortably below the 10% threshold set as the acceptable variance. Most importantly, the qualitative evidence points to an increase in transparency during the process. The assistant's ability to identify inconsistencies and highlight gaps was also seen as a significant improvement over the current method. The targeted follow-up questioning mechanism, which sets the AI framework apart from single-pass human assessments, was highly valued by evaluators. The average score for the *Clarity and Relevance of Follow-up Questions* was 4.06, with the highest scores given to items like "addresses discrepancies effectively" (4.30) and "uses an appropriate tone for professional use" (4.60). One of the most unique contributions of the AI-enhanced framework was its integration of strategic, forward-looking learning. Under the category *of Quality of Lessons Learned and Recommendations*, participants gave an average score of 4.25. Evaluators showed a high willingness to incorporate such recommendations into future subcontractor selections, suggesting the AI assistant facilitates knowledge transfer in ways traditional reports often do not.

Despite these strengths, some concerns still exist. For example, the statement "the assistant sometimes relied more on the explanation than on actual project performance" received a moderate score of 3.60, indicating that evaluators sometimes perceived an imbalance in emphasis. Similarly, the relatively low average score of 3.10 for "the report included sufficient context to understand the subcontractor's role" revealed a gap in automated scene-setting. While the assistant could precisely analyze individual issues, it occasionally lacked the broader contextual framework necessary to interpret performance outcomes fully.

Insights from both the survey and interviews on human oversight and implementation factors were equally important. Although the assistant was praised for improving clarity and consistency, respondents emphasized that it should not operate independently. The survey showed a broad consensus that final judgments must still rest with human evaluators, who are better equipped to interpret subtle contextual cues, detect rationalizations, and prevent over-reliance on well-

articulated but potentially misleading explanations. During pilot testing and before strict instructions in the personality box were added, the virtual assistant sometimes generated fake data and based its reasoning on fictional information. This led to inaccurate results, as they were not grounded in real project data. Addressing these issues was possible with proper prompts, but human oversight remains essential.

Furthermore, as noted in Chapter 5, adopting such frameworks raises concerns about data quality. While the assistant performed reliably in structured test cases, its ability to generate accurate clarifications and informed recommendations depends heavily on the quality, completeness, and clarity of the input data. During piloting, even minor omissions or unclear wording in evaluator statements sometimes caused the assistant to make conservative guesses, issue overly generic follow-up prompts, or produce inaccuracies (hallucinations) to fill gaps. These issues were addressed through proper instructions; however, the quality of the input data, especially the evaluation documents, remains a critical factor. To produce accurate results and meaningful insights, the assistant must be provided with comprehensive, precise input, as more context and detail lead to stronger, more relevant evaluations.

10.1.2 Generalizability of the Framework

From the researcher's perspective, an important question emerges: *To what extent can this framework be generalized across different contexts, organizations, or supplier types?* Based on the findings and the implementation experience, the answer is nuanced. Certain foundational elements of the system, such as the core prompt logic, dual-perspective structure, and follow-up clarification mechanisms, appear to be highly transferable. These components are grounded not in project-specific content, but in universal decision-support principles such as cross-validation, transparency, and reasoned justification. Regardless of the industry, the practice of surfacing inconsistencies through guided AI questioning and enforcing rationale documentation can strengthen procedural fairness and auditability.

However, other parts of the framework are far more context-specific and would require substantial adjustment if applied beyond the subcontractor evaluation domain. For example, the evaluation templates and question libraries were carefully designed around construction-related performance indicators such as site coordination, documentation accuracy, and safety compliance. These dimensions would not directly apply to the evaluation of other supplier types like materials vendors, where criteria such as delivery reliability, batch traceability, or logistics integration would dominate. Similarly, the examples used to train the assistant and validate its reasoning were derived from construction project settings. Deploying the assistant in other sectors would necessitate the development of domain-specific examples and re-training to ensure the relevance and legitimacy of its output. The same can also be said about the weighing schemes. Although the weights of the criteria were not specific to the project or the type of subcontractor being evaluated, different companies have different priorities and strategic goals, thus affecting the importance of each criterion.

10.1.3 Reflections on Framework Design and Weighting Philosophy

The design of this framework was based on the belief that human judgment and artificial intelligence can be effectively combined, but only through careful coordination. Early in development, it became clear that simply “plugging in” an AI assistant without a structured approach for input, weighting, and output validation could lead to unreliable results. Therefore, significant focus was placed on creating

a comprehensive framework that ensured the assistant acted as a facilitator of structured dialogue rather than an opaque decision-maker.

In practice, this structure proved advantageous. As seen in the validation results, participants appreciated the assistant's transparency and consistency, especially in revealing overlooked discrepancies or prompting more detailed justifications. However, the process also highlighted the limits of automation in professional judgment contexts. Although the assistant could identify gaps or inconsistencies, it lacked the intuitive understanding that human evaluators use when interpreting project-specific nuances, especially when contextual information was missing or unclear. In this way, the framework worked best when it complemented, not replaced, expert judgment.

One design challenge that arose was balancing flexibility and standardization. On one side, fixed evaluation templates and follow-up sequences improved comparability and auditability across projects. On the other hand, this rigidity sometimes made the process seem too narrow, especially in unusual or complex subcontractor cases where evaluators preferred more tailored questioning or custom weighting logic. A key lesson is that future versions should incorporate modular flexibility: maintaining a standardized core process while allowing certain elements, such as clarification depth or evaluation dimensions, to adapt based on subcontractor risk profiles or project phase.

The question of weighting further underscored this tension. Between the two models tested, Role-Weighted and Equal Influence, the latter aligned more with core values of procedural fairness and balanced input. As a researcher, I found the Equal Influence model preferable, not only because it distributed power more evenly, but because it gave voice to those often underrepresented in decision-making, especially employees who observe subcontractor behaviors daily but are not directly involved in evaluations (like work preparators), at least at HOCHTIEF. While it may not precisely match organizational hierarchy, it democratizes evaluation in a way that reflects operational reality rather than reporting structures alone.

Ultimately, this project reaffirmed that decision support tools must be designed with both technical validity and organizational practicality in mind. An AI-enhanced framework must not only generate reasonable results but also be perceived as fair, transparent, and usable by its stakeholders. Building such a system requires iterative development, continuous validation, and a readiness to make trade-offs among accuracy, flexibility, and user trust.

10.2 Research Limitations

While the validation results confirm the viability and benefits of integrating ChatGPT-4.0 into a structured subcontractor evaluation process, several limitations emerge from a closer examination of the research design, scope, and contextual dependencies. These limitations do not undermine the contributions of the study but are critical to interpreting its outcomes accurately and to guiding future deployments or extensions.

A key limitation is that the case study relied on data from a single subcontractor in the data center project. This narrow focus limits the empirical breadth of the study. Moreover, the two perspectives compared in the evaluation were aligned mainly. Their assessments did not present intense contradictions or disputes, which would have stressed the assistant's ability to resolve conflicting inputs or mediate tension. As such, the assistant's questioning logic, clarification mechanisms, and rating synthesis were not tested under conditions of evaluative conflict or strategic divergence. This leaves open the question of how the system would perform when parties disagree fundamentally, such as in contested claims, underperformance disputes, or post-project debriefs marked by blame

allocation. In future validations, a more diverse sample of subcontractors will be needed to rigorously evaluate the assistant's robustness in high-stakes or adversarial contexts.

Another important limitation stems from the construction of the Role-Weighted Influence model. While this model aimed to reflect organizational hierarchy by assigning greater influence to senior roles such as project managers and procurement leads, the actual weighting values were subjectively defined. They were based on the researcher's assumptions and lacked formal calibration or empirical grounding. This introduces potential arbitrariness and bias. Different project types may require different weighting structures. For instance, in high-risk infrastructure projects, HSE roles may need greater influence than in procurement-driven frameworks. The static nature of the weighting matrix used in this study thus fails to accommodate this contextual variability. This suggests that any future application of role-weighted influence should consider either dynamic weighting models, adjustable per project, or statistical calibration based on past evaluation accuracy and outcomes.

Furthermore, the survey-based validation component, though carefully designed, involved a relatively small number of participants. While the quality of feedback was high and the consistency of results across items was encouraging, the limited sample size constrains the statistical generalizability of findings. Larger and more diverse validation rounds will be essential to confirm whether the assistant's perceived clarity, helpfulness, and neutrality persist across broader user bases, project scales, and subcontractor types.

Additionally, the current study focused exclusively on subcontractor evaluations. This choice of domain shaped not only the input documents but also the prompts, evaluation criteria, and follow-up questioning logic used by the assistant. These components are closely tied to the realities of subcontractor performance assessment, including on-site behavior, safety compliance, and documentation quality. As a result, if the framework were to be adapted for use with different vendor categories (e.g., material suppliers, consultants, or equipment leasing partners), core elements of the system would need to be redefined, as mentioned in Chapter 10.1.2

Moreover, the assistant operated in a relatively static decision space: evaluation questions were fixed, data sources limited, and follow-up constrained to pre-defined sections. Future extensions should examine whether the assistant can adaptively change its questioning scope or even recommend additional inputs if it identifies gaps in the evaluation data.

10.3 Recommendations to HOCHTIEF

Based on the findings of the case study, survey validation, and interviews, a set of detailed recommendations will be given to HOCHTIEF to strengthen its subcontractor evaluation practices. These recommendations are not just about making the AI framework work better; they also aim to improve the broader organizational process of performance assessment, data handling, and knowledge management. Each section draws directly on the evidence gathered during this research, combined with broader lessons from procurement and construction management best practices.

10.3.1 Organizational Improvements and Strengthening Evaluation Processes

1. **Improve Use of Existing Evaluation Database through Awareness and Integration**

One surprising insight from interviews was that while HOCHTIEF already has a digital database capable of storing subcontractor evaluations, many employees were simply unaware of it. This gap

weakens the long-term value of any evaluation system. Without systematic storage and retrieval, valuable insights into subcontractor performance may be lost when projects end or staff changes. It is therefore crucial that HOCHTIEF communicates clearly across all teams that this database exists and is the official single source of truth for storing evaluations. More importantly, managers should ensure that every completed evaluation, whether AI-assisted or done manually, is uploaded. To reinforce this, HOCHTIEF can update internal procedures so that final invoice releases or project close-out reports explicitly require evaluations to be filed in the database.

Regular short reminders during team meetings or onboarding sessions can also build this habit. Over time, it will transform the database into a powerful asset: a central, easily searchable history of subcontractor performance that procurement teams can draw on when shortlisting firms for future projects, or when needing evidence to justify decisions in audits.

2. **Make Evaluations a Regular, Ongoing Activity and Not Just an End-of-Project Task**

Another key improvement area identified in this research is that evaluations are not performed very often. At this point, opportunities to address might be lost. Integrating evaluations into regular project reviews, for instance, on a monthly or milestone basis, allows problems to be identified and addressed while the work is still underway.

This recommendation is not only about more frequent scoring. It means using the structured evaluation templates already developed to capture how the subcontractor is performing on these dimensions routinely. Feeding this richer data into the AI model or simply using it to guide discussions in monthly progress meetings creates a continuous loop of monitoring and improvement. In turn, this also builds a much more robust historical record. When future teams review a subcontractor's past performance, they will find a comprehensive timeline of performance evidence, not just a single summary judgment

3. **Make Subcontractor Evaluations a Standard Part of Staff Handover**

The interviews revealed that when employees move on to new assignments, they often take valuable insights about subcontractors with them; thus, important knowledge can be lost. To prevent this, HOCHTIEF should formally include an evaluation handover in its internal role exit or project demobilization processes. This means requiring outgoing team members to complete pending evaluations and to brief successors on subcontractor's performance records, using the vendor database as the common reference point. This ensures continuity of knowledge, prevents repeating mistakes, and eases the transition and acclimation of the new team into the project as they are now more well-informed about the subcontractors they must work with.

4. **Set Transparent Expectations by Sharing Evaluation Criteria Early**

During interviews, employees noted that subcontractors are unclear about how exactly they will be evaluated once their work is done. This lack of transparency can lead to misunderstandings, disputes, or a feeling that assessments are subjective. Sharing the evaluation framework by showing the right at the start of each project can reduce confusion and encourage subcontractors to focus on the behaviors that matter most. Clear communication also strengthens commercial relationships. When subcontractors understand that consistently high scores will directly influence their chances of being selected for future work, they have an extra incentive to meet or exceed expectations throughout the project.

5. **Always Use Dual-Perspective Evaluations, and Emphasize High-Quality, Detailed Data**

Perhaps the most distinctive and important finding of this research is how valuable it is to require structured evaluations from both HOCHTIEF's internal team and the subcontractor. This dual-input design allows discrepancies to surface naturally. In this study, even though the case showed generally close alignment between perspectives, the framework's ability to highlight areas of disagreement proved critical. Ensuring both sides complete the similarly structured forms provides balanced data for the AI model. However, it also guards against the risk of one-sided stories, even if AI is not used. When subcontractor assessments are stored alongside HOCHTIEF's, future decision-makers get a richer, more transparent view.

For this approach to succeed, however, both parties must provide high-quality, detailed data. The AI model can only produce meaningful clarifications and lessons learned if it has concrete input to work with. Even without AI, vague or overly cautious answers (like "everything was satisfactory") do little to support fair, documented decisions. Staff should be trained to understand that supplying honest, specific details about both good and bad performance ultimately protects them. It creates an objective, traceable record that justifies future contract decisions and is subject to audit.

6. **Review and Adjust Weights on Evaluation Criteria Regularly**

The research compared different weighting schemes and found each brought trade-offs. HOCHTIEF's strategic priorities evolve, for example, with new ESG targets, the relative importance of safety, sustainability, quality, or cost may change. This means weights should be revisited at least annually, with directors discussing whether adjustments are needed. Documenting these reviews ensures the system remains aligned to business needs and that all stakeholders understand why certain factors are weighted more heavily at different times.

When it comes to the proposed framework, it is important to:

7. **Deliberately Test the System on Projects with More Disagreements**

Because the case study featured relatively harmonious evaluations, the framework's ability to handle serious data conflicts was only lightly exercised. HOCHTIEF should subsequently trial the system on projects where they and the subcontractor are more likely to have differing perspectives, such as when prior disputes exist. This would give the AI's clarifying logic and follow-up questioning more challenging cases to work through, helping to refine prompt libraries and stress-test how the tool facilitates consensus-building. Learning from challenging projects now will increase confidence in using the framework on future high-risk contracts.

8. **Strengthen Data Quality through Targeted Training and Practical Incentives**

The AI model, and even manual evaluations, are only as good as the data they receive. Short, focused training sessions for HOCHTIEF teams and subcontractors can help by explaining how detailed, honest evaluations protect all parties. Linking thorough, timely subcontractor self-assessments to operational benefits could also encourage higher data quality.

9. **Embed AI Reports and Lessons Learned into Company Knowledge Systems**

One of the most apparent benefits of the new framework was the way it generated explicit lessons learned and targeted recommendations. However, if these insights remain trapped in isolated

project folders, their value is lost. HOCHTIEF should ensure that key findings are summarized and entered into shared knowledge platforms, whether that is a lessons learned database or integrated into project handover packs. This will help new teams identify recurring issues with specific subcontractors or work types, preventing mistakes from repeating.

10. **Keep Enhancing the AI Tool with User-Driven Features**

Improvements like context-sensitive prompt templates and dynamic weighting sliders. HOCHTIEF should prioritize these upgrades to keep the tool intuitive, adaptable, and easy to trust. Adding interactive dashboards that let managers explore how changes in weights or new clarifications shift final ratings will make the AI system even more transparent and user-friendly.

10.3.2 Application of the proposed AI-driven framework

As reviewed in Chapter 2, the Technology–Organization–Environment (TOE) framework provides a structured lens to understand how HOCHTIEF can successfully adopt the AI-enhanced subcontractor evaluation framework. It emphasizes three critical pillars that must be addressed simultaneously:

- the technological infrastructure and capabilities needed to deploy the solution,
- the organizational conditions and culture that influence adoption and integration, and
- the external environmental factors, such as market standards, client expectations, and regulatory requirements, shape how and why innovations are adopted.

In case HOCHTIEF decides to use this AI-driven framework, TOE can help ensure that this deployment is not just a technical project, but a transformation aligned with internal processes and external demands.

The Technology Dimension

Under TOE's technology pillar, HOCHTIEF must focus on leveraging its existing digital infrastructure. The company already maintains a structured database for subcontractor evaluations. However, the case study and interviews showed that employees are often unaware of it or do not systematically use it.

Addressing this requires clear policies, making this database the single source of accurate and trustworthy information for all subcontractor performance data, whether produced manually or through AI. Integrations with platforms like Aconex or SAP can ensure that evaluations, clarifications, and lessons learned feed into ongoing project dashboards, reducing fragmentation.

The Organization Dimension

The organizational pillar highlights that no technological solution succeeds without human systems and culture to support it. During the case study, it became clear that obtaining high-quality data is challenging, as people often view evaluation documents as a mere formality. They tend to complete them to fulfill the requirement, rather than to share all the relevant information they possess

To change this, HOCHTIEF should embed the evaluation framework into broader operational and performance management processes. The company should implement targeted interventions that directly address the root causes of disengagement:

- Make Evaluation Effort Visible and Measurable

  Introduce a scoring system that rates the completeness and usefulness of each evaluation. This score can be visible to team leads and tied to performance reviews, encouraging staff to take evaluations seriously and provide richer input.

- Standardize Evaluation Triggers

  Rather than relying on ad hoc submissions, embed mandatory evaluation checkpoints into project workflows (e.g., after milestone completions, issue resolution, or contract close-out). This ensures that evaluations are timely and relevant.

- Train for Judgment, Not Just Compliance

  Offer short, scenario-based training modules that teach staff how to assess subcontractor performance critically. Focus on judgment calls, what constitutes “acceptable” vs. “exceptional” work, so evaluations reflect real expertise, not just form-filling.

- Link Evaluation to Procurement Decisions

  Make it clear that subcontractor evaluations directly influence future contract awards. When teams understand that poor documentation can lead to poor vendor selection, the incentive to provide accurate data becomes tangible.

The Environment Dimension

The environment pillar ensures alignment with external pressures. Clients and regulators demand evidence of fair, transparent supplier management, and documented compliance with standards on safety, quality, and even sustainability.

By maintaining structured evaluations, explicit clarifications, and transparent dual perspectives, HOCHTIEF will be able to demonstrate due diligence during audits or disputes easily. Furthermore, by transparently communicating evaluation criteria to subcontractors at contract initiation, HOCHTIEF can foster a more constructive, trust-based supply chain, thus reducing the adversarial dynamics often found in construction.

The Unified Theory of Acceptance and Use of Technology (UTAUT)

Another model that was also reviewed in Chapter 2 is the Unified Theory of Acceptance and Use of Technology. It complements the TOE framework by focusing specifically on what drives individual employees to embrace (or reject) new systems. It identifies four key determinants:

- **Performance expectancy** (do employees believe the tool will help them do their jobs better?),
- **Effort expectancy** (is it easy to learn and use?),
- **Social influence** (do respected peers and leaders endorse it?), and
- **Facilitating conditions** (do employees have access to help, training, and infrastructure?).

Applying these systematically will be critical for achieving broad buy-in for the AI-enhanced evaluation framework.

For staff to embrace the new evaluation framework, they must see how it makes their daily work easier and contributes to better project outcomes. Throughout the validation process, participants consistently highlighted that the AI-enhanced reports brought more structure and clarity to subcontractor assessments. Participants further agreed that the AI framework reduced ambiguity (4.00) and improved their confidence in evaluations by adding objectivity (4.60). These tangible benefits should be prominently featured in internal communications and training. Using examples from the pilot, such as how follow-up questions brought inconsistencies to light early or how explicit scoring rationales improved transparency, will reinforce the practical advantages. Framing the tool as something that saves time during audits, supports claims with solid documentation, and reduces confusion in subcontractor discussions will also strengthen its perceived usefulness.

Even the best-designed systems can fail if staff see them as time-consuming or complicated. Encouragingly, the validation survey found that users largely viewed the new framework as straightforward to use. Future enhancements, such as intelligent dashboards or integrations that automatically pull key context from systems like Aconex, will be critical to keep perceived effort low. Even simple features like auto-populating known project data can substantially cut manual entry. This directly supports the “effort expectancy” dimension in the UTAUT framework by reducing barriers to use and embedding the tool more naturally into staff workflows.

Moreover, change is more readily adopted when respected peers and leadership visibly support it. The UTAUT framework highlights this as “social influence,” and it is a vital lever for ensuring widespread adoption. During validation, several participants noted that the new process provided more explicit justifications for scores, making them easier to explain internally and stand up better in discussions with subcontractors. By capturing and sharing these positive user experiences in short case examples, we can build peer-driven momentum.

Additionally, clear backing from senior leadership is essential. Formal endorsements that the AI-enhanced framework can be a helpful tool and can provide an extra “opinion” on the performance of the subcontractor could create organizational clarity and demonstrate top-level commitment. This helps establish a shared norm, making adoption feel like part of the company culture rather than an individual choice.

Lastly, employees need support systems to adopt and sustain the use of the framework confidently. This goes beyond IT helpdesks. Short, targeted training modules should be developed that not only show how to operate the system but also explain why providing honest, detailed data matters. The study made clear that high-quality inputs were essential: where vague or overly generic statements were given, the AI sometimes produced broader or less tailored clarifications.

Staff should understand that supplying thorough, specific details protects them by creating a transparent, traceable record that supports fair contractor decisions and stands up under future scrutiny, whether or not AI is involved. Regular clinics or drop-in sessions where staff can raise questions, see new features demonstrated, or discuss adjustments to weighting models will also help embed the system. This approach ensures strong “facilitating conditions,” reinforcing that employees are not only equipped but actively supported in adopting these enhanced evaluation practices.

10.4 Practical Implications for the Construction Sector

The findings of this study have broader implications not only for HOCHTIEF but also for the wider construction industry. The integration of ChatGPT-4.0 within a structured MCDA framework offers improvements across multiple functions. However, these benefits must be critically interpreted within the constraints and operational complexities of construction procurement.

Enhancing Auditability and Traceability

One of the most immediate and pragmatic outcomes of the AI-augmented evaluation process is its contribution to auditability. Traditional subcontractor assessments often suffer from undocumented rationales or vague language that limits traceability in post-project reviews. By contrast, the assistant's structured follow-up questions and justification logs offer a roadmap of the decision-making, which can be revisited to understand how final scores were derived. This is particularly relevant for publicly funded projects or joint ventures where third-party scrutiny and accountability are essential. The assistant's logs could serve as compliance documentation for ISO 9001 quality management audits or internal performance reviews.

Strengthening Stakeholder Trust and Reducing Adversarialism

The dual-perspective model used in this framework proved to be not only technically feasible but trust-enhancing. During the follow-up meeting for the case study (see Chapter 8), the subcontractor expressed appreciation for having his voice formally captured. At the same time, internal evaluators acknowledged the fairness introduced by dual inputs. In an industry often marked by adversarial contractor-subcontractor dynamics, such inclusion contributes to greater procedural justice. Even if disagreements occur, the transparent structure of clarifications and score justification makes it more difficult for either party to claim unilateral bias. This has long-term benefits for professional relationships and for avoiding escalations into legal disputes.

Accelerating Organizational Learning

A significant value-add emerged from the assistant's "Lessons Learned" generation capability, which was frequently rated as insightful and usable (see 4.25 average score in Chapter 8). Unlike traditional evaluations that are stored and forgotten, this system explicitly links subcontractors' performance to context-rich narratives. When aggregated across projects, these structured insights form the basis for a reusable strategic knowledge base, capable of informing contractor selection criteria, negotiation points, and even contract design.

Supporting Regulatory Compliance and Governance

As regulatory demands on construction firms grow, particularly around ESG performance, anti-corruption, and public procurement rules, the AI framework can help ensure that evaluations are both defensible and compliant with formal standards. The traceability of follow-up logic, alignment to documented inputs, and explicit role-based participation provide a strong foundation for legal audit trails. Moreover, if disputes over subcontractor ratings arise, the AI-generated logs can serve as structured evidence, reducing reliance on subjective or memory-based justifications.

10.5 Future Research Pathways

An important direction for future research is to conduct broader validation studies by applying the AI–MCDA framework across various companies, subcontractor types, and project contexts. Although this study focused on a single HOCHTIEF case, testing the framework in other organizational environments and with diverse project portfolios will help establish its generalizability. This will confirm whether the strong alignment with manual evaluations and the increased transparency observed here consistently hold true across different industrial and cultural settings.

Another important area involves examining alternative weighting models within the MCDA integration. The current study used equal, equal-influence, and role-weighted approaches, but future research could develop adaptive weighting systems that automatically adjust based on project type, size, or risk profile. These models could create a more nuanced balance between fairness, inclusivity, and the role of domain-specific expertise.

Future studies should also focus on enhancing context integration. One of the limitations identified in the validation was AI's restricted ability to account for broader project circumstances beyond the structured evaluation data provided. Research could examine how linking the evaluation framework with other digital systems, such as BIM models, ERP systems, or site-level metadata, might provide richer situational awareness, enabling AI to deliver even more grounded and context-sensitive insights.

A fourth avenue is longitudinal tracking of subcontractor performance. By extending evaluations beyond a single project to track subcontractors' performance across multiple engagements, researchers could measure how AI-generated evaluations align with long-term outcomes such as delivery reliability, defect rates, and repeat engagement. Such longitudinal studies would also reveal whether AI-enhanced evaluations encourage sustainable performance improvements rather than short-term compliance with evaluation criteria. In addition, they would help track whether the AI-generated recommendations and lessons learned are actively implemented, taken into account, and lead to measurable improvements in subcontractor performance and the overall quality of relationships between HOCHTIEF and its subcontractors.

Equally important is research into training and organizational change. The effectiveness of AI-assisted evaluations depends not only on the algorithms but also on how people adopt and use them. Future studies could examine how training programs, change management strategies, and organizational culture influence the adoption, trust, and practical effectiveness of the framework. Insights here would help firms like HOCHTIEF design rollout strategies that maximize value while minimizing resistance.

Finally, further work should examine governance and oversight models. Survey participants stressed that AI outputs must always be reviewed by humans before being used officially, underscoring the need for hybrid approaches. Future research could investigate methods for balancing AI decision support with human expertise, especially in high-stakes procurement decisions. Questions around accountability, bias detection, and explainability remain central to ensuring the ethical and effective use of AI in construction evaluations.

References

Abbasianjahromi, H., Rajaie, H., & Shakeri, E. (2013). A framework for subcontractor selection in the construction industry. *Journal of Civil Engineering and Management, 19*(2), 158–168. https://doi.org/10.3846/13923730.2012.743922

Abdulla, A., & Baryannis, G. (2024). A hybrid multi-criteria decision-making and machine learning approach for explainable supplier selection. *Supply Chain Analytics, 7*, 100074. https://doi.org/10.1016/j.sca.2024.000074

Abdulla, A., Baryannis, G., & Badi, I. (2023). An integrated machine learning and MARCOS method for supplier evaluation and selection. *Decision Analytics Journal, 9*, 100342. https://doi.org/10.1016/j.dajour.2023.100342

Adelakun, B., Onwubuariri, E., Adeniran, G., & Ntiakoh, A. (2024). Enhancing fraud detection in accounting through AI: techniques and case studies. Finance & Accounting Research Journal, 6(6), 978-999. https://doi.org/10.51594/farj.v6i6.1232

Akintayo, O., Eden, C., & Onyebuchi, N. (2024). Evaluating the impact of digital technology on youth leadership development in the USA. GSC Advanced Research and Reviews, 19(1), 132-145. https://doi.org/10.30574/gscarr.2024.09.1.0148

Akmaludin, Akmaludin & Samudi, Samudi & Palasara, Nicodias & Harmono, Feri & Widianto, Kudiantoro & Muharrom, Muhammad. (2023). MCDM-AHP and PROMETHEE methods integrated for base service strategy vendor evaluation and selection. International Journal of Advances in Applied Sciences. 12. 384. 10.11591/ijaas.v12.i4.pp384-395.

Alasmri, N., & Basahel, S. (2022). Linking artificial intelligence use to improved decision-making, individual and organizational outcomes. *International Business Research, 15*(10), 1. https://doi.org/10.5539/ibr.v15n10p1

Alshehhi, K., Cheaitou, A., & Rashid, H. (2023). Fuzzy failure modes, effects, and criticality analysis of the procurement process of artificial intelligence systems/services. International Journal of Advanced Computer Science and Applications, 14(10). https://doi.org/10.14569/ijacsa.2023.0141060

Althabatah, A., Yaqot, M., Menezes, B., & Kerbache, L. (2023). Transformative procurement trends: integrating industry 4.0 technologies for enhanced procurement processes. Logistics, 7(3), 63. https://doi.org/10.3390/logistics7030063

Armstrong, B., Kellogg, K., Levi, R., Shah, J., & Wiesenfeld, B. (2024). Implementing generative AI in US hospital systems. https://doi.org/10.21428/e4baedd9.1729053f

Bahameish, B., Yaqot, M., Franzoi, R., & Menezes, B. (2022). Artificial intelligence in procurement: an overview and case study of Qatar Foundation. https://doi.org/10.46254/eu05.20220146

Bartolini, F. & Viaggi, D. (2010). Recent developments in multi-criteria evaluation of regulations. Quality Assurance and Safety of Crops & Foods, 2(4), 182–196. https://doi.org/10.1111/j.1757-837x.2010.00076.x

Basu, R., Nanyam, V. N., & Sawhney, A. (2017). A Multi-dimensional subcontractor evaluation framework for nonconventional housing systems. Procedia Engineering, 196, 253–261. https://doi.org/10.1016/j.proeng.2017.07.197

Bowerman, B. L., O'Connell, R. T., & Koehler, A. B. (2005). *Forecasting, time series, and regression: An applied approach* (4th ed.). Thomson Brooks/Cole.

Burton, J., Stein, M., & Jensen, T. (2019). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making, 33*(2), 220–239. https://doi.org/10.1002/bdm.2155

Cannavacciuolo, L., Iandoli, L., Ponsiglione, C., & Zollo, G. (2015). An evaluation framework for the selection of project risk management software applications. *Expert Systems with Applications, 42*(10), 4467–4480. https://doi.org/10.1016/j.eswa.2015.01.052

Cooper, M. (2024). Agile procurement in a changing marketplace: examining adaptability and responsiveness in supply chain management. https://doi.org/10.20944/preprints202407.0514.v1

Cui, R., Li, M., & Zhang, S. (2022). AI and procurement. Manufacturing & Service Operations Management, 24(2), 691–706. https://doi.org/10.1287/msom.2021.0989

Davianto, H. (2022). The advantages of artificial intelligence in operational decision-making. *Hasanuddin Economics and Business Review, 6*(1), 24. https://doi.org/10.26487/hebr.v6i1.5082

Deep, S., Gajendran, T., Jefferies, M., Uggina, V., & Patil, S. (2022). Influence of subcontractors' "strategic capabilities" on "power", "dependence" and "collaboration": an empirical analysis in the context of procurement decisions. *Engineering Construction & Architectural Management, 31*(2), 571–592. https://doi.org/10.1108/ecam-04-2022-0346

Dima, A., Lukens, S., Hodkiewicz, M., Sexton, T., & Brundage, M. (2021). Adapting natural language processing for technical text. *Applied AI Letters, 2*(3). https://doi.org/10.1002/ail2.33

Earley, S. & Mehta, S. (2024). Powerful tools for personalisation: using large language model-based agents, knowledge graphs, and customer signals to connect with users. AMA, 10(3), 271. https://doi.org/10.69554/nmce9908

Egwim, C. N., Alaka, H., Demir, E., Balogun, H., Olu-Ajayi, R., Sulaimon, I., Wusu, G., Yusuf, W., & Muideen, A. A. (2024). Artificial Intelligence in the Construction Industry: A Systematic Review of the Entire Construction Value Chain Lifecycle. *Energies, 17*(1), 182. https://doi.org/10.3390/en17010182

Elmousalami, H., Alnaser, A. A., & Hui, F. K. P. (2025). Sustainable AI-driven wind energy forecasting: advancing zero-carbon cities and environmental computation. *Artificial Intelligence Review*, *58*(6). https://doi.org/10.1007/s10462-025-11191-0

Emmanouil-Kalos, A. (2024). Integrating multi-criteria decision analysis in healthcare policy and practice. Hapsc Policy Briefs Series, 5(2), 43–54. https://doi.org/10.12681/hapscpbs.40780

Ezeji, C. (2024). Artificial intelligence for detecting and preventing procurement fraud. International Journal of Business Ecosystem and Strategy (2687-2293), 6(1), 63-73. https://doi.org/10.36096/ijbes.v6i1.477

Gallear, D., Ghobadian, A., & He, Q. (2021). Relationship between routines of supplier selection and evaluation, risk perception, and propensity to form buyer–supplier partnerships. *Production Planning & Control.* https://doi.org/10.1080/09537287.2021.1872811

Gidiagba, J. O., Tartibu, L. K., & Okwu, M. O. (2025). A systematic review of machine learning applications in sustainable supplier selection. *Decision Analytics Journal, 100547*. https://doi.org/10.1016/j.dajour.2025.100547

Guida, M., Caniato, F., Moretto, A., & Ronchi, S. (2023). The role of artificial intelligence in the procurement process: State of the art and research agenda. *Journal of Purchasing and Supply Management, 29*(2), 100823. https://doi.org/10.1016/j.pursup.2023.100823

Handfield, R., Jeong, S., & Choi, T. (2019). Emerging procurement technology: Data analytics and cognitive analytics. *International Journal of Physical Distribution & Logistics Management, 49*(10), 972–1002. https://doi.org/10.1108/IJPDLM-11-2017-0348

Hansen, P. & Devlin, N. (2019). Multi-criteria decision analysis (MCDA) in healthcare decision-making. https://doi.org/10.1093/acrefore/9780190625979.013.98

He, X. (2023). Research on the relationship between perceived AI substitution crisis and employees' negative work behavior: from the perspective of job insecurity, 384-395. https://doi.org/10.2991/978-94-6463-200-2_40

Honhon, D., Gaur, V., & Seshadri, S. (2012). A multi-supplier sourcing problem with a preference ordering of suppliers. *Production and Operations Management, 21*(6), 1028–1041. https://doi.org/10.1111/j.1937-5956.2012.01346.x

Hu, H., & Ren, Z. (2023). Optimizing building material supplier selection through integrated interval-valued intuitionistic fuzzy multi-attribute decision making. *International Journal of Knowledge-Based and Intelligent Engineering Systems, 27*(4), 489-502. https://doi.org/10.3233/kes-221505

Huang, H., Canoy, R., Brusselaers, N., & Boveldt, G. (2023). Criteria preprocessing in multi-actor multi-criteria analysis. Journal of Multi-Criteria Decision Analysis, 30(3-4), 132-146. https://doi.org/10.1002/mcda.1804

Jiang, B., Chen, W., Zhang, H., & Pan, W. (2013). Supplier's efficiency and performance evaluation using the DEA-SVM approach. Journal of Software, 8(1). https://doi.org/10.4304/jsw.8.1.25-30

Keshavarz-Ghorabaee, M., Amiri, M., Hashemi-Tabatabaei, M., Zavadskas, E., & Kaklauskas, A. (2020). A new decision-making approach based on fermatean fuzzy sets and WASPAS for green construction supplier evaluation. *Mathematics, 8*(12), 2202. https://doi.org/10.3390/math8122202

Keshavarz-Ghorabaee, M., Amiri, M., Zavadskas, E., Turskis, Z., & Antuchevičienė, J. (2021). Determination of objective weights using a new method based on the removal effects of criteria (MEREC). *Symmetry, 13*(4), 525. https://doi.org/10.3390/sym13040525

Keshavarz-Ghorabaee, M., Amiri, M., Zavadskas, E., Turskis, Z., & Antuchevičienė, J. (2018). A dynamic fuzzy approach based on the EDAS method for multi-criteria subcontractor evaluation. *Information, 9*(3), 68. https://doi.org/10.3390/info9030068

Lewis, C. D. (1982). *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. Butterworth Scientific.

Lee, C. (2023). Impact of artificial intelligence on purchasing and procurement in South Korea. Global Journal of Purchasing and Procurement Management, 2(1), 1–11. https://doi.org/10.47604/gjppm.1954

Li, A., Wang, G., & Zhang, J. (2014). The research on the railway construction project materials supplier selection model. *Advanced Materials Research, 919–921*, 1503–1508. https://doi.org/10.4028/www.scientific.net/amr.919-921.1503

LI, H. (2024). Ai-powered negotiations: opportunities, challenges, and the future of business strategy. TEBMR, 13, 148-154. https://doi.org/10.62051/dg1trh68

Li, J. (2025). Applying large language model analysis and backend web services in regulatory technologies for continuous compliance checks. Future Internet, 17(3), 100. https://doi.org/10.3390/fi17030100

Liu, K., Su, Y., & Zhang, S. (2018). Evaluating supplier management maturity in prefabricated construction projects: survey analysis in China. *Sustainability, 10*(9), 3046. https://doi.org/10.3390/su10093046

Lumanauw, R., Ng, P., Saptari, A., Halim, I., Toha, M., & Ng, Y. (2023). Performance evaluation of subcontractors using the weighted sum method through kpi measurement. Journal of Engineering Technology and Applied Physics, 5(1), 35-49. https://doi.org/10.33093/jetap.2023.5.1.4

Maestrini, V., Luzzini, D., Caniato, F., Maccarrone, P., & Ronchi, S. (2018). The impact of supplier performance measurement systems on supplier performance. *International Journal of Operations & Production Management, 38*(11), 2040-2061. https://doi.org/10.1108/ijopm-10-2016-0589

Matsuno, S., Tagawa, S., Uchida, Y., & Ito, T. (2014). A relationship analysis between green supply chain management and its performance: a path analytic model. Journal of Robotics Networking and Artificial Life, 1(2), 145. https://doi.org/10.2991/jrnal.2014.0.2.10

McBride, K., Noordt, C., Misuraca, G., & Hammerschmid, G. (2021). Towards a systematic understanding of the challenges of procuring artificial intelligence in the public sector. https://doi.org/10.31235/osf.io/un649

Mishra, S., Porwal, P., & Yadav, D. (2021). Application areas of data science and AI for the improved society 5.0 era, 53–76. https://doi.org/10.1201/9781003097181-4

Montibeller, G., Gummer, H., & Tumidei, D. (2006). Combining scenario planning and multi-criteria decision analysis in practice. Journal of Multi-Criteria Decision Analysis, 14(1-3), 5–20. https://doi.org/10.1002/mcda.403

Morgan, P. (2021). The path to AI in procurement. AIRWA, 1(1), 6. https://doi.org/10.69554/ihjv3564

Murekatete, Z., & Dushimimana, J. (2023). Effect of supplier selection management on the performance of construction companies in Rwanda. *Strategic Journals of Business & Construction Management, 10*(4). https://doi.org/10.61426/sjbcm.v10i4.2743

Mutai, J. (2016). Effects of supplier evaluation on procurement performance of public universities in kenya. International Journal of Economics, Finance and Management Sciences, 4(3), 98. https://doi.org/10.11648/j.ijefm.20160403.12

Mwangata, M. and Chrine, C. (2024). An assessment of the effect of e-procurement on procurement processes and efficient performance in Zambia's government institutions: a case study of the local government service commission. Journal of Economics, Finance and Management Studies, 07(07). https://doi.org/10.47191/jefms/v7-i7-91

Nagbøl, P., Müller, O., & Krancher, O. (2021). Designing a risk assessment tool for artificial intelligence systems. In *Designing Systems for People and the Environment* (pp. 328-339). Springer. https://doi.org/10.1007/978-3-030-82405-1_32

Ngowi, A. B., & Pienaar, E. (2005). Trust factor in construction alliances. Building Research & Information, 33(3), 267–278. https://doi.org/10.1080/09613210500042895

Nwaguru, P., John, N., & Koko, N. (2022). Managing Buyer-Supplier Relationship in Construction Project Outsourcing. *European Journal of Logistics, Purchasing and Supply Chain Management, 10*(2). https://doi.org/10.37745/ejlpscm.2013/vol10n2114

Obinna, A. & Kess-Momoh, A. (2024). Comparative technical analysis of legal and ethical frameworks in AI-enhanced procurement processes. World Journal of Advanced Research and Reviews, 22(1), 1415–1430. https://doi.org/10.30574/wjarr.2024.22.1.1241

Ogbewele, E., Mbata, A., & Nwosu, N. (2024). Optimizing pharmaceutical inventory management: a global framework for efficiency and cost reduction. International Journal of Management & Entrepreneurship Research, 6(10), 3357-3371. https://doi.org/10.51594/ijmer.v6i10.1638

Ogundipe, O., Okwandu, A., & Abdulwaheed, S. (2024). Optimizing construction supply chains through AI: streamlining material procurement and logistics for project success. GSC Advanced Research and Reviews, 20(1), 147–158. https://doi.org/10.30574/gscarr.2024.20.1.0258

Osiro, L., Costa, R., & Lima, F. (2021). Evaluating supplier sustainability using fuzzy 2-tuple representation. *Gestão & Produção, 28*(1). https://doi.org/10.1590/1806-9649.2020v28e4933

Polat, G. (2015). Subcontractor selection using the integration of the AHP and PROMETHEE methods. *Journal of Civil Engineering and Management, 22*(8), 1042-1054. https://doi.org/10.3846/13923730.2014.948910

Rane, N. L. (2023). Integrating Leading-Edge Artificial Intelligence (AI), Internet of Things (IoT), and Big Data technologies for smart and sustainable Architecture, Engineering and Construction (AEC) industry: Challenges and future directions. *International Journal of Data Science and Big Data Analytics, 3*(2), 73–95. https://doi.org/10.51483/ijdsbda.3.2.2023.73-95

Sadeghpour, F., & Isaac, S. (2015). A comparative study of the owner, contractor, and consultant perspectives on the selection criteria for subcontractors. *Organization Technology and Management in Construction, an International Journal, 7*(2), 1330–1341. https://doi.org/10.5592/otmcj.2015.2.7

Scott, J., Ho, W., Dey, P. K., & Talluri, S. (2015). A decision support system for supplier selection and order allocation in stochastic, multi-stakeholder, and multi-criteria environments. *International Journal of Production Economics, 166*, 226–237. https://doi.org/10.1016/j.ijpe.2014.01.008

Silva, A., Seleme, R., Silva, W., Zattar, I., Nara, E., Canciglieri, O., … & Benitez, L. (2022). Evaluation and choice criteria of sustainable suppliers in the construction industry: a comparative study in Brazilian companies. *Sustainability, 14*(23), 15711. https://doi.org/10.3390/su142315711

Singh, A., Dwivedi, A., Agrawal, D., & Singh, D. (2023). Identifying issues in the adoption of AI practices in construction supply chains: towards managing sustainability. *Operations Management Research, 16*(4), 1667–1683. https://doi.org/10.1007/s12063-022-00344-x

Spreitzenbarth, J. (2021). AI methods in procurement. https://doi.org/10.6084/m9.figshare.16641049

Stratil, J., Baltussen, R., Scheel, I., Nacken, A., & Rehfuess, E. (2020). Development of the who-integrate evidence-to-decision framework: an overview of systematic reviews of decision criteria for health decision-making. Cost Effectiveness and Resource Allocation, 18(1). https://doi.org/10.1186/s12962-020-0203-6

Sundquist, V., Hulthén, K., & Gadde, L. (2018). From Project Partnering Towards Strategic Supplier Partnering. Engineering Construction & Architectural Management, 25(3), 358-373. https://doi.org/10.1108/ecam-08-2016-0177

Tai, P., Chiu, M., & Wei, C. (2023). Developing a green supplier risk assessment system applying natural language processing and life cycle assessment: an empirical study. https://doi.org/10.3233/atde230632

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: challenges and a path forward. *California Management Review, 61*(4), 15–42. https://doi.org/10.1177/0008125619867910

Tatini, P. (2025). Transforming Sourcing and Supply Chain Management: The Evolution of AI Agents in Modern Procurement. International Journal of Scientific Research in Computer Science Engineering and Information Technology, 11(1), 1219–1226. https://doi.org/10.32628/cseit251112131

Tornatzky, L. G., Fleischer, M., & Chakrabarti, A. K. (1990). The processes of technological innovation. New York, NY. The Free Press

Uddin, J., Elliott, G., & Parvin, S. (2021). Impact of company and country antecedents on b2b buyer perceived supplier performance. Journal of Business and Industrial Marketing, 37(9), 1835-1851. https://doi.org/10.1108/jbim-04-2021-0217

Upadhyaya, J. K., Biswas, N., & Tam, E. K. L. (2021). Using qualitative indicators in infrastructure assessment using the Functionality–Resiliency–Sustainability Framework. Frontiers in Sustainable Cities, 3. https://doi.org/10.3389/frsc.2021.746537

Wang, H., Zhang, F., & Mu, C. (2025). One for All: A General Framework of LLMs-based Multi-Criteria Decision Making on Human Expert Level. ArXiv.org. https://arxiv.org/abs/2502.15778

Wilson, G. (2024). The role of machine learning in predictive analytics for supply chain management. https://doi.org/10.20944/preprints202408.0343.v1

Wilson, J. H., & Keating, B. (2009). *Business forecasting with forecastX* (6th ed.). McGraw-Hill/Irwin.

Yin, S., Li, B., Dong, H., & Xing, Z. (2017). A new dynamic multi-criteria decision-making approach for green supplier selection in construction projects under time sequence. *Mathematical Problems in Engineering, 2017*(1). https://doi.org/10.1155/2017/7954784

Appendix A: Questionnaire

Subcontractor Evaluation Interview Questionnaire

Introduction

Thank you for participating in this interview. The purpose is to understand Hochtief's post-project subcontractor evaluation process to improve it based on your insights.

Instructions

- Confidentiality: All responses are confidential and used solely for research.
- Duration: Approximately 60 minutes.
- Format: Answer questions to the best of your ability and feel free to elaborate.]

Respondent Information

- Name:
- Position in the company:
- Date:
- Contact details:

Section 1: Roles, Experiences, and Responsibilities

General Role and Experience

1. Briefly describe your role in the company.
2. What are your key responsibilities within your department or team?
3. How long have you been working in this role?

Subcontractor Evaluation Process

1. How does subcontractor evaluation fit into your overall responsibilities?
2. How long have you been involved in subcontractor evaluations at Hochtief?
3. What specific tasks do you perform during subcontractor evaluations?

Section 2: Current Evaluation Process

1. Please describe the current subcontractor evaluation process:
    - Key steps involved after a project is completed.
2. Who oversees the evaluation process?
    - Which departments participate, and what are their contributions?
    - Are subcontractors involved in receiving feedback? If so, how?
3. What criteria are currently used to evaluate subcontractors?
    - How were they determined?
    - Are subcontractors aware of the criteria in advance?
4. Can the result of the evaluation of the subcontractors lead to elimination from consideration for future projects?
5. How is the evaluation documented and utilized?
    - Is there a standardized report format for recording evaluations?
    - Are results shared across projects to guide future subcontractor selection?

Section 3: Challenges in Evaluation

1. What are the biggest challenges in the current subcontractor evaluation process?
    - Are there recurring issues that impact objective assessments?
    - Do external factors complicate evaluations?
2. How do these challenges affect project outcomes?
    - Can you describe a situation where a practical subcontractor evaluation positively impacted project results?
    - Have there been instances where poor evaluations led to issues in subsequent projects? Please elaborate.
3. Have any solutions been implemented to address these challenges?
    - If so, what actions have been taken, and how effective were they?
    - If no actions have been taken, what would you suggest?

Section 4: Ranking and Weighing Evaluation Criteria

Instructions

- Rank the following criteria from 1 (most important) to N (least important), where N is the number of relevant criteria.
- Assign percentage weight based on importance. The total must equal 100%.
- If a criterion does not apply to your role, mark it with a dash (-).

Criterion	Rank (1-N)	Weight (%)
Quality of Workmanship		
Timeliness and Schedule Adherence		
Cost Management		
Safety and Compliance		
Communication and Collaboration		
Reliability and Repeat Engagement		
Technical Ability and Qualifications		
Financial Stability		
Sustainability Practices		
Innovation and Problem-Solving		

Example: If Quality of Workmanship is the most critical, rank it 1 and assign a weight based on importance.

Section 5: AI Integration in Subcontractor Evaluation

AI in Procurement

1. Have you ever used AI in subcontractor evaluations or procurement?
    - [ ] Yes
    - [ ] No
2. If yes, what AI tools have you used, and how did they improve the process?
3. If no, how comfortable are you with AI integration? (Scale: 1 – Not Comfortable, 5 – Very Comfortable)

AI Benefits and Challenges

1. What aspects of subcontractor evaluation do you think AI could improve the most?
2. What concerns do you have regarding AI-based evaluations?
3. If you have used AI before, what challenges did you encounter, and how were they managed?

Closing Statement

Thank You for Your Participation!

I sincerely appreciate the time and effort you have taken to share your insights. Your responses are very important for this research.

If you have any additional comments or thoughts after this interview, please feel free to reach out

Thank you again for your valuable contribution!

Appendix B: Evaluation Documents

Questionnaire for Internal Team

Subcontractor Evaluation		
Thank you for participating. All responses are confidential. Please answer the relevant sections according to your role. Proceed to each section of the questionnaire, responding to the questions based on your experience with this subcontractor. Use the rating scale where provided (1 for Very Poor to 5 for Very Good). Make sure to provide explanations, comments, and detailed observations as requested in each section to ensure a comprehensive evaluation. Estimated time for the entire document: 45 minutes.		
Respondent Information		
Name(s)	(Answer)	
Company	(Answer)	
Position(s) in the company	(Answer)	
Role(s) in the project	(Answer)	
Date	(Answer)	
Contact details (email)	(Answer)	
Quality of Workmanship		
Questions	**Responses**	**Comments**
Did the subcontractor meet the agreed-upon quality standards and specifications?	(Yes or No)	(If no, please provide details)
What percentage of work completed required rework or corrections?	(Answer)	(Provide any additional comments)
How many snags have been raised against the subcontractor?	(Answer)	(Provide any additional comments)
How many handover dates were delayed due to issues with workmanship?	(Answer)	(Provide any additional comments)
How would you rate the overall quality of the subcontractor’s workmanship?	(Rate 1 through 5)	(Provide an explanation)
Has the subcontractor made necessary improvements (if/where needed) during the project to meet the required quality standards?"	(Yes / No / Not Needed)	(Please provide details)

Timeliness and Schedule		
Questions	**Responses**	**Comments**
What was the initial schedule agreed in the contract? (Contract Duration)	(Answer)	(Provide any additional comments)
What was the final duration from project start to completion ? (Contract Duration plus Delay)	(Answer)	(Provide any additional comments)
How many days behind (or ahead of) schedule was the subcontractor's portion of the job?	(Answer)	(Provide any additional comments)
What portion of the delay can be attributed specifically to the subcontractor?	(Yes or No)	(If yes, please provide details)
Which were the reasons for the changes in the schedule?	(Answer)	(Provide any additional comments)
Did the subcontractor provide timely updates on their schedule and progress?	(Yes or No)	(Provide any additional comments)
How often (number of times) did the subcontractor request schedule adjustments from their side?	(Answer)	(Provide any additional comments)
How many formal revisions to the project schedule were required because of them?	(Answer)	(Provide any additional comments)
How effective was the subcontractor in adhering to the project timeline?	(Rate 1 through 5)	(Provide an explanation)
Cost Management		
Questions	**Responses**	**Comments**
What type of contract was signed with the subcontractor?	(Answer)	(Provide any additional comments)
What was the agreed-upon contract price?	(Answer)	(Provide any additional comments)
Which were the actual expenses?	(Answer)	(Provide any additional comments)
What was the overall percentage increase or decrease from the initial estimate which was agreed as a budget for this package?	(Answer)	(Provide any additional comments)

What percent of the change was covered by the client ?	(Answer)	(Provide any additional comments)
Were there additions to the scope? If yes, then why	(Answer)	(Provide any additional comments)
Was any part of the initial scope taken out? If yes, then why	(Answer)	(Provide any additional comments)
How was the quality of the substantiation of charges?	(Rate 1 through 5)	(Provide an explanation)
How effective was the subcontractor in reporting on project budgets?	(Rate 1 through 5)	(Provide an explanation)
In case of re-measurable contract, how effective was the subcontractor's forecast?	(Rate 1 through 5)	(Provide an explanation)
When applying for payments did they follow the agreed procedure and did they request the correct amount?	(Answer)	(Provide any additional comments)
How did the subcontractor cooperate in agreeing for an acceptable solution due to the cost overruns caused by them?	(Answer)	(Provide any additional comments)
How would you rate the subcontractor’s cost management?	(Rate 1 through 5)	(Provide an explanation)
Safety and Compliance		
Questions	**Responses**	**Comments**
Did the subcontractor possess the VCA certification?	(Yes or No)	(Please provide details)
Was the VCA certification valid during the entire project duration?	(Yes or No)	(Please provide details)
Did the subcontractor have certification under the Safety Culture Ladder?	(Yes or No)	(Please provide details)
Was the Safety Culture Ladder certification valid throughout the project timeline?	(Yes or No)	(Please provide details)
Did the subcontractor have the ISO 45001 certification?	(Yes or No)	(Please provide details)

Was the ISO 45001 certification valid for the whole duration of the project?	(Yes or No)	(Please provide details)
Were there any safety incidents or accidents involving the subcontractor's team?	(Yes or No)	(If yes, please provide details)
How many safety violations or accidents occurred on-site?	(Answer)	(Provide any additional comments)
Did the subcontractor pass all applied audits?	(Answer)	(Provide any additional comments)
How would you rate the subcontractor's safety awareness?	(Rate 1 through 5)	(Provide an explanation)
To what extent does the subcontractor prioritize safety and participation in safety meetings?	(Rate 1 through 5)	(Provide an explanation)
Did the subcontractor adhere to the required Personal Protection Equipment (PPE)?	(Yes or No)	(If yes, please provide details)
Did the subcontractor comply with all relevant health and safety instructions? (For example behaviour and work methods)	(Yes or No)	(Please provide details)
How did the subcontractor follow up on safety incidents or violation remarks?	(Answer)	(Provide any additional comments)
How would you rate the subcontractor's adherence to safety protocols and regulations?	(Rate 1 through 5)	(Provide an explanation)
Communication and Collaboration		
Questions	**Responses**	**Comments**
How would you rate the communication between the subcontractor and project stakeholders?	(Rate 1 through 5)	(Provide an explanation)
How responsive was the subcontractor to inquiries and feedback?	(Rate 1 through 5)	(Provide an explanation)
How effective and proactive was the subcontractor in providing regular progress updates?	(Rate 1 through 5)	(Provide an explanation)
Did the subcontractor proactively initiate any communication or coordination meetings?	(Yes or No)	(Please provide details)
Was the frequency and the quality of the reporting by the subcontractor up to the expectations?	(Yes or No)	(Please provide details)

Did you find that escalation meetings were necessary to ensure the work was completed on time, with good quality, and safely?	(Yes or No)	(Please provide details)
How would you rate the subcontractor's communication and collaboration?	(Rate 1 through 5)	(Provide any additional comments)
Reliability and Repeat Engagement		
Questions	**Responses**	**Comments**
How consistently did the subcontractor meet deadlines and deliver on commitments?	(Answer)	(Provide any additional comments)
Did the subcontractor have any unexpected challenges?	(Answer)	(Provide any additional comments)
How did the subcontractor handle unexpected challenges or changes in the project scope?	(Rate 1 through 5)	(Provide any additional comments)
How often did the subcontractor require follow-ups or reminders to fulfil their responsibilities?	(Answer)	(Provide any additional comments)
How would you rate the subcontractor's reliability and repeat engagement?	(Rate 1 through 5)	(Provide any additional comments)
Technical Ability and Qualifications		
Questions	**Responses**	**Comments**
Does the subcontractor qualify in the necessary technical skills and expertise for the project?	(Yes or No)	(Please provide details)
Was the subcontractor prequalified with the necessary certifications?	(Answer)	(Provide any additional comments)
Did the subcontractor complete most of his scope of works internally, or were they subcontracted to third parties?	(Answer)	(Please provide details including percentage)
How relevant and up-to-date are the subcontractor’s certifications and qualifications?	(Answer)	(Provide any additional comments)
How would you rate the subcontractor’s technical abilities and qualifications?	(Rate 1 through 5)	(Provide an explanation)
Financial Stability		

Questions	Responses	Comments
How promptly does the subcontractor deliver their financial reports and documentation (if applicable)?	(Answer)	(Provide any additional comments)
How did the finance department of the subcontractor interact on requests for payment criteria?	(Answer)	(Provide any additional comments)
Were there any financial issues that affected the subcontractor's performance during the project? If yes, how were they handled?	(Yes or No)	(If yes, please provide details)
Was the subcontractor's average credit rating or financial score below average?	(Answer)	(Provide any additional comments)
Were there any bank guarantees or any other securities issued by the subcontractor?	(Answer)	(Provide any additional comments)
How would you rate the subcontractor's financial stability?	(Answer)	(Provide any additional comments)
Sustainability Practices		
Questions	**Responses**	**Comments**
To what extent did the subcontractor fulfil its contractual obligation regarding sustainable practices?	(Rate 1 through 5)	(Provide an explanation)
What percentage of materials used in the project were certified as eco-friendly or sustainable (e.g., LEED-approved)?	(Answer)	(Provide any additional comments)
Did the subcontractor implement any energy- or water-saving practices on-site?	(Answer)	(Provide any additional comments)
Does the subcontractor have any sustainability-related certifications (such as ISO 14001, CO2 performance letter)?	(Answer)	(Provide any additional comments)
How would you rate the subcontractor's sustainable practices?	(Answer)	(Provide any additional comments)
Innovation and Problem-Solving		
Questions	**Responses**	**Comments**
How proactive was the subcontractor in introducing innovative solutions to project challenges?	(Rate 1 through 5)	(Provide an explanation)
How adaptable was the subcontractor to changes in project scope or requirements?	(Rate 1 through 5)	(Provide an explanation)

To what extent did the subcontractor contribute to improving project processes or outcomes?	(Answer)	(Provide any additional comments)
How would you rate the subcontractor's overall problem-solving capabilities?	(Rate 1 through 5)	(Provide an explanation)
Overall Evaluation		
Questions	**Responses**	**Comments**
What were the key strengths of this subcontractor during our collaboration?	(Answer)	(Provide any additional comments)
Did working with this subcontractor present any new opportunities that differentiate them from other subcontractors?	(Answer)	(Provide any additional comments)
What were the key weaknesses of this subcontractor during our collaboration?	(Answer)	(Provide any additional comments)
Did this subcontractor introduce any risks to the project that we should consider for future projects?	(Answer)	(Provide any additional comments)
Would you want to work with this subcontractor again?	(Yes / No / Maybe)	(Provide any additional comments)
Overall Comments		
(Response)		

Questionnaire for Subcontractor

Evaluation of HOCHTIEF		
Thank you for participating. All responses are confidential. For ranked questions, use a scale from 1 (Very Poor) to 5 (Very Good). Proceed to each section of the questionnaire, responding to the questions based on your experience with HOCHTIEF. Use the rating scale where provided (1 for Very Poor to 5 for Very Good). Make sure to provide explanations, comments, and detailed observations as requested in each section to ensure a comprehensive evaluation. Estimated duration: 45 minutes		
Respondent Information		
Name	(Answer)	
Company	(Answer)	
Position in the company	(Answer)	
Role in the project	(Answer)	

Date	(Answer)	
Contact details (email)	(Answer)	
Quality of workmanship		
Questions	**Responses**	**Comments**
How would you rate the quality of the work carried out by your company?	(Rate 1 through 5)	(Please provide an explanation)
How would you rate HOCHTIEF's standards and support in achieving high-quality workmanship?	(Rate 1 through 5)	(Please provide an explanation)
How did HOCHTIEF support you in maintaining quality standards?	(Answer)	(Provide any additional comments)
Timeliness and Schedule		
Questions	**Responses**	**Comments**
How effective was your company in completing all the agreed tasks according to the schedule?	(Rate 1 through 5)	(Provide any additional comments)
How did HOCHTIEF manage and communicate project timelines?	(Answer)	(Provide any additional comments)
How effective was HOCHTIEF in adhering to project schedules and providing everything on time?	(Rate 1 through 5)	(Please provide an explanation)
What challenges have you encountered regarding timeliness, and how were they addressed?	(Answer)	(Provide any additional comments)
Cost Management		
Questions	**Responses**	**Comments**
How effective was your company in adhering to the initial contract price?	(Rate 1 through 5)	(Please provide an explanation)
How satisfactory is HOCHTIEF's approach to managing project costs and budget allocations?	(Rate 1 through 5)	(Please provide an explanation)
How effectively did HOCHTIEF manage changes or variations to the contract?	(Rate 1 through 5)	(Please provide an explanation)
How good was HOCHTIEF at meeting the agreed payment term?	(Rate 1 through 5)	(Please provide an explanation)
How clear were the financial expectations and support provided by HOCHTIEF?	(Answer)	(Provide any additional comments)

How did HOCHTIEF handle financial disputes or unexpected costs?	(Answer)	(Provide any additional comments)
Safety and Compliance		
Questions	**Responses**	**Comments**
How effective is HOCHTIEF in maintaining safety standards and ensuring regulatory compliance?	(Rate 1 through 5)	(Please provide an explanation)
How would you rate the safety of the construction site?	(Rate 1 through 5)	(Please provide an explanation)
What safety protocols and training did HOCHTIEF provide?	(Answer)	(Provide any additional comments)
How did HOCHTIEF respond to safety incidents or compliance issues?	(Answer)	(Provide any additional comments)
Communication and Collaboration		
Questions	**Responses**	**Comments**
How effective was the communication and collaboration with HOCHTIEF throughout the project?	(Rate 1 through 5)	(Please provide an explanation)
What measures were taken to ensure effective communication between your company and HOCHTIEF?	(Answer)	(Provide any additional comments)
How receptive was the HOCHTIEF to subcontractor feedback and suggestions?	(Answer)	(Provide any additional comments)
Reliability and Repeat Engagement		
Questions	**Responses**	**Comments**
How reliable was HOCHTIEF in meeting its commitments and supporting your company?	(Rate 1 through 5)	(Please provide an explanation)
What factors contribute to HOCHTIEF's reliability (or lack of) from your perspective?	(Answer)	(Provide any additional comments)
Technical Ability and Qualifications		
Questions	**Responses**	**Comments**
How would you assess HOCHTIEF's technical support and qualifications relevant to your work?	(Rate 1 through 5)	(Please provide an explanation)
What technical support or training did HOCHTIEF provide to your company?	(Answer)	(Provide any additional comments)

How does HOCHTIEF facilitate the resolution of technical challenges?	(Answer)	(Provide any additional comments)
Financial Stability		
Questions	**Responses**	**Comments**
How stable and dependable was HOCHTIEF from a financial perspective in supporting projects?	(Rate 1 through 5)	(Provide any additional comments)
Did HOCHTIEF ensure timely payments and financial transparency?	(Yes or No)	(Provide any additional comments)
How would you rate the invoicing procedure?	(Rate 1 through 5)	(Provide any additional comments)
What financial support did HOCHTIEF offer you during challenging project phases?	(Answer)	(Provide any additional comments)
Sustainability Practices		
Questions	**Responses**	**Comments**
How would you rate HOCHTIEF's efforts to reduce environmental impact on site (e.g., waste management, energy use, noise and dust control)?	(Rate 1 through 5)	(Please provide an explanation)
Did HOCHTIEF clearly communicate their sustainability goals and expectations at the beginning of the project?	(Yes or No)	(Provide any additional comments)
Did HOCHTIEF provide clear guidance on sustainable construction practices (e.g., material reuse, recycling, energy efficiency)?	(Yes or No)	(Provide any additional comments)
Innovation and Problem-Solving		
Questions	**Responses**	**Comments**
How supportive was HOCHTIEF in fostering innovation and assisting with problem-solving?	(Rate 1 through 5)	(Please provide an explanation)
How effectively did HOCHTIEF leverage technology or external expertise to overcome technical or process-related obstacles?	(Rate 1 through 5)	(Please provide an explanation)
How did HOCHTIEF encourage and implement innovative solutions?	(Answer)	(Provide any additional comments)
What resources or support did HOCHTIEF provide to address unexpected challenges or technical issues?	(Rate 1 through 5)	(Provide any additional comments)
Overall Evaluation		

Questions	Responses	Comments
Please share your overall impression of working with HOCHTIEF based on your experience throughout the project.	(Answer)	(Provide any additional comments)
What were the key strengths of HOCHTIEF during your collaboration?	(Answer)	(Provide any additional comments)
What areas do you believe HOCHTIEF could improve upon?	(Answer)	(Provide any additional comments)
How would you rate your overall satisfaction with HOCHTIEF as a main contractor?	(Rate 1 through 5)	(Please provide an explanation)
Would you be willing to work with HOCHTIEF on future projects?	(Yes / No / Maybe)	(Provide any additional comments)
Overall Comments		
Do you have any additional comments, suggestions, or feedback regarding your experience? (Fill in this box)		

Appendix C: Personality Boxes of Virtual Assistant

Personality Box Version 1

You are a subcontractor evaluation assistant. Your task is to review two PDF evaluation reports — one from an internal team, so employees of Hochtief and one from the subcontractor — and provide a structured performance assessment report also in pdf format. Keep the format of the pdf simple. Your process is transparent, data-driven, and relies on structured question-and-answer analysis, not just sentiment or keyword scanning. You work in a professional, logical, and unbiased manner.

Follow these steps:

Step 1: Read the Reports

First of all, welcome the user and provide a small summary of the process that will be followed, as described below

Afterwards, you will be given two PDF files:

- One is the internal team's evaluation of the subcontractor.
- The other is the subcontractor's self-evaluation.

Your job is to read and extract the content of both documents. Focus on:

- Structured question–answer pairs (e.g., "Were there any delays? Yes – significant delays…")
- Numerical values (e.g., number of schedule changes, days delayed, rating scores)
- Yes/No answers
- Open-ended justifications or comments

You will combine both reports into a single knowledge base to evaluate from.

Step 2: Identify Criteria-Based Inputs

There are 10 evaluation criteria:

- Safety and Compliance
- Quality of Workmanship
- Timeliness and Schedule Adherence
- Technical Ability & Qualifications
- Cost Management
- Reliability and Repeat Engagement
- Financial Stability
- Communication and Collaboration
- Innovation and Problem-Solving
- Sustainability Practices

For each criterion:

- Locate all questions and answers related to it across both documents.
- Use the full context of responses, including numbers, explanations, and yes/no inputs.
- Combine the structured feedback to form a complete picture of the subcontractor's performance on that topic.
- Do not rely only on sentiment analysis. Instead, assess performance based on the facts and data in the answers.

Step 3: Detect Gaps and Inconsistencies

Before scoring:

- Check for discrepancies between internal and subcontractor responses (e.g., contradictory claims, mismatched ratings, different delay durations).
- Check for missing answers or vague, low-information explanations.
- Cross reference the answers of the internal team and the subcontractor to identify differences in similar questions and criteria and then ask for explanations to get both perspectives.

If any of the above are found and before moving on to step 4, ask user if they would like you to generate follow-up questions for clarification.

If you are asked to generate follow up questions:

- Make sure that the answers to the question do not already exist in the evaluation documents.
- Keep in mind that these questions will be asked to the subcontractor and the employees during face-to-face meetings, so they need to be accurate, based on their responses and based on the comparison between their responses and the ones of the other side.
- Make a distinction between the questions intended for the subcontractor and the ones intended for Hochtief's employees.

Examples:

"You mentioned delays — how many days behind schedule was the subcontractor?"
"The internal team rated technical ability as 2/5, but no justification was provided. Please elaborate."
Await or incorporate clarification answers before proceeding to scoring.

Step 4: Weight Selection

Ask the user whether to use:

Equal weighting: 10% for each criterion

OR

Predefined industry weighting:

- Safety and Compliance – 31.23%
- Quality of Workmanship – 14.64%
- Timeliness and Schedule Adherence – 11.34%
- Technical Ability & Qualifications – 8.90%
- Cost Management – 7.32%
- Reliability and Repeat Engagement – 6.92%
- Financial Stability – 6.07%
- Communication and Collaboration – 5.09%
- Innovation and Problem-Solving – 4.45%
- Sustainability Practices – 4.03%

OR

Role-Weighted Influence Model:

- Safety and Compliance – 30.12%
- Quality of Workmanship – 13.69%
- Timeliness and Schedule Adherence – 12.76%
- Technical Ability and Qualifications – 9.18%
- Communication and Collaboration – 7.31%
- Cost Management - 7.04%

- Reliability and Repeat Engagement - 6.22%
- Financial Stability – 4.99%
- Innovation and Problem-Solving - 4.55%
- Sustainability Practices - 4.03%

Use the chosen weight values in the final rating calculation. The evaluation criteria are the same for each option

Step 5: Score Each Criterion

For each of the 10 criteria:

- Analyse the related Q&A content holistically:
- Consider numerical inputs, free-form justifications, yes/no responses and comments left for every answer
- If provided, use the answers to the follow-up questions for further clarification and information
- Include performance ratings if present but assign your score based on your own reasoning without being biased (e.g., "Adherence to timeline: 2").

Use intelligent reasoning to assign a score from 1.0 to 5.0, where:

- 5 = Excellent
- 4 = Good
- 3 = Acceptable
- 2 = Poor
- 1 = Unacceptable

Generate a clear explanation for each score, citing data from the answers. The explanation of each rating should be based on facts and data from the evaluation document and, if available, the answers to the follow-up questions.

Example: “Subcontractor was 60 days late, made 3 formal schedule revisions, and failed to communicate consistently — resulting in poor adherence to project timelines (score: 2.0).”

Step 6: Calculate Final Rating

Use the Weighted Sum Method to calculate the final rating
Multiply each criterion score by its assigned weight.
Sum the weighted scores.

Step 7: Generate the Evaluation Summary

Generate a pdf report, which will include all the details of the subcontractor and will be structured as follows:

Subcontractor Name: [insert name]

Final Rating: [1.0 to 5.0]
Scores Breakdown:
[Criterion]: [Score] – [Explanation]
Summary:

An extensive summary of the overall performance, describing how the collaboration between Hochtief and the subcontractor went by explaining in detail both perspectives and by providing and objective assessment of the situation.

Lessons Learned:

Provide a list of the takeaways from the engagement while using specific examples based on the evaluation documents and, if available, the follow up answers.

Recommendations:

Actionable advice or next steps for the evaluator
Provide insights of how Hochtief can improve in general
Provide advice of how to improve collaboration with the specific subcontractor in case of future collaboration

Personality Box Version 2

STRICT DATA GROUNDING POLICY:

- You must only generate outputs—findings, follow-up questions, summaries, and recommendations—that are strictly based on explicit, factual content (ratings, comments, numbers, answers, narrative etc.) present in the submitted evaluation documents or direct user follow-up answers.
- Never invent, interpolate, generalize, extrapolate, fill narrative gaps, or use hypothetical/template/example scenarios unless they are direct quotations or paraphrases from the supplied material.
- If a field, justification, or explanation is blank or absent, state: “No explanation/data provided in the submitted reports.”
- This override policy applies globally and at every step in the process, and supersedes any learned/default/template behaviour.

SUBCONTRACTOR EVALUATION ASSISTANT STRICT DATA/OUTPUT POLICY

- Never output templates, outlines, process flows, or section headers without immediately populating each with actual, explicitly extracted data, ratings, or comments from the submitted evaluation documents. If information is missing, write only: “No explanation/data provided in the submitted reports.”
- Do not include any placeholder, generic, or summary fields such as [Details], [Summary], or [Section]. Output in every section must always be specific and factual. If data does not exist, state: “No explanation/data provided in the submitted reports.”
- Always present actual extracted answers, ratings, and comments for each criterion from each report in the order they appear, before any interpretation, scoring, or summary. Do not provide process explanations, outlines, or summaries unless specifically requested by the user.
- If any section lacks explicit content to extract, immediately notify the user and do not proceed with outlines or process explanations. Write only: “No explanation/data provided in the submitted reports.”
- All explanations, summaries, and follow-up questions must be directly based on explicit, factual evidence from the uploaded files or direct user follow-up. No extrapolation, generalization, or assumptions are permitted.
- Do not show hidden structure, planned headers, or “what will be done” statements. Only generate headings and sections when actual, report-based content is available to fill them.
- Only present follow-up questions or clarifications when a specific fact, rating, comment, or omission is present or missing in the provided data. Do not invent scenarios or questions.
- The user must always see and confirm the concrete extracted data for each criterion before any scoring or summary is generated.
- All output must be sanitized to ASCII-only. If any data cannot be shown in ASCII without loss of meaning, halt output and notify the user.
- Under no circumstances invent, interpolate, generalize, fill narrative gaps, or use hypothetical or example content—no matter how typical it seems—unless directly present in the supplied documents or written user follow-up.

🧠 Personality: Subcontractor Evaluation Assistant

You are a subcontractor evaluation assistant. Your task is to review two PDF evaluation reports—one from an internal team (employees of Hochtief) and one from the subcontractor—and provide a structured performance assessment report (also in PDF format). Your processing is transparent, data-driven, and relies on structured question-and-answer analysis—not sentiment or keyword scanning. You always work in a professional, logical, precise, and unbiased manner.

You must only use facts, numerical ratings, explanations, and comments that are explicitly present in the provided evaluation documents or supplied via direct user follow-up answers. Do not invent, generalize, interpolate, extrapolate, or assume justifications or statements, even for the sake of a complete narrative, to "fill gaps," or to represent 'typical' reporting practices. You must NOT use examples, hypothetical statements, or generalizations, even if they appear typical or logical, unless these are explicitly present in the submitted documents or user's written follow-up. Do NOT use "for example," "such as," "generally," or similar language unless directly quoted from the source. If any field, justification, or explanation is blank or unsupported, write: "No explanation/data provided in the submitted reports." If data or explanation is absent or blank, state: "No explanation/data provided in the submitted reports." This rule applies globally throughout the entire evaluation and reporting process.

IMPORTANT ABOUT EXAMPLES:

- The "Examples" shown below serve strictly to illustrate the structure, tone, and detail level you should use for follow-up questions and clarifications, NOT as templates to be used or as triggers for questions when no such issue/discrepancy/ambiguity exists in the supplied documents.
- You may only generate a follow-up question if a directly supporting field, phrase, rating, ambiguity, contradiction, or omission is present in the documents or user clarifications.
- NEVER inject a hypothetical or 'just for illustration' question into gap analysis if it is not solidly present in the working documents.

Examples (for style/tone/structure ONLY—not to be used as templates):
"You mentioned 3 schedule revisions, which were initiated by Hochtief and which by the subcontractor?"
"The internal team rated collaboration as 2/5 but gave no detail, please explain."
"The subcontractor was delayed by 2 months, but you still gave them a 4 – why is that?"
"You rated technical quality as 3/5, but provided no comments, could you explain the issues behind this rating?"
"You gave a 5 for scheduling, but the project was completed 6 weeks late, was this delay justified?"
"You rated your company's safety compliance as a 4 but our manager rated it as 2 and mentioned 24 reported safety incidents—how do you explain that?"
"You mentioned that 95% of the work required rework or corrections, but rated the quality of the subcontractors work as 5. Could you please explain?"
"The subcontractor claims all certifications were valid, but you flagged issues with documentation. What specifically was missing?"
Special Character and Encoding Control
Immediately after extracting any text from the source documents and before any internal processing, analysis, or storing to variables, all content must be sanitized for ASCII-only compliance. This includes all intermediate and temporary text, not just that intended for final output or export.
Before generating any PDF or exporting text for reports, always sanitize all output text to ensure it contains only printable ASCII characters.
Replace or remove the following characters:

- All Unicode currency signs (€, £, ¥, etc.) → replace with "EUR", "GBP", "YEN", "USD", or the correct ASCII label
- Em dash (—) and en dash (–) → replace with regular dash (-)
- Curly apostrophes/quotes (‘ ’ “ ”) → replace with normal straight apostrophe (') or quote (")
- Accented letters (é, ö, ñ, etc.) → replace with plain ASCII equivalents (e, o, n, etc.)

- Any other non-ASCII punctuation or special symbols → replace with nearest ASCII equivalent, or delete if not critical to meaning
- Convert or replace any special formatting, mathematical signs, or table characters that cannot be rendered in standard ASCII. Note any conversions to the user.
- For every sanitization event, keep a log that details each replaced or removed non-ASCII character, its replacement, and its original location or context if identifiable. Provide this log to the user on request, or whenever a replacement alters a key meaning or field.
- Never allow unsupported characters to be included in PDF or text exports—pre-validate and clean all content before export.
- Before any export (including PDF or plain text), perform a systematic scan of all content for any remaining non-ASCII or unprintable characters. If detected, halt the process and clearly report the issue, showing both the affected text and proposed correction.
- If you detect unsupported characters in provided source material, flag them, correct them, and note the correction in the conversation.
- If you encounter a character that cannot be sanitized or replaced without loss of essential technical or contractual meaning, halt the process immediately, notify the user, and await further instruction before proceeding.
- Regularly run sample text blocks with high-risk characters through the sanitization routine before processing actual evaluation files, to verify error-free export performance under all expected project conditions.

Appendix: Common Non-ASCII Characters and Replacements

Curly quotes/apostrophes ‘ ’ “ ” → ' or "
Em dash, en dash — – → -
Currency symbols: €, £, ¥ → EUR, GBP, YEN
Accented letters: é, ö, ñ → e, o, n
Mathematical/engineering: √, ±, ≤ → sqrt, +/-, <=
All others: Replace with nearest ASCII text, or flag for user

📥 **Step 1: Read the Reports**

First, welcome the user and provide a summary of the process.

You will then receive two PDF files:

- One from Hochtief's internal team
- One from the subcontractor’s self-evaluation

Extract and display, verbatim and criterion by criterion, the structured answers, ratings, and comments from BOTH evaluation reports (Hochtief & subcontractor).
Ensure this is purely what exists in the documents so no summaries, no invented data, no interpretation, just the actual provided content, clearly labelled by source.

Read and extract:

- Structured question–answer pairs
- Numerical data (e.g., delay days, schedule revisions)
- Yes/No answers
- Explanatory comments
- Combine both into a single knowledge base.

🔎 **Step 2: Identify Criteria-Based Inputs**

Evaluate these 10 criteria:

- Safety and Compliance
- Quality of Workmanship
- Timeliness and Schedule Adherence
- Technical Ability & Qualifications
- Cost Management
- Reliability and Repeat Engagement
- Financial Stability
- Communication and Collaboration
- Innovation and Problem-Solving
- Sustainability Practices

For each:

- Collect all relevant Q&As from both reports
- Use full context (numbers, text, explanations)
- Don't rely on sentiment alone, instead analyse facts

❗ **Step 3: Detect Gaps and Inconsistencies**

Before scoring:

- Immediately upon receiving both evaluation documents, systematically compare all responses, ratings, yes/no answers, and comments for every criterion and relevant field.
- For each criterion, cross-check the internal team's and the subcontractor's answers for differences in data, ratings, narrative explanations, and completeness.
- Find discrepancies or contradictions
- Detect missing, vague, or incomplete answers
- Compare both parties' responses to the same topics

Ask the user:

"Would you like me to generate follow-up questions based on the detected gaps and inconsistencies?"

If yes:

- Generate specific, contextual follow-up questions based on identified gaps and differences between the responses of Hochtief's employees and subcontractors
- Only generate follow-up or clarification questions when a specific fact, rating, answer, or statement is present or clearly missing in the provided documents.
- If data or explanations are missing from a clearly expected field (e.g., unanswered rating, blank comment, or unexplained score), generate a targeted question asking for the missing information, explicitly stating which field or criterion is incomplete.
- If a field is marked as "N/A" (not applicable), do not generate a follow-up question for that field, as no clarification is required.
- If information is ambiguous (e.g., marked "occasionally" or "sometimes" without specifics), generate a follow-up question requesting quantification or detail, using the exact wording from the document.
- Only reference facts, ratings, numerics, or comments that explicitly appear in the supplied documents, never infer, generalize, or assume data.
- Every question must cite the specific field, rating, or phrase so the respondent knows what clarification is needed.
- Do not skip or gloss over any differences; always err on the side of raising questions when data or explanation is incomplete, ambiguous, or not mutually supported.
- Before generating the questions double-check the documentation and confirm all follow-up questions are directly and solely based on explicit data or comments from the provided reports.

- Automatically present the side-by-side findings and follow-up questions to the user as soon as gap analysis is complete.

⚖️ **Step 4: Weight Selection**

Ask the user to select:

Equal weighting: 10% per criterion

OR

Equal Influence Model:

- Safety and Compliance - 31.23%
- Quality of Workmanship - 14.64%
- Timeliness and Schedule Adherence - 11.34%
- Technical Ability and Qualifications - 8.90%
- Communication and Collaboration - 7.32%
- Cost Management - 6.92%
- Reliability and Repeat Engagement - 6.07%
- Financial Stability - 5.09%
- Innovation and Problem-Solving - 4.45%
- Sustainability Practices - 4.03%

OR

Role-Weighted Influence Model:

- Safety and Compliance – 30.12%
- Quality of Workmanship – 13.69%
- Timeliness and Schedule Adherence – 12.76%
- Technical Ability and Qualifications – 9.18%
- Communication and Collaboration – 7.31%
- Cost Management - 7.04%
- Reliability and Repeat Engagement - 6.22%
- Financial Stability – 4.99%
- Innovation and Problem-Solving - 4.55%
- Sustainability Practices - 4.03%

Explain each option to the user:

- The Equal Weights model is the simplest and most neutral. It treats all evaluation criteria as equally important.
- The Equal Influence Model ensures that all of the HOCHTIEF employees who participated in the ranking of the criteria have the same level of input in the final weights, and thus in the final score.
- The Role-Weighted Influence Model adjusts the influence of each employee in the determination of the final weights based on their position in HOCHTIEF.

Use the selected weights for scoring.

📊 **Step 5: Score Each Criterion**

For each criterion:

- Analyse all related Q&As
- Include follow-up answers if available
- Consider ratings only if justified by evidence

- Assign a score from 1 to 5

Each score must include a factual explanation:

If underlying data or justification is missing, write: "No explanation/data provided in the submitted reports."

Step 6: Calculate Final Rating

Multiply each score you generated by its weight depending on the selected method
Add up the results to produce a final score
Present the final score with two decimal precision (e.g., 3.44 / 5)

Step 7: Generate the Evaluation Summary (PDF)

Generate a PDF report that evaluates the subcontractor's performance. Use Hochtief's responses only as an input source, not as a subject of evaluation.

PDF Report Formatting Guidelines

You must only use facts, numerical ratings, explanations, and comments that are explicitly present in the provided evaluation documents or supplied via direct user follow-up answers.
Do not invent, generalize, or assume justifications or statements, even for the sake of a complete narrative or "typical" reporting practices.
Every explanation in the summary, lessons learned, recommendations, and criterion justifications must be traceable to a specific, verifiable source in the supplied data.
If necessary, data or explanation is missing, leave the section blank or state: "No explanation/data provided in the submitted reports."
If requested, show extracted source snippets for every factual statement included in the report.
Always populate the Subcontractor Name field accurately. Use the name from the evaluation documents or ask the user if not found.
All text, tables, and sections must fit within the standard width of an A4 page when generating the PDF.
Adjust cell/multicell widths, font size, and line breaks to ensure NO text or section overflows or is cut off from the printable page margins.
Insert at least one blank line of vertical space (e.g., PDF's .ln()) before and after each section header (e.g., "Summary," "Lessons Learned," "Scores Breakdown," "Recommendations for Hochtief," etc.).
Insert a blank line before and after each criterion in all score breakdowns.
Ensure the document layout is clearly spaced and not condensed; avoid crammed or dense sections, and break up long paragraphs for readability.
If possible, keep major section headers on a new page (if space is tight), but ensure criteria breakdowns are always easy to read and visually separate each criterion.
Reject excessively long lines or long blocks of text—wrap all text to fit the PDF page cleanly.
Always extract, sanitize, and fully consider the responses from the "Overall Evaluation," "Overall Comments," or equivalent final reflection/conclusion sections in both the Hochtief internal and subcontractor self-evaluation reports.

Use these sections—regardless of their structure or narrative style—as essential sources when drafting:

The project summary (integrating key commentary, collaboration themes, and overview insights)
Lessons learned (pulling concrete problems, successes, or "what to change next time" items)
Both sets of recommendations (base actionable points and advice on facts, reflections, and issues highlighted in these responses)
Never omit these "Overall" answers, even if the content is unstructured prose or generalist. Always cite, paraphrase, or quote directly for maximum relevance and data fidelity.

Structure the report as follows:

Subcontractor Name: Get the name from the evaluation document
Final Rating: [e.g., 3.44 / 5]
Scores Breakdown:
[Criterion]: [Score] and then [Explanation citing subcontractor's performance]
…

Summary:

Provide a comprehensive, factual summary of the subcontractor's performance across all 10 criteria.

The summary must:

Reference specific events, patterns, and decisions
Include specific facts and data points
Use clear, direct language and cite multiple evaluation criteria (not general trends)
Include examples
Reflect both perspectives, but clearly describe how the subcontractor performed
Be approximately 250-300 words, not a brief paragraph

🧠 Lessons Learned

List project-specific lessons that arose directly from the subcontractor's performance. Use real facts and events from the documents or follow-up answers.

Each lesson should:

Explain what happened
Identify why it happened
Suggest what should be done differently next time
Aim for concise, high-impact lessons that are directly supported by the evaluation content.

✅ Recommendations

Split into two sections:

Recommendations for Hochtief:

Provide tailored and specific insights on how Hochtief can better manage or coordinate with the subcontractor in future projects.

Recommendations for the Subcontractor:

Provide tailored, specific suggestions for this subcontractor based on their behaviour in this project.

For both parties:

Use as reference points as much data as possible from the evaluation documents and, if available, from the answers to the follow-up questions to justify your recommendations and to make them more specific to the collaboration and to the project

Avoid generic phrasing like "improve communication", instead, say:

"Assign a full-time site coordinator to handle reporting and client check-ins twice weekly."
"Include a contractual clause requiring the subcontractor to submit updated schedules every two weeks with variance explanations for any deviation beyond 3 days."
"Request digital submission of all required safety certifications two weeks before site access is granted and schedule a pre-mobilization audit."
"Require that the first completed unit of work (e.g., the first installed ceiling section) undergo a detailed quality walkthrough with both parties to establish a shared standard for the rest of the project."

“Organize a biweekly review meeting with the subcontractor’s site manager and HOCHTIEF’s project engineer to track coordination, open issues, and any interpersonal frictions.”
“Use a shared cloud folder for all documentation with weekly automatic reminders sent to responsible contacts and escalation to HOCHTIEF’s project coordinator if submissions are overdue by more than 3 days.”
“Provide the subcontractor with a scope clarification workshop during the project onboarding phase, especially for contracts involving multiple interfaces or design changes.”

🧠 Behaviour Notes

Be specific, evidence-based, and impartial
Do not invent or assume data, ask for clarifications
Justify every score using clear references
Be consistent in format and tone throughout the report
Be objective in your evaluation
The final report should be based on facts and data provided to you. Include examples and be as accurate and specific as possible
Avoid producing generic responses and be precise and use data to back up the content you generate

Appendix D: Evaluation Documents from Case Study

Internal Teams' Evaluation of Subcontractor's Performance

Quality of Workmanship	**To be completed by:**	
Questions	**Responses**	**Comments**
Did the subcontractor meet the agreed-upon quality standards and specifications?	Yes	
What percentage of work completed required rework or corrections?	1%	Faced, Snags and Cannepy
How many snags have been raised against the subcontractor?	14	Only 2 reworks
How many handover dates were delayed due to issues with workmanship?	0	
How would you rate the overall quality of the subcontractor's workmanship?	5	good work
Has the subcontractor made necessary improvements (if / where needed) during the project to meet the required quality standards?"	No	
Timeliness and Schedule	**To be completed by:**	
Questions	**Responses**	**Comments**
What was the initial schedule agreed in the contract? (Contract Duration)	Our Schedule	Jan 2024 - July 2025
What was the final duration from project start to completion? (Contract Duration plus Delay)	See Schedule	18 months
How many days behind (or ahead of) schedule was the subcontractor's portion of the job?	non	
What portion of the delay can be attributed specifically to the subcontractor?	no	
Which were the reasons for the changes in the schedule?	Scope changes	
Did the subcontractor provide timely updates on their schedule and progress?	Yes	Weekly
How often (number of times) did the subcontractor request schedule adjustments from their side?	Occasionally	
How many formal revisions to the project schedule were required because of them?	Occasionally	
How effective was the subcontractor in adhering to the project timeline?	4	The try hard to do
Cost Management	**To be completed by:**	
Questions	**Responses**	**Comments**
What type of contract was signed with the subcontractor?	Lumpsum	n/a
What was the agreed-upon contract price?	€1.687.080,12	We commenced with the works prior to contract in place. A conditional award in the amount of €118,875 was granted to enable the sub-contractor to proceed. This did not affect the overall initial price.
Which were the actual expenses?	€1.618.972,89	This is the amount spent to date

What was the overall percentage increase or decrease from the initial estimate which was agreed as a budget for this package?	0%	Their contract price is their budget
What percent of the change was covered by the client ?	100%	Our price towards client is fixed
Were there additions to the scope? If yes, then why	Yes	Delays and disruption due to chroom 6, co-ordination, works that did not form part of base scope and additional requests from client
Was any part of the initial scope taken out? If yes, then why	Yes	Although it was a lumpsum contract there were certain scope allowed for that did not realise and or transferred to others. They are willing to credit these works
How was the quality of the substantiation of charges?	3	Their cost breakdown was detailed and good. On submission of the costs, it lacked proper substantiation. This we had to clarify and asked for information afterwards.
How effective was the subcontractor in reporting on project budgets?	4	Costs submitted timeously, some followed after works completed but majority prior to works commencing.
In case of re-measurable contract, how effective was the subcontractor's forecast?	n/a (lumpsum contract)	n/a
When applying for payments did they follow the agreed procedure and did they request the correct amount?	Yes	We verified their payment applications of which there were differences. These differences not to be considered incorrect.
How did the subcontractor cooperate in agreeing for an acceptable solution due to the cost overruns caused by them?	Very co-operative	We were more reliant on their expertise to come up with a solution
How would you rate the subcontractor's cost management?	4	n/a
Safety and Compliance	**To be completed by:**	
Questions	**Responses**	**Comments**
Did the subcontractor possess the VCA certification?	Yes	
Was the VCA certification valid during the entire project duration?	Yes	
Did the subcontractor have certification under the Safety Culture Ladder?	Yes	level 4
Was the Safety Culture Ladder certification valid throughout the project timeline?	Yes	
Did the subcontractor have the ISO 45001 certification?	Yes	
Was the ISO 45001 certification valid for the whole duration of the project?	Yes	
Were there any safety incidents or accidents involving the subcontractor's team?	No	

How many safety violations or accidents occurred on-site?	0	
Did the subcontractor pass all applied audits?	Yes	
How would you rate the subcontractor's safety awareness?	4	
To what extent does the subcontractor prioritize safety and participation in safety meetings?	4	
Did the subcontractor adhere to the required Personal Protection Equipment (PPE)?	Yes	
Did the subcontractor comply with all relevant health and safety instructions? (For example behaviour and work methods)	Yes	
How did the subcontractor follow up on safety incidents or violation remarks?	NA	
How would you rate the subcontractor's adherence to safety protocols and regulations?	4	
Communication and Collaboration	**To be completed by:**	
Questions	**Responses**	**Comments**
How would you rate the communication between the subcontractor and project stakeholders?	5	
How responsive was the subcontractor to inquiries and feedback?	4	
How effective and proactive was the subcontractor in providing regular progress updates?	5	
Did the subcontractor proactively initiate any communication or coordination meetings?	5	
Was the frequency and the quality of the reporting by the subcontractor up to the expectations?	4	
Did you find that escalation meetings were necessary to ensure the work was completed on time, with good quality, and safely?	No	
How would you rate the subcontractor's communication and collaboration?	5	
Reliability and Repeat Engagement	**To be completed by:**	
Questions	**Responses**	**Comments**
How consistently did the subcontractor meet deadlines and deliver on commitments?	good	
Did the subcontractor have any unexpected challenges?	Some	
How did the subcontractor handle unexpected challenges or changes in the project scope?	4	
How often did the subcontractor require follow-ups or reminders to fulfil their responsibilities?	Non	
How would you rate the subcontractor's reliability and repeat engagement?	5	
Technical Ability and Qualifications	**To be completed by:**	
Questions	**Responses**	**Comments**
Does the subcontractor qualify in the necessary technical skills and expertise for the project?	No	
Was the subcontractor prequalified with the necessary certifications?	Yes	

Did the subcontractor complete most of his scope of works internally, or were they subcontracted to third parties?	Most Internally	
How relevant and up-to-date are the subcontractor's certifications and qualifications?	Up to date	
How would you rate the subcontractor's technical abilities and qualifications?	5	
Financial Stability	**To be completed by:**	
Questions	**Responses**	**Comments**
How promptly does the subcontractor deliver their financial reports and documentation (if applicable)?	Promptly	Reports referring to their application for payment. They complied with our payment certification process.
How did the finance department of the subcontractor interact on requests for payment criteria?	Good	They invoiced according to our payment certificates
Were there any financial issues that affected the subcontractor's performance during the project? If yes, how were they handled?	No	n/a
Was the subcontractor's average credit rating or financial score below average?	Not sure	n/a
Were there any bank guarantees or any other securities issued by the subcontractor?	No	n/a
How would you rate the subcontractors financial stability?	4	n/a
Sustainability Practices	**To be completed by:**	
Questions	**Responses**	**Comments**
To what extent did the subcontractor fulfil its contractual obligation regarding sustainable practices?	all	
What percentage of materials used in the project were certified as eco-friendly or sustainable (e.g., LEED-approved)?	not applicable	
Did the subcontractor implement any energy- or water-saving practices on-site?	No	
Does the subcontractor have any sustainability-related certifications (such as ISO 14001, CO2 performance letter)?	1. CO2 Performance ladder level 5 2. ISO 14001	1. Subcontractor X has level 5 on the CO2 Performance ladder 2. Subcontractor X is ISO 14001 certified
How would you rate the subcontractor's sustainable practices?	4	Subcontractor X educates on all levels
Innovation and Problem-Solving	**To be completed by:**	
Questions	**Responses**	**Comments**
How proactive was the subcontractor in introducing innovative solutions to project challenges?	4	
How adaptable was the subcontractor to changes in project scope or requirements?	5	
To what extent did the subcontractor contribute to improving project processes or outcomes?	Helpful	
How would you rate the subcontractor's overall problem-solving capabilities?	5	
Overall Evaluation	**To be completed by:**	
Questions	**Responses**	**Comments**

What were the key strengths of this subcontractor during our collaboration?	Collaboration	Practical and open for ad hoc adjustments
Did working with this subcontractor present any new opportunities that differentiate them from other subcontractors?	Flexibility in time management	day by day operations open for quick response
What were the key weaknesses of this subcontractor during our collaboration?	No significant weaknesses	
Did this subcontractor introduce any risks to the project that we should consider for future projects?	Schedule	Priorities were not always clear
Would you want to work with this subcontractor again?	YES	This opinion is supported by Project management
Overall Comments		
Great company to work with		

Subcontractor's Evaluation for HOCHTIEF and for their own performance

Quality of workmanship		
Questions	**Responses**	**Comments**
How would you rate the quality of the work carried out by your company?	4	
How would you rate HOCHTIEF's standards and support in achieving high-quality workmanship?	4	
How did HOCHTIEF support you in maintaining quality standards?	Via Aconex	
Timeliness and Schedule		
Questions	**Responses**	**Comments**
How effective was your company in completing all the agreed tasks according to the schedule?	3	
How did HOCHTIEF manage and communicate project timelines?	In a too large group and unstructured	
How effective was HOCHTIEF in adhering to project schedules and providing everything on time?	3	
What challenges have you encountered regarding timeliness, and how were they addressed?	Our workers where on site and scheduled where needed, working by the principle first things first	
Cost Management		
Questions	**Responses**	**Comments**
How effective was your company in adhering to the initial contract price?	4	
How satisfactory is HOCHTIEF's approach to managing project costs and budget allocations?	3	
How effectively did HOCHTIEF manage changes or variations to the contract?	3	
How good was HOCHTIEF at meeting the agreed payment term?	3	
How clear were the financial expectations and support provided by HOCHTIEF?	Towards the end of the project things look to go	

	more difficult (everything took longer).	
How did HOCHTIEF handle financial disputes or unexpected costs?	It took too long, because of disagreement	
Safety and Compliance		
Questions	**Responses**	**Comments**
How effective is HOCHTIEF in maintaining safety standards and ensuring regulatory compliance?	4	
How would you rate the safety of the construction site?	4	
What safety protocols and training did HOCHTIEF provide?	Various meetings periodically and if required	
How did HOCHTIEF respond to safety incidents or compliance issues?	4	
Communication and Collaboration		
Questions	**Responses**	**Comments**
How effective was the communication and collaboration with HOCHTIEF throughout the project?	3	
What measures were taken to ensure effective communication between your company and HOCHTIEF?	Not only the use of Aconex but also regular e-mail	
How receptive was the HOCHTIEF to subcontractor feedback and suggestions?	Where open to changes and alternatives	
Reliability and Repeat Engagement		
Questions	**Responses**	**Comments**
How reliable was HOCHTIEF in meeting its commitments and supporting your company?	3	
What factors contribute to HOCHTIEF's reliability (or lack of) from your perspective?	A lot of changes and surprises	
Technical Ability and Qualifications		
Questions	**Responses**	**Comments**
How would you assess HOCHTIEF's technical support and qualifications relevant to your work?	4	
What technical support or training did HOCHTIEF provide to your company?	If asked there was support, no training	
How does HOCHTIEF facilitate the resolution of technical challenges?	By agreement, a solution is found for every challenge	
Financial Stability		
Questions	**Responses**	**Comments**
How stable and dependable was HOCHTIEF from a financial perspective in supporting projects?	4	
Did HOCHTIEF ensure timely payments and financial transparency?	Yes	
How would you rate the invoicing procedure?	3	
What financial support did HOCHTIEF offer you during challenging project phases?	We each had our own financial statements which caused disagreement	
Sustainability Practices		
Questions	**Responses**	**Comments**

How would you rate HOCHTIEF's efforts to reduce environmental impact on site (e.g., waste management, energy use, noise and dust control)?	4	
Did HOCHTIEF clearly communicate their sustainability goals and expectations at the beginning of the project?	No	
Did HOCHTIEF provide clear guidance on sustainable construction practices (e.g., material reuse, recycling, energy efficiency)?	Yes	
Innovation and Problem-Solving		
Questions	**Responses**	**Comments**
How supportive was HOCHTIEF in fostering innovation and assisting with problem-solving?	4	
How effectively did HOCHTIEF leverage technology or external expertise to overcome technical or process-related obstacles?	4	
How did HOCHTIEF encourage and implement innovative solutions?	Where necessary	
What resources or support did HOCHTIEF provide to address unexpected challenges or technical issues?	4	
Overall Evaluation		
Questions	**Responses**	**Comments**
Please share your overall impression of working with HOCHTIEF based on your experience throughout the project.	More work preparation and total overview	
What were the key strengths of HOCHTIEF during your collaboration?	Positivity and patience	
What areas do you believe HOCHTIEF could improve upon?	Work preparation and scheduling	
How would you rate your overall satisfaction with HOCHTIEF as a main contractor?	4	
Would you be willing to work with HOCHTIEF on future projects?	Yes	
Overall Comments		
For scheduling, meetings should first be conducted in smaller groups to finalize 70-80% of the plan before engaging the entire team. Large group meetings often consume two to three hours, which can be inefficient, especially if topics relevant to particular scopes are addressed late in the meeting. Despite disagreements throughout the project, there was a strong sense of mutual respect and honesty on site. Challenges were addressed collaboratively, with all parties willing to move past differences to continue effective collaboration. Also, there should be a model for the whole building from the beginning. All of the drawings of the subcontractors need to be in that model. Everybody needs to work on that model. If there is a new phase the engineers should start scanning the entire building quite early. The measurements need to be in the model early.		

Appendix E: AI-generated Follow-up Questions

Follow-Up Questions

1. Safety and Compliance

For the Hochtief internal team:

“You rated safety awareness, prioritization of safety meetings, and adherence to protocols all as 4/5 instead of 5/5, despite confirming no safety incidents or violations and full certification validity. Can you explain specifically what prevented a top score in these areas?”

There are different ways of thinking safe if you check the Safety Culture Ladder

For the subcontractor:

“You mentioned periodic meetings for safety protocols/training and rated site safety 4/5. Could you specify if there were any safety concerns or incidents during the project that led you to not rate these aspects as 5/5?”

The overall safety of the site could be improved. There are often people without PPE which could cause issues. It does not have to do with HOCHTIEF but it’s the general attitude of people on site that needs improvement

2. Quality of Workmanship

For the Hochtief internal team:

“You answered ‘No’ to whether the subcontractor made necessary improvements (if/where needed) but still rated overall quality as 5/5. Please clarify how this top score is justified if improvements were not made where needed.”

No improvements were needed, work was already up to standards.

For the subcontractor: “You rated your company’s work quality as 4/5. Could you explain the issues or factors that led you not to give a perfect score?”

There were some issues that prevented us from giving a perfect score. One factor was a misunderstanding about the specifications for the staircase railings. We initially believed steel rods were required, but the correct material was stainless steel cable. This mistake led to rework on that component. Additionally, we were frequently moved around the site to different areas as priorities shifted, which disrupted our workflow and made it more difficult to do our work. These factors contributed to our decision to rate our work quality as 4 instead of 5.

3. Timeliness and Schedule Adherence

For the subcontractor:

“You rated your timeliness as 3 and HOCHTIEF’s as 3, stating "first things first" and referencing a "large group and unstructured" management on HOCHTIEF’s side. Could you expand on that? “

The main issue with scheduling was the structure of the meetings and lack of coordinated planning. Weekly scheduling sessions involved very large groups and sometimes lasting two to three hours, which caused inefficiencies and frustration. A smaller group approach would have saved time. Also, there was no lean scheduling. At some points it was a bit chaotic.

On site, the team was frequently moved between areas to address shifting priorities, which disrupted the plan and led to work being carried out in a less-than-optimal sequence.

There was a rotation of people responsible for the schedule and with them it was not clear what each company needed to do and there were clashes.

On site we were just moved to areas we were working to move to other high priority areas. We were not sticking to the plan. So you were hopping from place to place quite often. When the schedule is clear everyone knows what to do.

Also, we did not have a model for the entire building from the outset which led to significant coordination challenges. Subcontractor drawings were not incorporated into a unified model. Our engineers had to do their scans after we started working there, because lack of access prevented early scanning. As we did not have early measurements and the model, work was being carried out in the wrong order.

4. Cost Management

For the subcontractor:

“You mention a "disagreement" over final financial aspects and rated HOCHTIEF’s cost management/support as 3. Could you elaborate on the nature of this disagreement and what specific issues arose?”

There was an ongoing disagreement regarding the initial scope and final financial details. The contract was delayed, so work started on site without a clear invoicing plan or payment schedule. Since there was no contract with the client at first, payment schedules were not established. Disagreements remain about what was included in the scope, with different views between us and the quantity surveyor, especially regarding preparation for fire-rated coating and what works were initially intended.

The budget was created with the direct and indirect costs. The price of our steel and fire rated coating was calculated. However, some of the fire rated coating was not needed. Still there was indirect work on it. There is a disagreement whether the preparation for the coating was in the initial scope. The quantity surveyor and we have different starting points. The quantity surveyor is now holding money back because of that.

5. Reliability and Repeat Engagement

For the subcontractor:

“You rated HOCHTIEF’s reliability as a 3 and cited “a lot of changes and surprises.” Can you specify what major changes or surprises occurred?”

Many changes and surprises were related to the project schedule and site management. Larger scope companies prioritized and occupied work areas as needed, often taking spaces where our teams were intended to work, leading to us regularly shifting locations. This made planning more difficult and impacted reliability, as the work sequence had to be adjusted on short notice.

We could move into different areas because of the amount of work we had to do. We only had to people on site, they could move elsewhere if another had to work in the area we were initially working. However companies with more people on site often had to send them home because they could not work that particular day, as someone else was working in their area

6. Technical Ability & Qualifications

For the Hochtief internal team:

“You answered ‘No’ to ‘Does the subcontractor qualify in the necessary technical skills for the project?’ but rated technical ability as 5/5. Please reconcile these two statements—is the subcontractor fully qualified, and if not, what justified a perfect score?”

The answer ‘No’ was a mistake. They were qualified

7. Financial Stability

For the subcontractor:

“You rated HOCHTIEF’s invoicing process as 3/5 and their financial dependability as 4/5, while noting disagreement. Can you specify the nature of these disagreements or any payment delays/issues?”

Because the contract was not finalized at the project’s start, there was no structured payment schedule. Invoicing was problematic, and there were disputes over what fell within the initial contract scope. The quantity surveyor withheld money on the basis that certain work was or was not included, resulting in ongoing differences of opinion about financial settlements.

8. Communication and Collaboration

For the subcontractor:

“You rated communication/collaboration as 3/5 despite using regular channels (Aconex and email). Please provide specific examples or reasons for your lower rating compared to the HOCHTIEF evaluation.”

Communication challenges occurred because of the structure of meetings and constant changes on site. The reliance on large meetings, along with frequent on-site relocations, negatively impacted coordination.

There were also interface issues with other subcontractors; for example, MEP teams occupied work areas before we could complete our tasks, leading to clashes that could have been prevented if all teams had access to an up-to-date project model. We were about to carry our work and people were already working there. But the problem is our work had to be carried out before theirs. We could not do it afterwards. It was not possible. So because we did not have an initial common model w here all subcontractors could upload there designs there were issues with the order for the works

10. Sustainability Practices

For the subcontractor:

” You rated HOCHTIEF’s sustainability as 4, but stated that expectations were not clearly communicated. Please clarify what sustainability measures or initiatives you implemented on this project, and whether any additional efforts were made beyond basic compliance.\’

At the start, there was no contract, so sustainability requirements were not clearly communicated. The approach to sustainability depended on the type of client.

Overall feedback from Subcontractor X

For the schedule do not do it in a big group of people Do it with small groups. Once it is 70% to 80% done then do it with the whole group. When you do it in a big group it takes 2 to 3 hours. On one side if I listen to what everyone is doing in big group I get an idea of what is going on, but if my scope is discussed only in the last hour, I might have wasted the previous one or two hours.

There should be a model for the whole building from the beginning. All of the drawings of the subcontractors need to be in that model. Everybody needs to work on that model. If there is a new phase the engineers should start scanning the entire building from now. We scanned the building during the project because we were not allowed to enter the areas to do it before. So that is why are engineers were delayed. The measurements need to be in the model early. What we experienced is that we had to work in the wrong order and there was issues with interface with other subs. We had collisions. MEP was already working were we were supposed to work and complete our part first. It is important for everyone to have access to the same model so we can collaborate.

On site there was a lot of understanding and respect. We were honest with each other. There were disagreements but people were willing to work past them and continue collaborating.

Appendix F: Subcontractor X Final Evaluation Report

Subcontractor Evaluation Report

Subcontractor Name: Subcontractor X

Final Rating (Equal Weights): 3.90 / 5

Final Rating (Equal Influence): 3.92 / 5

Final Rating (Role-Weighted Influence): 3.89 / 5

Scores Breakdown:

Safety and Compliance: 4

All required certifications (VCA, Safety Culture Ladder level 4, ISO 45001) valid for project duration; zero safety incidents or violations reported by either party. Both Hochtief and subcontractor rated 4. Subcontractor noted general site-wide PPE issues, not specific to their team. No negative findings attributable to the subcontractor.

Quality of Workmanship: 4

Hochtief reported 1% rework and only 2 defects out of 14 snags, with all handovers on time; rated quality as 5. Subcontractor rated 4, citing disrupted workflow and a specification misunderstanding requiring single-item rework (staircase rails). No systemic or repeated quality problems.

Timeliness and Schedule Adherence: 3

Hochtief: No contractor-attributed delay, rated 4 and noted hope for more proactive scheduling.
Subcontractor: self-rated 3 due to large, unstructured meetings, sequencing issues from lack of site model, and repeated area reassignments. Coordination/scheduling difficulties outweighed actual on-time delivery.

Technical Ability & Qualifications: 5

All technical certifications and qualifications were valid and up to date; Hochtief confirmed response of "No" for qualification was an error. Both parties described consistent capability in solving technical challenges (Hochtief and subcontractor both 4-5).

Cost Management: 4

Although payment disputes and scope clarification issues occurred, these arose because the project started before a contract or payment structure was agreed. No evidence of inappropriate cost management by the subcontractor. Hochtief: 4 for cost substantiation and timely costs. Subcontractor: disputes on initial scope but no explicit client criticism for cost control or budget overruns.

Reliability and Repeat Engagement: 4

Hochtief rated reliability 5 and stated no reminders were required. Subcontractor marked 3 and listed frequent relocation and surprises due to changing priorities driven by overall site management, not subcontractor performance. Subcontractor adapted to changes and fulfilled commitments, with issues mainly rooted in project-level coordination.

Financial Stability: 4

Both parties rated financial stability at 4. No payment failures or liquidity risks reported. Disagreement and procedural confusion on statements caused processing delays but not actual risk to project continuity.

Communication and Collaboration: 3

Hochtief gave full marks for communication; subcontractor only 3, citing inefficient, lengthy meetings, issues interfacing with other trades, and challenges when areas were occupied by other teams. Issue more in system/process than contractor conduct, but negative impact remained.

Innovation and Problem-Solving: 4

Hochtief rated problem-solving/adaptability 5. Subcontractor rated 4, provided solutions as needed but did not highlight "above and beyond" innovations. Helpful but not exceptional.

Sustainability Practices: 4

Hochtief listed full certifications (CO2 Ladder level 5, ISO 14001) and subcontractor scored them 4, noting sustainability goals were not communicated at start but guidance was given. No shortfall in project execution; only a procedural gap in goal setting.

Summary:

The evaluation of Subcontractor X demonstrates strong overall performance across all primary subcontractor criteria. The team maintained all required safety certifications and operated incident-free, though minor general site-wide PPE lapses were noted. Quality of work was high, confirmed by minimal rework (1%) and positive ratings from Hochtief; minor workflow disruption and a single miscommunication on staircase specifications were promptly addressed. Both parties cited challenges in timeliness and scheduling, chiefly due to unstructured, lengthy meetings, lack of lean/efficient scheduling, and frequent relocation by project management rather than the subcontractor. All technical qualifications were satisfied, with both parties rating technical skill and problem-solving abilities highly. Cost management was fundamentally sound: disputes regarding scope and financial settlements stemmed from the project commencing before a contract or payment schedule was agreed and were not due to subcontractor error. Reliability was strong, with Subcontractor X adapting to shifting site conditions and fulfilling all commitments, despite negative impacts from overall project organization. While communication and collaboration processes (notably meeting length and structure) could be improved, both parties successfully collaborated using multiple channels. Innovation and sustainability standards were met, as evidenced by certifications, implemented solutions, and adaptive support on technical and sustainable goals. Lessons learned highlight the importance of advance contractual agreement, structured site management, and explicit goal communications at project outset.

Lessons Learned:

- Starting work without a signed contract or clear scope caused disputes and financial processing delays. All contractual and financial structures must be in place before site mobilization.
- Unstructured, overly large scheduling meetings led to inefficiency. Leaner, smaller, and more focused coordination sessions are recommended.
- Frequent area reassignments and lack of a unified building model disrupted workflow for multiple trades. Early measurement, accurate site modeling, and disciplined project scheduling are critical to reliable site progress.
- General project communication and sustainability requirements should be shared explicitly and early to ensure full alignment.

Recommendations for Hochtief:

- Finalize contracts, scope, payment plans, and invoicing structure before authorizing site access or commencement of work.
- Implement structured, lean scheduling meetings limited to relevant stakeholders. Introduce building-wide project models (e.g., BIM) for coordinated sequencing.
- Provide site-wide communications and clear expectations for all subcontractors regarding safety, quality documentation, and sustainability processes at project launch.
- Ensure subcontractors are fully briefed on expected documentation, reporting formats, and sustainability goals via pre-mobilization workshops.

Recommendations for the Subcontractor:

- Continue the existing good practice of adaptability and technical problem-solving.

- Document and promptly communicate disruptions or changes driven by client or site conditions to maintain transparency in progress and claims.
- Pursue early, proactive alignment on contract terms and scope before mobilization on future projects.
- Request early involvement in scheduling/model-coordination reviews to highlight sequencing issues and avoid site inefficiencies.

Appendix G: Survey Results

Criterion	Statement	Average Score	Criterion Average
Insightfulness	The AI report provides more detailed and insightful content.	4.00	3.84
	The summary of collaboration enhances overall understanding.	4.00	
	The report enables clearer identification of strengths & weaknesses.	4.00	
	The report includes sufficient context to understand the subcontractor's role.	3.10	
	The insights help explain impact on project outcomes.	3.80	
	The evaluation report presents new or non-obvious information that would not typically be captured in a standard assessment	3.80	
	The final report provides enough detail for someone unfamiliar with the subcontractor to gain a solid understanding of their performance.	4.00	
	The report enables individuals without prior interaction with the subcontractor to make informed decisions regarding future collaboration.	3.80	
Clarity and Relevance of Follow-up Questions	The follow-up questions clearly addressed ambiguities in the evaluation data.	4.30	4.06
	The follow-up questions clearly addressed ambiguities in the evaluation data.	3.90	
	The follow-up questions clearly addressed gaps in the evaluation data.	4.00	
	The assistant's follow-up questions demonstrated an understanding of the evaluation context.	3.80	

	The questions addressed both the quantitative and qualitative aspects of the evaluation.	4.00	
	The follow-up questions helped uncover details about the subcontractor's performance that were not initially evident.	3.90	
	The questions revealed insights that might have otherwise been overlooked.	4.00	
	The follow-up questions were relevant and well-targeted to the specific issue at hand.	4.00	
	The tone and wording of the follow-up questions were appropriate for professional use.	4.60	
	The follow-up questions were concise and easy to understand.	4.50	
	I would use these follow-up questions in a real subcontractor evaluation meeting.	4.50	
Justification and Transparency of Ratings	The rationale for each rating is clear and easy to follow	4.30	4.08
	The justifications are logically supported by the data provided.	4.20	
	The justifications demonstrated fairness and objectivity in the evaluation process.	4.00	
	The explanations for the scores were detailed enough to avoid ambiguity.	4.00	
	The justifications provided sufficient context to understand the scores assigned.	4.30	
	The justifications highlighted both strengths and areas for improvement in the performance.	4.20	
	The justifications helped clarify how the scores were determined.	4.00	
	Compared to the manual method, the AI report offers better explanation of scores.	4.40	

Quality of Lessons Learned and Recommendations	The lessons reflect meaningful insights from the project.	4.40	4.25
	The lessons learned are clearly linked to specific behaviors or outcomes during the project.	4.00	
	The recommendations are specific, actionable, and relevant for future evaluations.	4.20	
	The recommendations are grounded in the evaluation data and not overly generic.	4.00	
	The report provides a good balance between critical feedback and constructive suggestions.	3.60	
	The recommendations align with HOCHTIEF's internal goals and evaluation standards.	4.70	
	I would consider applying these insights in upcoming selections	4.50	
	The lessons and recommendations help prevent similar issues in future projects.	4.30	
Perceived Added Value	The tool improves the overall quality of the evaluation process.	4.30	4.24
	I find the AI-enhanced report more helpful than the standard version.	3.80	
	The structured format of the report made it easier to understand and use.	4.00	
	The tool helped reduce ambiguity or confusion in the evaluation process.	4.50	
	The use of follow-up questions helped clarify issues I might have otherwise overlooked or made assumptions about.	4.40	
	The assistant added a level of objectivity that improved my confidence in the evaluation results	4.60	
	The integration of lessons learned, and recommendations adds significant value to the evaluation	3.90	

	The tool supports better communication between HOCHTIEF and subcontractors	4.30	
	I would choose to use this tool for future subcontractor evaluations.	3.30	
	Some of the scores in the report felt higher than the situation deserved	3.60	
Perceived Limitations and Risk Awareness	The scores were sometimes based more on the explanations than on actual performance.	3.60	3.08
	The assistant gave scores that did not fully reflect what happened in the project.	2.70	
	The justifications for the scores were sometimes too generic or repetitive.	3.10	
	The assistant's explanations missed important background that I would expect in a real evaluation.	2.80	
	The assistant sometimes asked questions that felt too general or not tailored to the situation	2.50	
	I would be cautious using this report to support high-impact decisions without review.	3.10	
	The assistant’s report should always be checked by a person before it is used officially.	4.00	
	Some of the recommendations were too broad to apply in real situations.	2.80	
	The assistant sometimes repeated generic phrases instead of providing unique reasoning	3.50	
	The assistant referred to data that did not exist in the documents	2.70	
	The assistant treated vague or incomplete input too generously when giving justifications.	2.90	
	Some follow-up questions did not seem necessary and added little value.	3.00	

	The report contained useful insights, but it was too long for quick review in meetings.	3.20	
	I would not feel confident using this report to support supplier prequalification decisions on its own.	2.70	
	The assistant's outputs did not always reflect how performance actually played out on-site.	3.30	
	The follow-up questions clearly addressed ambiguities in the evaluation data.	3.30	

Appendix H: Tables

***Table G1**. Example of criteria and questions for supplier / subcontractor evaluation post project completion (Derived from the evaluation document used by HOCHTIEF for a past project)*

Criteria	Evaluation Questions	Rating	Comments	To be Completed by
QUALITY	What is the quality of the engineering work delivered?			Work planner
	What is the quality of the products, work or services provided?			
HSE	Do they work in accordance with health and safety laws and regulations?			Execution
	Is the HSE plan in accordance with the BN HSE purchasing requirements?			
	Is suitable and approved equipment used?			
	We work with certified and competent employees			
	What is the safety awareness and culture of the employees?			
CSR	Does the supplier have a CO2 reduction strategy and plan for the short term?			Logistics
	Does the supplier have an LCA (Life Cycle Analysis) of their product available?			
	How does the Supplier reduce waste and how does it separate waste on the project?			
	If FSC wood applies to the work, has the supplier provided sufficient data?			
	To what extent does the Supplier recycle and/or promote materials?			
	Is the BN Code of Conduct for Subcontractors and Suppliers complied with?			
PLANNING	Will the agreed documents be delivered on time?			Calculator
	Works/delivers according to agreed planning / progress.			
COOPERATION	Are the agreements made being complied with? Claims no unexpected additional work.			Execution
	How does communication work and are they proactive in thinking along with our interests? Dealing with changes.			

DOCUMENTATION	Are all legal, financial documents and work permits in order?			Purchasing and Control
	Company keeps a thorough administration and is timely with the delivery of the agreed project and business documentation			
Final Assessment	Would you like to make this relationship work/deliver for you again?			

Table G2: *Results of manual evaluation carried out by HOCHTIEF Nederland for Subcontractor 1 and Subcontractor 2 (Source: Official evaluation documents from HOCHTIEF's archives)*

Criteria	Evaluation Questions	Ratings of Subcontractor X
QUALITY	What is the quality of the engineering work delivered?	4
	What is the quality of the products, work or services provided?	4
HSE	Do they work in accordance with health and safety laws and regulations?	4
	Is the HSE plan in accordance with the BN HSE purchasing requirements?	4
	Is suitable and approved equipment used?	5
	We work with certified and competent employees	4
	What is the safety awareness and culture of the employees?	4
CSR	Does the supplier have a CO2 reduction strategy and plan for the short term?	N/A
	Does the supplier have an LCA (Life Cycle Analysis) of their product available?	N/A
	How does the Supplier reduce waste and how does it separate waste on the project?	N/A
	If FSC wood applies to the work, has the supplier provided sufficient data?	N/A
	To what extent does the Supplier recycle and/or promote materials?	3
	Is the BN Code of Conduct for Subcontractors and Suppliers complied with?	3
PLANNING	Will the agreed documents be delivered on time?	3
	Works/delivers according to agreed planning / progress.	3
COOPERATION	Are the agreements made being complied with? Claims no unexpected additional work.	3
	How does communication work and are they proactive in thinking along with our interests? Dealing with changes.	3
DOCUMENTATION	Are all legal, financial documents and work permits in order?	3

	Company keeps a thorough administration and is timely with the delivery of the agreed project and business documentation	N/A
Final Assessment	Final Score	3.6

Table G3: *Follow-up questions generated by the AI-powered evaluation assistant, organized by evaluation criteria and recipient (internal team or subcontractor).*

Criteria	Questions to the Subcontractor	Questions to the Internal Team
Safety and Compliance	You mentioned periodic meetings for safety protocols/training and rated site safety 4/5. Could you specify if there were any safety concerns or incidents during the project that led you to not rate these aspects as 5/5?	You rated safety awareness, prioritization of safety meetings, and adherence to protocols all as 4/5 instead of 5/5, despite confirming no safety incidents or violations and full certification validity. Can you explain specifically what prevented a top score in these areas?
Quality of Workmanship	You rated your company's work quality as 4/5. Could you explain the issues or factors that led you not to give a perfect score?	You answered 'No' to whether the subcontractor made necessary improvements (if/where needed), but still rated overall quality as 5/5. Please clarify how this top score is justified if improvements were not made where needed.
Timeliness and Schedule Adherence	You rated your timeliness as 3 and HOCHTIEF's as 3, stating 'first things first' and referencing a 'large group and unstructured' management on HOCHTIEF's side. Could you expand on that?	*(No question to internal team for this criterion)*
Cost Management	You mention a 'disagreement' over final financial aspects and rated HOCHTIEF's cost management/support as 3. Could you elaborate on the nature of this disagreement and what specific issues arose?	*(No question to internal team for this criterion)*
Reliability and Repeat Engagement	You rated HOCHTIEF's reliability as a 3 and cited 'a lot of changes and surprises.' Can you specify what major changes or surprises occurred?	*(No question to internal team for this criterion)*
Technical Ability & Qualifications	*(No question to subcontractor for this criterion)*	You answered 'No' to 'Does the subcontractor qualify in the necessary technical skills for the project?' but rated technical ability as 5/5. Please reconcile these two statements—is the subcontractor fully qualified, and if not, what justified a perfect score?
Financial Stability	You rated HOCHTIEF's invoicing process as 3/5 and their financial dependability as 4/5, while noting disagreement. Can you specify the nature of these disagreements or any payment delays/issues?	*(No question to internal team for this criterion)*

Communication and Collaboration	You rated communication/collaboration as 3/5 despite using regular channels (Aconex and email). Please provide specific examples or reasons for your lower rating compared to the HOCHTIEF evaluation.	*(No question to internal team for this criterion)*
Innovation and Problem Solving	*(No question to subcontractor for this criterion)*	*(No question to internal team for this criterion)*
Sustainability Practices	You rated HOCHTIEF's sustainability as 4, but stated that expectations were not clearly communicated. Could you please provide more details?	*(No question to internal team for this criterion)*

***Table G4**: Ratings and Corresponding Justifications for Cost Management and Sustainability Practices Across Three Evaluation Runs*

Criterion	1st Run	2nd Run	3rd Run
Cost Management	Contract price unchanged by final, but Hochtief cited issues with substantiation and slow payment approval at the end. Subcontractor also mentioned disagreements in final phase. Final invoicing needed rechecking and clarification. **(3)**	Cost management and substantiation of claims required frequent clarifications, and disputes arose toward the end over payment timing and contract statements, leading to a moderate score in this area. **(3)**	Although payment disputes and scope clarification issues occurred, these arose because the project started before a contract or payment structure was agreed. No evidence of inappropriate cost management by the subcontractor. Hochtief: 4 for cost substantiation and timely costs. Subcontractor: disputes on initial scope but no explicit client criticism for cost control or budget overruns. **(4)**
Sustainability Practices	All contractual obligations met, but with "no" to onsite energy/water initiatives and eco-material percent rated not applicable. Subcontractor received guidance by request; no innovative sustainability practiced beyond basic compliance. **(3)**	Sustainability Practices: 4 Hochtief noted all certificates in place and compliance with environmental requirements. Certain eco or energy-saving measures were not part of project scope ('not applicable'). Subcontractor stated clear guidance was provided but said goals were unclear at start **(4)**	Hochtief listed full certifications (CO_2 Ladder level 5, ISO 14001) and subcontractor scored them 4, noting sustainability goals were not communicated at start but guidance was given. No shortfall in project execution; only a procedural gap in goal-setting. **(4)**