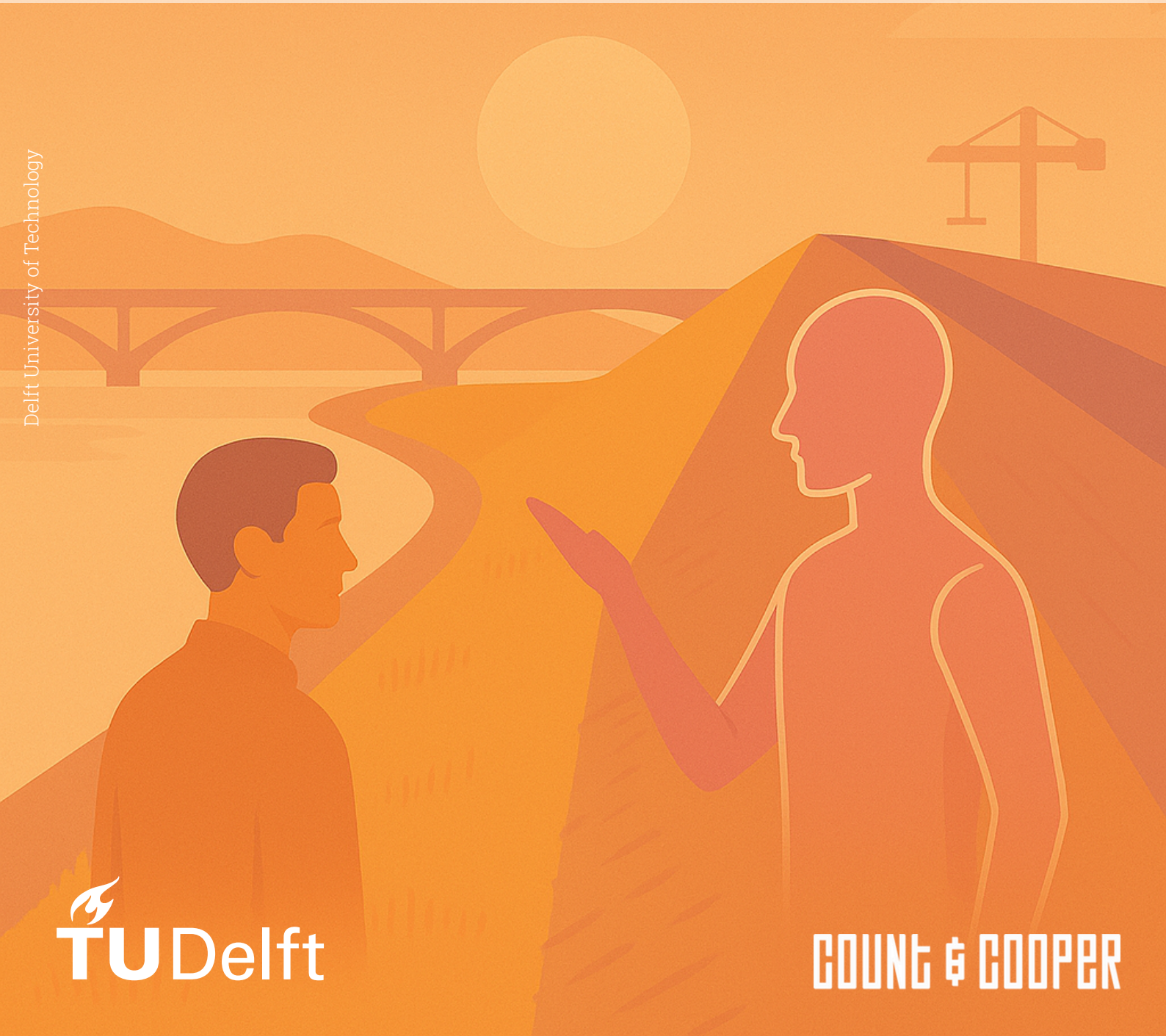
Generative AI in Infrastructure Tendering

A Study on Human-Generative AI Deliberation

Master Thesis - Construction Management and Engineering

Roos van Duuren



Delft University of Technology

TU Delft

COUNT & COOPER

Generative AI in Infrastructure Tendering

A Study on Human-Generative AI Deliberation

by

Roos van Duuren

Chair TU Delft:	Dr. ir. Marian Bosch-Rekveldt
First Supervisor TU Delft:	Dr. Tong Wang
Second Supervisor TU Delft:	Dr. ir. Ranjith Kuttantharappel Soman
Supervisor Count & Cooper:	ir. Niels Dusseldorp
Faculty:	Faculty of Civil Engineering and Geosciences, Delft

Cover: OpenAI (2025). Image generated by ChatGPT 4o with the prompt: *"An image of a human and AI deliberating beside a dike in a warm orange landscape"*. OpenAI.

TU Delft

COUNT & COOPER

Preface

This thesis has been a journey of discovery, learning, gaining experience, and overcoming challenges. During one of our first meetings, my supervisor Tong told me: *"After ten years, you'll be able to look back on the process in a positive way, because you'll realise how much you've learned"*. At that moment, I was shocked to think that the journey might be that tough. Luckily, I can already look back on the experience with a sense of positivity. Throughout the process, I have learned to let go more easily, to do what I can within the time available, and to accept that some things must be left for another day. Nonetheless, I am very curious to see what additional positive aspects I will discover ten years from now. ;)

I would like to express my grateful thanks to my committee from TU Delft. Without their constructive and forward-looking feedback, I would not have been able to deliver the thesis I am presenting today. I am particularly grateful to Dr. Tong Wang, with whom I regularly had meetings. Her guidance helped me to overcome the obstacles I encountered along the way. I would also like to thank Dr. ir. Marian Bosch-Rekveldt and Dr. ir. Ranjith Kuttantharappel Soman for their valuable insights and contributions during progress meetings and during other interactions. The support and input of my committee have been really valuable throughout this process.

I am also very thankful to Count & Cooper, and in particular to my supervisor and committee member, Niels. Each week, Niels made time in his schedule to support me throughout the process, offering valuable insights and helping me shape the design of my research. He also attended every session organised for data collection and provided assistance, for which I am truly grateful.
In addition, I would like to thank Count & Cooper in general for giving me the opportunity to do my thesis internship with them. I would also like to thank all the participants in the experiment and the experts who contributed to the evaluation sessions. Their input was essential to this research.

Lastly, I am very grateful for the support of my family and friends. I would especially like to thank my big brother, Luuk, who has been a huge help. Every Tuesday morning, we had a meeting, often helping me to prioritise tasks, think things through, ask the right questions, or just to have a laugh. My boyfriend, Remko, has also been a great support, not only mentally but also through his skills with Overleaf. Lastly, I would like to thank my parents for their constant support throughout this journey.

Finally, I hope that this thesis contributes to the effective implementation of generative AI in infrastructure tendering. Specifically, to support the creation of tenders that achieve higher quality.
And above all, I hope you enjoy reading this thesis!

Roos van Duuren
Delft, May 2025

Summary

This thesis explores the possibilities of using generative AI (GenAI) to assist in the review of open infrastructure tenders. It focuses on the role of the consultant, working together with contractors to provide a tender bid. The study aims to identify how individual interactions with GenAI shape tender reviews and to explore the underlying motivations for adopting or avoiding its use. Central to this is the concept of deliberation, defined as a reflective exchange of ideas between, in this case, an individual and GenAI, that aims to deepen understanding on both parties rather than to reach consensus. The aim is to understand how individuals experience such deliberation during the review processes, with the goal of answering the following research question:

"How do individuals experience human-generative Artificial Intelligence deliberation when reviewing open infrastructure tenders within project-based organisations?"

Methodology

This thesis followed a qualitative methodology aimed at exploring human–GenAI deliberation in a real-world professional context. The literature review provided the necessary background, and an experiment was conducted focusing on a realistic infrastructure tendering scenario. This structured experiment, followed by group discussions and evaluation sessions, was used to collect data. These steps were designed to gather both individual and collective perspectives on the interaction between individuals and GenAI. The findings from the literature, the experiment, the group discussion, and the evaluation sessions were brought together to draw conclusions and to develop an evaluation framework for researching human–GenAI interactions.

Literature Review

This literature review introduces the Activity Theory as the primary analytical framework, offering a lens to explore how GenAI operates within broader socio-technical systems. Central to this approach is the concept of contradictions, which are understood as tensions within or between components of an activity system that may either act as drivers or barriers for the adoption and use of GenAI.

The review also highlights a range of factors influencing user engagement with GenAI, with trust identified as a critical element. Trust is shaped by several conditions, including the reliability and explainability of GenAI outputs, the ease of use, and the level of scepticism among users. In addition, broader contextual factors such as privacy, organisational culture, and environmental concerns play a role in shaping user attitudes.

Following, the Human–GenAI Deliberation framework is introduced. This framework is designed to bridge the gap between humans and AI by directly involving users in the AI's decision-making process. It emphasises the importance of careful evaluation and aims to support more well-founded decisions through active human participation. The Human–GenAI Deliberation framework outlines clear, structured steps that facilitate effective deliberation between a human and a GenAI system. These steps form a key foundation for this study and are integrated into its design to guide meaningful interaction.

Experiment Design

The research involves the design of an experiment with participants from the consultancy firm Count & Cooper, each with varying levels of experience in tendering. First, participants and the GenAI tool individually reviewed a specific measure, a small section of a tender, using a structured questionnaire. This was followed by an interaction with a pre-instructed version of the GenAI tool, designed according to the Human–AI Deliberation Framework to facilitate a structured deliberative conversation. During this session, both participants and GenAI had the opportunity to revise their thoughts and update their ratings of the review.

Following the deliberation, a group discussion allowed participants to reflect on their experience and evaluate a series of pre-defined statements about GenAI. These discussions were transcribed and analysed to identify recurring themes. Finally, two evaluation sessions were held with industry experts to assess the relevance of the findings and to contribute additional insights.

Findings

The findings of the experiment show that while some participants were open to adjusting their opinions, others remained more committed to their initial views. In particular, the younger and less experienced individuals tended to be more open, possibly due to professional insecurities. In contrast, older, more experienced users seemed more resistant to GenAI's suggestions. Active engagement with the tool correlated with more significant shifts in opinion, particularly among those who were willing to ask questions and therefore seemed to trust the system more. However, a key issue was GenAI's overly servile behaviour: it rarely doubted the input of the individuals. This servility seemed to lower its perceived value.

Several contradictions were identified that could act as drivers and barriers to GenAI adoption. Clear rules of engagement, such as structured interaction guidance, could drive users to navigate the tool more confidently and formulate effective prompts. GenAI's perceived objectivity and lack of personal needs encourage reflection and more critical thinking, which users found beneficial. Efficiency, although not always directly experienced, was often expected and seen as a significant driver. Curiosity, especially when reinforced by peer encouragement or positive examples, also played a key role in stimulating interest and experimentation. On the other hand, barriers included scepticism about the usefulness of GenAI, difficulties in creating good prompts, and a general lack of awareness of its capabilities. The lack of explainability and transparency further undermined trust and usability, limiting deeper engagement. While privacy assurance was not identified as a driver in itself, its absence would significantly hinder adoption, highlighting the importance of maintaining this condition.

As a contribution, an evaluation framework has been developed to guide research on human–GenAI interaction. It is structured around the contradictions identified in the Activity Theory, which served as a lens for analysing empirical data.

Conclusion

This study highlights the importance of addressing drivers and barriers to GenAI adoption. A central theme that arises from the broader analysis is the role of trust in shaping users' willingness to adopt. Trust appears to be fostered not only by transparent and explainable system responses but also by factors such as peer-supported curiosity. Importantly, an observed knowledge gap about how to engage meaningfully with GenAI is a barrier to more widespread and effective use. This includes a lack of understanding of the potential of GenAI and prompt formulation. At the same time, addressing this gap by providing users with guidance and knowledge could serve as a key driver. To support this, offering targeted training and contextual support could help users gain confidence while also ensuring they do not become over-reliant on the tool.

Recommendations

Practical recommendations for organisations involved in infrastructure tendering may include supporting the adoption of GenAI by addressing the knowledge gap on how to engage with GenAI. This could entail targeted training on prompt formulation, encouraging informal peer learning, and providing structured guidance with practical examples. Stressing that GenAI is a collaborative partner, rather than a passive assistant, could help foster engagement and trust. Introducing a GenAI-assisted deliberation step close to the final submission of tenders seemed to support critical reflection. Thereby, developing a customised GenAI environment and ensuring confidentiality could further facilitate engagement.

Future research may consider scaling up the findings of this study using quantitative methods to validate the identified contradictions, drivers, and barriers. Further research into effective prompting techniques and exploring a deeper understanding of trust in this interaction appears important for wider adoption. In addition, extending the use of GenAI deliberation beyond consultancy firms and into other phases of project delivery may reveal further value. GenAI-to-GenAI interaction within the contracting party could also emerge as a relevant area for future investigation.

Contents

List of Figures

List of Tables

1

Introduction

This introduction provides an overview of the key aspects of generative AI in infrastructure tendering. It begins by defining the context in Section 1.1, followed by the exploration of the nature of the problem in Section 1.2. Next, the research scope is formulated (see Section 1.3). In Section 1.4, the research gap is identified, which in turn leads to the formulation of the research questions (see Section 1.5). The relevance of the research is then discussed in Section 1.6, and finally, Section 1.7 presents the outline of the remainder of the thesis.

1.1. Context

A shift is occurring in the tendering culture due to the introduction of Artificial Intelligence (AI) systems such as ChatGPT (Van der Linden, 2025). According to Van der Linden, winning tenders has become significantly easier with the help of AI, suggesting that AI can offer a significant advantage in the bidding process.

Tendering is a process in which a client invites bids for a project, allowing them to select a contractor whose proposal best aligns with their vision for project realisation (Ogunsemi and Aje, 2006). Contractors and consultants often work together in preparing bids, collectively forming the contracting party. When selecting the winning bid, the contracting party is chosen based on the economically most advantageous tender, referred to in Dutch as the economisch meest voordelige inschrijving (EMVI) (Overheid.nl, 2022). Under the EMVI framework, three criteria can be applied to determine the winner: the best price–quality ratio (beste prijs-kwaliteitverhouding or BPKV), the lowest price, and the lowest cost based on cost-effectiveness (Overheid.nl, 2022). Since 2012, these criteria have been nationally adopted, making the price–quality ratio particularly important. Contracting parties offering the most balanced combination of cost and quality have the highest chance of winning the tender (Overheid.nl, 2022). Therefore, for the contracting party, addressing the quality aspect of the tender is crucial to winning. As a result, it is important to explore ways to improve the quality of tender documents.

The use of Generative AI (GenAI) could serve as a valuable tool to support and improve the quality of tender documents. GenAI is a type of AI with the capabilities to generate content that is almost identical to human-created work (Bubeck et al., 2023). It involves computer technology methods that can create original, meaningful content, including text, images, and audio, by learning from existing data (Feuerriegel et al., 2023). Tools such as Dall-E, copilot (Feuerriegel et al., 2023), and ChatGPT (Nah et al., 2023) are examples of the GenAI. ChatGPT is a prime example as it is seen as a major breakthrough that has driven rapid progress in AI technology (Kanbach et al., 2024). ChatGPT was launched at the end of 2022 (OpenAI, 2022b) and quickly gained popularity due to its simple user interface, which allows almost any person to generate content in a short amount of time (Kanbach et al., 2024). Furthermore, Kanbach et al. describe ChatGPT as a breakthrough in the field of natural language processing, as it uses language in a manner that closely resembles humans. The conversational format enables ChatGPT to respond to follow-up questions, acknowledge its mistakes, challenge incorrect assumptions, and decline inappropriate requests (OpenAI, 2022b). Compared to traditional AI, which focuses

on analytical tasks for decision-making, GenAI introduced capabilities for creative content generation, blending analysis with creativity. With GenAI, the emphasis shifts to a human-AI interaction: when a user provides prompts, AI interprets the user's intent, generates responses, and anticipates potential follow-up prompts, creating a dynamic feedback loop between human and machine (Feuerriegel et al., 2023). This shift in AI towards an interaction between humans and AI enhances GenAI's potential to significantly impact various industries, highlighting its transformative capabilities for individuals, organisations, and society because of its wide range of possible applications (Banh and Strobel, 2023).

This study involves three key stakeholders in the tendering process: the client, the contractor, and the consultant. In this setup, the contractor and consultant collaborate as the contracting party, working together to prepare a tender in response to the client's request. Each party may interact with tools such as GenAI. This thesis focuses specifically on the dynamics of human–GenAI interactions, particularly how individuals on the consultancy side engage with GenAI tools. Traditionally, communication in tendering processes occurs between humans, either across parties or within a single party. However, the integration of GenAI introduces the possibility of individuals interacting directly with GenAI systems. By analysing such interactions, this research aims to uncover how consultants engage with GenAI and to explore its potential role in shaping and supporting the tendering process.
For this study, the collaborating company is the consultancy firm Count & Cooper. Count & Cooper points out that the consultant usually contributes most to the qualitative and process-oriented aspects of the tender document, while the contractor focuses more on the technical components. The company also emphasises its openness to innovation, recognising that it must show contractors its value in order to be chosen as a partner for collaborative tender bids. The need to highlight unique capabilities underscores the motivation and necessity for innovation, aligning with the implementation of GenAI.

1.2. Nature of the Problem

As a consultancy firm, specialised in assisting contractors with the preparation of high-quality tender bids in infrastructure projects, the company Count & Cooper is continuously seeking ways to innovate and improve. This drive for improvement aims to strengthen the firm's competitive position in the market. As outlined in Section 1.1, enhancing quality is essential for winning tenders.
In the academic field, ChatGPT is regarded as transformative, reshaping how people interact with technology (Haleem et al., 2022). ChatGPT has proven effective in enhancing efficiency in business communication, particularly in management, by improving decision-making processes through data analysis and insights that support more informed choices (Nugroho et al., 2023). Additionally, ChatGPT is noted as a valuable tool for research and innovation, accelerating research workflows and enabling quicker identification of new innovation opportunities (Nugroho et al., 2023). To stay competitive and adaptable in a rapidly changing landscape, companies must stay updated with AI advancements.
Count & Cooper values innovation and embraces new technologies and approaches within the construction sector. The use of ChatGPT is in line with Count & Cooper's ambition and commitment to innovation.
The problem, and therefore the reasoning behind conducting a research, lies in the need to gain a thorough understanding of how ChatGPT, and therefore GenAI, can effectively enhance Count & Cooper's workflows. Gaining this understanding requires exploring how GenAI is integrated into daily operations, examining its specific contributions, evaluating the reasons for its use or non-use, and identifying barriers to broader adoption.

1.3. Research Scope

While the broader field of tendering and its possibilities in the field of GenAI is being explored, this study narrows its focus to a size that can fit within the scope of a master's thesis. To define the scope of the research, the following elements are discussed: the focus on the consultant within the contracting party, the concept of project-based organisations, open infrastructure tenders, the concept of deliberation, the focus on the consultant, and the specific type of GenAI tool used in the research.

1.3.1. Project-Based Organisations

Project-based organisations are organisations for which most or all business activities are carried out in the form of projects (Bartsch et al., 2013, Thiry and Deguire, 2007). Miterev et al. (2017) define project-

based organisations as project-oriented organisations, adopting management by projects as a core strategy, conducting work through projects and programs as temporary entities. They manage a diverse portfolio of both internal and external projects, employing distinct processes for project, program, and portfolio management. Lastly, Miterev et al. note that these organisations embrace uncertainty, change, and collaboration, fundamentally viewing themselves as project-oriented.
Since the tendering process is part of the pre-construction phase in the construction project life cycle (Mohemad et al., 2010), it is considered a project in itself and thus aligns well with the concept of a project-based organisation. Thereby, since the number of tenders in organisations often exceeds the number of projects, and given the fact that tenders occur in an environment that is highly competitive, the ability to learn quickly and adapt to steep learning curves provides advantages (Uil, 2021). For this reason, tendering within project-based organisations fits the scope of this master's thesis.

1.3.2. Open Infrastructure Tenders

In the Netherlands, the two most common types of tendering are open tenders (65% of the tenders in 2024) and restricted tenders (20% of the tenders in 2024) (Pianoo, 2025). In an open tender, all interested contracting parties are allowed to submit a bid before the same deadline (Overheid.nl, 2022). This open tender approach is intended to promote transparency and foster competition (Ogunsemi and Aje, 2006). In contrast, a restricted tender consists of two stages: suppliers first apply and are screened, after which a selected group is invited to submit a bid (Overheid.nl, 2022). Given that open tenders are by far the most commonly used in the Netherlands (Pianoo, 2025), this research focuses specifically on open tendering procedures.

1.3.3. Review Tendering Stage

From the perspective of the contracting party, the general tendering process in the construction industry involves several steps prior to submission (Mohemad et al., 2010). A distinction can be made between the phase before the decision to proceed with the bid and the phase after the decision has been made: the pre-tendering stage and the tendering stage. The pre-tender stage involves assessing project feasibility (Oladokun et al., 2011), while the tendering stage includes the preparation and submission of the bid (Mohemad et al., 2010).
This thesis focuses specifically on the tendering stage within infrastructure tenders in the construction industry. The choice to concentrate on this phase stems from its complexity, uncertainty, and known challenges, such as bias and inconsistent decision-making often influenced by intuition, subjective judgment, or emotion (Mohemad et al., 2010). These challenges underline the need for improved methods and tools to support more effective and higher-quality tendering.

1.3.4. Deliberation

This thesis investigates the interaction between humans and GenAI through letting them have a deliberative conversation. Therefore, it is crucial to define the concept of deliberation clearly. According to CambridgeDictionary (2025), deliberation means considering or discussing something. A more detailed explanation comes from Kenter et al. (2016), who describe deliberation as the process of considering or evaluating information to improve understanding and develop values around previously unfamiliar topics. Elaborating further, Gracia (2003) defines deliberation as a process where everyone affected by a decision can share their perspectives and listen to the views of others. The goal, as explained by Gracia, is not necessarily to reach consensus but to enrich individual viewpoints and through this process of deliberation, decisions and perspectives become more thoughtful and well judged.
In summary, for this master's thesis, deliberation is defined as the process of considering or discussing information to enhance understanding and enrich individual perspectives. It involves parties sharing their views and listening to others, with the goal of fostering thoughtful and well-informed judgments rather than necessarily reaching consensus.

1.3.5. Consultant

As outlined in Section 1.1, within the contracting party, the consultant is responsible for the process-oriented aspects of the tender, while the contractor typically focuses on technical standards. In order to be selected as a partner in tender bids, the consultant must demonstrate added value to the contractor. Emphasising innovation, such as the use of GenAI, is a way to strengthen the consultant's position in

the tendering market. Due to their openness to adopting innovative tools such as GenAI, consultants are the primary focus of this research.
Moreover, as elaborated by Count & Cooper, consultants are generally more innovation-oriented than contractors, making them better positioned to adopt and implement new tools like GenAI effectively within the tendering process. For these reasons, this study focuses specifically on the consultant.

1.3.6. ChatGPT as GenAI Tool

GenAI refers to a range of tools within AI that utilise computer technology methods to create original, meaningful content by learning from existing data (Kanbach et al., 2024). As mentioned in Section 1.1, ChatGPT serves as a great example of GenAI, representing a significant breakthrough that has accelerated progress in AI technology (Kanbach et al., 2024).
For this thesis, ChatGPT is available as a tool for research. Count & Cooper has invested in a Team licence, a restricted enterprise version of ChatGPT, which ensures that data inputs from employees are not used for training purposes, thereby preventing information from leaving the organisation (OpenAI, 2022a). Due to the availability of this secure version, ChatGPT is used as the primary tool for this research. Within this restricted version of ChatGPT, several models are accessible, including GPT-4o. This version allows users to create their own GPTs, which are versions of ChatGPT that can be customised for specific tasks by combining instructions, knowledge, and capabilities (OpenAI, 2024b). For this reason, GPT-4o is used in this thesis, as it enables the development of a custom version suited to the research needs.

1.4. Research Gap

GenAI is relatively new, with advancements accelerating rapidly (Nah et al., 2023). This indicates that numerous developments will come and that substantial research is still needed to understand the impact of GenAI. As Feuerriegel et al. (2023) describe, GenAI has the potential to reshape industries where creativity, innovation, and knowledge processing are essential. However, realising this potential requires continued research. This section elaborates on the research gap, defining the different components and systems underlying GenAI that require further research, forming the foundation for this study. It highlights the need for research on the integration of social and technical aspects, more specifically on the interaction between humans and GenAI. Additionally, it addresses the gap in research on the use of GenAI in the tendering process within infrastructure projects, specifically from the consultant's perspective.

1.4.1. Socio-Technical

GenAI is seen not only as a technology, but as a part of a larger system that includes both social (human) and technical elements, also known as a socio-technical system (Feuerriegel et al., 2023). In socio-technical systems, a gap exists due to a lack of effective approaches for addressing both the technical and societal aspects (Zhang et al., 2024). The emphasis is primarily placed on the technical aspects, including those related to AI systems (Weidinger et al., 2023). However, since AI systems are socio-technical, it is crucial to also consider their societal implications (Selbst et al., 2019). Selbst et al. describe that this singular focus on technicalities creates a division between social and technical factors, which negatively impacts both the development and deployment of the socio-technical model. However, for other socio-technical systems, current research primarily emphasises societal components (Weidinger et al., 2023). Again, there is a notable gap in addressing both technical and societal elements necessary for ensuring that AI systems align with human values and societal norms (Zhang et al., 2024). Malone (2019) underlines the importance of research to both aspects, arguing that the most promising applications of AI will not replace humans but will instead involve humans and computers collaborating as "superminds" to accomplish cognitive and physical tasks that were previously impossible.

1.4.2. Human-GenAI

Human-GenAI can be considered as a socio-technical system as it addresses the interaction between humans and technology. Studies have shown that GenAI systems enable users to collaborate, receive assistance, take suggestions, and adjust recommendations (Shi et al., 2024). However, research on human-GenAI interactions often examines these aspects in isolation. Shi et al. describe, the need

for expanding research on human-GenAI interaction is essential to support advancements between humans and computers. The research of Manresa et al. (2024) explores the impact of GenAI on employee performance in the workplace, focusing on factors such as employee engagement, trust in GenAI, and attitudes toward its implementation. The study reveales that GenAI is still in its early stages of adoption within the companies surveyed, with more time needed to build confidence in its effectiveness.

1.4.3. GenAI on Client's and Contractor's side

In addition to the need for research combining social and technical perspectives, there is a notable gap in exploring GenAI as a tool for the contracting party in a tendering context. Much of the existing research about GenAI in tendering focuses on the client's side instead of the contractor's side, particularly in public procurement. Given that both contractors and clients operate within the same field, examining the potential of GenAI for contractors is a logical step. And given that contractors and consultants often collaboratively form the contracting party, this gap in research on the contractor's side also extends to consultants within infrastructure tendering.
Examples of research on AI implementation in public procurement include Berraida and El Abbadi (2024), who highlight areas where AI could improve procurement efficiency, and Rissanen (2024), who identifies GenAI as a usable solution for addressing inefficiencies by automating tasks, allowing employees to focus on more strategic work.

In conclusion, while GenAI has transformative potential, its rapid development requires thorough research to fully understand its impact. As a socio-technical system, GenAI connects human and technical elements, yet faces challenges in effectively researching both these components. Additionally, there is a gap in understanding GenAI's role for the contracting party in tendering processes, as existing studies focus on client's perspectives. Addressing these gaps aims to provide valuable insights into optimising GenAI's use, particularly in infrastructure tendering.

1.5. Research Questions

The research question of this thesis is designed to help address the problem outlined in the previous section, focusing on the deliberative interaction between humans and GenAI. The main research question is as follows:

"How do individuals experience human-generative Artificial Intelligence deliberation when reviewing open infrastructure tenders within project-based organisations?"

This research question aims to provide insights into individuals' experiences of human–GenAI deliberation in the context of reviewing infrastructure tenders. These experiences may encompass a range of aspects, including emotions, thoughts, perceptions, challenges, and expectations. Such elements reflect how individuals engage with the deliberation process, how they feel during the interaction, their reflections on the tool's input, the difficulties they encounter, and their expectations regarding the tool's usefulness and reliability.

The following sub-questions support the main research question and serve as a guide throughout this study:

1. "What factors should be considered when using human-generative Artificial Intelligence deliberation to review tenders?"

This sub-question explores the factors that should be considered when engaging in deliberation between humans and GenAI. Answering this question aims to provide guidance for the research by identifying key focus areas for further investigation on human-GenAI deliberation.

2. "How does human-generative Artificial Intelligence deliberation influence changes in individuals' opinions in the context of reviewing tenders?"

The second sub-question seeks to explore how human–GenAI deliberation influences an individual's opinion. The insights aim to deepen the understanding of how humans and GenAI interact in ways that may lead to shifts in perspective..

3. "What do individuals perceive as drivers and barriers to using human–generative Artificial Intelligence deliberation for reviewing tenders?"

The third sub-question aims to identify the drivers and barriers relevant to the future implementation of human–GenAI interaction in the context of reviewing tenders. It seeks to define the drivers, reasons to adopt GenAI in this setting, and the barriers, reasons for hesitation or rejection. The goal is to draw conclusions about how the specific characteristics of the tool influence its adoption.

1.6. Research Relevance

The findings of this thesis aim to offer insights to academia, bridging the gap between research and practical application. This is particularly relevant in the current era, where GenAI has the potential to significantly impact various industries and transform the capabilities of individuals, organisations, and society (Banh and Strobel, 2023). This potential remains, as AI technology is still in rapid progress (Kanbach et al., 2024). The complementary nature of human thinking and AI systems suggests that collaboration between the two can be more effective than either working independently (Feuerriegel et al., 2023). For this reason, the topic holds both academic and practical relevance. Academically, studying how humans can work with AI is essential, as there is still much to learn. Practically, optimising collaboration between humans and AI offers many opportunities to achieve greater effectiveness (Feuerriegel et al., 2023).

The practical relevance of this research also lies in its value to the collaborating company, Count & Cooper, which supported the development of this thesis. Throughout the research, people, knowledge, and other resources were made available to provide support. However, the conclusions drawn from this study are not only applicable to Count & Cooper. In the broader field of consultancy firms involved in infrastructure tendering, the findings of this thesis can be adapted and applied to enhance the implementation of GenAI as a tool for reviewing tenders.

1.7. Thesis Outline

Following this introduction (Chapter 1), the thesis begins with a literature review on the topics relevant to the research. This chapter, Chapter 2, concludes with a summary of the most important literature, which informs and shapes the rest of the research. The methodology is presented in Chapter 3, outlining the overall research structure and the reasoning behind the chosen approach. Next, the design of the experiment is described in Chapter 4, forming the foundation for the practical application of the theoretical insights. This chapter outlines the experiment, followed by a group discussion with the participants of the experiment. It also describes the expert evaluation sessions to assess and enrich the findings from both the experiment and the group discussion. Following, Chapter 5, presents the results derived from the experiment, the group discussion, and the expert evaluations. The discussion in Chapter 6 reviews the literature used in the study and considers both the validity and limitations of the research. Finally, the thesis concludes with Chapter 7, which addresses the research questions, outlines the contributions of the study and provides recommendations for both practical application and future research.

2

Literature Review

This chapter explores the key theoretical aspects relevant to this study. To conduct the review, search terms such as “Tendering strategy”, “Background Generative AI”, “Human–Generative AI interaction”, “Activity Theory AI”, “Deliberation AI”, and “Factors influencing AI use” were used. These terms were primarily searched via GoogleScholar (n.d.), and relevant literature was organised and accessed using Mendeley (n.d.). The list of factors influencing GenAI use was obtained from existing research. These factors have frequently been identified in studies exploring what might influence the use of GenAI, and so this list was created as part of this study.
The literature review starts with an elaboration on tendering in the construction industry, including the tender process and its associated interactions in Section 2.1. This is followed by an overview of the development of GenAI (Section 2.2) and the dynamics of human–GenAI interactions, described in Section 2.3. Next, in Section 2.4, the Activity Theory is introduced as a theoretical basis for investigating interactions between humans and technology. This is followed by a discussion of various influencing factors that affect the use of GenAI, written in Section 2.5. Following, an exploration of the Human–AI Deliberation framework, which has been applied in previous research on the dynamics of human–GenAI interactions, is done (Section 2.6). The section concludes with Section 2.7, in which the literature is described that form a basis for the design of the experiment.

2.1. Tendering

In the construction industry, two main types of tendering are commonly used: open tender and restricted tender. This master thesis primarily focuses on open tenders for the reason that open tenders are the most common in the Netherlands (for information on tendering, read Subsection 1.3.2). An open tendering approach allows any contractor, including new and unfamiliar ones, to enter the tender bidding (Ogunsemi and Aje, 2006). Ogunsemi and Aje states that through this transparent tendering process, equal opportunities are ensured, which leads to a highly competitive bidding process. As outlined in Section 1.1, this thesis focuses on the consultant, which, together with the contractor, forms the contracting party, as illustrated in Figure 2.1.

There are generally two main stages in the tendering process: the pre-tender stage and the tendering stage (explained in Subsection 1.3.3). Mohemad et al. (2010) describe the pre-tender stage as the phase during which the contracting party decides whether to participate in the bidding process. During this stage, the feasibility of the building project is evaluated (Oladokun et al., 2011). If the contracting party expresses interest and decides to proceed with bidding, the tendering stage begins. This involves receiving the tender documents and necessary information, preparing the bid, and submitting it before the deadline (Mohemad et al., 2010).
The focus of this thesis is on the perspective of the contracting party. However, because these parties have unique procedures and bidding strategies, which is often confidential information, there is limited information available publicly from the contractor’s side. To provide a broader understanding of the tendering process, an outline from the client’s perspective is presented below.

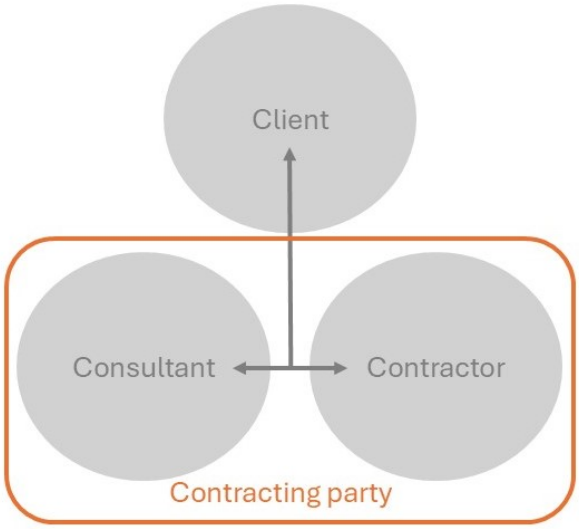


Figure 2.1: Interaction between parties in tendering (Created by the author)

2.1.1. Tendering from a Client's Perspective

All tenders are publicly presented online on the website TenderNed (n.d.). TenderNed is the tendering system of the Dutch government, consisting of two components: an announcement platform where public clients publish their tender announcements and an application that facilitates the complete digital handling of tenders (Rijksoverheid, n.d.).
The Aanbestedingswet 2012 presents the rules and regulations considering tendering from a client's perspective. In the Aanbestedingswet 2012, a client is referred to as the Dutch state, a province, a municipality, a water authority, or a public law institution, as well as a cooperative partnership of these governments or public law institutions (Overheid.nl, 2022). When the client is preparing and establishing a public contract, it sets only those requirements, conditions, and criteria for bidders and tenders that are reasonably proportionate to the subject matter of the contract. According to the Aanbestedingswet 2012, the client follows several steps when applying an open procedure. First, the client announces the public contract. Then, it checks whether a bidder falls under any disqualification criteria set by the client. For bidders who are not disqualified, the client assesses whether they meet the established suitability requirements. Following, the Aanbestedingswet 2012 examines whether the bids comply with the technical specifications, requirements, and standards set out by the client. Valid tenders are then evaluated based on the award criteria specified by the client. These criteria are used to determine the most economically advantageous tender, based on the best price–quality ratio (elaborated on in Section 1.1). Finally, the client prepares a report of the awarding process, communicates the award decision to the relevant parties, finalises the agreement if applicable, and announces the awarded contract (Overheid.nl, 2022).

Rijkswaterstaat as a Client

Rijkswaterstaat serves as a public commissioning authority and can be used as an example for the explanation of a "general" public client (Rijkswaterstaat, 2024b). Figures Table B.1, Table B.2, and Table B.3 in Appendix B show examples of a tender planning from Rijkswaterstaat. These planning examples are included in the tender guidelines of a tender project. The planning examples outline three main stages: the announcement, application phase, and evaluation phase. In the announcement, the tender is published along with the necessary documents to enable the preparation of the bid. In a Dutch tender, these tender documents must be requested by the contracting party no later than six days before the final submission deadline (Staatscourant, 2020).
Through TenderNed, contracting parties can request clarification of the tender documents (Rijkswaterstaat, 2024b). Questions must be clearly formulated by the contractor, and the client is required to respond within a short timeframe. This phase, in which questions can be answered, is limited and differs per project. Additionally, in some cases, it is possible to request one or more meetings to ask questions orally. Such requests must also be submitted within the specific time frame outlined in the

tender documents. The request should include a clear explanation of the questions to be asked and a justification of why answering these questions is directly relevant to the economic advantages of the tender. Following, the client prepares a clarification document, including additional information based on the questions submitted. In Dutch, this clarification document is referred to as the “nota van inlichtingen inschrijffase”. The clarification document will be published shortly after ending the phase in which clarification can be requested.
Finally, the application phase concludes with the deadline for submitting the bid. In addition to submitting the bid, the contractor must meet several conditions to be authorised for their documents to be evaluated. These conditions are outlined based on the tender guidelines for a project by (Rijkswaterstaat, 2024b). For example, the submission must be written in Dutch and exclusively uploaded via TenderNed. Additionally, the submission must be authorised, meaning for example that it must be submitted by an individual with the necessary authority within the contractor’s organisation. Furthermore, collaborations between companies are permitted, allowing them to form a contracting party and submit a bid jointly.

All of the above is written from the client’s perspective, whether based on legal regulations or the specific rules of a given client. However, this thesis focuses on the contractor. To prepare a strong bid, the contractor must, of course, consider the client’s perspective, particularly since the client is the party evaluating the submission. At the same time, within the bidding framework outlined by the client, contractors develop their own methods of working to meet these requirements. These methods of working, unique to contracting parties, will be explored through a case to gain insights into the contractor’s side of the tendering process.

2.2. Development towards Generative AI

The development towards GenAI began with the foundations laid by traditional AI. Figure 2.2 represents the progression from traditional AI towards GenAI. Banh and Strobel (2023) note that AI broadly refers to approaches that perform tasks requiring human intelligence, including natural language understanding, pattern recognition, and decision-making. Initial AI models were rule-based, focused on supporting decisions, but advancements in machine learning (ML) enabled algorithms to learn autonomously from data (Banh and Strobel, 2023). The development of deep learning (DL), a subset of ML, facilitated the way for more advanced models, particularly deep generative models (DGMs), which create new content by learning complex data distributions. The shift toward GenAI opened new possibilities, as generative models leverage high-dimensional probability distributions to generate content that closely resembles real-world data.

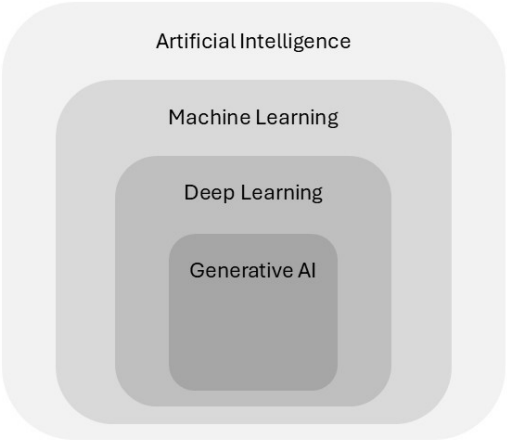


Figure 2.2: AI Concepts (Adapted from Banh and Strobel, 2023)

A GenAI model can be considered incomplete on its own, as it depends on learned patterns and knowledge that must be refined and adapted through user input and adjustments (Feuerriegel et al., 2023). Interactions between humans and AI play a crucial role in shaping an effective model, yet each interac-

tion is unique, leaving much to discover about how to engage with GenAI effectively. The capabilities of GenAI remain uncertain to both designers and users of the system (Shi et al., 2024), and understanding how to utilise this new technology is essential for unlocking its full potential.

2.3. Human-Generative AI Interaction

GenAI, as a new tool for human interaction, introduces uncertainties about how to effectively integrate such technologies into our digital society that is increasing. Despite these unknowns, GenAI is not the first computing tool that individuals must adapt to and learn to use effectively. This challenge falls within the domain of Human-Computer Interaction (HCI). HCI is an interdisciplinary field focused on designing, evaluating, and implementing computing systems for human use, as well as studying related phenomena (Hewett et al., 1995). Its core emphasis lies on the interaction between one or more humans and one or more machines. As Kocsis (2019) explains, integrating human and computer activities is crucial for organisations to fully implement computer systems. Effective HCI can significantly increase individuals' operational capabilities when using these systems (Mohammed and Karagozlu, 2021).
Also, for AI as a computer system, Fügener et al. (2022) explain that the diversity of human thinking allows humans to possess knowledge that complements AI systems. This complementarity suggests that collaboration between humans and AI can be more effective than either working alone, particularly when solving problems. Similarly, Memmert and Bittner (2022) highlight the advantages of combining humans and AI, as their respective strengths and weaknesses align to create a beneficial relationship. Together, they can achieve outcomes that would not be possible individually.
In human-GenAI interactions, human feedback plays an important role in improving GenAI products. Users can respond to the model's outputs, and this feedback is then used to further train and improve the model (Nah et al., 2023). GenAI has the capability to adapt to users' needs and intentions while creating human-generated content. For instance, in ChatGPT, users rate the responses provided by the system, and these ratings are used to improve its performance (Ouyang et al., 2022). Therefore, GenAI heavily relies on human interaction to continuously evolve and improve its effectiveness.

For this thesis, activities will be examined within the context of interactions between a human and GenAI. In the scenario where a contractor and a consultant collaborate on a project assigned by a client, all three parties have the potential to engage with a GenAI tool (illustrated in Figure 2.3). However, communication between these parties typically involves human-to-human interactions or group discussions (discussed in Section 1.1). Despite this, individuals within each party can interact independently with a GenAI tool. This research focuses specifically on one form of interaction: the interaction between a single human and GenAI.

As human feedback plays a critical role in improving GenAI products (Nah et al., 2023), making it essential to consider both the user's and the GenAI system's perspectives (Shi et al., 2024). Shi et al. (2024) examined 291 papers to elaborate on human-GenAI interactions, focussing on key dynamics such as control from humans to models, feedback from models to humans, and the varying levels of engagement. Additionally, Zhu et al. (2024) categorise human-GenAI collaboration into four types: situations where humans lead, where AI leads, where contributions are evenly shared, and where leadership is uncertain. Their findings indicate that 77.21% of participants perceived collaboration as human-led or even contributions, while 15.19% viewed it as AI-led. For this thesis, the focus will be on two specific collaboration types: human-led and AI-led interactions. These will be analysed using the concepts of control from humans to models and feedback from models to humans.
The feedback provided by a GenAI model to humans can vary in how the model generates its output, the methods it uses to produce answers, and the way the output is synchronised (Shi et al., 2024). Output formats can include text, audio, numerical data, and visuals such as 2D, 3D, or layout designs. Another perspective is the control of humans over the models, such as how users can influence models, focussing on methods for improving output, the specific objects or elements that can be controlled, and the mediums through which this control is exercised (Shi et al., 2024).
Identifying the type of interaction with an AI tool is essential for doing a study on the use of GenAI and understanding how to employ the tool effectively for its intended purpose.

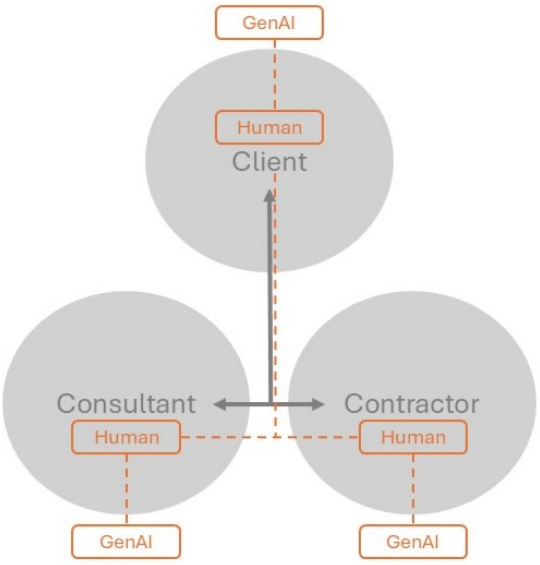


Figure 2.3: Interaction scheme (Created by the author)

2.4. Activity Theory

Section 2.3 showed that integrating human and computer activities is key for organisations to maximise the benefits of computer systems (Kocsis, 2019). Effective HCI can significantly enhance users' operational capabilities (Mohammed and Karagozlu, 2021), and therefore, investigating these interactions is key.

Using the Activity Theory framework to measure HCI as a foundation is valuable, as it supports a multi-level examination of human-technology interactions at individual, organisational, and societal levels (Nah et al., 2023). This approach aligns with the field of HCI shift towards understanding activities in a comprehensive, contextualised manner. Similar to HCI, the Activity Theory does not view activities as isolated or independent tasks, instead, they are seen as interconnected (Kuurti, 1995). This interconnectedness implies that changes or disruptions in one part of the system can impact other activities, leading to potential problems. Kuurti (1995) refers to these disruptions as contradictions within the Activity Theory. Although these contradictions may cause challenges, the Activity Theory considers these challenges as essential sources for growth and change. This makes researching the contradictions in the Activity Theory even more important. Crawford and Hasan (2006) highlight additional reasons for using the Activity Theory, emphasising its value as a framework for understanding and analysing across various domains within information systems research and practice. The Activity Theory accounts for the interaction between subjective, creative, and emotionally driven human behaviour and technological activities, offering a socio-technical perspective. Additionally, it adopts a holistic approach to human experience, assisting a deeper understanding of human factors (Crawford and Hasan, 2006).

Figure 2.4 illustrates the structure of an activity. It shows different triangles, based on the Activity Triangle Model of Engeström (1987). An activity can be described as a systemic whole, where all elements are interconnected (Kuurti, 1995). The initial triangle, consisting of the top three elements in Figure 2.4 (subject, object, and tool) is the original model introduced by Vygotsky (1978). This model later served as the basis for Engeström's Activity Triangle Model. The relationship between the subject, which could be an individual, and the object, representing the goal or purpose of the activity, can be stimulated by a tool (Kuurti, 1995). Tools are the instruments used by the subject to perform an activity, enabling the subject to transform objects into outcomes (Good and Omisade, 2019). Another triangle within Figure 2.4 connects the subject, rules, and community. Kuurti (1995) describes that the connection between the subject and the community, which refers to the group or social system within which the activity occurs, can be stimulated by rules, which may include laws or organisational policies. Lastly, the relationship between the object and the community is facilitated by the division of labour, which outlines how tasks are distributed to achieve the desired outcomes from the transformation process. This

activity framework, also described as “what people do”, reflects through actions as individuals interact with their environment (Good and Omisade, 2019).

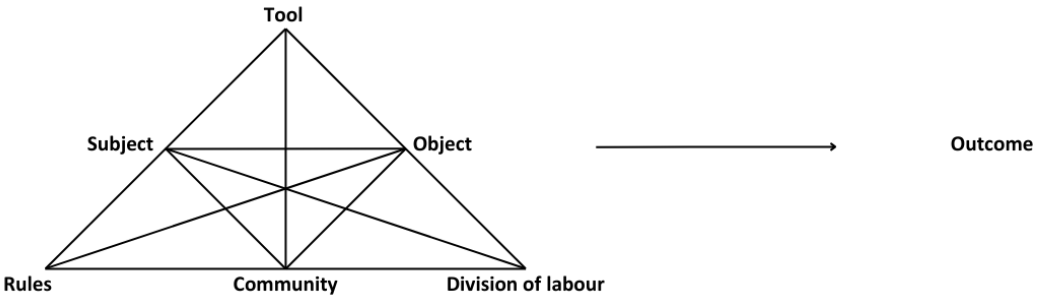


Figure 2.4: Basic structure of an activity (Adapted from Engeström, 1987)

According to Kuurti (1995), activities cannot achieve outcomes at once, as they require a process of several steps. This means an activity can be divided into different levels, as shown in Figure 2.5. The figure shows that activities are carried out through actions, which can be further divided into smaller tasks known as operations. Kuurti describes that an activity is driven by a motive and is completed through actions aimed at achieving a specific goal. The research further elaborates that actions are executed through operations, which consist of smaller, practical steps that support the action. While actions occur at the conscious level, operations are routine behaviours that happen subconsciously (Kuurti, 1995). Kuurti states that in order for these routines to occur, certain conditions must be present to support the action. If conditions change, an operation may “unfold” and, according to Kuurti, return to the conscious stage of the action. This means that when a necessary condition is absent, the action may be hindered or blocked from being successfully completed.

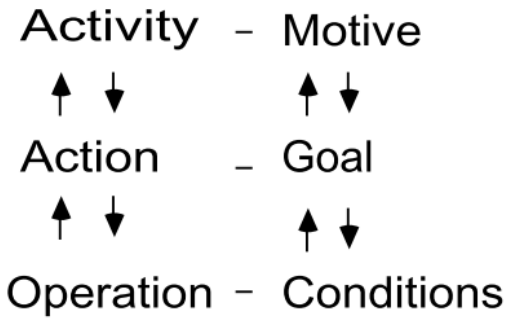


Figure 2.5: Hierarchical levels of an activity (Kuurti, 1995)

2.4.1. Contradictions

Given that Activity Theory views contradictions not only as potential sources of challenges but also as essential drivers of development and change (Kuurti, 1995), it is important to take them into account. Within an activity system, four levels of contradictions can be identified: primary, secondary, tertiary, and quaternary contradictions (Engeström, 1987). According to Engeström, primary contradictions occur within a single element of the activity system, for example, a conflict within the “tool” element. Secondary contradictions arise between different components of the same activity system, such as between the subject and the tool (Engeström, 1987). Engeström describes tertiary contradictions to emerge when a more advanced version of the activity system is introduced, resulting in tensions between the new model and the original one. Finally, as Engeström explains, quaternary contradictions occur when neighbouring activity systems, each with their own rules, communities, and objectives, come into contact. While these systems operate separately, their interaction can cause contradictions due to their mutual influence. It is essential to address these contradictions, as unresolved tensions can hinder the achievement of the outcome (Engeström, 1987).

In Figure 2.6, these potential contradictions are visually represented by dotted arrows. Arrows between two different components indicate secondary contradictions, while those within individual components reflect primary contradictions.

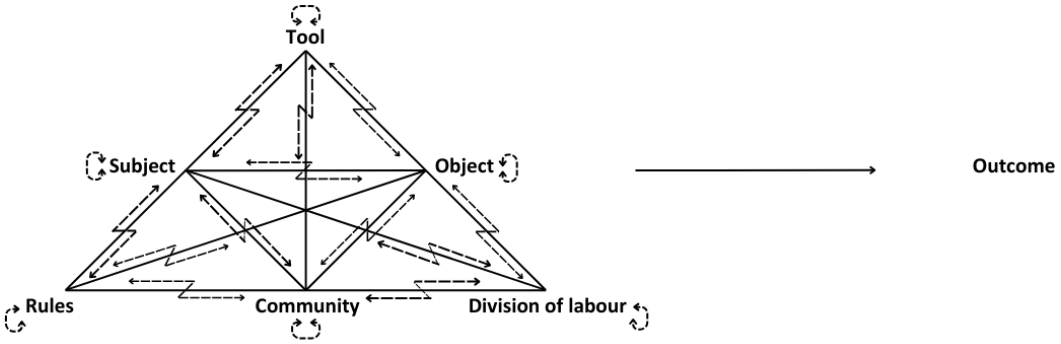


Figure 2.6: Primary and secondary contradictions in the basic structure of an activity (Adapted from Engeström, 1987)

As contradictions must be addressed within an activity system to achieve its intended outcome (Engeström, 1987), it is important to recognise their dual nature. Contradictions can be interpreted both negatively and positively within the context of an activity system. On the one hand, contradictions are often described as disruptions or conflicts that hinder progress and obstruct the achievement of outcomes (Kuurti, 1995). From this perspective, they represent obstacles and are therefore negative. On the other hand, contradictions can represent challenges within the system that can enable a focus on transformation efforts (Engeström et al., 1996). In this light, Kerosuo (2006) adds that contradictions function as drivers of development, learning, and innovation. Therefore, contradictions should be recognised as both negative and positive influences, viewing them not only as challenges to overcome but also as opportunities for growth and improvement within the activity system.

2.5. Factors Influencing GenAI Use

To study an activity system through the lens of the Activity Theory, it is useful to examine the factors that may influence the execution of the activity. This section explores various factors that could affect the use of GenAI, beginning with the factor of reliance, followed by trust, explainability, and transparency. It then addresses privacy, scepticism, and company culture, and concludes with ease of use and sustainability.

2.5.1. Reliance

Reliance is described by de Fine Licht and Brülde (2021) as a concept where agent A depends on agent B to perform action C because A believes that B will actually make C happen. Reliance can apply both to people and to objects (de Fine Licht and Brülde, 2021). Kakolu et al. (2024) further adds that reliability also concerns a system's ability to operate stably over a longer period without the need for constant bug fixes. Thus, if a user believes the model performs well and is helpful, their reliance on it increases.

However, when it comes to reliance, individuals can both over-rely and under-rely on an AI system. As Gao et al. (2023) highlight, there are potential pitfalls associated with both tendencies. A study by Klingbeil et al. (2024) found that people often over-rely on an AI, even when it contradicts their understanding of the situation. This over-reliance arises when users accept the recommendations of an AI without question (Zhai et al., 2024). Such over-reliance can lead to outcomes that may lack depth (Gao et al., 2023). Over-reliance often occurs when individuals increasingly favour solutions that are fast and seemingly optimal, choosing them for practical reasons over slower alternatives (Gao et al., 2023). Zhai et al. (2024) support this by noting that people often prefer efficient cognitive shortcuts.

On the other hand, Gao et al. (2023) state that under-reliance may prevent users from fully engaging with or utilising the capabilities of an AI system and therefore might hinder users from fully benefiting from it.

2.5.2. Trust

Trust, compared to reliance, is a more complex concept (de Fine Licht and Brülde, 2021). As explained by de Fine Licht and Brülde, trust occurs when agent A relies on agent B to achieve C because A believes that B feels a moral responsibility to do C. Therefore, to trust something or someone, all aspects of reliance must be met, with the addition of a moral obligation (de Fine Licht and Brülde, 2021). The reliability of an AI system plays a crucial role in building user trust, as users are more likely to trust a system that performs consistently and effectively across different scenarios without any loss of functionality or performance (Tahmasbi et al., 2025). Furthermore, Grodzinsky et al. (2010) explain that trust involves risk: the less information the agent has about the object to be trusted, the higher the level of trust required.

2.5.3. Explainability and Transparency

Due to the complex data distributions in GenAI, it can be challenging to trace specific outputs back to their sources (Banh and Strobel, 2023). This makes it harder for users to understand how the AI generates its outputs (Tahmasbi et al., 2025). As a result, explainability and transparency are crucial for building trust with users. Shin (2021) notes that most AI users have limited knowledge or visibility into how decisions are made by the model, making explainability a critical factor in creating trust. This concept aligns with transparency, which Kakolu et al. (2024) define as the extent to which users are informed about the decision-making process of the GenAI system. Transparent AI models that clearly explain their reasoning can significantly enhance user trust in the model's capabilities and accuracy (Alonso and Puente, 2018).

2.5.4. Privacy

Privacy is one of the main ethical considerations associated with the use of GenAI (Al-kfairy et al., 2024). This concern comes from a general lack of trust in GenAI, as Al-kfairy et al. (2024) highlight the constant threat of privacy breaches. Such breaches pose risks to individuals and institutions, potentially damaging their identity or reputation. To fully utilise GenAI, users must trust that their personal and organisational information will remain safe during their interactions with the system (Tahmasbi et al., 2025). However, Al-kfairy et al. (2024) argue that addressing these privacy challenges cannot rest solely on the shoulders of individual organisations. Instead, they support rules and regulations to take the lead in upholding the core values associated with the responsible use of GenAI.

2.5.5. Scepticism

Scepticism is defined as the doubt that something is true or useful (CambridgeDictionary, n.d.) and can significantly influence how users interact with a technology. When users are sceptical, it often limits users' willingness to adopt or trust the system (Adepoju et al., 2024). To mitigate such scepticism, Adepoju et al. note that it is essential that the technology is developed and implemented in a way that aligns with the needs. Several background factors can influence the level of scepticism towards tools like GenAI. For example, generational differences, cultural background and gender have been identified as important factors that shape users' scepticism (Adepoju et al., 2024).

2.5.6. Company culture

Organisational climate plays an important role, as a climate that promotes innovation, learning, and related values is more likely to encourage employees to adopt and make use of tools (Venkatesh, 2022).This is confirmed by Chaudhuri et al. (2024), who found that an organisational data-driven culture significantly influences product innovation and process improvement.

2.5.7. Ease of Use

Ease of use refers to how much a person believes that using a particular system will be free of effort (Davis, 1989). "Ease" is defined as being not too difficult or not needing too much effort (Davis, 1989). When technology is easy to learn and user-friendly, it is more likely that users will adopt it and continue using it (Topsakal, 2024). Therefore, the ease of use of a technology such as GenAI is a key factor in the adoption of the technology.

2.5.8. Sustainability

Sustainability may be a factor influencing the use of GenAI, as carbon emissions are generated during both the training and usage of GenAI models (Chien et al., 2023). As these models continue to grow in complexity, and as the number of users and requests per user increases, the environmental impact associated with GenAI also rises (Chien et al., 2023). In the field of AI, the focus on achieving the best possible results often overshadows other considerations, such as the environmental impacts of AI development (Korolev and Mitrofanov, 2023). Deep learning, a form of AI that facilitates content generation by learning from data (Banh and Strobel, 2023), has been rapidly developing (Thompson et al., 2021). This means the training compute required for training models is been doubling every 8 to 17 months since 2015 (Korolev and Mitrofanov, 2023). Training compute refers to the number of operations needed to train an AI model, and has increased significantly due to the demand for new models to beat existing models. This growth is influenced by the number of GPUs (Graphical Processing Units) used during training (Lacoste et al., 2019). Lacoste et al. (2019) explain that beating existing models often involves training on more GPUs, using larger datasets, and running computations for larger periods. As training requirements grow, so does the energy consumption needed to power these models. Schwartz et al. (2020) describe this trend as Red AI, which prioritises improving accuracy through high computational power while neglecting the associated environmental costs, such as increased carbon emissions. In contrast, they propose Green AI, which promotes approaches that do trade-offs between the favourable performance and efficiency. Efficiency in this context refers to the amount of computational work required to train a model (Schwartz et al., 2020). This computational work contributes to carbon emissions through the use of electricity. The level of emissions depends on several factors, including the time required to run the parameters for generating an AI model and the number of parameters involved (Schwartz et al., 2020). Additionally, the geographic region where the model is trained and the time of day also play a significant role (Dodge et al., 2022). Carbon emissions are correlated to the amount of electricity consumed, which is highly influenced by the local electricity infrastructure (Schwartz et al., 2020).

The amount of carbon emitted by AI usage is influenced by multiple factors. While users may not have the ability to select the geographic region where AI is processed, they can make choices about the type of AI model they use. For instance, the computational power required by a model significantly impacts electricity consumption, and therefore, the carbon emissions associated with its use. To promote awareness of these differences, several researchers have developed tools to help users understand the environmental impact of various AI models. For example, Lacoste et al. (2019) introduced the Machine Learning Emissions Calculator, which considers key details involved in training a Machine Learning model, such as the geographical location of the server, training duration, and type of GPU used. By combining this information, it estimates the approximate amount of CO2-equivalent emissions produced. Beyond calculating emissions, the Machine Learning Emissions Calculator highlights the factors contributing to carbon output and offers actionable recommendations for individuals and organisations to minimise their environmental impact.

However, the environmental impact of GenAI is not solely negative. GenAI can play an important role in promoting sustainability by improving forecasting and optimisation for renewable energy sources (Arunkumar et al., 2024). For instance, ChatGPT can assist with data analysis, modelling, and forecasting of solar energy, helping to optimise solar panel placement and predict energy production (Rane, 2024). Additionally, ChatGPT has been shown to support wind energy management by aiding in wind turbine control, maintenance scheduling, and wind speed predictions (Rane, 2024).

As both the carbon footprint from the training and use of GenAI models has been growing, with expectations that this trend will continue (Schwartz et al., 2020), the environmental impacts cannot be overlooked in this study.

2.5.9. Conclusion Factors Influencing GenAI Use

In summary, multiple factors influence how users engage with GenAI, with trust emerging as a central theme. This trust depends on conditions such as the reliability and explainability of GenAI outputs, as well as the ease of use of the system and the user's sceptical mindset. Beyond these considerations, other factors such as privacy concerns, organisational culture and environmental impact also shape user attitudes. Together, these elements form a set of factors that need to be considered when evaluating the human-GenAI interaction.

2.6. Human-AI Deliberation Framework

Ma et al. (2024) developed a novel framework, the Human-AI Deliberation framework, to promote human reflection and discussion in scenarios where humans and AI conflict in opinions during decision-making. In multiple previous studies, the study of Ma et al. (2024) found that individuals often either over-rely on incorrect AI suggestions or dismiss correct ones. The study also highlights that people's analytical thinking is not sufficiently activated when interacting with AI systems. Therefore, the research of Ma et al. (2024) introduces an AI system that can identify different viewpoints, engage in deliberation, and adapt its suggestions during discussions. Deliberation, as described in Section 1.3, is understood as the process of engaging in discussions and critically assessing information to deepen understanding and shape values (CambridgeDictionary, 2025; Kenter et al., 2016), not with the goal of reaching agreement, but to broaden their own viewpoints, leading to more informed and well elaborated decisions (Gracia, 2003). Deliberation in AI systems goes beyond traditional AI-assisted decision-making, which typically involves humans agreeing or disagreeing with AI suggestions without sufficient reasoning from the system (Ma et al., 2024). Ma et al. describes that this lack of reasoning makes it challenging for humans to determine whether and to what extent they should adopt the AI's recommendation. The framework for Human-AI Deliberation is designed to promote deliberative processes involving both humans and AI, illustrated in Figure 2.7. In this figure, the left part (A) represents how the AI model makes decisions from a set of alternatives, and the right part (B) focuses on the deliberation between the human and AI. Part A presents a decision-making process. According to Ma et al. (2024), a general decision-making process includes several steps: identifying the problem, collecting information, identifying alternatives, weighing evidence, making a decision, implementing an action, and evaluating the result. Among these steps, weighing evidence is a critical component of the framework. The concept of the weight of evidence is widely applied across numerous decision-making domains (Alvarez-Melis et al., 2021). Ma et al. (2024) define the weight of evidence as the most important to a final decision based on one's interpretation. AI systems perform this step independently, determining whether to present the evaluated information to the user.

The framework proposed by Ma et al. (2024) contains four interconnected activities, illustrated in Figure 2.7, explained as follows. The first activity, "elicitation of thoughts", involves both the AI system and the human presenting their reasoning. During this step, the AI explains its rationale for its decisions, while the human is encouraged to clarify their reasoning and ideas. The second activity, "alignment of human-AI thoughts", focuses on highlighting potential differences between the user's and the AI system's ideas. In this stage, the weight of evidence plays a significant role, as assessing and presenting the evidence can guide the human in navigating these differences. The third activity, "discussion", includes the exchange of opinions through communication. This involves further clarification of the decisions made, including the evidence and the rationale behind the weighting of evidence. The success of this discussion depends on the input provided by both parties and factors such as who leads the conversation. The final activity, "update of thoughts", focuses on revising the perspectives of both the human and the AI based on the discussion. For the AI, this step serves as feedback, improving its mechanisms to assist users more effectively in the future. For humans, this process can refine their evaluation of evidence, enhance their decision-making, and improve their interaction with the AI system by updating their understanding of its interface.

The Human-AI Deliberation framework, with its four interlinked activities, serves as an effective tool for examining interactions between humans and AI in decision-making processes. Understanding this interaction could help organisations identify strategies to enable their employees to optimise their engagement with AI systems, ultimately enhancing productivity and decision-making outcomes.

In the study by Ma et al. (2024), an exploratory user study was done to understand the impact of the Human-AI Deliberation framework on AI-assisted decision-making. The study used a graduate admission task as an illustrative example to explore the framework's potential effects. A custom deliberative AI system was created and compared to traditional, explainable AI and human decision-making without AI assistance. With the deliberative AI, participants followed the Human-AI Deliberation framework, sharing their thoughts before receiving AI recommendations. In contrast, the explainable AI system provided its ideas before the participants shared their opinions, followed by a view on AI's recommendations, and then made a final decision. The human-alone condition involved participants making decisions independently without AI input.

The study's primary task involved participants forming opinions on four admission cases, followed by

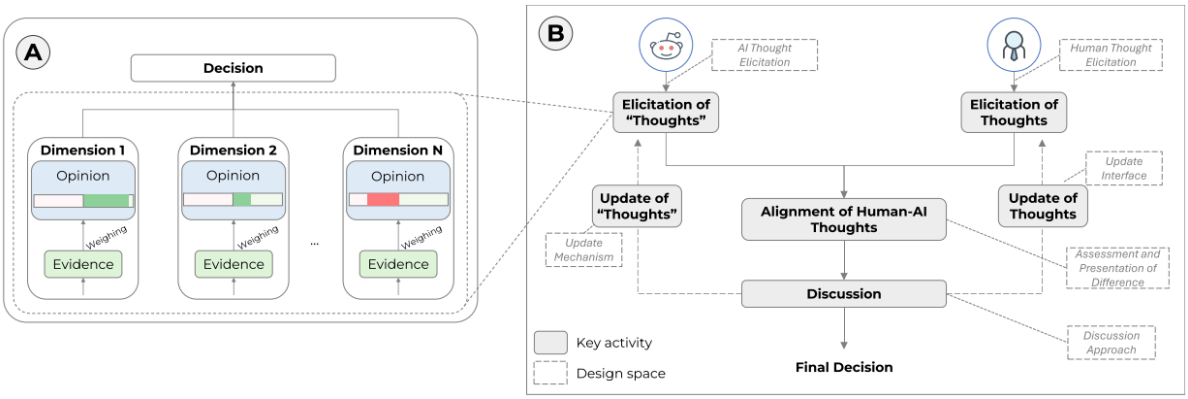


Figure 2.7: The framework for Human-AI deliberation (Ma et al., 2024)

receiving AI feedback, engaging in discussions with the AI, and finalising their decisions. This process allowed the researchers to evaluate changes in participants' opinions. In addition, the participants' experiences and feedback were collected to gain further insights. The experiment on Human-AI deliberation was evaluated across six areas: task performance, reliance, behaviours, perceptions of AI, user experience, and open-ended feedback. The first three aspects were measured objectively, while the other three focused on subjective experiences.

The work of Ma et al. (2024) offers a key conceptual basis for this study, presenting a framework that encourages deliberation between humans and AI models. Unlike traditional AI systems that operate independently up to the point of decision-making, their framework integrates the human into the process, enabling more thorough consideration of evidence. This approach highlights the importance of reflective, evidence-based reasoning in making decisions.

2.7. Literature Shaping the Experiment Design

In this section, the key insights from the literature that inform the design of the experiment are outlined, providing the foundation for the remainder of the study.

2.7.1. Activity Theory

The Activity Theory provides the theoretical basis for the qualitative research for this study. Therefore, it is essential to apply the Activity Theory to the specific context of this research. Figure 2.8 presents the activity system, including the components associated with each of the elements of the Activity Theory. The activity system is outlined to represent the interaction between a human and GenAI during the review of a tender, within the context of infrastructure tenders. Each element of the Activity system of Figure 2.8 is described below.

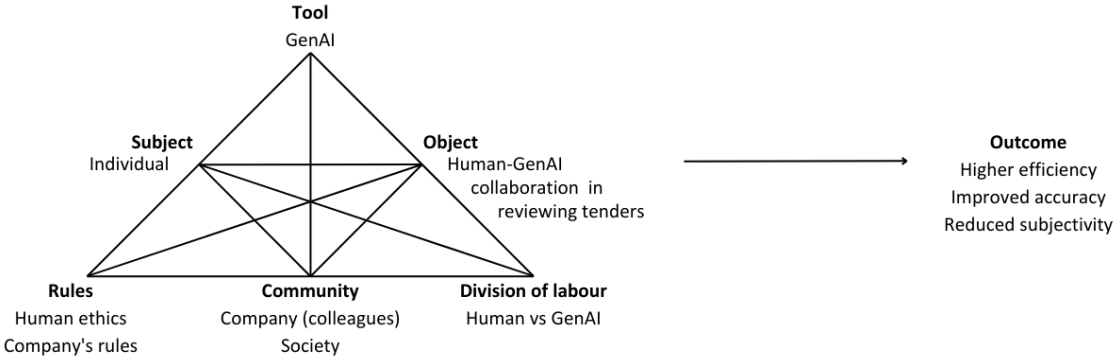


Figure 2.8: Activity system human-GenAI deliberation (Adapted from Engeström, 1987)

Subject
The subject of the activity refers to the individual or group performing the activity. In this case, focusing on the interaction between a single human and GenAI, the subject is the human.

Object
The object of the activity is its goal or purpose, meaning what drives the activity. In this context, the driving aim is the implementation and improvement of Human–GenAI collaboration in reviewing tenders. This object reflects the broader objective of enhancing the use of GenAI across various sectors, including infrastructure tendering.

Tool
The tool or instrument mediating the interaction is GenAI.

Outcome
The outcome refers to the expected results once the object has been achieved through the activity system. These outcomes include increased efficiency in reviewing tenders, improved review accuracy through the assistance of the GenAI tool, and reduced subjectivity. These benefits align with Count & Cooper's initial motivation for introducing a GenAI tool into the tendering workflow.

Rules
Rules include the norms, laws, and policies that guide the interaction within the activity system. These include general regulations and laws, ethical considerations influencing whether or not to use the tool, and company-specific policies that may encourage or regulate the use of GenAI by employees.

Community
The community involves other individuals or groups engaged in the activity. In other words, the community is the social context within which the subject operates. In this case, it includes colleagues, managers, and others within the organisation involved in infrastructure tendering. Additionally, it includes the broader societal context: both the infrastructure tendering community and the wider public that engages with or holds opinions about the use of GenAI.

Division of Labour
The division of labour refers to how tasks are distributed to achieve the objective. In this case, focusing on the individual interaction between a human and GenAI, it concerns the division of tasks between the human and the GenAI tool.

2.7.2. Contradictions
A key concept within Activity Theory is that of contradictions, which play a crucial role in shaping outcomes. These contradictions influence the results of actions and are therefore fundamental to fostering growth and transformation within activity systems (Kuurti, 1995). Therefore, it is essential to identify such contradictions. In the context of this thesis exploring the interaction between humans and GenAI in infrastructure tendering, both primary and secondary contradictions are involved. Primary contradictions are internal tensions within a single element of the activity system, whereas secondary contradictions arise between different elements within the same system (Engeström, 1987). These types of contradictions are illustrated in Figure 2.6. Recognising these contradictions within the activity system is crucial to increasing the likelihood of successfully achieving the desired outcome (Engeström, 1987). Therefore, the concept of contradictions is important to consider when analysing the data resulting from this study.

2.7.3. Drivers and Barriers
As contradictions must be addressed to achieve an activity system's intended outcome (Engeström, 1987), this study categorises them as either drivers or barriers to make them more tangible. While these specific terms are not directly used in the literature, they are grounded in existing theory. Contradictions are often described as disruptions that hinder progress (Kuurti, 1995), referred to here as barriers. However, contradictions can also facilitate transformation and development (Engeström et al., 1996), which justifies their framing as drivers when they promote learning and change (Kerosuo, 2006).

2.7.4. Factors Influencing GenAI Use

The literature review in Section 2.5 outlines several factors that may influence the use of a GenAI tool. These relevant factors are summarised in Table 2.1, with a particular focus on trust, which plays a central role. Trust is shaped by factors such as the system's reliability, explainability, and ease of use, as well as the user's level of scepticism. The factors listed in Table 2.1 should be carefully considered when applying human–GenAI deliberation in the context of reviewing tenders and are therefore essential to integrate throughout the remainder of the study.

Factor	Relevant Literature
Reliance	de Fine Licht and Brülde (2021) and Kakolu et al. (2024)
Over-reliance	Gao et al. (2023), Klingbeil et al. (2024), and Zhai et al. (2024)
Under-reliance	Gao et al. (2023)
Trust	de Fine Licht and Brülde (2021), Grodzinsky et al. (2010), and Tahmasbi et al. (2025)
Explainability	Shin (2021)
Transparency	Alonso and Puente (2018) and Kakolu et al. (2024)
Privacy	Al-kfairy et al. (2024) and Tahmasbi et al. (2025)
Scepticism	Adepoju et al. (2024)
Company Culture	Chaudhuri et al. (2024) and Venkatesh (2022)
Ease of Use	Topsakal (2024)
Sustainability	Chien et al. (2023) and Schwartz et al. (2020)

Table 2.1: Factors in Human–GenAI Deliberation (Adapted by author; see references in table)

2.7.5. Human-AI Deliberation Framework

The study by Ma et al. (2024) provides another relevant foundation for this experiment, introducing a framework designed to bridge the gap between humans and AI models. Their framework focuses on deliberation, which supports a greater weighing of evidence. Typically, AI systems perform steps leading up to decision-making independently, however, through deliberation, the human is actively involved in the AI's decision-making process, thereby supporting the consideration of the evidence (Ma et al., 2024). The study emphasises that the careful weighing of evidence is crucial for making well-founded decisions.

As a result, the human–AI deliberation framework presented in Figure 2.7 forms an important basis for the remainder of the study. All the steps outlined in the figure, the elicitation of thoughts, the alignment of human–AI reasoning, the discussion, and the updating of thoughts, are important steps be incorporated into the design of this study as they provide clear steps for the human-GenAI interaction.

2.8. Conclusion

In conclusion, the literature review chapter has established a foundational basis for this research by offering background information on tendering processes, GenAI, and the specific focus of this study: human–GenAI interaction. It introduced the Activity Theory as a guiding analytical lens for understanding the dynamics within activity systems. The concept of contradictions within the Activity Theory is positioned as a key perspective for examining contradictions, either as drivers or barriers to GenAI use. Furthermore, the review has identified various factors that influence the adoption and application of GenAI, which will inform the analytical focus of future empirical work. Lastly, the Human–AI Deliberation framework offers a structural perspective that may guide the design and interpretation of the research as it develops.

3

Methodology

This chapter presents the research methodology of the study. The first section provides an overview of the research design in Section 3.1. Next, Section 3.2 describes the methods used, followed by a discussion of the types of data and data collection techniques in Section 3.3. The analytical techniques applied are outlined in Section 3.4. Finally, Section 3.5 outlines the intended outcomes of the study.

3.1. Research Methodology

To provide an overview of the structure of the study, the research methodology is illustrated in Figure 3.1. The process begins with a literature review to build a foundational understanding, followed by the development of the experiment design framework to guide both the experiment, group discussion, and evaluation sessions. These form the basis of the data collection. The findings are analysed thematically to draw conclusions and develop an evaluation framework that can be used in similar contexts. An iterative approach ensures ongoing alignment with the study's objectives. All aspects of the research methodology are described below.

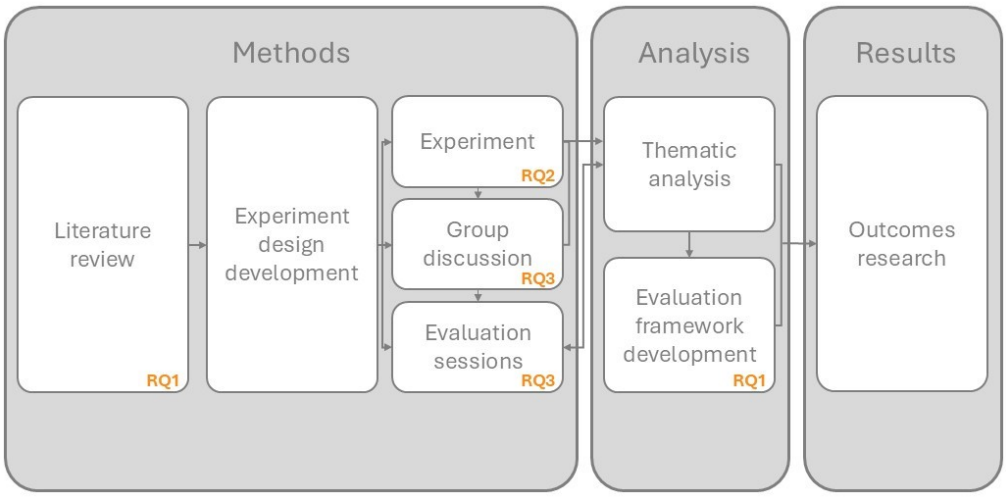


Figure 3.1: Research Methodology (Created by the author)

3.2. Methods

This section on methods includes the literature review, the design of the experiment, and the implementation of an experiment, a group discussion, and evaluation sessions, as shown in Figure 3.1.

3.2.1. Literature Review

A literature review serves as the foundation for all types of research, providing essential understanding across academic disciplines (Snyder, 2019). The purpose of this literature review has been to provide a knowledge base and to identify relevant practices and a theoretical framework that can be applied to research human–GenAI interactions in infrastructure tendering. It has also provided the basis for contextualising the experiment design and has helped to answer the first research question introduced in Section 1.5: *"What factors should be considered when using human-generative Artificial Intelligence deliberation to review tenders?"* Based on the literature, a preliminary list of relevant factors has been established to outline factors to consider when applying human–GenAI deliberation in this context. This list has been developed further throughout the study and has finally formed the basis of the evaluation framework that has been constructed in the later stages of the research.

3.2.2. Development Experiment Design

Flyvbjerg (2006) argues that case studies are essential for developing a rich, detailed understanding of real-life situations. Case studies not only enable researchers to explore complex phenomena in context, but also serve as powerful tools for the researcher's own learning, helping to refine research skills through practical engagement (Flyvbjerg, 2006). Flyvbjerg highlights the often underestimated "force of example", which refers to the value of situated examples in generating meaningful insights.
This "force of example" has been central to the design of this research. An experiment has been conducted with the primary aim of creating a realistic environment in which a human and a GenAI tool could interact directly. The focus was placed on the interaction between the individual and GenAI. This was followed by a group discussion to deepen the understanding of participants' experiences, perceptions, and attitudes towards the tool. Lastly, evaluation sessions have been held to validate and expand upon the research findings.

Experiment

The setup of the experiment has been designed to generate factual insights into how individuals engage with GenAI during the process of reviewing a tender. Through these interactions, information has been collected on the nature of the interaction, such as whether participants allowed the GenAI tool to assist them in the review process, and what form that interaction took. The Human-AI Deliberation framework proposed by Ma et al. (2024) offered an approach for creating a situation to encourage and study deliberation between humans and GenAI. As illustrated in Figure 2.7, the framework outlines four key steps for deliberation, providing a clear structure for setting up a human–GenAI interaction in a deliberative form. Although this thesis focuses on GenAI, the Human–AI Deliberation framework primarily addresses deliberative AI. Nevertheless, deliberative AI is highly relevant to this study. Similar to GenAI, deliberative AI supports users by facilitating brainstorming and encouraging reflection and discussion, particularly in decision-making scenarios (Ma et al., 2024). This overlap underscores the value of adopting the framework of Ma et al. for the study of human–GenAI interaction in the tendering context. However, the framework of Ma et al. has been adapted to fit the scope and the size of this master's thesis. The experiment aims to answer the second research question, introduced in Section 1.5: *"How does human-generative Artificial Intelligence deliberation influence changes in individuals' opinions in the context of reviewing tenders?"*
For taking part in the experiment, a group of five participants has been selected. The participants needed to have a background in infrastructure tendering with differing amounts of years of experience in the field, which ensures a diverse range of perspectives on tendering practices. Thereby, there needed to be an almost even distribution of male and female participants. This variety ensures greater diversity, providing a better representation of an average group of tendering professionals within a consultancy firm. The number of participants was limited to five due to practical constraints within the collaborating company. In addition to the five participants, one extra person was involved in the testing phase to trial and refine the experiment setup.

Group Discussion

After the experiment, a group discussion was held with the five participants. This enabled a deeper investigation into the factual outcomes and the subjective, emotionally influenced behaviours that impact engagement with the GenAI tool. A key focus was to understand why individuals choose to accept or reject GenAI's arguments, or whether they are willing to use such tools at all.

The third research question, introduced in Section 1.5, will be addressed through the group discussion, and the following evaluation sessions, and states: *"What do individuals perceive as drivers and barriers to using human–generative Artificial Intelligence deliberation for reviewing tenders?"*
Focus group discussions are a widely used qualitative method for gaining an in-depth understanding of participants' experiences and reasoning (Nyumba et al., 2018). In the context of this thesis, this approach is particularly useful for exploring the subjective and interpretive factors that influence human–GenAI interactions. The discussion enabled participants to engage in dialogue, reflect critically on their own perspectives and reconsider their reasoning in light of others' viewpoints. This created opportunities for new insights to emerge throughout the conversation. Before the group discussion, the participants rated a number of statements and explained their choices. These statements were adapted from the measurements used in the study by Ma et al. (2024), which analyses the influence of deliberation on decision-making. Due to the smaller scale and limited timeframe of this thesis, only the most relevant measurements were selected and refined. Additional adaptations were made based on the list of factors identified in the literature review. Some of these factors overlapped with the measures proposed by Ma et al. (2024), while others were adapted to align with the aims of the research and were supplemented with specific statements to reflect the context of this study.

Evaluation Sessions

Two evaluation sessions were organised to serve two purposes: first, to critically assess the findings from the experiment and group discussion, and second, to refine and expand upon these findings. These sessions, along with the group discussion, are designed to address the third research question: *"What do individuals perceive as drivers and barriers to using human–generative Artificial Intelligence deliberation for reviewing tenders?"*
A total of seven infrastructure tendering professionals participated in the sessions. The participants had varying levels of experience and offered a diverse range of industry perspectives. To ensure that each expert had sufficient opportunity to explain their reasoning, the group was split into two, with each conducting a separate session. This format enabled more in-depth contributions from each participant.
The first objective of the sessions was to evaluate the validity of the research and establish whether the findings were consistent with professional viewpoints. Validity refers to the extent to which a study accurately measures what it intends to measure and the extent to which the results can be regarded as truthful (Joppe, 2000). As outlined by Joppe, researchers typically evaluate validity by posing key questions and examining how these are addressed within the research itself and in relation to existing literature. In addition to validation, the sessions served to refine and extend the findings. The input of the experts contributed to a more comprehensive interpretation of the results. The study by Van den Berg et al. (2021), for example, directly integrated its findings into the research process to further refine its research. While this approach would be highly valuable, it is less feasible within the limited scope of this master's thesis. Therefore, the evaluation sessions were adapted to serve as a testing ground for the conclusions drawn from the experiment and the group discussion.
Instead of using a questionnaire format, which would have limited the opportunity for follow-up questions and real-time interaction, the sessions were structured as discussions. This format encouraged open conversation, enabling experts to respond to one another, provide clarifications, and contribute to a deeper understanding, as described by (Nyumba et al., 2018).

3.3. Data Types and Collection

The research in this thesis is based on different types of qualitative data, collected through the experiment, the group discussion, and evaluation sessions (described in Subsection 3.2.2). Data collection and storage did follow the upfront prepared Data Management Plan (Appendix F.1), approved by TU Delft (Appendix F.3). This plan outlines data handling procedures following university regulations. For example, all data has been anonymised and only shared with the thesis committee when necessary. Additionally, all participants have completed an informed consent form, explaining how their personal data would be managed (Appendix F.2). The data types and collection methods are described below.

Questionnaires

Several questionnaires have been completed by the participants, including a background information form, a review questionnaire has been completed before and after the deliberative conversations, and a set of statements that have been rated, each to be accompanied by the participant's reasoning.

The questionnaires were created using Microsoft Forms, and the data was collected, anonymised, and used for analysis. The ratings of the statements, including reasoning, were collected using sticky notes, which were anonymised and documented by the researcher.

Deliberative Conversations Human-GenAI

ChatGPT has been used as the GenAI tool for the experiment. As described in Section 1.3, a ChatGPT Team licence is provided through Count & Cooper. This Team licence ensures safeguarding any sensitive information that has been shared during the experiment (OpenAI, 2024a). For the deliberative part of the experiment, GPT-4o has been used for the reason that the creation of a customised GPT is only supported by GPT-4o (OpenAI, 2024b), which is been needed in this research. The interactive conversations between participants and ChatGPT were held on the participants' ChatGPT accounts and have only been shared with the researcher. As a result, only the participant involved in the interaction and the researcher have had access to the conversation. All data has been collected and stored within the secured ChatGPT Team environment, also Microsoft Word has been used to add notes to these conversations.

Group Discussion

The group discussion was stimulated by the presentation of statements on paper. Data from the group discussion consisted of a recording that was made via Microsoft Teams. Following, the recording was transcribed and anonymised using Microsoft Word.

Evaluation Sessions

For the evaluation sessions, the participating experts have been presented with the findings from the experiment and group discussion through a PowerPoint presentation. Similar to the group discussion, these evaluation sessions have been recorded via Microsoft Teams. The recordings have been transcribed and anonymised using Microsoft Word.

3.4. Analysis

The research methodology in Figure 3.1 presents the analytical part in two key stages: first, a thematic analysis of the collected data, and second, the development of the evaluation framework.

3.4.1. Thematic Analysis

Once the data from the experiment and group discussion wasa collected, the first thematic analysis was conducted. This analysis was based on factors identified in both the literature review (Section 2.5) and the study by Ma et al. (2024). These factors were used to establish the initial themes that guided the design of the data collection. During the analysis, additional factors were identified and categorised into emerging themes.

Following this initial analysis, the findings were validated and expanded during the evaluation sessions. The additional data from these sessions was integrated with the existing dataset and analysed further. The data coding was conducted using Atlas.ti, a qualitative data analysis tool that supports the identification of relevant patterns (ATLAS.ti, n.d.). The coding process was guided by the previously identified factors and refined by subthemes that emerged during the analysis. This reflects an iterative approach, allowing the framework to evolve as new insights were uncovered during the analytical process.

3.4.2. Evaluation framework development

The findings from the literature, the experiment, the group discussion, and the evaluation sessions have formed the foundation for the evaluation framework. These were analysed using the Activity Theory by Engeström (1987), which focuses on contradictions. The factors identified in the literature influencing GenAI use guided the development of the framework. The study by Ma et al. (2024) contributed evaluation measures that shaped both the design of participant rating statements and the analytical approach. All empirical findings were systematically mapped to corresponding types of contradictions within the activity system. These factors were then organised into a framework that offers an approach for analysing and understanding contexts similar to human–GenAI interaction in infrastructure tendering.

Finally, the study by Van den Berg et al. (2021) served as a methodological reference for applying the

Activity Theory in practice. Their approach to identifying contradictions in real-world organisational contexts informed how this study interpreted empirical data. Furthermore, their demonstration of translating theoretical insights into actionable outcomes directly supported the framework's design.

3.5. Outcomes Research

Lastly, the outcomes of the research aim to answer the sub-questions posed in this research: identifying the factors to consider when using human–GenAI deliberation in infrastructure tendering, understanding how this form of deliberation influences changes in individual opinions, and exploring what individuals perceive as drivers and barriers to its use. In doing so, the study addresses the main research question concerning individuals' experiences with human–GenAI deliberation.

The deliverable of this study is the evaluation framework developed through the analysis of the results. The study also aimed to provide insights and suggestions for improving the implementation of human–GenAI interaction within the context of infrastructure tendering.

4

Experiment Design

The outline of this chapter, in broad terms, begins with the setup of the experiment (Section 4.1), followed by the procedure of the experiment described in Section 4.2. This section on procedure outlines all aspects related to the experimental design of this research.

4.1. Setup Experiment

This section elaborates on the preparatory steps required before conducting the experiment with participants.

4.1.1. Tender

To design the experiment for this research, a general understanding of Count & Cooper's approach to tendering was required. Therefore, research is conducted into the tendering strategy of Count & Cooper, as presented in Appendix C. For more general details on tendering, read Section 2.1.
To select a suitable tender for the experiment, it was essential to ensure that the tender followed an open procedure and related to an infrastructure project. Following an evaluation of available tenders, one was selected. This tender consisted of several documents, each with a structure similar to that of the client's documents, which made it easier to align and compare the content from both parties. To design an experiment suitable for a short time frame, the scope of the tender was narrowed. A full tender typically includes multiple documents on different subjects, such as technical aspects or sustainability. For this study, one document was selected, focusing on collaboration. Collaboration was chosen because it is generally easy to understand and engage with, even for individuals without a background in the subject. Unlike technical topics, it requires little prior knowledge, making it accessible to participants with only a basic understanding.
After narrowing the scope to collaboration, the focus was further refined to the evaluation of a single measure. Tenders are often structured around multiple measures to be implemented during the execution of the project. Each proposal in the full tender contains several measures developed using a consistent format, allowing them to be reviewed independently. Focusing on one measure allowed the experiment to fit the available time and avoided unnecessary repetition. Therefore, evaluating only two measures was sufficient without compromising the validity of the results.

A framework is developed as a basis for the entire experiment, as shown in Figure 4.1. This figure illustrates all the steps involved in the experiment. In the following sections, each of these steps is outlined in detail, along with an explanation of all the documents that are created for the experiment and are required during the experiment. The titles of subsections beginning with a letter correspond to the numbered documents shown in Figure 4.1.

a. Guidelines Proposal

Due to confidentiality restrictions, both the client's guideline for the proposal and the tender proposal itself are not included in this thesis. However, the contents of these documents can be described to

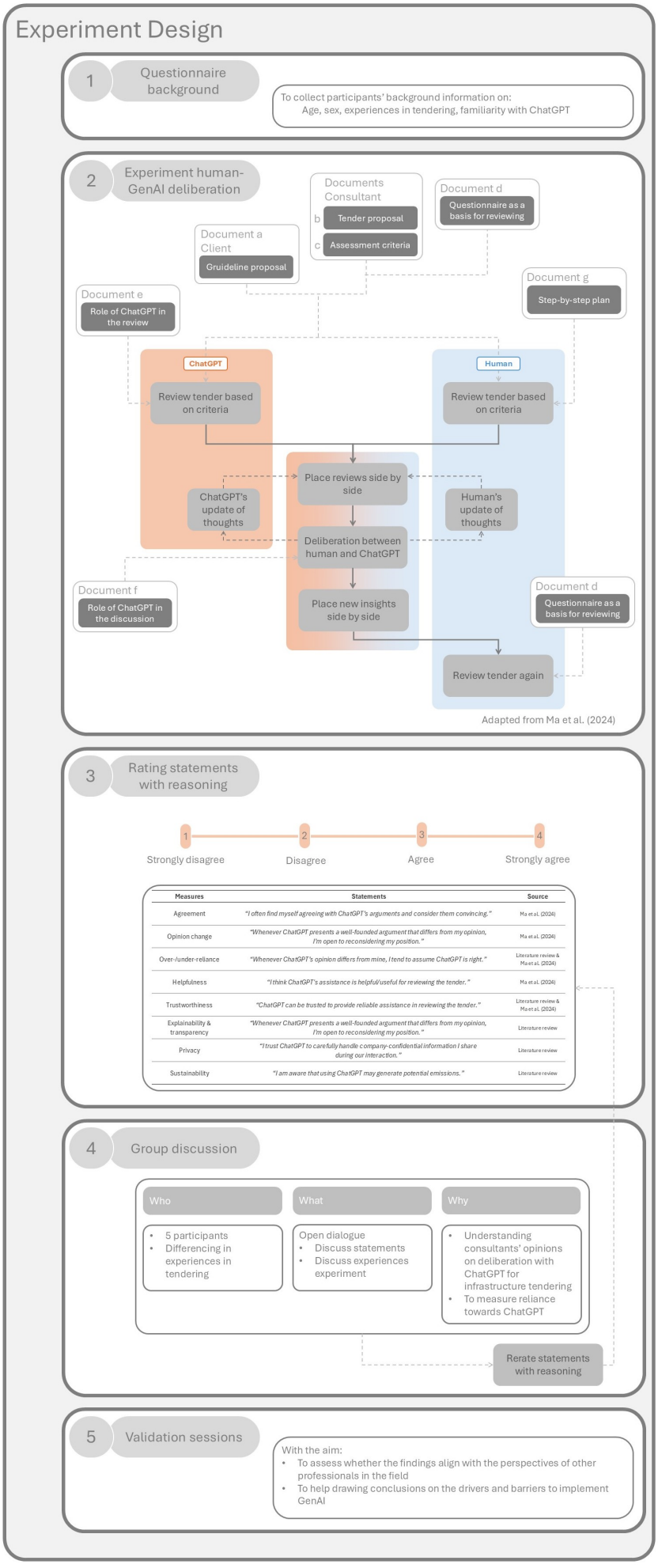


Figure 4.1: Framework experiment design (Created by the author)

give an idea of the type of information they contain.
The guideline proposal outlines the client's expectations and highlights the elements the client considers important. For the experiment, a shortened version of the guideline was used to ensure participants could read and understand the content within the limited time frame. Only the information relevant to the review of the selected measure was included, resulting in a two-page summary. These two pages contained the measure on collaboration, including criteria for how proposals can earn awards and the corresponding evaluation method. The collaboration section outlines the client's objective of achieving successful collaboration with the contractor, supported by defined success factors. It specifies the required information that must be included in the proposal, which often forms the structure of the tender, and presents the evaluation framework that assesses how well the tender proposal contributes to the objective, addresses the success factors, and demonstrates a realistic, effective, and efficient approach. The submission format, which outlines formal requirements such as page limits, was excluded from the shortened version, as the focus of the experiment is on content. The guideline also includes a rating scale ranging from 2 ("insufficient") to 10 ("excellent"), where a top score is awarded for proposals that fully meet the criteria, add clear value, and show no shortcomings. Lower scores reflect decreasing quality in addressing the evaluation criteria.

b. Tender Proposal

The client's guideline for the proposal forms the foundation for the content of the actual tender proposal. The contracting party, including both the contractor and the consultant, uses this guideline as a basis. However, during the proposal writing process, the contracting party typically has several opportunities to have a conversation with the client to gather crucial information. These conversations allow them to ask questions regarding the client's priorities and expectations. This process is described in more detail in Section 2.1. The insights gained from these interactions also support the development of a proposal that is better aligned with the client's objectives.
A tender proposal is structured around a series of measures, each designed to contribute to the overall goals set by the client. These measures provide structure to the proposal and follow a format known as the context–measure–effect–justification structure, ensuring consistency in the evaluation of multiple tenders (van Infrastructuur en Waterstaat, 2025). According to the structure, each measure begins with a clear description of the context, including the potential consequences if no action is taken. This is followed by the measure itself, outlining the proposed action. The expected effect of the measure is then explained, and finally, a justification is provided. For a strong justification, references are often made to previous projects carried out by the contracting party. This structure is further elaborated in Subsection 4.1.2.

4.1.2. Approach for Reviewing

Next to the guideline provided by the client, the consultant follows a standard procedure for reviewing measures, as described in the assessment criteria below.

c. Assessment Criteria

Count & Cooper uses a standard table for the review of measures, as shown in Appendix D.3. This review table follows a structure based on context–measure–effect–justification, and also includes the SMART table. This structure requires each measure to include a clear description of the context, including the potential consequences if no action is taken. The measure itself must be written according to the SMART principles: Specific, Measurable, Achievable, Relevant, and Time-bound (Bjerke & Renger, 2016). This includes that a measure should clearly state what the measure entails, how it will be implemented, who is responsible, and when the implementation will occur. Following this, the expected effect of the measure must be described, including how it contributes to achieving the intended objective. Lastly, the measure should be supported with a justification, ideally by referencing similar or previous projects. Demonstrating good comparisons makes the measure more realistic.

4.1.3. Human-GenAI Deliberation Setup

To ensure the process ran smoothly, a customised version of ChatGPT was created. This version, referred to as a GPT in Subsection 1.3.6, was created to support the part of the experiment involving deliberation between a human and ChatGPT. It allowed multiple instructions to be pre-programmed.

The actual setup for the human–GenAI deliberation experiment is shown in Figure 4.1, particularly in part 2. Here, the Human-AI Deliberation framework of Ma et al. (2024), shown in Figure 2.7, was used as a basis. The four interconnected steps were implemented, where the "elicitation of thoughts" has become "review tender based on criteria", the "alignment of thoughts" has become "place reviews side by side", and so on. One additional step was added, saying "place new insights side by side", to give the participant a clear overview of the updates resulting from the deliberation.
The first step, "review tender based on criteria", involves both ChatGPT and the human reviewing the tender independently. All the documents leading to this step in Figure 4.1 were included to provide both the human and ChatGPT with the necessary information. Documents a, b, and c, the documents provided by both the client and the consultant, have been discussed in previous subsections, while the remaining documents are discussed below.

d. Questionnaire as a Basis for Reviewing

A questionnaire was used as the basis for the review. It is structured according to the assessment criteria, following the context–measure–effect–justification format, along with the SMART table, as presented in Subsection 4.1.2. Eight questions were developed to reflect all elements of the assessment criteria, assessing how well the context is described and how each component of the SMART framework is addressed. Each question required participants to (1) assign a score on a scale from one to ten and (2) provide reasoning for that score. A 10-point scale was selected because it is relatively easy to use and allows participants to express nuanced opinions (Taherdoost, 2019). Providing reasoning was essential to allow both the human and ChatGPT to align with each other's perspectives or not. In addition, two final questions were included to capture the overall strengths and weaknesses of the measure. The full questionnaire can be found in Appendix D.4.

e. Role of ChatGPT in the Review

The document that guides ChatGPT in conducting the review is shown in Appendix D.5. This document provides detailed instructions on how ChatGPT should review the measure of the tender. It provides an outline of the situation, and it sets out the specific "rules" that ChatGPT must follow. For example, it states that responses must be presented in a structured table with the prepared columns. Additionally, the document stresses that all responses from ChatGPT should be based solely on the information contained within the provided documents. External sources are not permitted to support the arguments. To develop this prompt effectively, multiple test runs are conducted to ensure that ChatGPT delivers the output. These tests are performed using different prompts on various measures taken from several tenders written by Count & Cooper. This is done to ensure the prompt is applicable across a range of measure types.

f. Role of ChatGPT in the Discussion

The role of ChatGPT in the discussion is introduced at the start of the deliberation phase, illustrated as "deliberation between human and ChatGPT" in the framework in Figure 5.10. This document is shown in Appendix D.6, which includes a summary of the review process to guide ChatGPT, as well as a step-by-step set of instructions that it must follow during the deliberation with the participant. These instructions specify, for example, how ChatGPT should initiate the conversation. When the participant uploads their review, ChatGPT is instructed to generate a table with predefined columns displaying both its own and the participant's scores and reasoning, without changing any of the submitted content. If the participant has not submitted their review when the conversation begins, ChatGPT is prompted to request it. ChatGPT is also instructed to maintain the flow of the discussion by consistently asking the participant for their reasoning throughout the deliberation. The instruction document includes example questions to help facilitate the dialogue and serve as inspiration. Finally, when the participant indicates that the conversation has ended by typing "done," ChatGPT is directed to generate a second table summarising all new insights gained during the discussion. Crucially, ChatGPT is explicitly instructed to include only insights that were clearly mentioned in the conversation, and not to introduce any content that was not actually discussed.
However, during the testing phase, it became clear that the instructions from the "Role of ChatGPT in the discussion" document were not always followed consistently by the participant, despite repeated testing with different inputs and versions of the instruction document. To address this, the most important instructions are also written directly into the GPT environment. This approach proved effective, as this ensures that ChatGPT follows the rules consistently each time a new conversation starts.

4.1.4. Participants

The experiment involved five employees from the consultancy firm Count & Cooper. The conditions for the selection of these participants are elaborated on in Subsection 3.2.2. Count & Cooper categorises work experience into five scales. One participant is selected from each scale, with an additional participant to take part in the testing phase (see Table 4.1).

Levels within organisation Count & Cooper		Participants
1	Project analyst	Participant 5
2	Consultant	Participant 3
3	Team lead	Participant 2
		Participant 4
4	Project lead	Test person
5	Senior Project lead	Participant 1

Table 4.1: Participants experiment (Created by the author)

4.1.5. Test Run

After the setup of the experiment, a test run was conducted with one of the employees of Count & Cooper, who is experienced in tendering. This test run provided valuable insights, as it became clear that the review process took longer than initially expected. As a result, the review section was shortened. Initially, participants were required to review two measures, but based on the findings from the test run, this was reduced to one. This adjustment was made because each measure can be reviewed independently, as Count & Cooper requires every measure to include given aspects, as described in Subsection 4.1.2.
Additionally, it became evident that the instructions given to ChatGPT regarding the discussion process needed some adjustments. The guideline on how ChatGPT should engage in discussions with participants was refined to provide clearer instructions, ensuring a more structured and effective deliberation process.

4.2. Procedure Experiment

With the setup complete, the experiment itself could start. As outlined in Figure 4.1, the experiment consists of four main parts: beginning with a background questionnaire, followed by the deliberation session, then the rating of statements with reasoning, and concluding with a group discussion and validation sessions. The entire process takes approximately two hours, a duration chosen to limit the time commitment on participants and reduce the risk of fatigue. The following subsections describe each part of the experiment in detail.

4.2.1. Questionnaire Background Information

Understanding the participants' backgrounds enables well-informed conclusions to be drawn from the experiment's results. Therefore, participants have completed a questionnaire to gather background information on their experience in tendering, the length of time they have been involved in tendering, their background and knowledge of AI, as well as their age and sex. The questionnaire is shown in Appendix D.1.

4.2.2. Human-GenAI Deliberation

The first part of the human-GenAI deliberation process in Figure 4.1 is outlined in Subsection 4.1.3. In this stage, ChatGPT is asked to review a measure from the written tender. Following this, the participant is asked to do the same as ChatGPT: to carry out the same review of the measure from the tender. The participants have been provided with the same documents as ChatGPT, including the guideline proposal (a), the proposal (b), the assessment criteria (c), and the questionnaire as a basis for reviewing (d). However, instead of receiving the document titled "The Role of ChatGPT in the Review", participants have received a step-by-step plan to guide them through the experiment, elaborated

on below.

g. Step-by-Step Plan for the Experiment

The step-by-step plan has been essential, as all five participants were required to complete the experiment simultaneously. The step-by-step plan is presented in Appendix D.2. The document offers a general overview of the steps participants needed to follow during the two-hour session, including the proposed time for each part. It also contains references to various hyperlinks necessary to complete the experiment. For instance, participants have been directed to the questionnaires they need to complete, and a link has been provided to take the participants directly to the appropriate deliberative ChatGPT environment via another clickable hyperlink. The document allowed the participants to complete the experiment independently. However, if they did have any remaining questions, they could ask for assistance.

The first step for the participants was to fill in the questionnaire, in which the review could be completed. Once the participant has filled in the questionnaire, the personal review has been sent back to them in PDF form. Following the step-by-step plan, they are then instructed to access the customised version of ChatGPT. If the participant had submitted their review in the ChatGPT environment, ChatGPT starts by presenting both reviews side by side. From that point, ChatGPT initiated the deliberation by asking the participant to justify their scores and reasoning. Similarly, ChatGPT has provided its own explanations, allowing the discussion to develop naturally. Both ChatGPT and the participant could propose points for deliberation, but ChatGPT was specifically instructed never to end the discussion on its own. ChatGPT was specifically instructed to ask participants for reasoning and to challenge their thinking. As a result, ChatGPT was also permitted to update its own reasoning during the conversation. Both ChatGPT and the participant had the opportunity to refine their perspectives throughout the deliberative exchange. They were allowed not only to reconsider their overall viewpoints but also to revise the scores they had initially assigned during the review stage. After a deliberation period of fifteen minutes, the participant was instructed to send “done” to ChatGPT. ChatGPT then summarised the conversation, presenting the original and revised insights and scores side by side.

Finally, the participant was asked to complete the questionnaire again (Subsection 4.1.3) to reassess the measure of the tender. This process enabled an evaluation of the extent to which interaction with ChatGPT influenced participants’ initial scores and opinions.

4.2.3. Rating Statements with Reasoning

Following the deliberation process, participants were presented with a series of statements, which they were asked to rate on a four-point scale ranging from ’strongly disagree’ to ’strongly agree’ (step 3 in Figure 4.1). A four-point scale helps reduce bias by removing the neutral option, thereby not including a midpoint encourages participants to make a decision on whether they agree or disagree (Chyung et al., 2017). Rating the statements has been done individually to prevent participants from influencing one another. In addition to assigning a rating, participants were asked to provide reasoning for their choices. They documented their explanations on sticky notes to ensure that their thoughts were captured. All statements have been displayed throughout the room in which the experiment took place. Those statements have been printed on A3 paper and spread out across the room. Participants were free to walk around the room and attach their written sticky notes to the statements.

The statements presented to the participants are shown in Table 4.2. This table includes a selection of subsets from the study by Ma et al. (2024), which presents a broad range of objective and subjective measures. However, due to the much smaller sample size in this study, only the most relevant measures were selected. In addition to measures from Ma et al. (2024), adaptations were made based on insights from the literature review, resulting in a finalised list of important factors for researching human–GenAI interaction, as shown in Table 2.1. Since it was not feasible to include all identified factors, a selection was made to ensure a differentiated and balanced set of measures. Following, all selected measures have been reformulated as statements, enabling participants to express their level of agreement.

Measures	Statements	Source
Agreement	*"I often find myself agreeing with ChatGPT's arguments and consider them convincing."*	Ma et al. (2024)
Opinion change	*"Whenever ChatGPT presents a well-founded argument that differs from my opinion, I'm open to reconsidering my position."*	Ma et al. (2024)
Over-/under-reliance	*"Whenever ChatGPT's opinion differs from mine, I tend to assume ChatGPT is right."*	Literature review & Ma et al. (2024)
Helpfulness	*"I think ChatGPT's assistance is helpful/useful for reviewing the tender."*	Ma et al. (2024)
Trustworthiness	*"ChatGPT can be trusted to provide reliable assistance in reviewing the tender."*	Literature review & Ma et al. (2024)
Explainability & transparency	*"Whenever ChatGPT presents a well-founded argument that differs from my opinion, I'm open to reconsidering my position."*	Literature review
Privacy	*"I trust ChatGPT to carefully handle company-confidential information I share during our interaction."*	Literature review
Sustainability	*"I am aware that using ChatGPT may generate potential emissions."*	Literature review

Table 4.2: List of statements (Adapted from Ma et al. (2024) and the literature review)

4.2.4. Group Discussion

After the sticky notes had been attached to the statements, the group discussion started (step 4 in Figure 4.1). To start, participants have been invited to walk around the room to read each other's notes. These statements served as the foundation for the group discussion. Each statement has been discussed one by one. The discussion encouraged participants to share their views on the statements and broader topics related to the experiment. Participants were also welcome to share their reflections on the deliberation process with ChatGPT and their experiences in reviewing the tender.
This group discussion has lasted approximately thirty minutes. Afterwards, participants were given the opportunity to revise their initial thoughts. They were asked to walk around the room once more and, if desired, update their scores and/or add new or revised reflections to the statements.

4.2.5. Evaluation Sessions

The evaluation sessions, Step 5 in Figure 4.1, do have two purposes: first, to validate the results of the experiment and group discussion, and second, to support the development of the study's conclusions. The sessions began with an explanation of the research focus: the concept of human–GenAI deliberation in the context of reviewing infrastructure tenders. Experts are asked to imagine themselves using GenAI (ChatGPT) during their review process. More specifically, as a final check just before submitting the final written tender for the final group review session.

In the first part of the session, the results are presented using a PowerPoint presentation. Each of the key themes that shape the research is introduced. Quotes and insights are integrated to help the experts relate to the findings. This structure is intended to spark discussion, allowing the experts to share their views and engage critically with the outcomes.
In the second part of the session, participants are asked to help fill in the themes that define the implementation of GenAI. They are asked to suggest drivers and barriers to implementation. This input provides information for drawing conclusions and identifying the factors that influence the adoption of GenAI in tender reviews. The PowerPoint presentation is available in Appendix E.

5

Findings

This chapter presents the findings of the research. The structure is as follows: first, the findings from both the experiment and the group discussion are presented, organised thematically as outlined in Section 5.1. This is followed by the findings from the evaluation sessions in Section 5.2, beginning with a discussion of the themes and then addressing the potential drivers and barriers to the implementation of GenAI. Next, Section 5.3 brings together the findings from the experiment, group discussion, and evaluation sessions, followed by an analysis of these findings through the lens of Activity Theory in Section 5.4. Finally, the chapter concludes with the presentation of the evaluation framework developed during this study in Section 5.5.

In this chapter, the findings are presented anonymously. However, to connect related statements made by the same individuals, participants are referred to by a number, either written out as Participant x or shortened to Px. To provide some background information on the participants, Figure 5.1 offers insight into the ages of the participants, while Figure 5.2 shows the number of tenders they have worked on.

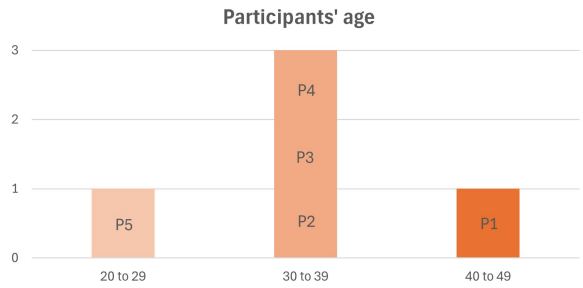


Figure 5.1: Participants' age (Created by the author)

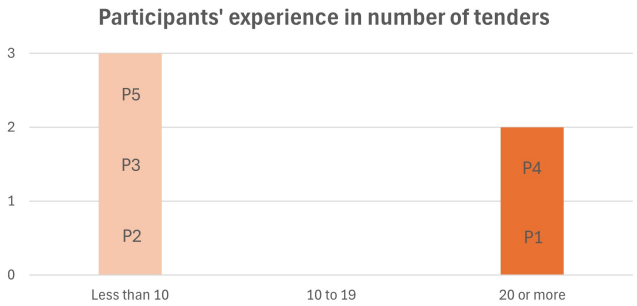


Figure 5.2: Participants' experience in number of tenders (Created by the author)

5.1. Findings Experiment and Group Discussion

This section presents the findings from the experiment and the group discussion. The results are organised thematically to provide a structured overview.

5.1.1. Agreement Level

For the level of agreement, it is crucial to investigate the adjustments made to the scores. Figure 5.3 presents the adjustments in scoring after the deliberative conversation between the participants and ChatGPT, broken down by participant. The figure illustrates how often participants partially adjusted, fully adjusted, or adjusted their scores in the opposite direction of ChatGPT's score. The figure also shows the situations where no change was made and where the initial scores already matched those of ChatGPT.

An observation from Figure 5.3 is that P5 ultimately rated all questions exactly the same as ChatGPT. While half of P5's initial scores already matched ChatGPT's, the other half were fully adjusted to align with ChatGPT's ratings after deliberation. The discussion revealed that P5 strongly agreed with ChatGPT's arguments, which led the participant to adopt the same ratings. A notable remark was made after deliberating with ChatGPT when P5 wrote the following:

> *"ChatGPT served as a validation for me because I'm not yet fully confident in my own assessment skills as a project analyst."* – P5

A project analyst is someone at the start of their career, positioned within the first group of experience levels at Count & Cooper (see Table 4.1).

On the other hand, P1 initially had no scores matching those of ChatGPT. For half of the scores, the participant made no adjustments at all. Only in three situations did the participant (partially) adjust their scores to align with ChatGPT's. However, in two of these cases, P1 wrote in the comments that the participant had not gained any new insights after engaging in a discussion with ChatGPT. Notably, in one case, the participant even adjusted the score in the opposite direction of ChatGPT's, explaining during the discussion that the participant surely did not always agree with ChatGPT's reasoning.

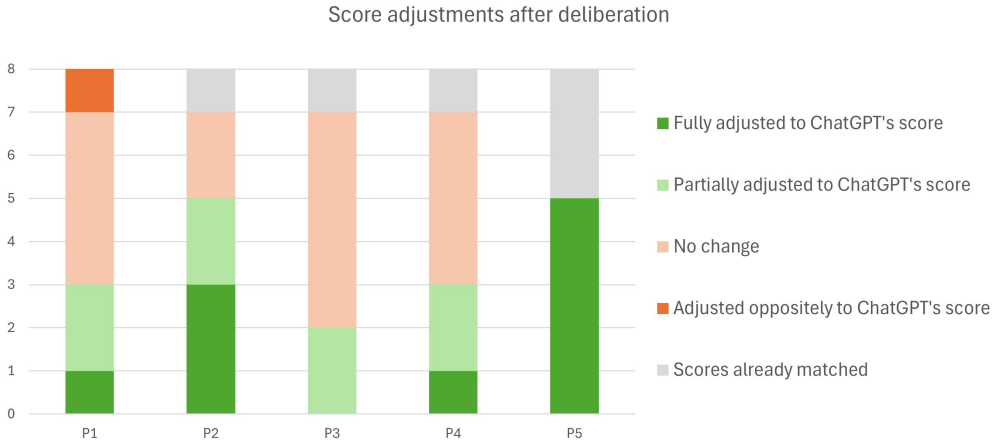


Figure 5.3: Score adjustments after deliberation (Created by the author)

Figure 5.4 displays the specific changes in scoring. This figure presents the scores both from before and after deliberation, structured by the different aspects used to evaluate the measure.

For P1, the figure confirms a varied approach, showing a mix of score adjustments, scores remaining the same, and even in one aspect of adjusting a score in the opposite direction of ChatGPT's score. This suggests no strong pattern in P1's scoring behaviour. P2 and P4 show a relatively balanced approach, with a combination of adapting their scores in line with ChatGPT and maintaining their original scores. P3 kept most scores unchanged. However, when P3 did adjust a score, the adjustment was by more than one point out of the score of ten. The figure confirms P5's adaptive attitude, as the scoring line for P5 consistently shifts towards ChatGPT's scores.

Figure 5.4: Score adjustments before and after deliberation (Created by the author)

Within the theme of agreement, a statement was also included to which participants could give a rating on whether they agreed, namely: *"I often find myself agreeing with ChatGPT's arguments and consider them convincing."* This statement was met with unanimity, as all participants agreed that they often do, but surely not always agree with ChatGPT. A key factor influencing their agreement was the quality of the model's reasoning. All participants, independently of one another, emphasised that well-elaborated arguments were essential for accepting ChatGPT's responses. P1 noted that if the arguments are well-elaborated, then ChatGPT is right. P5 expanded on this perspective, explaining to stay critical. But if there is no need to doubt ChatGPT because the reasoning is clear, the participant would not start another discussion "just to discuss". According to P3, ChatGPT is generally capable of producing strong and well-founded arguments.
Out of the reported deliberative conversations between the participants and ChatGPT, it became clear that there were several moments in which participants expressed agreement with the model, often replying with brief reactions such as *"agreed"* or *"exactly"*. However, consistent with the earlier observations, participants also showed a critical attitude. When ChatGPT's responses were unclear, they did not hesitate to ask for clarification, often asking questions to the model before responding to the prompts it had presented.
P4 stressed the importance of formulating prompts correctly when interacting with ChatGPT, bluntly stating: *"shit goes in, shit goes out"*. This participant highlighted that the likelihood of agreeing with ChatGPT largely depends on how the tool is prompted.
Additionally, Participant 1 noted that the interaction was not just about the opportunity to agree with ChatGPT. They explained that ChatGPT genuinely engaged with their input, incorporating their arguments into its reasoning and even adapting its thoughts during the conversation.

5.1.2. Trust

"The discussion with ChatGPT feels like a conversation where you can lay everything on the table, rather than one with a real person." - P5

Participant 5 clearly described ChatGPT to be trusted, out of a real human form of feeling, that this participant can totally open up about their ideas without being judged.
Participant 4 also demonstrated a form of trust in ChatGPT but framed it differently. According to P4, trust in the tool depends heavily on how the user interacts with it. P4 referred to this as the "rules" a user applies when interacting with ChatGPT and implies that trustworthiness is fully dependent on the input of the user.

In contrast, P3 took a more critical attitude, stating that ChatGPT is certainly not always correct, which in P3's view makes the tool less trusted.

5.1.3. Reliance

Statement: *"ChatGPT can be trusted to provide reliable assistance in reviewing the tender."*

All participants agreed with the statement. P5 highlighted ChatGPT's objectivity, while P2 stated to rely on it as a source of inspiration but remained critical. P4 noted that ChatGPT can be reliable if the prompt includes well-defined assessment criteria for the review. Participant 1 expressed a strong sense of reliability in ChatGPT's output, describing reliability as "knowing what you get". Based on this participant's experience with the tool, P1 feels confident in how ChatGPT formulates its responses and what to expect from it.

"ChatGPT is always a good discussion partner, you know what to expect." - P1

Despite this familiarity, P1 stressed that users should not forget to rely on their own intuition or gut feeling, which is important when evaluating ChatGPT's responses.
Participant 3 also highlighted this point, stating that ChatGPT is certainly not always correct:

"ChatGPT is not always correct and therefore less reliable." - P3

However, the discussion among participants revealed insights into how ChatGPT responded to their comments. It appeared that ChatGPT never directly disagreed with the participants' contributions. Instead, it constantly affirmed that the input of the participants contained good suggestions and never challenged their arguments. All participants agreed that this discovery changed their reliance on ChatGPT. As a result of this discussion, two participants (P4 and P5) changed their ratings. They found that ChatGPT was too agreeable to their input, leading them to downgrade their responses from agreeing with the statement to disagreeing with it. P4 further added that they need ChatGPT to provide more pushback to rely on its assistance more.

5.1.4. Over- and Under-Reliance

Building on the statement designed to assess agreement levels, an additional statement was presented to the participants to explore both over-reliance and under-reliance on ChatGPT: *"Whenever ChatGPT's opinion differs from mine, I tend to assume ChatGPT is right."* This statement focuses on whether people blindly follow ChatGPT's opinions. It became clear that no participant did so during the deliberation process, as everyone disagreed with the statement. In general, participants emphasised the need for well-reasoned arguments to change their opinions and remained critical of ChatGPT. Even P4 wrote that the participant felt "superior to AI". However, notably, after the group discussion, P4 updated their answer, agreeing more with the statement and writing down the reasoning that the opinion of this participant is not (always) "unquestionable".
It appeared that the youngest participant (P5), who also has the least experience in the workplace, was more open to allowing ChatGPT to explain its reasoning. While the other participants clearly expressed the need for strong argumentation, P5 stated to remain critical but actively continue to ask ChatGPT for explanations. This openness was further confirmed during the deliberation conversation with ChatGPT. P5's comments were the longest among all participants, allowing ChatGPT more space to elaborate, continuously asking follow-up questions, and frequently agreeing with ChatGPT's responses.

5.1.5. Helpfulness

The participants were asked whether they agreed with the statement about helpfulness: *"I think ChatGPT's assistance is helpful/useful for reviewing the tender."* All participants agreed with the statement: four participants agreed, and one participant (P4) fully agreed. During the group discussion, P4 expressed surprise at being the only one to fully agree with the statement. P4 explained their reasoning by stating that ChatGPT could always be helpful or useful, particularly for general adjustments in the review.
Participant 2 supported their agreement with the statement by writing the following:

"Stay critical, but good as an inspiration." - P2

This argument that ChatGPT is helpful as a source of inspiration was confirmed by the other participants

during the group discussion. However, also Participant 5 emphasised the importance of remaining critical of the answers it produces. Thereby, P2 noted during the discussion that ChatGPT could be helpful or useful after the internal review process, particularly when the feedback received is unclear. In such cases, ChatGPT could assist in interpreting the feedback by helping to explore what the reviewer might have meant.
Lastly, P1 appreciated the insights ChatGPT provided and the ease of use during the review process. Thereby, ChatGPT is seen as helpful in the role of a discussion partner, with both P1 and P5 stating that it could contribute valuable insights through discussion.

Objectivity

Throughout the discussion, there were frequent comments about ChatGPT providing an objective opinion, reinforcing the idea that no human reviewer can fully evaluate a tender with complete objectivity. Three out of five participants literally mentioned the objectivity of ChatGPT as a positive aspect. P1 commented: *"just to have a mirror held up to you"*. P5 found ChatGPT helpful due to its neutrality, noting that ChatGPT offers an objective perspective. Because in human conversations, individuals often have specific interests.

"Valuable discussion with ChatGPT, who has no direct interest." – P5

P4 also noted the benefit of ChatGPT's objectivity, highlighting that ChatGPT does not require users to adopt its opinion.

Reflectivity

Both Participant 1 and Participant 3 shared the view that using ChatGPT to review one or more measures encourages deeper thinking and reflection.

"Just to be confronted with my own thinking" - P1

P3 also noted that ChatGPT mirrors the participant's own thinking, forcing users to reflect on their reasoning. Which, in turn, helps to become more critical during the review process.

"ChatGPT helps me to sharpen my thinking" - P3

This view was confirmed by P1, who acknowledged the existence of blind spots in reasoning, both in others and in their own thinking. P1 reflected on a trend of having tunnel vision when reviewing their own written work and found ChatGPT to be a valuable tool in identifying those blind spots. Although P1 did not always agree with ChatGPT and therefore thinks it is necessary to critically assess the outcomes of the conversation.

Time Efficiency

P4 highlighted time efficiency as a benefit as P4 pointed out that ChatGPT can be especially valuable in the review process, as a review check by ChatGPT can be carried out in a short amount of time. This participant also pointed out that review sessions themselves often result in wasted time, with participants claiming time without necessarily adding valuable information. This ties back to the idea that ChatGPT, unlike human reviewers, does not have a personal interest in the discussion, whereas all other parties involved do.

"Especially because ChatGPT does not claim all the speaking time during a review session." - P4

In this case, Participant 4 referred to the replacement of review sessions with the ChatGPT review session. Responding to that, P3 suggested using ChatGPT primarily as a checking tool, rather than as a replacement for human reviewers. According to P3, incorporating ChatGPT as an additional step prior to human review could be beneficial, as it requires minimal time to allow ChatGPT to check the content.
P2 and P3, on the other hand, thought differently, experiencing that ChatGPT takes a lot of time compared to reviewing a measure on their own.

"It does take a lot of time, so I would mainly apply it to the things I doubt. If I am sure about something, or reasonably sure, I wouldn't use it." - P2

"It does take longer than just manually reviewing a measure." - P3

P3 also noted that it took considerable time before ChatGPT was convinced of their opinion.

Participant 4 brought up another time-related benefit, noting that ChatGPT does not have an agenda that needs to be considered when there is a need for a review session.

"What I think is another added advantage is that ChatGPT has no agenda." - P4

Interestingly, time efficiency was evaluated from different perspectives. Some participants found ChatGPT to be time-efficient due to its ease of use, while others considered it time-inefficient because they lacked the capabilities to use it quickly and effectively. Most participants viewed deliberating with ChatGPT as an addition to the usual review process. However, Participant 4 went further, comparing the use of ChatGPT for reviewing a tender to participating in a traditional review session, and concluded that ChatGPT required significantly less time.

ChatGPT Lacking Information

The uncertainty about how helpful a tool like ChatGPT can be for a review largely arises from the fact that ChatGPT is not provided with the full context of the assignment, addressed by P1. Participant 2 added that ChatGPT is not up to date with all the conversations held with the client, which are essential for fully understanding the assignment and therefore is less helpful for reviewing. However, this point was contradicted by Participant 4, who emphasised that the user is responsible for ChatGPT's output. If the user does not provide the tool with relevant information, it cannot generate the desired result.
Also, due to ChatGPT lacking information, Participant 2 believed it would be beneficial if users asked ChatGPT specific questions related to the aspects they were uncertain about and not to let ChatGPT do a full review.

To summarise, all participants agreed that ChatGPT is helpful, particularly as a source of inspiration when reviewing tenders. However, they consistently emphasised that the tool should be used as inspiration only, as the user must remain critical of the output it generates. ChatGPT's objectivity and reflectivity were also considered valuable. Both concepts were mentioned multiple times by multiple participants as reasons why the tool could be of support. Time efficiency, meanwhile, was evaluated from differing perspectives. Some participants found ChatGPT to be time-efficient in the context of reviewing. Others, however, felt it was time-consuming, often depending on how familiar they were with using the tool.
Finally, it was suggested that ChatGPT does not have enough information to be really helpful. Two participants mentioned concerns about the tool being up to date. However, this was countered by another participant, who argued that it is the user's responsibility to provide relevant context and input to ensure helpful responses.

5.1.6. Explainability and Transparency

Statement: *"Whenever ChatGPT presents a well-founded argument that differs from my opinion, I'm open to reconsidering my position."* During the discussion on this statement, one of the participants (P3) addressed that this statement indirectly questioned whether the participants were sceptical of ChatGPT. More specifically, whether they took ChatGPT seriously and actually considered its ideas. However, everyone agreed that they did take ChatGPT's input into account. P1 also explained this by writing: *"If the arguments are good arguments that I haven't thought of myself."* P3 further confirmed that ChatGPT helps becoming sharper by providing well-founded arguments. Participant 5 stated the following:

"I stay critical and keep on asking questions, but am willing to adjust my opinion." - P5

This critical attitude was confirmed in the deliberative conversation between P5 and ChatGPT, as seen in the number of questions asked by P5. The participant's willingness to reconsider their views was also confirmed during the interaction. P5 asked several questions to better understand the reasoning behind ChatGPT's responses and, multiple times, P5 expressed agreement with ChatGPT, resulting in adjusting their own answers to align with those of ChatGPT.
Similarly, P1 noted that ChatGPT's ability to explain its answers contributed to increased trust in the tool as a supportive discussion partner.

5.1.7. Privacy

There was variation among participants in how privacy concerns influenced their use of ChatGPT. The following statement was presented to them: *"I trust ChatGPT to handle company-personal information I share during our interaction."* Opinions on trusting ChatGPT with company-personal information varied, as shown in Figure 5.5. One participant (P5) strongly agreed with the statement, while another participant (P4) disagreed. The remaining participants agreed to some extent.

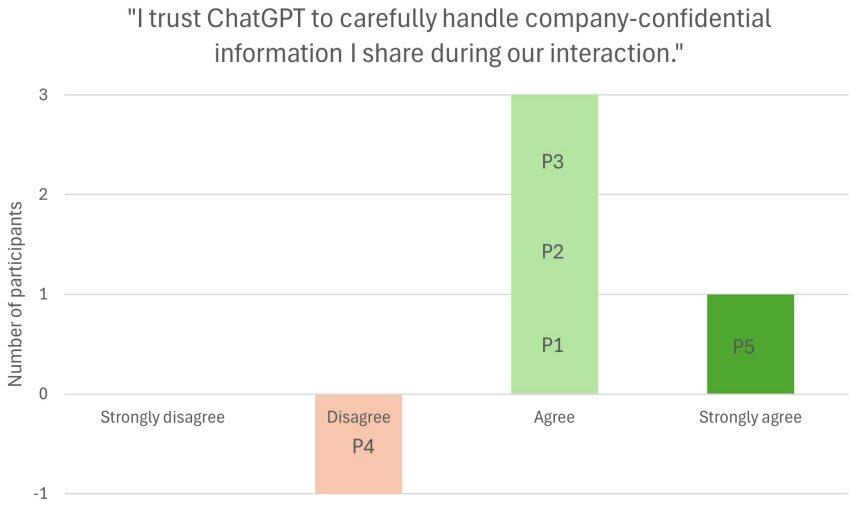


Figure 5.5: Privacy concerns company-confidential information (Created by the author)

P4, who disagreed with this statement, explained during the discussion that the participant was sceptical about sharing confidential information with ChatGPT. This participant raised concerns about trust in the company, referencing to the connections between Sam Altman (CEO of OpenAI, ChatGPT) and Donald Trump (current president of the United States), which contributed to a lack of confidence in how company-sensitive information might be handled. Furthermore, P4 questioned the reliability of the secured version of ChatGPT, in particular the claim that information shared is not used for the general training of the model. This participant stated not to believe the model to be fully trusted.

"I would never put company-sensitive information in ChatGPT anyway." - P4

While P4 was clear about not sharing company-confidential information with ChatGPT, this participant added to use ChatGPT just as a source of inspiration. On the other hand, P4 was open to the idea of sharing documents Created by the client, expressing less concern about the sensitivity of information outside the company.

The three participants who agreed with the statement about trusting ChatGPT to handle company-confidential information well were aligned, both in their scoring and in their reasoning. All three expressed a general level of trust in ChatGPT's handling of sensitive information. P1, for instance, shared the view that confidentiality is a core selling point of OpenAI. According to this participant, OpenAI must maintain confidentiality within its paid versions in order to keep its customer base. While P1 expressed some doubt about complete confidentiality, P1 still believed that the conversations held with ChatGPT are not directly used to train the model.
P2 also acknowledged concerns regarding privacy, particularly with respect to the future. P2 expressed feelings of insecurity, specifically mentioning concerns about political leadership in the United States under Donald Trump. Although P2 currently believes the information is safe, this participant expressed insecurities about whether this confidentiality will remain guaranteed over time. However, P2 noted: *"much of the information we share now will actually be outdated in the future"*, which is why this concern does not prevent the participant from using ChatGPT.
P3 shared their lack of technical knowledge on the subject and therefore hoped that the information would remain confidential.

However, the scepticism of the three participants only extends to doubting the security of confidential information rather than completely avoiding its use. As a result, all participants except P4, do share confidential information with ChatGPT.

P5 initially expressed full trust in ChatGPT's ability to handle company-personal information securely and did not question its confidentiality. This participant mentioned that there was no personal reason to distrust the tool, although this participant acknowledged that their opinion could change over time. Notably, P5 appeared to have taken on board some of the concerns raised by other participants regarding the safety of shared information. As a result, P5 adjusted their response from "strongly agree" to "agree" after the group discussion. P5 explained that the participant still did not fully understand the complexities of confidentiality but recognised and aligned with the concerns expressed by the others.

The privacy concerns expressed by participants are summarised in Figure 5.6, which illustrates the relationship between privacy concerns and the resulting hesitancy in using ChatGPT. Overall, trust in ChatGPT's handling of company-specific information is limited among some participants, with concerns about potential future data leaks contributing to their scepticism. However, despite these concerns, four out of five participants are not hesitant to use ChatGPT in their work.

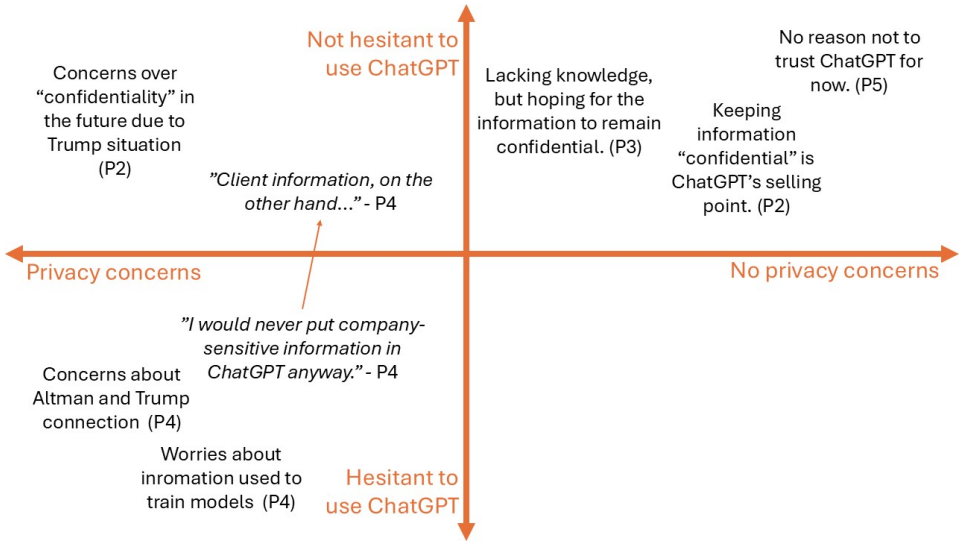


Figure 5.6: Privacy concerns influencing GenAI use (Created by the author)

5.1.8. Scepticism

During the group discussion, an important observation was made, which made participants look differently at the deliberative conversation with ChatGPT, which affected several participants' opinions. Participant 1 shared with the group that this participant had the impression that ChatGPT genuinely understood their reasoning, particularly because the model had changed its response after receiving their arguments, which had happened more than once. However, other participants agreed that ChatGPT tended to adjust its answers quickly to align with theirs, often without providing counterarguments. There was a general consensus that ChatGPT rarely challenged the participants' views. As the discussion progressed, the perception shifted from seeing ChatGPT as being understanding to concerns about it being too serving. This made Participant 4 raise a concern:

"ChatGPT did not really give much of a pushback, so there may be some doubts about whether it is reliable." - P4

In contrast, Participant 5 noted that, in their experience, ChatGPT tended to seek a compromise. It did so by incorporating arguments from both sides in its final responses.

After the group discussion, participants were allowed to revise their ratings of the statements and provide additional comments. The idea that ChatGPT was too serving or agreeable became a key reason

behind several changes in scoring. This was clearly reflected in participants' remarks:

"Yes, as far as I am concerned, a little too compliant. It could have been a bit more challenging." – P1

"ChatGPT could be a bit more critical." – P4

"The comment about 'compliance' lowers the score. Chat is serving and often just agrees with you." – P5

5.1.9. Sustainability

On the topic of sustainability, the following statement was presented to the participants: *"I am aware that using ChatGPT may generate potential emissions."* All participants acknowledged some kind of awareness of the potential emissions generated by using ChatGPT. However, how this awareness influences their behaviour varies among them. The differences in how participants limit their use of ChatGPT due to environmental concerns are shown in Figure 5.7.

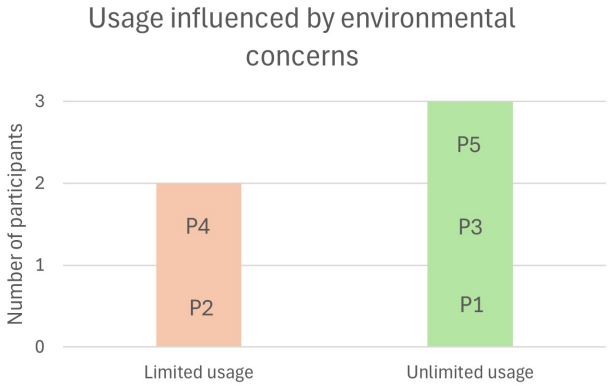


Figure 5.7: Environmental concerns influencing usage GenAI (Created by the author)

For two out of five participants, sustainability awareness leads them to limit their use of ChatGPT. Both shared that they are up to date on the fact that generating images costs a lot of energy, and they therefore avoid generating images with ChatGPT. P4 explained that this participant would not use ChatGPT for "useless" purposes and only generates images when necessary, and so does P2.
When diving a bit deeper, P2 mentioned not to hold back on the use of ChatGPT, as long as it does not involve the generation of images. P4, on the other hand, gives a different reason for not restricting the use of the tool other than for generating images. P4's reasoning reflects a conscious decision to overlook potential ethical concerns:

"I know, but I purposely ignore it because I already have enough climate guilt." - P4

Efficiency over Sustainability

The remaining three participants in Figure 5.7, on the other hand, do not restrict their use of ChatGPT at all. For them, "efficiency" is the key factor in determining their usage. Whether in their professional or personal lives, they do not limit their ChatGPT use. As long as it is time-efficient, which they believe it is, they use it to its full extent.

"For me, ChatGPT is something that brings efficiency." - P1

Participant 1 bases the use of a tool entirely on the amount of time it requires, stating that it is all about the level of efficiency it offers.

"In fact, I think ChatGPT is even more efficient than hours of googling." - P1

This participant is aware that using ChatGPT may consume a significant amount of energy, however, efficiency remains the primary driver in the decision to use the tool.
Participants 3 and 5 also do not place limitations on their use of ChatGPT and did put its energy consumption into perspective, noting that many other tools and activities also consume energy.

The environmental concerns influencing the use or non-use of ChatGPT are summarised in Figure 5.8. Participant quotes and perspectives are positioned within the figure according to how these concerns impact their use of ChatGPT.
Overall, there is little to no hesitancy among the five participants when it comes to using ChatGPT. The exception is Participants 2 and 4, who choose not to generate images with the tool. Both participants consider the potential environmental impact of ChatGPT's use, but this awareness does not lead to broader non-use beyond avoiding image generation. The remaining three participants expressed no environmental concerns and reported no hesitation in their use of ChatGPT. For these participants, efficiency is the main driver behind their use of ChatGPT.

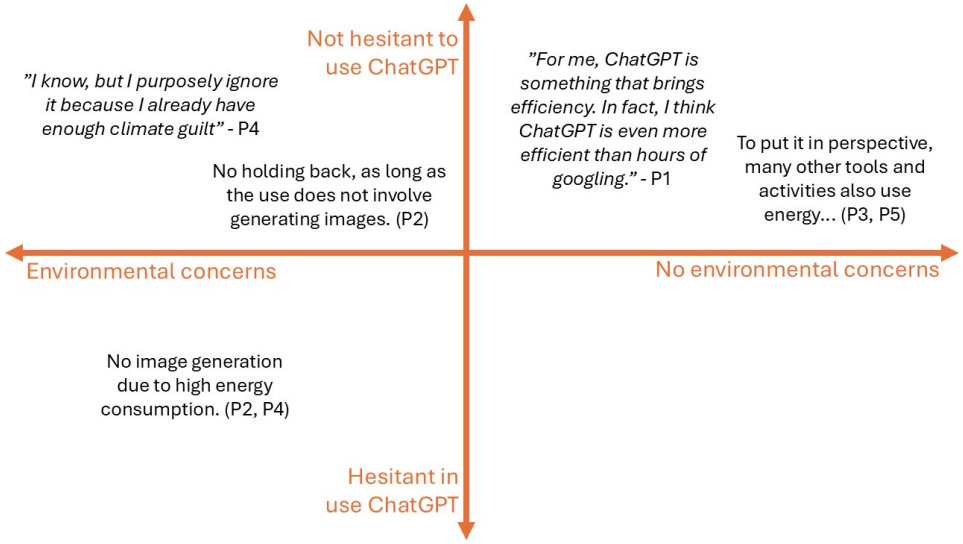


Figure 5.8: Environmental concerns influencing the GenAI use (Created by the author)

5.1.10. Community

The influence of community on the use of GenAI can be divided into two levels: the company and broader society.

Company

All participants noted that curiosity is the main driver behind exploring the use of ChatGPT, rather than any external pressure. While there is some encouragement, there is no pressure. For Participant 4, it is "curiosity about the added value", and for Participant 3, "curiosity driven by the urge to work more efficiently" is the main motivator for use. However, this curiosity is driven either by hearing from colleagues or by being shown examples of the added value of ChatGPT.
Participant 4 noted that the company helps stimulate usage through initiatives such as "fresh-up" sessions, which help to refresh ideas on how ChatGPT can be used. Additionally, the fact that the tool is provided and paid for by the company was seen as an encouraging factor. In contrast, Participant 5 did not find the company's initiatives stimulating. Instead, informal conversations and observing colleagues using ChatGPT were the only sources of encouragement, according to this participant. Participant 4 further highlighted the positive influence of colleagues actively using the tool. Nevertheless, all participants emphasised that using ChatGPT remains a personal choice, with no sense of pressure. Participant 5 also pointed out differences between teams, observing that in another team this participant works, the use of ChatGPT is more actively promoted.

Society

Regarding societal influence, all participants agreed that ChatGPT is rarely a topic of conversation in their social environments. Although they sometimes hear people mention it, they do not feel that these conversations take place very often and, moreover, do not influence their own use of the tool.

The participants generally do not feel pressured by either the company or society to use GenAI. Curiosity was identified as the main driver behind their engagement with the tool, and this is largely influenced by the company. This influence may come from enthusiastic colleagues sharing their experiences, from being shown the positive outcomes of using the tool, or from “fresh-up” sessions. However, the final responsibility for usage lies with the individual participant.

5.2. Findings Evaluation Sessions

Two evaluation sessions were used to validate and elaborate on the findings from both the experiment and the group discussion. This section presents the results from these evaluation sessions. Furthermore, the drivers and barriers to implementation that emerged from the evaluation sessions are discussed. As the participants of the experiment and group discussion are referred to as “participants” (Participant x or Px), the people who joined the evaluation sessions are referred to as “experts”.

5.2.1. Themes

The themes discussed in this subsection represent the findings from the evaluation session. As a result, not all of the themes previously outlined in Section 5.1 are included here.

Trust and Reliance

Participants were generally positive about ChatGPT’s potential to support the review of tenders, with all agreeing that it can be a helpful tool. However, this usefulness depends on the user’s ability to formulate a clear and effective prompt, saying that there must be trust that the prompt is strong enough to generate a reliable response.
At the same time, several experts agreed on the need for a critical approach. One expert pointed out that ChatGPT does not always deliver the desired output, even with a well-formulated prompt, according to their opinion. Another expert highlighted the importance of setting clear boundaries for the tool, as this helps determine whether the answer can be considered trustworthy. Building on this, a participant explained that breaking down complex questions into smaller parts often improves the quality of ChatGPT’s responses, as the tool appears to better grasp simpler, more focused questions.

Despite these strategies, two experts noted that ChatGPT can still produce incorrect or unreliable responses, regardless of how well the prompt is written. It was also acknowledged that, even when clear boundaries are set, there is no guarantee the tool will provide the expected answer for the fact that ChatGPT simply cannot generate everything requested.
Furthermore, experts expressed concern about the unknowns regarding ChatGPT’s internal workings. As a “black box” system, users have no insight into the reasoning or steps behind the generated responses, making it difficult to fully trust or verify the process.

Explainability and Transparency

Experts were divided on the question of whether ChatGPT is self-explanatory and transparent. Some expressed scepticism, observing that ChatGPT occasionally provides irrelevant information that may confuse the user. In response to this, one expert noted that this person often requests references from ChatGPT to verify the validity of its responses, which helps assess its transparency and reliability. Another expert remarked that the quality of ChatGPT’s explanations largely depends on the prompt, a view echoed by another expert who shared that they consistently ask the model to justify its answers. Another expert was confident that ChatGPT can offer strong arguments when rightly prompted. Despite the different perspectives, every expert agreed that the clarity and transparency of ChatGPT responses is highly dependent on the quality of the prompts given.

Helpfulness

Regarding the theme of helpfulness, the experts were asked to reflect on their agreement with the statements that ChatGPT can serve as a valuable source of inspiration and that it helps prevent tunnel vision or reveals blind spots in one’s thinking. One expert agreed, sharing an example from a tender process, where ChatGPT was asked to review a written tender against the client’s guidelines to identify any missing elements. The model noted a point that had been intentionally ignored earlier in the process. The expert added that in this case, ChatGPT was effective in finding missing parts of the document and that it can actually point out blind spots.

Other experts also agreed with the statements, based on their own experiences. One expert added that, while ChatGPT is indeed helpful, human judgement remains essential for filtering and evaluating the relevance of the information it provides.

Objectivity

The objectivity of ChatGPT and its potential helpfulness in reviewing documents, the tool's lack of direct interest, its neutrality, and the fact that it does not attempt to direct or influence decisions were seen as benefits of the experiment. All of the experts agreed that ChatGPT is objective in its responses.

Time Efficiency

Helpfulness in terms of time efficiency was experienced both positively and negatively during the experiment, depending on the individual. The evaluation groups recognised the potential of ChatGPT to improve time efficiency. However, its effectiveness in this context was seen to rely heavily on the user's ability to instruct ChatGPT properly. As a result, some participants were unable to determine whether ChatGPT was truly time-efficient in practice.
It was also found that an additional review step involving ChatGPT increases the overall time spent on the process. However, the experts agreed that if meaningful output is gained from the tool, then time efficiency can be achieved.
One expert responded specifically to a quote about how a human reviewer might take a lot of time during a review session. However, according to this expert, the time invested by a human is valuable and cannot be fully replaced by a GenAI tool. Another expert pointed out that, in their view, Count & Cooper colleagues are still way more capable than ChatGPT when it comes to performing the task effectively.

Reflectivity

Reflectivity, such as helping users think more critically or allowing them to mirror their own thinking, most experts fully agreed with the statement that ChatGPT is reflective and therefore helpful. Only two experts noted that they had not used the tool enough to form an opinion on its reflective qualities.

ChatGPT Lacking Information

Most of the experts disagreed with the statement that ChatGPT lacks the necessary information to be helpful in the review process. They argued that, in principle, all relevant information can be entered into ChatGPT to fully update the tool on the case. In fact, several experts supported the view of one participant who stated that it is the user's responsibility to provide ChatGPT with the right input in order to receive the desired output.

Privacy

To assess alignment with opinions on privacy, the axes representing the level of scepticism about privacy security and its relationship to hesitation to use ChatGPT were shown (illustrated in Figure 5.6). Interestingly, all experts in the evaluation sessions reported no hesitation in using ChatGPT, even though they were sceptical about whether information would be kept confidential. Many experts mentioned that they felt comfortable using the tool because ChatGPT claims to keep information secure. They were also reassured by the fact that Count & Cooper had purchased a team licence and that employees were free to use the tool in their work. Although some remained doubtful that privacy could be fully guaranteed, this did not affect their use.
One expert made a notable comment that this person was not hesitant, but still sceptical. The expert explained that the tool is trained on a large amount of data. These models work by "regression to the mean", which means that they tend to give answers that are average. Therefore, the expert believes that when the model uses the data, it would not be a problem.

Servility

The experts were unsure whether they agreed with the idea that ChatGPT was too servile. One expert put this down entirely to prompting, explaining that ChatGPT appears to be serving when it does not clearly understand what is being asked. To avoid this, the expert stressed the importance of well-structured prompts, meaning breaking the request down into clear steps that ChatGPT can follow. Another expert agreed with this point and added that ChatGPT needs a clearly defined framework in order to operate effectively.

Sustainability

For the topic of sustainability, the axed in Figure 5.7, were presented which illustrate how environmental concerns relate to hesitancy in the use of ChatGPT. This visual representation allowed the experts to position themselves in relation to the quotes and findings from the experiment and group discussion. Most experts indicated that they do not actively think about energy consumption when using ChatGPT and, as a result, do not feel hesitant to use it. In fact, some had never even thought about the energy consumption of ChatGPT.

One expert acknowledged concerns around energy use, specifically when generating images, and therefore avoids generating images. Another expert stressed the efficiency that ChatGPT can offer, prioritising efficiency over environmental concerns. A third expert reflected on energy consumption, noting that as technology improves, it could become both more efficient and use less energy. This would ultimately reduce its environmental impact. A fourth expert reflected on energy consumption, noting that as technology improves, it could become both more efficient and use less energy. This would eventually reduce its energy consumption.

Finally, one expert commented that the energy use associated with a single person using a GenAI tool like ChatGPT is relatively insignificant on an individual level. Although this expert is informed about the energy consumption, this expert considers the personal impact too small to be a reason for hesitation and therefore continues to use the tool without hesitation.

Community

The community within the company was described as stimulating, particularly due to organised sessions aimed at inspiring colleagues and demonstrating the possibilities of using GenAI, creating curiosity. These sessions made people feel encouraged to explore the tool. In contrast, the environment outside the company was not seen as stimulating or supportive.

5.2.2. Drivers and Barriers to Implementation

This subsection lists the drivers and barriers for the implementation of GenAI that were identified together with the experts during the evaluation sessions. These are listed in Table 5.1. The drivers or barriers that have been written down the most are listed at the top.

Drivers to implementation	Barriers to implementation
Supporting creativity	Too serving / Doubt on helpfulness
Providing time efficiency / Increasing speed	Lack of specificity
Provided thoughtful prompt sets	Lack of transparency
Review capacity / Comparative reflection capacity	Doubting reliability / Non-accurate output
Helping to see other perspectives	Lacking information
Chat is an expert on all fronts	Reasoning issues
Confidentiality reassurance	Prompting problems
Providing reflectiveness	Lack of practical knowledge of GenAI
Providing objectivity	Review reliance issues
Helping structuring	Effort for reliability checks
Identifying quick wins	Effort for good prompts
Enriching the plan	

Table 5.1: Drivers and barriers to implementation according to the evaluation groups (Created by the author)

5.3. Structuring Findings

This section serves to structure the findings presented earlier in the chapter, with the aim of contributing to the answers of the sub-questions. The findings are organised into two categories: those from the experiment, and those from the group discussion and evaluation sessions. This distinction is made to align with the structure of the research and to support answering the sub-questions of the study. The categorised findings are analysed through the lens of the Activity Theory to identify contradictions underlying the use of GenAI, which are then interpreted as drivers and barriers. Finally, the evaluation framework is presented as a contribution to this research, summarising the insights gained throughout the study.

5.3.1. Experiment

The findings of the experiment aim to answer research question 2: *"How does human-generative Artificial Intelligence deliberation influence changes in individuals' opinions in the context of reviewing tenders?"* The experiment explored whether engagement in human–GenAI deliberation led to changes in participants' opinions.

A key indicator of such opinion changes was the shift in participants' agreement levels before and after the deliberation, as illustrated in Figure 5.4. Participants showed varying levels of openness to adjusting their opinions. Some showed significant shifts, while others mostly maintained their original opinions, making only small adjustments. These variations suggest that deliberation with a GenAI system like ChatGPT does not result in uniform shifts in opinion, but rather interacts with underlying personal characteristics such as critical thinking, openness, engagement, and the level of trust participants have in GenAI.

Age and Work Experience

Age and work experience appear to influence participants' openness to GenAI's reasoning. The youngest participant (P5) seemed to be more vulnerable to being influenced by GenAI's reasoning, showing greater openness and curiosity about its way of thinking. This younger participant expressed insecurities about being new to the field, mentioning feeling safe and unjudged when interacting with GenAI, suggesting a degree of professional insecurity. In contrast, the oldest, way more experienced participant (P1) showed greater hesitation to change their opinions, sometimes even adjusting their scores in the opposite direction to the GenAI system's suggestions. This participant did not engage by asking the system questions but rather provided additional information to the tool. Overall, the experienced participant seemed less willing to adopt information from the GenAI system, whereas the least experienced participant was more open to doing so (see differences in age and experience in Figure 5.1 and Figure 5.2).

Trust in GenAI

Participants with higher levels of trust in GenAI, such as P5, approached deliberation by asking questions and using GenAI as a validation tool. Trust was reflected both in viewing the GenAI system as a "safe" deliberation partner, someone who would not judge (as P5 noted), and in trusting the tool to reliably deliver expected outputs (as P1 highlighted). In contrast, P2 and P3 frequently questioned GenAI's output and showed minimal changes to their original scores. Overall, higher trust appears to correlate with greater opinion shifts, while scepticism is associated with limited change.

Engagement in the Deliberation

Participants who were more actively engaged in the deliberative conversation, particularly those who asked clarifying questions or encouraged the GenAI system to elaborate, were more likely to shift their judgments. For instance, P5 asked numerous clarifying questions and adjusted their ratings considerably towards the system's ratings. Conversely, participants who approached the GenAI system with scepticism (P3) or who maintained a more passive role in the conversation (P4) showed limited opinion change (see the "no change" bar in Figure 5.3).

Perception of GenAI Behaviour

Interestingly, despite their varying attitudes towards GenAI, all participants noted in the group discussion that the system tended to be highly servile, meaning that it quickly agreed with user inputs rather than challenging them. This observation made participants reconsider the value of GenAI, and this seems to reduce users' willingness to rely on it as a valuable tool for reviewing tenders.

In conclusion, participants' openness to GenAI was shaped by trust, experience, and engagement. Those who were younger, more trusting, and actively involved showed greater opinion shifts. However, GenAI's high servile nature raised doubts about its reliability for critical evaluations.

5.3.2. Group Discussion and Evaluation Sessions

As both the group discussion and the evaluation sessions were open discussions that encouraged participants to freely express their opinions, there is a lot of overlap in the data gathered from them. The findings reflect the perspectives of the participants involved in the experiment and the group discussion, as well as the experts who contributed to the evaluation sessions. Table 5.2 summarises the key

findings from both sources and organises them thematically for clarity and comparison. These findings aim to serve as a stepping stone toward answering the third research question: *"What do individuals perceive as drivers and barriers to using human–generative Artificial Intelligence deliberation for reviewing tenders?"*

In Table 5.2, some findings have been highlighted: findings that are important or prominent are indicated in orange, and findings in italic represent those that appeared in both the group discussion and the evaluation sessions. The results suggest that the way individuals use GenAI plays an important role in how they experience the interaction. This is based on the many references to prompting, which appeared to influence how useful, trustworthy and efficient the tool is perceived to be.
Participants generally perceived deliberating with GenAI as valuable, associating it with benefits such as inspiration and objective input. The perceived objectivity of the tool, arising from its lack of personal interest, combined with its reflective responses, was seen to encourage users to think more critically and reflect on their own reasoning. However, their willingness to adopt the tool seemed to depend on the level of trust they placed in it. This trust appears to grow when GenAI offers clear, transparent and explainable answers. Several participants expressed a desire to understand the reasoning behind the tool's suggestions, what steps it took and how it reasoned. In the context of explainability, some individuals noted that GenAI's explainability is good enough, depending on the prompt, while others described the explanations as unclear. The concept of a 'black box' was raised in this context, reflecting a desire for greater insight into how the tool operates.
Prompting was often discussed as a key influence. Experts, in particular, often pointed to the way prompts are phrased as a factor that could influence GenAI's responses. Additionally, the complexity of formulating effective prompts seems to be a factor of influence. Some individuals noted that it is not always clear how they should interact with the tool effectively.
Scepticism also emerged with regard to the tool's behaviour, particularly when it was described as being too servile, confirming input from the user without question. Some saw this tendency as limiting the tool's usefulness and influencing how much users would rely on it.
Helpfulness, by contrast, was discussed more consistently. Participants described how ChatGPT could encourage broader thinking, identify information gaps, and challenge assumptions. However, these benefits could only be realised if the user remained in control and continued to evaluate the tool's suggestions.
Opinions varied on the topic of time efficiency. Participants commonly referred to time efficiency as achieving the same outcome with less time and effort invested through the use of the tool. While some participants shared that the tool helped them speed up their work, others were less positive. They reported that it did not improve their efficiency, often highlighting a lack of knowledge about how to use the tool effectively. Once again, it seems that prompting plays a key role in shaping perceptions of the tool's efficiency.
Curiosity also emerged as a factor contributing to GenAI use. Participants mentioned their own curiosity as a key motivator for using GenAI. However, encouragement from others, colleagues or the company environment, to explore the capabilities of the tool is also a contributing factor. This kind of external influence seems to stimulate people's curiosity and influence their willingness to use the tool.

Overall, it seems that there is a lack of knowledge regarding the effective use of GenAI tools. Trust emerged as a key theme across these reflections, often influencing other themes and being influenced by them. This suggests that, while trust cannot be guaranteed, it may grow through clearer interaction and more confident use.

Theme	Findings group discussion	Findings evaluation sessions
Trust	*Source of inspiration* *Discussion partner* Trust on own intuition required when deliberating	*Source of inspiration* *Discussion partner* Trust depends on way of prompting
Reliance	*Reliability depends on user interaction, way of prompting* Less reliable, not always correct Reliance lowered due to ChatGPT's excessive agreement	*Reliability linked to user's ability of prompting*
Explainability and Transparency	Need for strong arguments to increase trust Openness to reconsider opinions if well-founded arguments are presented Divided opinions about explainability	Explainability depends on prompting Some find ChatGPT explainable when asking for justifications Others sceptical about occasional irrelevant information Concerns about ChatGPT being a "black box" system
Helpfulness	*Good for inspiration, critical assessment needed* *Helpful to prevent tunnel vision and blind spots*	*Human judgement remains essential for evaluating relevance* *Helpful to prevent tunnel vision and blind spots* Effective in finding missing parts
Objectivity	*No direct interest* *Neutrality*	*No direct interest* *Neutrality*
Time efficiency	Time efficiency depends Not time-efficient due to lack of knowledge	Time efficiency depends on prompting
Reflectivity	*Helping users think more critically* *Mirror thoughts*	*Helping users think more critically* *Mirror thoughts* Some could not reflect on this due to lack of experience
Lacking information	Context is missing, therefore not helpful	Lacking information depends on providing the right information
Privacy	Confidentiality as a selling point of GenAI *Confidentiality is trusted currently* Privacy concerns about future	*Confidentiality is trusted*
Scepticism	High servility, less trusted	Servility is the result of bad prompting
Sustainability	*No environmental concerns, not hesitant* *Efficiency over sustainability* Hesitant in energy consumption, not generating images	*No environmental concerns, not hesitant* *Efficiency over sustainability* Individual energy consumption is relatively insignificant
Community	*Stimulated by the company, but curiosity main driver*	*Feeling encouraged to explore the tool by the company* *Activating curiosity*

Table 5.2: Structured findings from the group discussion and evaluation sessions per theme (Created by the author)

5.4. Structure Findings through Activity Theory

The Activity Theory provides a lens for analysing an activity. Therefore, the data from the research is structured following this theory. In the literature review, it is stated that contradictions are important aspects influencing the outcome of an activity and can be interpreted both negatively and positively (Subsection 2.4.1). To specifically identify the drivers and barriers needed to answer the third research question, *"What do individuals perceive as drivers and barriers to using human–generative Artificial Intelligence deliberation for reviewing tenders?"*, it is important to first identify all the contradictions within the Activity Theory framework. In addition, the conditions that must be present to enable successful implementation are also considered.

5.4.1. Contradictions

There are multiple contradictions identified either within elements (primary contradictions) or between different elements (secondary contradictions) of the Activity Theory framework. All of these contradictions can affect the achievement of the outcome of the activity. These contradictions within the Activity Theory framework are visualised in Figure 5.9 and elaborated on below.

Subject – Subject

A primary contradiction within the subject itself is a lack of curiosity among individuals to engage with the GenAI tool (Figure 5.9a). Although some participants and experts suggested that curiosity could be stimulated by colleagues, overall, curiosity does not appear strong enough to effectively encourage engagement with the GenAI tool.

Subject – Tool

Several secondary contradictions are found between the subject (the human) and the tool (the GenAI system), Figure 5.9b. These include the lack of explainability from the GenAI tool to the human, GenAI's issue of not being critical due to its servility, the failure of the interaction to be deliberative, the lack of explainability and transparency from GenAI's side and the tension between the tool's objectivity and reflectivity versus the biases introduced by the human user. Additionally, prompting issues from humans towards the GenAI tool are observed.

Tool – Object

A secondary contradiction (Figure 5.9c) appeared between the tool and the object: the GenAI tool sometimes lacks sufficient information to enable effective human–GenAI interaction according to participants.

Subject – Object

Between the subject and the object appeared the efficiency assurance problem (Figure 5.9d). Participants differ in their perceptions of whether the use of GenAI is truly efficient, and this uncertainty contributes to the hesitation to use the tool.

Division of labour – Subject

A contradiction between the division of labour and the subject is also identified in Figure 5.9e. There seems to be a knowledge gap between the two elements, which indicates that individuals are not up to date with the possibilities offered by human–GenAI deliberation to assist in reviewing tenders.

Community – Subject

Finally, a contradiction is found between the community, specifically the company culture, and the subject (Figure 5.9f). Although some individuals feel encouraged to use GenAI, others believe that the community's stimulation efforts are insufficient to promote adoption. This refers to the encouragement provided by peers, colleagues, or the company itself, specifically aimed at supporting the task of reviewing a tender.

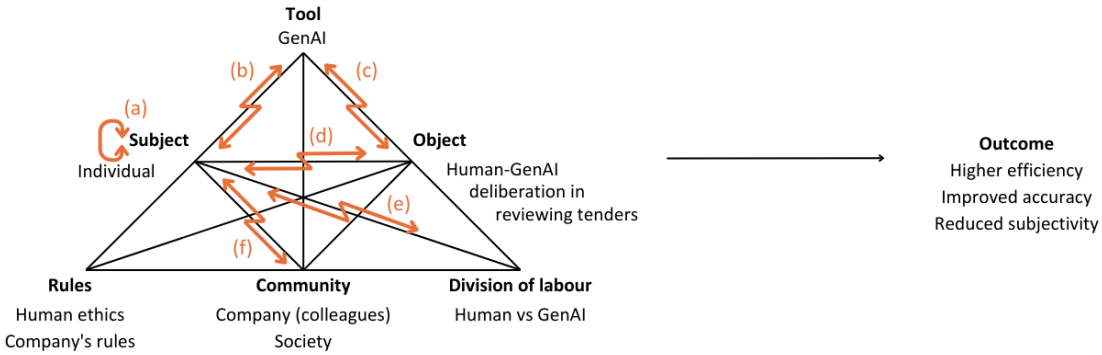


Figure 5.9: Concluding contradictions human-GenAI deliberation (Adapted from Engeström, 1987)

In this study, many of the identified contradictions seem to result from a single underlying issue. These include contradictions between the subject and the tool, (b) in Figure 5.9, the tool and the object (c), the subject and the object (d), and the division of labour and the subject (e). Examples include the tool's lack of explainability and limited transparency, its perceived servility, and prompting difficulties from the individual's side.

5.4.2. Conditions

In addition to the contradictions identified within the Activity Theory framework that shape the outcome of an activity, a condition is also identified from the findings. As mentioned in Section 2.4, conditions do not act as active drivers when present but become barriers when absent. An important condition identified in this research relates to privacy, specifically, the assurance that confidentiality is maintained. The findings show that, in this case, employees trust the company's assurance that using the GenAI tool is safe. As a result, they use the tool without fear of losing confidentiality or exposing sensitive information.

5.4.3. Drivers and Barriers

The findings from the group discussion and the evaluation sessions, structured through the lens of the Activity Theory, form the basis for answering the third research question: *"What do individuals perceive as drivers and barriers to using human–generative Artificial Intelligence deliberation for reviewing tenders?"* To address this question, Table 5.2 presents the most frequently mentioned outcomes from both the group discussion and the evaluation sessions. In addition, the identified contradictions, mapped within the Activity Theory framework in Figure 5.9, help to trace the origins of these findings within the system. Based on this analysis, the key drivers and barriers influencing the implementation of GenAI in tendering processes are outlined and explained in Table 5.3.

Several of these contradictions seem to result from one key aspect: the user's limited understanding of how to interact effectively with GenAI. This knowledge gap is a significant barrier that includes the smaller contradictions identified earlier in Subsection 5.4.1. Addressing this gap could serve as a key driver in improving the functioning of the activity system. Rather than relying on users to explore the tool independently, offering support may help facilitate a smoother integration of GenAI into existing workflows.

Drivers to Implementation	Barriers to Implementation
Clear rules of engagement Clear guidelines on how to interact with GenAI reduce uncertainty and help users formulate prompts correctly to receive the desired response.	**Perceived servility of GenAI** Doubts about the potential of GenAI due to its perceived servility lead to scepticism amongst users.
GenAI's objectivity and reflectivity GenAI's lack of personal interest encourages users to think more critically and reflect on their reasoning.	**Insufficient explainability and transparency** Difficulty in understanding how GenAI systems reach conclusions undermines trust and usability.
Increased efficiency Individuals do not commonly experience increased efficiency, yet its potential is highlighted as a key driver for GenAI use.	**Challenges in prompting** The difficulty in formulating effective prompts affects individuals' ability to extract meaningful results from GenAI.
Stimulation of curiosity When an individual's curiosity is stimulated, they seem to be more likely to engage with the tool. In addition, encouragement and success stories from colleagues can further motivate adoption.	**Limited Awareness of GenAI Capabilities** Individuals are not fully aware of the current possibilities for using GenAI to support reviewing tenders.

Table 5.3: Drivers and barriers to implementation (Created by the author)

5.5. Evaluation Framework

A framework has been developed to evaluate human–GenAI interaction within the specific context of this research. Figure 5.10 presents the resulting evaluation framework, which has been developed throughout this study. The framework is based on the Activity Theory, using the concept of contradictions as introduced by Engeström (1987) as a guiding lens. Within this evaluation framework, contradictions are categorised by type and linked to the corresponding elements of the activity system. Each element is then associated with specific contradicting factors. These factors have been derived from the literature (presented in Table 2.1) and further confirmed and expanded through the empirical findings, as explained in the methodology of this study (Chapter 3). The sources of the contradicting factors are visually distinguished by colour in Figure 5.10, differentiating between factors derived from empirical findings, the literature, and specifically the study by Ma et al. (2024). Alongside the development of the evaluation framework, a list of search codes was created during the data analysis to provide guidance for the analytical process. This coding approach is also described in the methodology of this research (Chapter 3). The complete list of codes is presented in Appendix A.

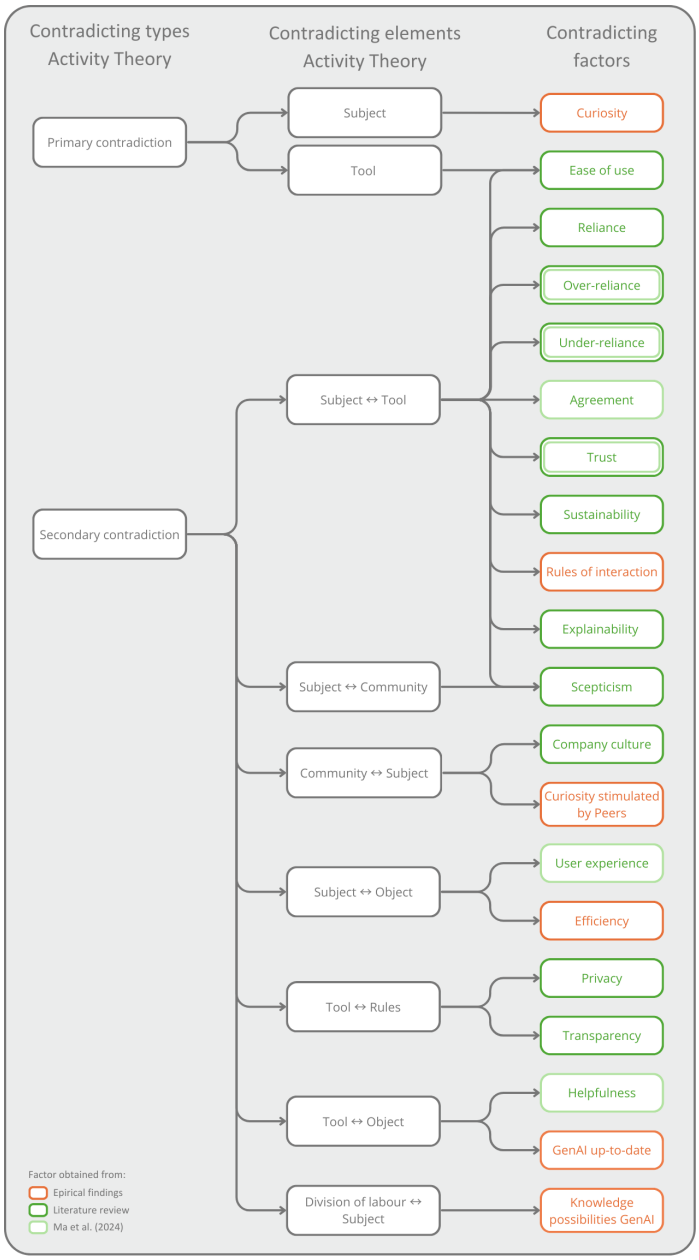


Figure 5.10: Evaluation framework (Created by the author)

6

Discussion

The chapter on discussion provides the link between the literature and the results of the empirical part of the study. The discussion begins with an interpretation of the findings in Section 6.1, followed by an assessment of this research's validity in Section 6.2, and concludes with a reflection on its limitations in Section 6.3.

6.1. Interpretation of Findings

6.1.1. Knowledge Gap Based on Multiple Contradictions

Contradictions, as highlighted in Activity Theory by Engeström (1987), are central to shaping activity systems. While such contradictions can act as barriers, overcoming them can also serve as important drivers for growth and innovation (Kuurti, 1995). In this study, many of the contradictions identified appeared to result from an underlying issue: the user's limited understanding of how to interact effectively with GenAI. This knowledge gap forms a key barrier, and addressing it could significantly enhance the functioning of the activity system. This finding reinforces the value of the Activity Theory framework as outlined in the literature. It supports the view of Kuurti (1995), who argues that overcoming contradictions can act as a driver for growth and innovation. By addressing the knowledge gap as the main barrier, the functioning of the activity system could be substantially improved.

6.1.2. Trust

Trust emerged repeatedly throughout this study as a critical factor influencing participants' use of GenAI. Similarly, the literature highlights trust as a central concept linked to multiple factors that affect the adoption of GenAI tools. First, reliance forms a requirement for trust: users must rely on the AI's performance before they can trust its intentions (de Fine Licht and Brülde, 2021). Second, explainability and transparency are essential for building trust, as emphasised by Shin (2021) and Kakolu et al. (2024). Participants in this study expressed a strong preference for well-reasoned arguments, which increased their trust in ChatGPT's outputs. However, some participants were sceptical about ChatGPT's limited ability to clarify or explain when challenged. This highlights a current limitation of GenAI systems in meeting user expectations for explanation and transparency. Following, participants generally trusted the GenAI system to handle data securely, largely due to Count & Cooper's adoption of a Team licence of ChatGPT. This finding supports the argument by Al-kfairy et al. (2024) that not having privacy concerns in GenAI tools is crucial for their full utilisation. Furthermore, company culture played an important role in encouraging the use of GenAI. The participants' willingness to engage with ChatGPT reflects a stimulated, innovation-oriented organisational environment. This aligns with the findings of Venkatesh (2022) and Chaudhuri et al. (2024), who emphasise that organisational support for digital transformation is a key driver of AI adoption.

6.1.3. Servility

The concept of servility emerged during the group discussion. Servility was generally perceived as something negative, as the GenAI consistently agreed with almost all the arguments presented by

the participants. Participants felt that the tool was not critical enough and therefore less trustworthy. However, it is important to consider an alternative view of servility. From a more positive perspective, servility can be seen as a way for the tool to show a supportive attitude. The GenAI's consistent agreement could be interpreted as a sign that it is always open to engaging with the user's argument and is willing to help. This openness could contribute to a more welcoming user experience. If the tool were to adopt a more critical attitude by frequently questioning the user's input, it might risk discouraging the user. The study by Jones-Jang and Park (2022) found that perceived controllability by a human is a key psychological factor: individuals tend to assess AI failures more harshly when they believe the AI had greater control, whereas they are more tolerant of unfavourable outcomes when the AI is perceived to have had less control. This suggests that a critical attitude, especially when it leads to outputs the user finds unfavourable, can negatively affect the user's perception of GenAI.

6.1.4. Over-Reliance on GenAI

Research suggests that people often over-rely on AI systems, even when the AI's suggestions contradict their own understanding of the situation (Klingbeil et al., 2024). This often occurs when users do not question the AI's recommendations (Zhai et al., 2024). Furthermore, Ma et al. (2024) found that individuals tend to either over-rely on incorrect AI suggestions or dismiss correct ones.
This over-reliance is also reflected in the empirical findings of this study. One participant has stated that the GenAI tool served as a form of validation due to this participant's own insecurities as a newcomer to the work field. This statement suggests a strong reliance on the tool, which may indicate over-reliance. Such reliance is aligned with trends described in the literature, where users often over-rely on an AI.

6.2. Validity of the Research

6.2.1. Adoption of Existing Frameworks

The Activity Theory framework by Engeström (1987) played a significant role in this study. It served as a lens for both structuring and interpreting the context of human–GenAI deliberation, as well as for analysing the research findings. Using an established framework helps to focus a study, guide the research design, and ensure that data collection and analysis are aligned with recognised theoretical constructs (Miles et al., 2014). In this case, the use of the Activity Theory contributed to the validity of the research by providing a structured and tested foundation for both the design and the analysis. Additionally, it helped to narrow the scope of the study. In designing the research, the Activity Theory played a key role in outlining the study's context. By applying the Activity Theory framework to the specific setting of this research, a clear picture was formed of the activity system under investigation. Additionally, the concept of contradictions, as introduced by Engeström (1987), served as a central focus of the study, guiding the identification of both drivers and barriers that influence the adoption of GenAI in infrastructure tendering. These contradictions not only shaped the analytical focus but also provided a structure for presenting the research findings.
The use of another existing framework from Ma et al. (2024) strengthened the methodological foundation of this study and ensured alignment with prior research. This choice also saved time that could be dedicated to other aspects of the research. The framework's previous application to a large participant group supports its credibility and contributes to the internal validity of this study. The contrast in prompting methods between the two studies introduced variation, which may affect the consistency and comparability of outcomes. However, the repetition of key themes, such as user scepticism, across both studies reinforces the reliability of the findings of this study. While scepticism appeared in both cases, its presence did not prevent users from engaging with GenAI. This consistency provides evidence for the external validation of the results.

6.2.2. Combining Research Methods

This research used a combination of methods to gain an understanding of human-GenAI deliberation. The experiment provided a controlled, individual method of data collection. In addition, the group discussion allowed participants to share their reflections and engage with each other's perspectives, facilitating collaborative insights. By combining an experiment with a group discussion, the study provides a richer perspective and deeper understanding of both individual behaviours and collaborative reflections. In addition, the inclusion of evaluation sessions added a layer of validation to the findings and further enriched the data by incorporating expert perspectives.

6.2.3. Variation in Participant Group
For the experiment, group discussion, and evaluation sessions, a diverse group of participants and experts was selected. For the experiment and group discussion, participants were chosen from all different layers of the company, representing a range of working experience. This ensured a well-rounded reflection of the team involved in tendering at Count & Cooper. The research has shown that differences in working experience do indeed influence perspectives on human–GenAI deliberation. The same approach was applied when selecting the expert group, ensuring a variety of professional backgrounds and experience levels were represented.

6.2.4. Pre- and Post-Deliberation Comparison
Despite the small sample size used in this research, it was still possible to compare participants' ratings before and after the deliberative conversation. This allows for a comparison between the scores given by humans alone and those given following human–GenAI deliberation. The ability to make this comparison adds depth and value to the study's findings.

6.3. Limitations of the Research
6.3.1. Data Saturation
In the current study, there may not have been enough iterations or reflection during the experiment and the group discussion. Some additional insights were gained after the group discussion, such as allowing participants to walk around again and add or change their scores and reasoning. But given that only one group discussion was conducted, it was not possible to achieve data saturation. Several group discussions would have been beneficial to reach saturation. Repeating the experiment or discussion session could have provided an opportunity to refine the focus or make improvements based on previous findings. For example, Van den Berg et al. (2021) incorporated their findings directly into the ongoing research process, allowing them to immediately adjust and sharpen the focus of the study. This approach could also have strengthened the current research. Although such iteration was not applied during the experimental phase, some of this reflective improvement occurred later, during the evaluation session, so in a sense, a simplified form of this approach was implemented after all.

6.3.2. Interpretation of Terminology
Interpretation of terminology may have influenced the results of this research. At various stages of the study, differences in understanding or interpretation of certain terms could have led to different responses or conclusions. These differences in interpretation could have come from the researcher, the participants in the experiment and group discussion, or during the evaluation sessions.
An example of such complexity is the interpretation of the themes of trust and reliance. These concepts are related but not identical: trust generally includes reliance, but one can rely on something without necessarily trusting it, but not the other way around. Distinguishing between the two was difficult. The difficulty was further complicated by the fact that the research was conducted in Dutch. In Dutch, there is no clear distinction between trust and reliance, as both are often translated as a single term. This language limitation made it more difficult to distinguish between the two and, therefore, could have resulted in a higher variation due to different interpretations of the terminology.

6.3.3. GenAI as a Deliberative Partner
An important point for discussion is whether GenAI has truly functioned as a deliberative partner, or whether it is more correctly described as a tender review tool. The original aim of this research was to develop a tool that would enable individuals to engage in deliberation and thereby assist them in reviewing tenders. However, the experiment, which was designed to assess whether human-GenAI deliberation influenced opinion change, revealed limited interaction, making it difficult to draw general conclusions about the deliberative impact of the tool. As a result, in the absence of much deliberative exchange, the tool developed in this study may be better characterised as a tender review tool rather than a true deliberative partner. In practice, this difference is further supported by a certain amount of scepticism among users about fully relying on GenAI for tender evaluation. Given this hesitation, it is likely that users will engage with GenAI as a supportive review tool, one that offers suggestions and perspectives, rather than as a deliberative partner.

In addition to discussing whether GenAI has truly functioned as a deliberative partner, it is important to consider the possibility that the results of this research regarding deliberative conversations may have been influenced by participants' general experiences of using GenAI. During the group discussion, many participants shared personal experiences relating to their own use of GenAI. Consequently, the focus of the discussion was not solely on the deliberative interaction that occurred during the experiment. These personal reflections may have influenced the research findings, potentially affecting the clarity and specificity of the results related to the deliberative process itself.

6.3.4. Programming Deliberative AI

The Ma et al. (2024) study used a custom-built AI specifically designed to initiate and maintain discussion, thereby ensuring that deliberation would proceed. In contrast, this experiment used ChatGPT, where the responsibility for initiating and maintaining the interaction lies entirely with the human user. While it would have been possible to develop a custom AI system for the experiment, this was outside the scope of the project. In addition, because of Count & Cooper's access to a secure team licence for ChatGPT, which also ensured the protection of confidential information. However, ChatGPT was instructed to keep the discussion going, which it generally did. Although a custom-built AI could have provided more opportunities to guide and control the model's behaviour. This difference can be seen as a limitation of the current experiment, as it placed more responsibility on the participants to keep the conversation going themselves.

6.3.5. Challenges in Prompting GenAI

A key challenge in the experiment was how ChatGPT was instructed. Given the many possible ways to prompt the model, even small variations in wording can lead to significantly different results. The outcome of the deliberative conversations was likely influenced by the two types of prompting used in the study. Firstly, the initial instructions given by the researcher were crucial in guiding ChatGPT to behave deliberatively. For example, if the model had been prompted to provide more critical pushback, the conversations themselves and outcomes of the conversations might have been very different. Second, the prompts given by the participants during their interactions also played a crucial role in shaping ChatGPT's responses. Both types of prompting created a certain amount of variability in the results.

Another issue was that ChatGPT may not have received enough contextual information, particularly in terms of abbreviations and specific terminology used in the tender documents. In at least one case, this led to confusing ChatGPT, resulting in a significantly different interpretation of a question than that of the participants. This highlights the need for clear and complete input when using GenAI for detailed and context-specific assessments.

An additional limitation was the absence of the dialogue phase, which typically takes place in real bidding processes. These conversations, both intern and with the client, are key to understanding the client's needs. Leaving them out of the experiment may have left ChatGPT without essential context, which could have affected the quality and accuracy of the review. On the other hand, participants were also not provided with the results of the dialogue phase, which sort of balanced this out.
While the inclusion of insights from the customer dialogues would have enriched the experiment, it would have significantly increased its scope and complexity. Importantly, in practice, it is the responsibility of the user to provide ChatGPT with relevant context when interacting with ChatGPT.

6.3.6. Rapid Development GenAI

Another limitation of this research lies in the ongoing and rapid development of GenAI and other large language models (LLMs). This rapid development brings two major challenges. Firstly, by the time the research is completed, it risks being outdated due to continual advancements in the tool. This creates a dilemma: whether to regularly update the research direction to reflect the latest developments, potentially affecting uniformity, or to remain committed to a particular version of the tool, knowing that it may no longer reflect the current status of the technology.
Secondly, rapid development leads to variation in user experience. While some individuals are already experts in the field in implementing such a tool, others may only be beginning to explore GenAI. This creates a considerable gap between experienced and less experienced users, even within the same professional field, which can influence both the research outcomes and their generalisability.

However, this study included participants and experts with varying levels of experience with GenAI. Differences in experience were not a barrier to participation in the experiment, which makes the findings relevant and valuable across differences in GenAI user experience.

6.3.7. Uncertainty in Decision-Making

The selection of a suitable tender was based on specific criteria to ensure its relevance and usability in the experiment. The tender had to follow an open procedure, relate to an infrastructure project, and include both strong and weaker aspects to support a balanced and critical evaluation by participants. Clear and concise client feedback was also desirable to guide the assessment process. However, most tenders lacked detailed feedback, often offering only brief comments on entire multi-page documents. As a result, identifying a clear 'right' answer for a specific section proved difficult. This means that in the deliberative conversations between participants and GenAI during the experiment, there was no definitive 'right' answer to the questions discussed. This contrasts with the study by Ma et al. (2024), from which the Human–GenAI Deliberation Framework was adapted, where a correct answer to the deliberated subject was available. In this study, however, the decisions made by both participants and GenAI could not be evaluated against a known correct outcome. This introduced a degree of uncertainty and added complexity to the experimental design.

A counterargument to this limitation, however, is that this uncertainty reflects the nature of real-world tendering, where interpreting limited feedback and navigating unclear expectations is often part of the process. In practice, definitive 'right answers' are rarely available. Therefore, the absence of clear answers in the experiment increased its realism.

6.3.8. Limited Number of Involved Parties and Participants

For this research, a limited number of parties and participants were involved in both the experiment and the group discussion, as elaborated below.

Limited Number of Parties

The study focused specifically on the consultant and did not include the contractor or the client in the experiment and the group discussion. It would have been valuable to compare the consultant's perspective with that of the other contracting parties. Including and comparing the views of both the contractor and the client could have offered a more complete understanding of the dynamics between all parties involved.

Limited Number of Participants

The experiment and the group discussion were conducted with a relatively small sample size of five participants. The company where the thesis internship was done did not allow for a much larger sample size. While the study of Ma et al. (2024) had a very large number of participants, this research was adjusted for the smaller group by decreasing the number of measurements. However, the limited number of participants may not have been sufficient to draw solid or generalisable conclusions.

As a result of the small sample size, the type of participants involved may have had a disproportionate influence on the study's outcomes. Consequently, the presence of specific participant profiles may have shaped the results in ways that are not broadly representative.

6.3.9. Human-GenAI Entanglement

Human-GenAI interaction can be seen as a form of HCI, which involves the evaluation and implementation of computer systems for human use (Hewett et al., 1995). When humans interact with GenAI, these actors intra-act, a concept that describes how the ongoing interaction leads to mutual transformation, with both the human and the GenAI changing in response to each other (Frauenberger, 2019). In this way, working with GenAI can influence how people think and behave, just as it can influence how the GenAI performs over time. This close and evolving relationship is often referred to as entanglement, and it carries certain risks (Frauenberger, 2019). For example, Frauenberger notes that users may lose the ability to critically question systems with which they have become entangled. While this thesis does not explore these potential effects, they are very important when considering the long-term implications of the use of GenAI.

To reflect on whether the human–GenAI entanglement influenced the findings of this research, it can be stated that, in terms of entanglement during the deliberation itself, this was likely not the case. The

deliberation sessions in the experiment lasted approximately fifteen minutes, which does not seem to be sufficient time for a deep level of entanglement to develop. However, participants who engaged more intensively in the deliberative conversation, particularly those with prior experience using GenAI outside the experiment, may have experienced some degree of entanglement. In such cases, both the human and the GenAI could have been influenced by the interaction. This observation also relates to the issue of possible over-reliance, as discussed in Subsection 6.1.4. As one participant, for example, showed signs of over-relying on GenAI, which could also be interpreted as a form of entanglement in the decision-making process.

7

Conclusion

The concluding chapter begins by addressing the three sub-questions that structured the thesis in Section 7.1, followed by the answer to the main research question (Section 7.2). Next, Section 7.3 outlines the theoretical contributions of the study, followed by the practical contributions in Section 7.4. Recommendations for practice are then discussed in Section 7.5, with Section 7.6 concluding the chapter by presenting suggestions for future research.

7.1. Answers Sub-Questions

This section provides answers to all the sub-questions formulated in this study. These answers are based on insights from the literature review and the empirical findings from the experiment, group discussion and evaluation sessions.

7.1.1. Sub-Question 1

For answering the first research question, the literature is used and is elaborated upon by the empirical findings of this research. This research question states: *"What factors should be considered when using human-generative Artificial Intelligence deliberation to review tenders?"*

When using human–GenAI deliberation for tender reviews, several factors should be considered, as they shape how individuals interact with and perceive the tool. All factors that play a role are shown in the evaluation framework, presented in the findings of this research (Figure 5.10). One key factor within the list of factors is trust, which forms the basis of many other factors. Trust can be strengthened or weakened depending on the system's explainability and transparency, and the level of data confidentiality it offers. A lack of explainability and transparency can make it difficult to understand how GenAI reaches its conclusions, which could make the tool less trustworthy. Scepticism and user experience are closely tied to this. When individuals doubt the system's relevance or have had negative experiences with it, they could be less likely to adopt it. Similarly, experiences of over- or under-reliance on the tool could create problems during use. While reliance on GenAI is necessary to benefit from its potential, users could also seek a balance in which human judgement remains essential. Positive agreement between human and GenAI outputs may encourage further adoption. Curiosity could also be another influential factor. When users are curious about the tool, particularly when stimulated by peers through shared experiences or support, they could be more likely to explore and adopt it. However, clear rules of interaction could be necessary to translate curiosity into meaningful engagement. These rules could reduce uncertainty and help users understand how to communicate effectively with GenAI. Practical elements could also affect usage. The system's ease of use could lower barriers to entry, while perceived efficiency, such as time savings, could make it more attractive. Furthermore, individuals could be more likely to engage with GenAI if they believe that it is up to date and capable of providing relevant information. This belief is reinforced when users are aware of GenAI's capabilities and limitations, enabling them to use it more accurately and productively. Wider organisational and ethical considerations also play a role. A company culture that encourages innovation and experimen-

tation could support adoption, while concerns about sustainability, such as the environmental impact of AI tools, could have a negative influence. Lastly, users' willingness to engage could also be influenced by the general helpfulness of GenAI.

7.1.2. Sub-Question 2

A conclusion can be drawn from the findings of the experiment, which serve as the foundation for addressing the second research question: *"How does human-generative Artificial Intelligence deliberation influence changes in individuals' opinions in the context of reviewing tenders?"*

Participants showed varying degrees of openness to changing their opinions after engaging with a GenAI system. Some shifted significantly, while others remained largely unchanged. This suggests that factors such as critical thinking, openness, trust and engagement influence how people interact with GenAI. Younger and less experienced individuals were generally more open and curious, partly due to professional insecurity, with the potential risk of over-reliance on GenAI output. In contrast, older, more experienced individuals were more resistant to GenAI's suggestions. Trust in the system was a key factor: those who trusted it were more likely to ask questions and adjust their views, whereas sceptical users made minimal changes. Active engagement in the conversation also correlated with greater shifts in opinion. Despite these differences, all participants noted that GenAI tended to be highly servile, being too agreeable and rarely challenging input, which reduced their confidence in its value for critical tasks.

7.1.3. Sub-Question 3

The findings from the group discussion and the evaluation sessions provide an answer to the third research question: *"What do individuals perceive as drivers and barriers to using human–generative Artificial Intelligence deliberation for reviewing tenders?"*

Several aspects were identified as potential drivers for the implementation of GenAI in the context of tender reviews, which are listed in the findings, Table 5.3. First, clear rules of engagement, such as structured guidance on how to interact with GenAI, can help reduce uncertainty and assist users in formulating effective prompts, and would therefore be a key driver. Second, GenAI's objectivity, lack of personal interest, and reflectivity can encourage more critical thinking and self-reflection in users, potentially enriching their decision-making process. Third, although not always directly experienced, increased efficiency is often perceived as a benefit of using GenAI. The expectation that the tool can save time contributes to its perceived driver. Finally, curiosity and peer support play an important role in encouraging engagement. If users are curious about the tool's capabilities, they are more likely to explore its potential. This effect is reinforced by positive examples from colleagues, which can motivate users to embrace the tool and experiment with it more confidently.

At the same time, several barriers to the implementation of GenAI in tender reviews were identified. First, doubts about the potential of GenAI arise from its perceived servility, which can lead to scepticism among users regarding its usefulness. Second, a lack of explainability and transparency makes it difficult for users to understand how GenAI arrives at its conclusions. This undermines both trust and usability. Third, the difficulty in formulating effective prompts limits users' ability to extract meaningful and relevant responses. Finally, many individuals are not fully aware of the current possibilities for using GenAI to support tender review processes, which restricts the tool's perceived value and limits experimentation with its potential applications.
A key barrier that arises from the findings is the limited user understanding of how to interact effectively with GenAI. This knowledge gap appears to underlie several smaller contradictions and contributes to multiple contradictions within the activity system. By addressing this barrier, there is potential to activate a key driver. Instead of expecting users to investigate everything on their own, providing structured support could help reduce contradictions and drive the integration of GenAI into existing workflows.

Alongside the key drivers and barriers, the research identified privacy assurance as a crucial condition for GenAI usage. Although its presence does not actively encourage its use, its absence would create a significant barrier. Therefore, it is important to address this issue. Most employees were able to use the tool without hesitation or concern, as the company's confidentiality guaranteed the users.

7.2. Answers Research Question

Based on the study's findings, a conclusion can be drawn in response to the main research question of this research: *"How do individuals experience human-generative Artificial Intelligence deliberation when reviewing open infrastructure tenders within project-based organisations?"*

Participants generally viewed GenAI deliberation as potentially valuable, particularly due to its ability to provide objective and reflective input, offer inspiration, and potentially enhance time efficiency. However, their willingness to adopt the tool was closely tied to the level of trust they placed in it. Trust appeared to be higher when GenAI's responses were transparent and explainable. However, when this transparency and explainability are lacking, this could act as a barrier. Trust can also be encouraged through the stimulation of curiosity, often triggered by colleagues or shaped by the company culture. Stimulating curiosity was identified as an important driver in encouraging the adoption of GenAI tools. A key barrier to GenAI implementation is user scepticism, particularly in response to the tool's perceived servile behaviour. More generally, many of the contradictions identified in this study appear to result from a central issue: the user's limited understanding of how to engage effectively with GenAI. This knowledge gap includes difficulties with prompt formulation and a general lack of awareness of GenAI's potential as a supportive resource. As such, the knowledge gap represents a significant barrier that also includes several smaller contradictions. Addressing this barrier could, in turn, act as a key driver in improving the overall functioning of the activity system. Rather than relying on users to explore the tool independently, providing support may help facilitate a smoother and more effective integration of GenAI into existing workflows, while also enabling individuals to experience their interaction with GenAI more positively. And since trust has been identified as a key theme, it is essential to provide users with the necessary knowledge to use GenAI effectively to build trust and to possibly overcome related barriers. At the same time, guidance is important to prevent the risk that individuals may over-rely on the tool, as over-reliance could undermine critical thinking and independent judgement.

7.3. Theoretical Contributions

7.3.1. Activity Theory in Human-GenAI Context

This research contributed to a new application of Activity Theory in the specific context of human-GenAI deliberation in infrastructure tendering. This application expands the applicability of the Activity Theory to this field and demonstrates its relevance for the analysis of socio-technical interactions.
Furthermore, applying the Activity Theory as a lens to structure and interpret the findings proved valuable for drawing conclusions. Several contradictions were identified within the activity system. Notably, many of these contradictions occurred between the tool and the subject, that is, between the GenAI system and the individual. This suggests that, when studying similar contexts, particular attention should be given to this secondary contradiction, as it appears to play a central role in shaping the interaction.

7.3.2. Role of GenAI in Contracting Parties

This study explores the use of GenAI from the perspective of a consultant working in collaboration with a contractor, known as the contracting party. This viewpoint has been underrepresented in existing research, which has mainly focused on the client's side. By focusing on the consultant's role, this study contributes to addressing the knowledge gap surrounding the application of GenAI within contracting parties.

7.3.3. Evaluation Framework

The evaluation framework developed through this study offers an approach for researching interactions and structuring analysis. Findings from all stages of the study, including the literature review, the experiment, the group discussion, and the evaluation sessions, provided insights into the key factors to consider when researching human–GenAI interaction. These findings were examined through the lens of the Activity Theory, which supported the design of the framework. The evaluation framework has been specifically developed to study human–GenAI interaction in the context of reviewing infrastructure tenders, with a particular emphasis on deliberative conversations. For research conducted in similar settings, it offers a clear structure to support the interpretation of qualitative data. However, the frame-

work could also be applied to analyse other forms of interaction between individuals and GenAI, when the underlying activity system is likely to share significant overlap. If used in different contexts, some contextual adaptations may be necessary.

7.4. Practical Contributions

7.4.1. Tender Review Process Optimisation

This study proposes the integration of a human–GenAI deliberation step into the existing tender review workflow. This addition represents a practical contribution to improving the review of tenders. By incorporating structured interactions with GenAI before submission, organisations can encourage critical reflection and help prevent tunnel vision.

The contradictions identified through the Activity Theory framework highlight both drivers and barriers to adopting human–GenAI interaction. By referring to the overview of these drivers and barriers presented in this study, companies such as Count & Cooper can proactively address relevant organisational contradictions. In doing so, users can gain a better understanding of how to interact with the tool, curiosity can be fostered, and efficiency may be improved. These insights provide a foundation for developing targeted change strategies supporting organisations in the effective integration of GenAI tools into their existing workflows.

7.4.2. Documents Describing Role of GenAI

Another practical contribution of this study is the development of two guiding documents that define the role of GenAI, specifically ChatGPT, when conducting a review and when facilitating a deliberative conversation about that review. Both documents were created to ensure that GenAI operates according to the rules for each task: reviewing and deliberating.

The document outlining GenAI's role in the review process includes the relevant review guidelines, instructions for incorporating supporting documents specific to the context, rules that GenAI should follow during the review, and a clear structure for how responses should be formulated.

The document defining ChatGPT's role in deliberative conversations provides a step-by-step guide for facilitating such interactions. It includes instructions on how GenAI should address the user, what types of questions it should ask, and how it should engage to foster a discussion.

These documents can serve as foundational templates for organisations aiming to design similar tools. By adapting these documents, organisations can create their own GenAI-supported review or deliberation tools designed to meet their specific needs.

7.5. Recommendations for Practice

7.5.1. Human-GenAI Deliberation as Extra Step in Review Process

The experiment in this research specifically created a setting to explore the use of human–GenAI deliberation as an additional step in the review of tenders. Within this setting, it may be worth considering the introduction of human–GenAI deliberation as an additional step in the real tender process, similar to the approach taken in the experiment, without replacing any existing elements. This step could be particularly valuable shortly before final submission, serving as an extra layer of review to encourage reflection on key aspects of the tender.

The potential value of GenAI in this role lies in its objectivity and capacity to support reflective thinking. As GenAI has no direct interest in the outcome, it may help users think more critically and examine their own reasoning more thoroughly. While participants saw GenAI as a useful tool for reflection and inspiration, it was not seen as a standalone review tool capable of replacing a human.

Organisations may want to explore the use of GenAI as a complementary support tool in the tender review process. For this to be effective, it might be helpful to provide users with structured support, such as clear guidance, practical examples, and training. Rather than expecting users to explore the tool independently, offering this support could help integrate GenAI more smoothly into existing workflows.

7.5.2. Supporting GenAI Adoption

To support the effective use of GenAI, it could be valuable to address the knowledge gap between users and the tool. The findings show that individuals recognise potential in using GenAI to support the review of tenders and to help break through tunnel vision. However, many of the contradictions

identified in this research originate from a mismatch in understanding, particularly in how users engage with GenAI. Organisations may consider providing targeted training that focuses on developing prompt generation skills and understanding how GenAI generates its responses. This could help to increase explainability, reduce scepticism and build trust in the tool.
In addition, peer learning could be encouraged to allow users to share successful experiences and practical tips, which could help to stimulate curiosity and support exploration. Addressing perceptions of GenAI's servility by positioning it as a collaborative partner, rather than a passive assistant, could also encourage greater engagement.
Finally, organisations could consider providing clear and structured guidance, examples of which are outlined in Subsection 7.5.1, to support users in interacting effectively with GenAI systems. Such guidance may help users develop more productive and informed ways of working with the tool.

7.6. Recommendations for Future Research

7.6.1. Up-Scaling Study

This study was based on qualitative research, offering small-scale insights. The limited sample size is seen as an aspect limiting the study, restricting the generalisability of the conclusions. To strengthen and validate these findings, future research should adopt large-scale quantitative methodologies. Conducting experimental studies with broader and more diverse populations across different organisations would make it possible to statistically test the drivers, barriers, and mechanisms identified in this research. Finally, quantitative validation would allow for clearer conclusions about the effectiveness of human–GenAI deliberation in tender reviews, supporting broader implementation across the construction sector.

7.6.2. Enriching Prompting Techniques

The importance of effective prompting was a central observation in this study. However, further research is needed to investigate how prompting strategies influence GenAI performance during tender reviews. This could include the development of validated prompt templates specifically suited for reviewing tenders, or a study on splitting the prompt into pieces, which has been addressed by one of the experts. Setting up proven prompting techniques would contribute to making GenAI support more reliable and efficient for practical use in tendering processes.

7.6.3. Standardised Review Tool

Future research could design a standardised chatbot tool specifically for infrastructure tendering. Such a tool would offer consultants a consistent way to use GenAI in reviews, adapted to the specific needs of the sector. Investigating how such a tool fits into workflows would help ensure practical adoption. As part of this study, documents were developed that elaborate on the role of GenAI in reviewing a tender and in leading a deliberative conversation. These documents could be further enriched through research aimed at programming a standardised chatbot tool based on their content.

7.6.4. Addressing Contradictions

It is recommended that future research explores how to address the contradictions found between or within elements of the activity system, as conceptualised by the Activity Theory. Investigating how these contradictions can be meaningfully addressed may provide valuable insights for improving Human–GenAI collaboration. While this study has highlighted key drivers and barriers to human–GenAI collaboration, further investigation could offer valuable insights into how these drivers might be strengthened and how barriers could be mitigated.
Special attention may be given to trust: how trust is built, how it is damaged, and what specific mechanisms can lower trust-related barriers. Since trust is a complex and critical element, also in human–GenAI interaction, understanding and influencing its dynamics could support broader adoption.

7.6.5. Hesitancy GenAI Use

In the findings, no individuals expressed hesitancy to use GenAI while also holding privacy or environmental concerns. It appears that all other combinations were represented: individuals who were hesitant due to such concerns, and individuals who were not hesitant, both with and without concerns. This

absence is notable and raises the question of whether such individuals were genuinely absent from the participant group or whether this reflects a limitation in data collection or participant diversity. It is important to consider the possibility that individuals with privacy or environmental concerns may still be hesitant to use GenAI tools, and that this perspective may not have been fully captured in the study.

7.6.6. Broadening Application Human-GenAI Deliberation

Human–GenAI deliberation methods could be explored further in other areas than consultancy firms within the infrastructure tendering sector. For example, this approach could be expanded to include contractors or other parties involved in the broader construction tendering sector. It could also potentially be applied to later project phases, such as project planning, stakeholder analysis, and sustainability evaluations. Broadening the application of human–GenAI deliberation in these areas could reveal greater potential for GenAI assistance.

7.6.7. Exploring GenAI-GenAI Interaction

The interaction not considered in this study is the potential interaction between the consultant's GenAI and the contractor's GenAI. While collaboration between consultants and contractors typically focuses on human-to-human interaction, and this study focuses on the interaction between GenAI and an individual within the consulting organisation, it could be interesting to explore the interaction between the GenAI systems of both parties.

At present, such interaction does not seem feasible due to current practices at Count & Cooper, where a restricted version of GenAI is used to ensure that confidential information does not leave the company. Furthermore, as consultants often work across multiple clients, maintaining consistent GenAI-to-GenAI communication across different organisations poses an additional challenge.

However, this may become more viable in the future. In this study, a GPT-based environment was established using a tailored version of ChatGPT. Looking ahead, it may be both possible and beneficial for consultants and contractors to each operate their collaboratively customised GenAI within a shared, secure environment.

References

Adepoju, O. D., Tijani, B., & Karera, S. (2024). Artificial intelligence skepticism in career domains. *International Journal for Digital Society*, *15*, 1880–1888. https://doi.org/10.20533/ijds.2040.2570.2024.0236

Al-kfairy, M., Mustafa, D., Kshetri, N., Insiew, M., & Alfandi, O. (2024). Ethical challenges and solutions of generative ai: An interdisciplinary perspective. *Informatics*, *11*, 58. https://doi.org/10.3390/informatics11030058

Alonso, V., & Puente, P. D. L. (2018). System transparency in shared autonomy: A mini review. https://doi.org/10.3389/fnbot.2018.00083

Alvarez-Melis, D., Kaur, H., Iii, H. D., Wallach, H., & Vaughan, J. W. (2021). From human explanation to model interpretability: A framework based on weight of evidence. www.aaai.org

Arunkumar, A., Pandiyarajan, P., Akshaya, S. B. L., Archana, A., & Archana, A. (2024). Renewable energy consumption on solar and wind energy prediction using deep learning and generative ai. *2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS)*, 770–777. https://doi.org/10.1109/ICICNIS64247.2024.10823172

ATLAS.ti, S. S. D. G. (n.d.). ATLAS.ti. https://atlasti.com/

Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, *33*. https://doi.org/10.1007/s12525-023-00680-1

Bartsch, V., Ebers, M., & Maurer, I. (2013). Learning in project-based organizations: The role of project teams' social capital for overcoming barriers to learning. *International Journal of Project Management*, *31*, 239–251. https://doi.org/10.1016/j.ijproman.2012.06.009

Berraida, R., & El Abbadi, L. (2024). *The Artificial Intelligence and Public Procurement* (tech. rep.). National School of Applied Sciences. https://doi.org/10.1109/logistiqua61063.2024.10571429

Bjerke, M. B., & Renger, R. (2016). Being smart about writing SMART objectives. *Evaluation and Program Planning*, *61*, 125–127. https://doi.org/10.1016/j.evalprogplan.2016.12.009

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. http://arxiv.org/abs/2303.12712

CambridgeDictionary. (n.d.). Skepticism. https://dictionary.cambridge.org/dictionary/english/skepticism#google_vignette

CambridgeDictionary. (2025). Cambridge dictionary. https://dictionary.cambridge.org/dictionary/english/deliberation#google_vignette

Chaudhuri, R., Chatterjee, ·. S., Vrontis, D., & Thrassou, A. (2024). Adoption of robust business analytics for product innovation and organizational performance: The mediating role of organizational data-driven culture. *Annals of Operations Research*, *339*, 1757–1791. https://doi.org/10.1007/s10479-021-04407-3/Published

Chien, A. A., Lin, L., Nguyen, H., Rao, V., Sharma, T., & Wijayawardana, R. (2023). Reducing the carbon impact of generative ai inference (today and in 2035). *2nd Workshop on Sustainable Computer Systems, HotCarbon 2023*. https://doi.org/10.1145/3604930.3605705

Chyung, S. Y. Y., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-Based Survey Design: The Use of a Midpoint on the Likert Scale. *Performance Improvement Journal*, *56*(10), 15–23. https://doi.org/10.1002/pfi.21727

Crawford, K., & Hasan, H. M. (2006). *Demonstrations of the Activity Theory Framework for Research in IS* (tech. rep.). School of Economics and Information Systems, University ofWollongong. https://ro.uow.edu.au/commpapers/276

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information.

de Fine Licht, K., & Brülde, B. (2021). On defining "reliance" and "trust": Purposes, conditions of adequacy, and new definitions. *Philosophia (United States)*, *49*, 1981–2001. https://doi.org/10.1007/s11406-021-00339-1

Dodge, J., Prewitt, T., Combes, R. T. D., Odmark, E., Schwartz, R., Strubell, E., Luccioni, A. S., Smith, N. A., Decario, N., & Buchanan, W. (2022). Measuring the carbon intensity of ai in cloud instances. *ACM International Conference Proceeding Series*, 1877–1894. https://doi.org/10.1145/3531146.3533234

Engeström, Y. (1987). *Learning by Expanding*. Orienta-Konsultit.

Engeström, Y., Virkkunen, J., Helle, M., Pihlaja, J., & Poikela, R. (1996). The change laboratory as a tool for transforming work.

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2023). Generative AI. *Business & Information Systems Engineering*, *66*(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7

Flyvbjerg, B. (2006). Five misunderstandings about Case-Study research. *Qualitative Inquiry*, *12*(2), 219–245. https://doi.org/10.1177/1077800405284363

Frauenberger, C. (2019). Entanglement hci the next wave? *ACM Transactions on Computer-Human Interaction*, *27*. https://doi.org/10.1145/3364998

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2022). Cognitive challenges in human–artificial intelligence collaboration: Investigating the path toward productive delegation. *Information Systems Research*, *33*, 678–696. https://doi.org/10.1287/isre.2021.1079

Gao, J., Cao, J., Yeo, S., Choo, K. T. W., Zhang, Z., Li, T. J.-J., Zhao, S., & Perrault, S. T. (2023). Impact of human-ai interaction on user trust and reliance in ai-assisted qualitative coding. http://arxiv.org/abs/2309.13858

Good, A., & Omisade, O. (2019). Linking activity theory with user centred design: A human computer interaction framework for the design and evaluation of mhealth interventions. *Studies in Health Technology and Informatics*, *263*, 49–63. https://doi.org/10.3233/SHTI190110

GoogleScholar. (n.d.). Google scholar. https://scholar.google.com/

Gracia, D. (2003). Ethical case deliberation and decision making.

Grodzinsky, F. S., Miller, K. W., & Wolf, M. J. (2010). Developing artificial agents worthy of trust: "Would you buy a used car from this artificial agent?" *Ethics and Information Technology*, *13*(1), 17–27. https://doi.org/10.1007/s10676-010-9255-1

Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of chatgpt as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, *2*. https://doi.org/10.1016/j.tbench.2023.100089

Hewett, T., Baecker, R., Card, S., Carey, T., Gasen, J., & Mantei, M. (1995). *Curricula for Human-Computer Interaction* (tech. rep.). NY, us, Association for Computing Machinery.

Jones-Jang, S. M., & Park, Y. J. (2022). How do people react to AI failure? Automation bias, algorithmic aversion, and perceived controllability. *Journal of Computer-Mediated Communication*, *28*(1). https://doi.org/10.1093/jcmc/zmac029

Joppe, M. (2000). The Research Process. https://www.torontomu.ca/~mjoppe/rp.htm

Kakolu, S., Pipelines, B., & Faheem, M. A. (2024). Building trust with generative ai chatbots: Exploring explainability, privacy, and user acceptance. https://doi.org/10.13140/RG.2.2.27825.60006

Kanbach, D. K., Heiduk, L., Blueher, G., Schreiter, M., & Lahmann, A. (2024). The genai is out of the bottle: Generative artificial intelligence from a business model innovation perspective. https://doi.org/10.1007/s11846-023-00696-z

Kenter, J. O., Reed, M. S., & Fazey, I. (2016). The deliberative value formation model. *Ecosystem Services*, *21*, 194–207. https://doi.org/10.1016/j.ecoser.2016.09.015

Kerosuo, H. (2006). Boundaries in action an activity-theoretical study of development, learning and change in health care for patients with multiple and chronic illnesses. www.yliopistopaino.fi

Klingbeil, A., Grützner, C., & Schreck, P. (2024). Trust and reliance on ai — an experimental study on the extent and costs of overreliance on ai. *Computers in Human Behavior*, *160*. https://doi.org/10.1016/j.chb.2024.108352

Kocsis, D. (2019). A conceptual foundation of design and implementation research in accounting information systems. *International Journal of Accounting Information Systems*, *34*. https://doi.org/10.1016/j.accinf.2019.06.003

Korolev, V., & Mitrofanov, A. (2023). Carbon footprint of artificial intelligence in materials science: Should we be concerned? https://doi.org/10.26434/chemrxiv-2023-zctn1

Kuurti, K. (1995). *Activity Theory as a Potential Framework for Human-Computer Interaction Research*. https://doi.org/10.7551/mitpress/2137.003.0006

Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). Quantifying the carbon emissions of machine learning. http://arxiv.org/abs/1910.09700

Ma, S., Chen, Q., Wang, X., Zheng, C., Peng, Z., Ming, Y., & Ma, X. (2024). Towards human-ai deliberation: Design and evaluation of llm-empowered deliberative ai for ai-assisted decision-making. *2022-March*. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Malone, T. W. (2019). *How can Human-Computer "Superminds" develop business strategies?* https://doi.org/10.1007/978-3-030-20680-2\{_}9

Manresa, A., Sammour, A., Mas-Machuca, M., Chen, W., & Botchie, D. (2024). Humanizing GenAI at work: bridging the gap between technological innovation and employee engagement. *Journal of Managerial Psychology*. https://doi.org/10.1108/jmp-05-2024-0356

Memmert, L., & Bittner, E. (2022). Complex Problem Solving through Human-AI Collaboration: Literature Review on Research Contexts. *Proceedings of the 55th Annual Hawaii International Conference on System Sciences/Proceedings of the Annual Hawaii International Conference on System Sciences*. https://doi.org/10.24251/hicss.2022.046

Mendeley. (n.d.). Mendeley. https://www.mendeley.com/

Miles, M. B., Huberman, A. M., Saldaña, J., et al. (2014). Qualitative data analysis: A methods sourcebook.

Miterev, M., Mancini, M., & Turner, R. (2017). Towards a design for the project-based organization. *International Journal of Project Management*, *35*, 479–491. https://doi.org/10.1016/j.ijproman.2016.12.007

Mohammed, Y. B., & Karagozlu, D. (2021). A review of human-computer interaction design approaches towards information systems development. *BRAIN. BROAD RESEARCH IN ARTIFICIAL INTELLIGENCE AND NEUROSCIENCE*, *12*, 229–250. https://doi.org/10.18662/brain/12.1/180

Mohemad, R., Hamdan, A. R., Othman, Z. A., & Noor, N. M. M. (2010). Decision Support Systems (DSS) in construction tendering processes. *International Journal of Computer Science Issues*, *7*(2), 35–45. https://www.ijcsi.org/papers/7-2-1-35-45.pdf

Nah, F., Cai, J., Zheng, R., & Pang, N. (2023). An Activity System-based Perspective of Generative AI: Challenges and Research Directions. *AIS Transactions on Human-Computer Interaction*, *15*(3), 247–267. https://doi.org/10.17705/1thci.00190

Nugroho, S., Sitorus, A. T., Habibi, M., Wihardjo, E., & Iswahyudi, M. S. (2023). The role of chatgpt in improving the efficiency of business communication in management science. *Jurnal Minfo Polgan*, *12*, 1482–1491. https://doi.org/10.33395/jmp.v12i1.12845

Nyumba, T. O., Wilson, K., Derrick, C. J., & Mukherjee, N. (2018). The use of focus group discussion methodology: Insights from two decades of application in conservation. *Methods in Ecology and Evolution*, *9*, 20–32. https://doi.org/10.1111/2041-210X.12860

Ogunsemi, D. R., & Aje, I. O. (2006). A model for contractors' selection in Nigeria. *Journal of Financial Management of Property and Construction*, *11*(1), 33–44. https://doi.org/10.1108/13664380680001078

Oladokun, M. G., Oladokun, A. A., & Odesola, I. A. (2011). Accuracy of Pre-Tender Cost Estimates of Consultant Quantity Surveyors in Nigeria. *Journal of International Real Estate and Construction Studies*, *1*(1).

OpenAI. (2022a). ChatGPT Enterprise. https://openai.com/chatgpt/enterprise/

OpenAI. (2022b). Introducing ChatGPT. https://openai.com/index/chatgpt/

OpenAI. (2024a). ChatGPT Team. @misc%7Bopenai-2024-gpt,%20author%20=%20%7BOpenAI%7D,%20title%20=%20%7B%7BCreating%20a%20GPT%20Enterprise%7D%7D,%20year%20=%20%7B2024%7D,%20url%20=%20%7Bhttps://help.openai.com/en/articles/8554397-creating-a-gpt%7D,%20%7D

OpenAI. (2024b). Creating a GPT Enterprise. https://help.openai.com/en/articles/8554397-creating-a-gpt

Ouyang, L., Wu, J., Jiang, Z., Almeida, D., Wainwright, C. L., & Mishkin, P. (2022). Training language models to follow instructions with human feedback. http://arxiv.org/abs/2203.02155

Overheid.nl. (2022, January). Aanbestedingswet 2012. https://wetten.overheid.nl/BWBR0032203/2022-03-02#Deel2

Pianoo. (2025, January). TenderNed: Jaarstatistieken 2024. https://www.pianoo.nl/nl/actueel/nieuws/tenderned-jaarstatistieken-2024?utm_source=chatgpt.com

Rane, N. L. (2024). Contribution of chatgpt and other generative artificial intelligence (ai) in renewable and sustainable energy. *Journal of Advances in Artificial Intelligence*, *2*, 1–26. https://doi.org/10.18178/JAAI.2024.2.1.1-26

Rijksoverheid. (n.d.). Over TenderNed | TenderNed. https://www.tenderned.nl/cms/nl/over-tenderned

Rijkswaterstaat. (2024a). *Aanbestedingsleidraad voor de aanbesteding van de opdracht met zaaknummer 31187297 voor het oplossen van storingen en correctief & klein preventief onderhoud vaste vaarwegmarkering (Basis Onderhouds Contract) Project "Zeker Zichtbaar".* (tech. rep.).

Rijkswaterstaat. (2024b). *Aanbestedingsleidraad voor de aanbesteding van de opdracht met zaaknummer 31197102 voor het engineeren en uitvoeren van suppletiewerken langs de Nederlandse kust* (tech. rep.).

Rijkswaterstaat. (2024c). *Aanbestedingsleidraad voor de aanbesteding van de opdracht met zaaknummer 31199586 voor het Verkeerstellingen 2025-2026 MN en WNN* (tech. rep.).

Rissanen, T. (2024). *The Use of Generative Artificial Intelligence in Public Procurement* (tech. rep.). Metropolia University of Applied Sciences. https://www.theseus.fi/bitstream/handle/10024/855072/Rissanen_Toni.pdf?sequence=2

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green ai. *Communications of the ACM*, *63*, 54–63. https://doi.org/10.1145/3381831

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and Abstraction in Sociotechnical Systems. *Association for Computing Machinery*. https://doi.org/10.1145/3287560.3287598

Shi, J., Jain, R., Doh, H., Suzuki, R., & Ramani, K. (2024). An HCI-Centric Survey and Taxonomy of Human-Generative-AI Interactions. *Department of Electrical and Computer Engineering, Purdue University, USA*.

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable ai. *International Journal of Human Computer Studies*, *146*. https://doi.org/10.1016/j.ijhcs.2020.102551

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339.

Staatscourant. (2020). *Aanbestedingsreglement Werken 2016* (tech. rep. No. nr. 30657).

Taherdoost, H. (2019). What is the best response scale for survey and questionnaire design; review of different lengths of rating scale / attitude scale / likert scale.

Tahmasbi, N., Rastegari, E., & Truong, M. (2025). *A conceptual model of trust in generative ai systems*. https://hdl.handle.net/10125/109690

TenderNed. (n.d.). TenderNed, hét online marktplein voor aanbestedingen. https://www.tenderned.nl/cms/nl

Thiry, M., & Deguire, M. (2007). Recent developments in project-based organisations. *International Journal of Project Management*, *25*, 649–658. https://doi.org/10.1016/j.ijproman.2007.02.001

Thompson, N., Greenewald, K., Lee, K., & Manso, G. F. (2021). The computational limits of deep learning; the computational limits of deep learning.

Topsakal, Y. (2024). How familiarity, ease of use, usefulness, and trust influence the acceptance of generative artificial intelligence (ai)-assisted travel planning. *International Journal of Human-Computer Interaction*. https://doi.org/10.1080/10447318.2024.2426044

Uil, S. (2021). Construction management and engineering improving tender performance: Creating awareness by steepening the learning curve.

Van den Berg, M., Voordijk, H., & Adriaanse, A. (2021). Bim uses for deconstruction: An activity-theoretical perspective on reorganising end-of-life practices. *Construction Management and Economics*, *39*, 323–339. https://doi.org/10.1080/01446193.2021.1876894

Van der Linden, T. (2025). Met AI is een aanbesteding winnen een koud kunstje. https://www.nrc.nl/nieuws/2025/01/15/met-ai-is-een-aanbesteding-winnen-een-koud-kunstje-a4879602?t=1745418343

van Infrastructuur en Waterstaat, M. (2025, April). Inkoopbeleid beste prijs kwaliteit verhouding. https://www.rijkswaterstaat.nl/zakelijk/zakendoen-met-rijkswaterstaat/inkoopbeleid/aanbesteden/economische-meest-voordelige-inschrijving

Venkatesh, V. (2022). Adoption and use of ai tools: A research agenda grounded in utaut. *Annals of Operations Research*, *308*, 641–652. https://doi.org/10.1007/s10479-020-03918-9

Vygotsky, L. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Library of Congress Cataloging in Publication Data.

Weidinger, L., Rauh, M., Marchal, N., Manzini, A., Hendricks, L. A., Mateos-Garcia, J., Bergman, S., Kay, J., Griffin, C., Bariach, B., Gabriel, I., Rieser, V., & Isaac, W. (2023). Sociotechnical safety evaluation of generative ai systems. https://arxiv.org/abs/2310.11986

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on ai dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, *11*. https://doi.org/10.1186/s40561-024-00316-7

Zhang, Z., Bai, F., Wang, M., Ye, H., Ma, C., & Yang, Y. (2024). Incentive compatibility for ai alignment in sociotechnical systems: Positions and prospects. https://arxiv.org/abs/2402.12907

Zhu, G., Ong, Y. S., Sudarshan, V., & Kow, J. F. (2024). Human-generative ai collaborative problem solving who leads and how students perceive the interactions.

A

Thematic Analysis Search Codes

Contradicting factors	Search codes
Ease of use	Easy to use
	Difficult to use
	Frustration with GenAI
	Experience in GenAI use
Over-reliance	Blind trust in GenAI
	GenAI decides for user
	Feeling superior to GenAI
	Lack of human intervention
	GenAI makes final decision
Under-reliance	Distrust in GenAI
	GenAI output is double checked
	Preference for manual method
	GenAI use is minimised
	Rejection of GenAI recommendations
Agreement	Alignment with GenAI output
	Confirming ideas
	GenAI used as second opinion
	Disagreement with AI outputs
	Human overrides GenAI
	If answer is well-explained
	GenAI understands input
Trust	(Dis)trust in GenAI output
	(Dis)trust in organisation GenAI
	Preference for human review
	Trust varies by context
	Trust in GenAI with human verification
	Peer trust in GenAI
Sustainability	Reduction of GenAI use
	User emphasises sustainability
	Ignorance
	Not effecting use
	Unknowing
	Sustainability overrides efficiency
	Efficiency overrides sustainability
Explainability	Output lacks reasoning
	Confusion due to missing explanation
	Explanation not available
Company culture	Learning culture
	Stimulation
	Innovation culture
	Space for trial
Scepticism	About GenAI accuracy
	About GenAI being too serving
	Lack of evidence
	Distrust due to past errors
	About added value
	Peer scepticism
User experience	Perceived usefulness
	Satisfaction
	Frustration
	Reliance due to past experiences
	Past conflicts with GenAI
	Expertise overrules GenAI
	GenAI being demanding
	Forces you to think
Privacy	Uncertainty about data use
	Concerns about data storage
	Privacy risks
	Lack of control over data
	No privacy concerns
	Conflict between efficiency and privacy
Transparency	Output (un)explained
	Decision process (un)clear
	(Un)predictable outcomes
Helpfulness	Makes task easier
	Provides guidance
	Supports reviewing
	Lacks contextual information
	Discussion partner
	Inspiration source
	Objective
	Time-effient
	Not time-efficient

Table A.1: Search codes for thematic analysis per contradicting factor (created by the author)

B

Tendering by Rijkswaterstaat (Dutch)

Activiteit	**Datum**
Verzenden aankondiging door publicatie op www.tenderned.nl	15-07-2024
Inschrijvingsfase	
Inlichtingenbijeenkomst verstrekken nadere inlichtingen	03-09-2024
Uiterste datum indienen verzoek om nadere inlichtingen	24-09-2024
Publicatie nota van inlichtingen inschrijvingsfase	15-11-2024
Uiterste datum ontvangst van de inschrijvingen	29-11-2024
Beoordelingsfase	
Openen digitale kluis in TenderNed met inschrijvingen	02-12-2024
Verzenden gunningsbeslissing	16-12-2024
Uiterste datum rechtsbeschermingstermijn	21-01-2025
Verzenden opdracht	05-02-2025

Table B.1: Tender planning example 1 (Rijkswaterstaat, 2024a)

Activiteit	**Datum/weeknr**
Verzenden aankondiging door publicatie op www.tenderned.nl	23-09-2024
Inschrijvingsfase	
Uiterste datum indienen verzoek om nadere inlichtingen	07-10-2024
Publicatie nota van inlichtingen inschrijvingsfase	21-10-2024
Uiterste datum ontvangst van de inschrijvingen	18-11-2024 15:00 uur
Beoordelingsfase	
Openen digitale kluis in TenderNed met inschrijvingen	19-11-2024
Verzenden gunningsbeslissing	31-01-2025
Uiterste datum rechtsbeschermingstermijn	20-02-2025
Verzenden opdracht	06-03-2025

Table B.2: Tender planning example 2 (Rijkswaterstaat, 2024b)

Activiteit	Start
Verzenden aankondiging door publicatie op www.tenderned.nl	6 november 2024
Indienen vragen via TenderNed	6 november 2024
Publicatie nota van inlichtingen inschrijvingsfase	29 november 2024
Uiterste datum ontvangst van de inschrijvingen	13 december 2024
Openen digitale kluis in TenderNed met inschrijvingen	14 december 2024
Verzenden gunningsbeslissing	20 januari 2025
Uiterste datum rechtsbeschermingstermijn	21 januari 2025
Verzenden opdracht	11 februari 2025

Table B.3: Tender planning example 3 (Rijkswaterstaat, 2024c)

C

Tender Strategy Count & Cooper

The research on human–GenAI deliberation in infrastructure tendering, with a focus on the specific context of Count & Cooper, requires a deeper understanding of the activities involved in the tender process. This appendix outlines the tender process as followed by Count & Cooper. It outlines the company's tender strategy, identifies the key stakeholders involved in the tendering process, introduces the tender team members, describes the various phases to be completed, and explains the standard meetings that have been established.

Tender Strategy

Count & Cooper follows a general approach to evaluating tenders, based on the guidelines set out in the Aanbestedingswet 2012 (Overheid.nl, 2022). These rules determine whether a bid meets the specifications, requirements, and standards set by the client. The criteria are aimed at selecting the most economically advantageous tender (MEAT). In Dutch, this is referred to as the economisch meest voordelige inschrijving (EMVI), which is based on the beste prijs-kwaliteitverhouding (BPKV). The BPKV means the best price–quality ratio, which ensures that the tender evaluation is not only focused on the lowest price. It places strong focus on quality factors, which can include technical aspects, sustainability, design and project management. Each of these factors is assessed according to specific scoring rules and is assigned a score. A higher score on any of these quality factors results in a discount on the bid price. In this way, high quality bids can outbid lower priced bids, meaning that the winning bid is not necessarily the cheapest.
Count & Cooper's tender strategy is built around this BPKV approach and therefore forms the foundation for how tenders are written.

Stakeholders

There are three main parties involved in the tendering process: the client, the contractor, and the consultant. In this context, the consultant is Count & Cooper. The roles of the client and the contractor are explained below.

Client

The client is the project owner who initiates the tendering process. Without a request from the client, no bid would be prepared. When the client announces a tender for a construction project, contractors are invited to submit a proposal in order to win the contract. The client must stick with the BPKV criteria to ensure fair and even competition when selecting the winning bid, thus avoiding any preferential treatment.

Contractor

The contractor is the main party responding to the client's request to prepare and submit a tender. In the case of Count & Cooper, the contractor hires a consultant, Count & Cooper, to assist in the

development of the bid. While the contractor and the consultant together form the contracting party, it is the contractor who is the direct client of the consultant.

Tender Team

A Count & Cooper tender team consists of different roles with clearly defined responsibilities, including a tender manager, delivery manager, product leads for technology and process, specialists, and reviewers. Count & Cooper, as a consultancy firm, always operates on behalf of, and in close collaboration with, the contractor. The number of people involved in the tender team will varies depending on the needs of the contractor, as well as the size and complexity of the tender. The various roles within the team and their respective responsibilities are outlined below.

Tender Manager

The tender manager reports to the steering group and organises the internal decision-making process involving executive boards, partnering, pricing, and procurement. This person also shares responsibility for client contact, together with the delivery manager. The tender manager may be either a consultant or an employee of the contractor.

Delivery Manager

The delivery manager, also known as the EMVI coordinator, organises the overall tender process and ensures the delivery of the required product. This person is responsible for coordinating the review and challenging the process, maintaining quality assurance across all deliverables. Like the tender manager, the delivery manager also shares responsibility for communication with the client.

Product Leads Technology and Process

The product leads for technology and process are jointly responsible for delivering the final product. They define the outline and depth of content, determine what input is needed, and pose relevant research questions to specialists. To gather this input, they organise sessions and interviews to ask questions, structure information, simplify complex issues and challenge views when necessary. They are also responsible for writing the product and creating visuals to support the product.

Specialists

The specialist team analyses the assignment in depth and provides answers to the questions posed by the product leads. They are given sufficient time to make an active and meaningful contribution to the process.

Reviewers

The role of the challenge and review team is to provide critical insights and share relevant expertise to strengthen the overall quality of the tender.

Tender Phases

The Count & Cooper tender process consists of three main phases: Bronze, Silver and Gold. Around these core phases are the pre-tender phase, which marks the beginning of the process, and the submission phase at the end.

Pre-Tender

During the pre-tender phase, the project team is put together, agreements are made with the contractor, and a planning schedule is established. This planning includes both tender planning and the planning on the product level, the concrete tender.

Bronze

The Bronze phase focuses on a full analysis of the project requirements. This includes developing a tender strategy, interpreting and defining the scope of the assignment, and setting objectives and approaches in line with the BPKV. Potential constraints and challenges are also identified to gain a clear understanding of the project's complexities and opportunities.

Silver
During the Silver phase, different solution alternatives are developed. A trade-off matrix is created to evaluate and compare these options. This phase is characterised by detailed consideration of different actions to determine the most effective and feasible solution.

Gold
The Gold phase involves refining the chosen solution out of the Silver phase and drafting the necessary documentation for the tender submission. This includes finalising all required materials for the bid.

Submission
Finally, at the submission phase, the proposal is checked for validity. This includes checking that the bid meets all the stated requirements, such as payment terms, technical specifications, insurance obligations and subcontractor conditions.

Meetings
Throughout the week, various meetings are held to ensure coordination and alignment within the tender team. These are outlined below.

Weekly Stand-up Whole Tender Team
Each week, the full tender team, including both contractor and consultant representatives, attends a short stand-up meeting. This meeting usually lasts between half an hour and an hour. During the meeting, updates on the tender or process are shared. In addition, the team reviews and discusses the planning in relation to the content of the tender, checking that the planning is still up to date and feasible.

Weekly Meeting Core Team
In the same week, a meeting is held for the core members of the tender team. This team includes the Tender Manager and Delivery Manager, as well as relevant managers from the contractor's side, depending on the structure of the team. As mentioned above, the delivery manager may represent either the contractor or the consultant. During this meeting, the team discusses and notes any questions that are intended for the client. In most tenders, there is an opportunity to raise questions with the client throughout the process, up to a set deadline. This meeting ensures that such questions are carefully considered and formulated well.

Weekly Meeting Count & Cooper Team
An optional weekly meeting is held exclusively for members of the Count & Cooper team involved in the tender. This meeting focuses primarily on the BPKV elements of the bid and general updates or issues can be discussed.

D

Experiment Documents

D.1. Questionnaire Background Information (Dutch)

1. Vragenlijst achtergrond

1. Binnen welke categorie valt jouw leeftijd?

- ◯ 20-29 jaar
- ◯ 30-39 jaar
- ◯ 40-49 jaar
- ◯ 50-59 jaar
- ◯ 60-69 jaar

2. Wat is jouw geslacht?

- ◯ Man
- ◯ Vrouw
- ◯ Zeg ik liever niet

3. Aan hoeveel tenders heb jij meegewerkt? (MISSCHIEN NOG KLEINER SCHALEN...?)

- ◯ Minder dan 10
- ◯ 10-19
- ◯ 20 of meer

4. Hoeveel jaar ervaring heb jij met tenderen?

- ◯ Minder dan 5 jaar
- ◯ 5-9 jaar
- ◯ 10-14 jaar
- ◯ 15-19 jaar
- ◯ 20 of meer jaar

5. Hoe ervaren ben je met het werken met ChatGPT (of andere vormen van AI)?

- ◯ Geen ervaring (ik heb ChatGPT nog nooit gebruikt)
- ◯ Weinig ervaring (Ik heb het een paar keer geprobeerd)
- ◯ Enige ervaring (Ik gebruik het soms, maar nog niet vaak)
- ◯ Veel ervaring (Ik gebruik het regelmatig en weet hoe het werkt)
- ◯ Zeer ervaren (Ik gebruik ChatGPT vaak en ken de mogelijkheden goed)

6. Wat vind je het grootste voordeel (of voordelen) van ChatGPT?

7. Wat weerhoudt jou om ChatGPT vaker te gebruiken? Loop je tegen beperkingen of problemen aan?

8. In hoeverre speelt de mogelijke klimaatimpact van ChatGPT voor jou een rol bij de keuze om ChatGPT te gebruiken?

D.2. Step-by-Step Plan for the Experiment

1. Review the tender

The following documents are in front of you:

i. Tender
ii. Guideline Client
iii. Review Table

a. Review Measure 2 from the tender using the guideline and review table. Do this by filling in the form via the following link: **Form Tender Review**.
b. Take **15 minutes** for this step (this includes both reading and filling in the form).

2. Discussion with ChatGPT

a. Download the PDF of your review, which you can find in your inbox.
b. Access the correct ChatGPT environment via the following link: **ChatGPT Link**. If the link does not open directly, copy and paste it into Chrome.
c. Upload your review document into the chat by clicking on + and start the conversation by pressing ↑. Your discussion will now begin!
d. Important: keep asking questions and sharing your opinion so that ChatGPT receives enough input. Use ChatGPT's responses as input or start a discussion on your own.
e. Take **15-20 minutes** for this discussion step.

3. Is your discussion with ChatGPT finished?

a. Send **"done"** in the ChatGPT conversation. ChatGPT will then provide you with a summary of the new insights from your discussion. This will help you complete step 3c more easily.
b. Share your ChatGPT conversation by clicking **'Share'** in the top right corner of ChatGPT and emailing the link to the organiser.
c. Review the measure one last time using the form via the following link: **Form Tender After Discussion**.

4. Group Discussion

D.3. Review Table

Context – Measure – Effect – Justification – Structure including SMART

	Short description measure
Context	Description of the background situation: - What is the situation if no measure is taken?
Measure	Specific description of the measure: - What does the measure specifically involve? - What additional actions are we taking beyond the contract? Measurable, Achievable, Time-bound measure - How exactly will we implement it? (What is the execution phase? How often? With what resources?) - When will we implement the measure? - Who is responsible? - How will we measure and demonstrate it? (Measurable/observable conditions)
Effect	Effect of the measure + Relevant expectations: - What is the impact of the measure? - How does it contribute to achieving the objective?
Justification	Reference projects: The measure is realistic because it has been successfully implemented in project X. - Why is the situation comparable? - How did it work, and what was the result?

Questions to define SMART:

Specific	Is the measure **clear** and **unambiguous**?
Measurable	What is the result, and can it be **measured**?
Achievable	Is the measure realistic and **achievable**?
Relevant	Why is this beneficial for the client or project? How does it **meet the requirements**?
Time-bound	What is the deadline? When should the objective be reached?

D.4. Questionnaire as a Basis for Reviewing (Dutch)

2. Review tender

Zou je de de maatregelen van de tender willen beoordelen aan de hand van de volgende vragen? Gebruik hiervoor de leidraad en de review tabel (inclusief de aanleiding-maatregel-effect-structuur en de SMART-tabel).

* Required

1. Wat is jouw naam?

2. Maatregel 2: Op vlieghoogte blijven *

Hoe is de *aanleiding* uitgewerkt?

- ◯ 0 (zeer slecht)
- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5
- ◯ 6
- ◯ 7
- ◯ 8
- ◯ 9
- ◯ 10 (uitmuntend)

3. Maatregel 2: Op vlieghoogte blijven *

Waarom heb jij de *aanleiding* zo beoordeeld?

4. Maatregel 2: Op vlieghoogte blijven *

In hoeverre is de maatregel *specifiek* beschreven?

- ◯ 0 (zeer slecht)
- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5
- ◯ 6
- ◯ 7
- ◯ 8
- ◯ 9
- ◯ 10 (uitmuntend)

5. Maatregel 2: Op vlieghoogte blijven *

Waarom heb jij de maatregel zo beoordeeld op '*specifiek*'?

6. Maatregel 2: Op vlieghoogte blijven *

In hoeverre is de *meetbaarheid* van de maatregel beschreven?

- ◯ 0 (zeer slecht)
- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5
- ◯ 6
- ◯ 7
- ◯ 8
- ◯ 9
- ◯ 10 (uitmuntend)

7. Maatregel 2: Op vlieghoogte blijven *

Waarom heb jij de maatregel zo beoordeeld op '*meetbaar*'?

8. Maatregel 2: Op vlieghoogte blijven *

In hoeverre is de maatregel *acceptabel*?

- ◯ 0 (zeer slecht)
- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5
- ◯ 6
- ◯ 7
- ◯ 8
- ◯ 9
- ◯ 10 (uitmuntend)

9. Maatregel 2: Op vlieghoogte blijven *

Waarom heb jij de maatregel zo beoordeeld op '*acceptabel*'?

10. Maatregel 2: Op vlieghoogte blijven *

In hoeverre is de maatregel *realistisch*?

- ◯ 0 (zeer slecht)
- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5
- ◯ 6
- ◯ 7
- ◯ 8
- ◯ 9
- ◯ 10 (uitmuntend)

11. Maatregel 2: Op vlieghoogte blijven *

Waarom heb jij de maatregel zo beoordeeld op '*realistisch*'?

12. Maatregel 2: Op vlieghoogte blijven *

Hoe is de *tijdgebondenheid* van de maatregel beschreven?

- ◯ 0 (zeer slecht)
- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5
- ◯ 6
- ◯ 7
- ◯ 8
- ◯ 9
- ◯ 10 (uitmuntend)

13. Maatregel 2: Op vlieghoogte blijven *

Waarom heb jij de maatregel zo beoordeeld op '*tijdgebonden*'?

14. Maatregel 2: Op vlieghoogte blijven *

Hoe is het *effect* van de maatregel uitgewerkt?

- ◯ 0 (zeer slecht)
- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5
- ◯ 6
- ◯ 7
- ◯ 8
- ◯ 9
- ◯ 10 (uitmuntend)

15. Maatregel 2: Op vlieghoogte blijven *

Waarom heb jij het *effect* zo beoordeeld?

16. Maatregel 2: Op vlieghoogte blijven *

Hoe is de *onderbouwing* van de maatregel uitgewerkt?

- ◯ 0 (zeer slecht)
- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5
- ◯ 6
- ◯ 7
- ◯ 8
- ◯ 9
- ◯ 10 (uitmuntend)

17. Maatregel 2: Op vlieghoogte blijven *

Waarom heb jij de *onderbouwing* zo beoordeeld?

18. Algemene sterke punten

19. Algemene zwakke punten

D.5. Role of ChatGPT in the Review

As a company, we are participating in a tender (procurement process) for a construction project. To win this tender, it is crucial that we thoroughly understand the client's requirements and present a clear and concrete plan. The plan is nearly finalised, and we want you, ChatGPT, to take on the role of a critical reviewer who evaluates the content of the tender.

To conduct this review effectively, we have provided the following attachments:

- The tender document, titled '*plan…*'
- The guideline, which outlines the client's expectations and the required content for the plan.
- Review Tender Form, which you must use as the basis for structuring your review.

As a company, we find it important that all measures have a clear structure in which the context, the measure, and the effect are clearly described. Additionally, each measure must be written in a SMART way. The context-measure-effect-justification structure and SMART writing are explained in the table below.

	Short description measure
Context	Description of the background situation: - What is the situation if no measure is taken?
Measure	Specific description of the measure: - What does the measure specifically involve? - What additional actions are we taking beyond the contract? Measurable, Achievable, Time-bound measure - How exactly will we implement it? (What is the execution phase? How often? With what resources?) - When will we implement the measure? - Who is responsible? - How will we measure and demonstrate it? (Measurable/observable conditions)
Effect	Effect of the measure + Relevant expectations: - What is the impact of the measure? - How does it contribute to achieving the objective?
Justification	Reference projects: The measure is realistic because it has been successfully implemented in project X. - Why is the situation comparable? - How did it work, and what was the result?

Questions to define SMART:

Specific	Is the measure **clear** and **unambiguous**?
Measurable	What is the result, and can it be **measured**?
Achievable	Is the measure realistic and **achievable**?

Relevant	Why is this beneficial for the client or project? How does it **meet the requirements**?
Time-bound	What is the deadline? When should the objective be reached?

I would like to ask you (ChatGPT) to assess the tender based on the guideline, the context-measure-effect-justification structure, and the SMART writing of the measure.

Below are some important points to follow:

1. Use the 'Questionnaire as a basis for reviewing' PDF as the basis for the structure of the review. Also, use the points from this document as evaluation criteria.
2. Assign a score to each section per measure on a scale of 0 to 10, as shown in the 'Questionnaire as a basis for reviewing' PDF. A score of 0 means the measure is very poorly developed, while 10 means it is excellently developed. Focus on the content of the text.
3. Create a table with the following columns: Score ChatGPT, Reasoning ChatGPT, Improvement Points ChatGPT.
4. Ensure that the context does not yet describe any effect or solution; only the cause should be stated here.
5. Ensure that in the effect section, the solution to the context is clearly stated and that a connection is made to the objectives from the guideline.
6. Ensure that each measure includes at least one success factor from the guideline.
7. For each measure, identify at least two strong points and two areas for improvement. These points should be solely based on content. Do not assess formulation or formatting. Describe the areas for improvement clearly and avoid unnecessarily long explanations.
8. Base your reasoning only on the information from the attachments, and do not use external sources to support your arguments.

D.6. Role of ChatGPT in the Discussion

The goal is for you (ChatGPT) to engage in a discussion with a participant. I have asked both you and the participant to assess various aspects of a tender based on the guideline provided by the client, whether the context-measure-effect-justification structure is followed, and whether the measures are written in a SMART way. This structure and SMART are described in the table below.

	Short description measure
Context	Description of the background situation: - What is the situation if no measure is taken?
Measure	Specific description of the measure: - What does the measure specifically involve? - What additional actions are we taking beyond the contract? Measurable, Achievable, Time-bound measure - How exactly will we implement it? (What is the execution phase? How often? With what resources?) - When will we implement the measure? - Who is responsible? - How will we measure and demonstrate it? (Measurable/observable conditions)
Effect	Effect of the measure + Relevant expectations: - What is the impact of the measure? - How does it contribute to achieving the objective?
Justification	Reference projects: The measure is realistic because it has been successfully implemented in project X. - Why is the situation comparable? - How did it work, and what was the result?

Questions to define SMART:

Specific	Is the measure **clear** and **unambiguous**?
Measurable	What is the result, and can it be **measured**?
Achievable	Is the measure realistic and **achievable**?
Relevant	Why is this beneficial for the client or project? How does it **meet the requirements**?
Time-bound	What is the deadline? When should the objective be reached?

I would like to ask you (ChatGPT) to lead the discussion and ensure that all aspects of the tender review are addressed. This means that when your assessment and that of the participant do not match, a discussion should take place about why you have evaluated that particular aspect in a certain way.

Follow the steps below to guide the discussion:

1. Ensure that when a document is uploaded, you start following the instructions from the 'Role of ChatGPT in the Discussion' file.
2. First, explain to the participant that you (ChatGPT), just like the participant, have conducted a review of Measure 2 and that you will now discuss the different reviews. Explain that you will first present the evaluations from both reviews in a table.
3. Then, create a table with the following columns: Score ChatGPT, Reasoning ChatGPT, Improvement Points ChatGPT, Score [name], and Reasoning [name]. Do not shorten the text from the files; the full text must be copied into the table. Also, number the points and rows to make it as easy as possible for the participant to reference specific aspects.
4. If you have not yet received a PDF titled 'Review [name]', first ask the participant to upload it.
5. ChatGPT asks a specific question to the participant to initiate the discussion.
    a. Examples of questions that ChatGPT can ask:
        i. "Why do you give that specific argument? Could you explain it more clearly?"
        ii. "What do you think about my argument on that specific point?"
6. Ensure that you keep the conversation going. Continue to provide input about your evaluation of the review's aspects and ask follow-up questions about the participant's reasoning. Do not stop this; keep engaging.
7. Also, discuss ChatGPT's improvement points. For example, pick one that you strongly believe will significantly enhance the writing of the measure, and ask the participant what they think about this point.
8. Ask the participant if they would like to discuss a specific aspect themselves.
9. To provide better arguments, base your responses on the other uploaded documents, including:
    a. Tender: This is the document where Measure 2 is described, which you have reviewed in Review ChatGPT.
    b. Guideline: This document comes from the client and contains the requirements they expect to see in the entire tender.
    c. Review ChatGPT: Your previously conducted assessment of Measure 2.
10. Respond to the participant’s answer.
    a. ChatGPT may agree or disagree but must provide clear arguments.
    b. If the participant provides a strong argument, ChatGPT may adjust its assessment.
11. Summarise the conclusion per section.
    a. Is there agreement, or does a difference remain? Why?
    b. This ensures that the discussion does not end without a resolution but actually leads to a conclusion.

12. If the participant sends "done" to ChatGPT, create a table with new insights from the discussion. Use the exact same structure as the previous table but only include new insights that were actually discussed between you and the participant. Do not add your own additional points—only mention what was discussed.

Additional Important Points to Follow:

1. Conduct the discussion immediately after presenting the table.
2. Always provide clear arguments.
3. Base your reasoning only on the information from the attachments—do not use external sources to support your arguments.
4. Stick to facts.

E

Evaluation Sessions

E.1. Presentation Evaluation Sessions (Dutch)

1

Ik vertrouw erop dat ChatGPT hulp kan bieden bij het reviewen van tenders

- Betrouwbaar als inspiratiebron
- Vertrouwen op je eigen intuïtie is belangrijk bij het gebruiken van ChatGPT
- "ChatGPT is altijd goede discussie partner. Weet wat je krijgt."
- "ChatGPT is niet altijd correct en daarom minder betrouwbaar."

- Vertrouwen op ChatGPT ligt bij jezelf: "shit goes in, shit goes out"

2

Ik vertrouw dat ChatGPT bedrijfsvertrouwelijke informatie waarborgt

3

ChatGPT produceert sterke en goed onderbouwde argumenten

- Goed uitgewerkte argumenten zijn essentieel voor het aannemen van dat ChatGPT juist is.

4

Ik vind het gebruik van ChatGPT nuttig bij het reviewen van tenders

- "Goed voor inspiratie, daarop voort borduren."
- Doorbreken tunnelvisie/blinde vlek

5

ChatGPT is objectief en daarom nuttig tijdens het reviewen

- "Waardevolle discussie met Chat, die geen direct belang heeft."
- ChatGPT probeert jou niet over te halen, is neutraal

6

ChatGPT is tijdsefficiënt en daarom nuttig tijdens het reviewen

Positief
- "Vooral ook omdat ChatGPT niet alle spreektijd opeist tijdens een reviewsessie."
- "En wat ik nog wel extra voordeel vind is dat ChatGPT geen agenda heeft."
- Makkelijk te gebruiken

Negatief
- "Het kost wel veel tijd, maar ik zou het dus ook vooral toepassen bij de dingen waar ik dan zelf over twijfel."
- "Het helpt scherper te maken. Maar het kost nu nog meer tijd."

7

ChatGPT is reflectief en daarom nuttig tijdens het reviewen

- "Het helpt me scherper te worden."
- "Een spiegel voorgehouden krijgen."

8

ChatGPT is niet genoeg geïnformeerd om hulp te bieden bij het reviewen

- ChatGPT mist de volledige context van de opdracht
  - Gesprekken opdrachtgever etc.
  - Interne gesprekken
- Review intern
- "Shit goes in, shit goes out"

9

ChatGPT is té dienend om mij goed te helpen bij het reviewen

- Té dienend

10

Het energieverbruik van ChatGPT speelt mee in mijn overweging om het te gebruiken

11

Ik word gestimuleerd door mijn omgeving om ChatGPT te gebruiken

- Het bedrijf
- Samenleving

12

Waar zou jij ChatGPT voor gebruiken met reviewen?

- Hulp bij persoonlijke review (tunnelvisie)
- Vervanger voor traditionele review sessies (groep)
- Extra check bovenop huidige review manier
- Sparringspartner
- Anders..?

13

CUES TO IMPLEMENTATION VS BARRIERS TO IMPLEMENTATION

De/Het … van ChatGPT

Betrouwbaarheid
Nuttigheid
Objectiviteit
Tijdsefficiëntie
Reflectiviteit
Ontbrekende informatie
Transparantie
Privacy aspect
Dienendheid
Duurzaamhelds aspect

14

F

Data Management

F.1. Data Management Plan

Plan Overview

A Data Management Plan created using DMPonline

Title: Master Thesis Roos van Duuren

Creator: Roos van Duuren

Principal Investigator: Roos van Duuren

Data Manager: Roos van Duuren

Project Administrator: Roos van Duuren

Affiliation: Delft University of Technology

Template: TU Delft Data Management Plan template (2025)

Project abstract:

This research aims to research the deliberation between humans and generative AI and how generative AI influences peoples' opinions about judging tenders in project-based organisations for infrastructure tendering.

The data will be collected through a study where humans interact with AI. This will be combined with interviews afterward to understand the participants' thoughts. Around 15 participants will beinvolved, recruited through the collaborating company, Count & Cooper, for this master thesis. If time allows, other parties working in infrastructure tendering may also be approached as an extension to the research.

The objective of this thesis is to explore the collaboration between humans and generative AI to provide organisations with insights for encouraging and optimising the use of generative generative AI systems.

ID: 168947

Start date: 23-09-2024

End date: 31-05-2025

Last modified: 29-01-2025

Master Thesis Roos van Duuren

0. Adminstrative questions

1. Provide the name of the data management support staff consulted during the preparation of this plan and the date of consultation. Please also mention if you consulted any other support staff.

Lora Armstrong, Data Steward at the faculty of Civil Engineering and Geosciences, has reviewed this DMP on 29-01-2025.

2. Is TU Delft the lead institution for this project?

- Yes, leading the collaboration – please provide details of the type of collaboration and the involved parties below

TU Delft is leading, but I am doing an internship at Count & Cooper. Therefore Count & Cooper is the involved party.

I. Data/code description and collection or re-use

3. Provide a general description of the types of data/code you will be working with, including any re-used data/code.

Type of data/code	**File format(s)**	**How will data/code be collected/generated?** *For re-used data/code: what are the sources and terms of use?*	**Purpose of processing**	**Storage location**	**Who will have access to the data/code?**
Tender	pdf	Through the company Count & Cooper I will get access to one (or multiple) tenders done by the company. This tender is done in the past.	Needed in order to set up the experiment of humans and AI discussing over the judgement on an infrastructure tender.	OneDrive TU Delft and SharePoint of Count & Cooper	Me, Count & Cooper
Contact information participants	pdf	Information of the participants of the company provided by the company of Count & Cooper.	To contact participants for the study and set up the experiment including interviews.	OneDrive TU Delft	Me
Informed consent forms	pdf	Before the experiment and interview.	Digital permission to use data	OneDrive TU Delft and after graduation, my supervisor (Dr. Tong Wang) will get access and store it, using Project Data Storage	Me, supervisor
Survey background information participants	pdf	Through an online survey, MS forms	Survey to get background information of the participants on their experiences in tendering, experiences in the use of AI, the age, sex. Done to eventually draw conclusions or explanations on the results of the experiment.	OneDrive TU Delft	Me
Survey participants' judgement	pdf	Online survey, MS forms	Let the participant judge a tender with the information on what the client would like to see in the tender submissions. Through the survey I get to know how the particpant rates the tender according to assessment points. This is done through a scale from 1-7 from fully disagree to fully agree. Following, the participants are asked to explain their ratings.	OneDrive TU Delft	Me
Comparison participants' and AI's judgment	pdf	Data on the comparison between the participants' and AI's judgement, done by me manually.	To analyse the differences in judgement, how much the human and AI agree/disagree. This information is the input for the actual interaction between human and AI.	OneDrive TU Delft	Me
Interaction human-AI	pdf	Data of the converstion between humans and AI. Done in a restricted version of ChatGPT purchased by Count & Cooper.	Data for insights into the interaction	OneDrive TU Delft	Me
Qualitative interview recordings	mp3	Recordings of the interview sessions. Recordings made on telephone. As soon as possible transferred to oneDrive and deleted from phone. Transcripts are made through the restricted version of ChatGPT, payed for by Count & Cooper.	To understand the interaction between humans and AI in opinion changes. The participants will be confronted with statements and have to explain how much they agree with the statements in order to get to know how they value the interaction with AI.	OneDrive TU Delft	Me
Qualitative interview transcripts	pdf	Transcripts of the recording of the interview sessions. I will anonymise the transcripts by hand.	To understand the interaction between humans and AI in opinion changes	OneDrive TU Delft	Me, interview participants
Anonymised interview transcriptions	pdf	Anonymous transcripts of the recording of the interview sessions	To understand the interaction between humans and AI in opinion changes and to use in the report	OneDrive TU Delft	Me, interview participants

II. Storage and backup during the research process

4. How much data/code storage will you require during the project lifetime?

- < 250 GB

5. Where will the data/code be stored and backed-up during the project lifetime? (Select all that apply.)

- TU Delft OneDrive

III. Data/code documentation

6. What documentation will accompany data/code? (Select all that apply.)

- Data – Methodology of data collection

IV. Legal and ethical requirements, code of conducts

7. Does your research involve human subjects or third-party datasets collected from human participants?

If you are working with a human subject(s), you will need to obtain the HREC approval for your research project.

- Yes – please provide details in the additional information box below

Interviews and doing a small experiment with the same participants.
And I intend to apply for ethical approval from the Human Research Ethics Committee, but have not yet done so.

8. Will you work with personal data? (This is information about an identified or identifiable natural person, either for research or project administration purposes.)

- Yes

9. Will you work with any other types of confidential or classified data or code as listed below? (Select all that apply and provide additional details below.)

If you are not sure which option to select, ask your Faculty Data Steward for advice.

- Yes, confidential data received from commercial, or other external partners

The tender itself is confidential information from the company. I will be using the information out of the tender for research, but I will not publish this information.

10. How will ownership of the data and intellectual property rights to the data be managed?

For projects involving commercially-sensitive research or research involving third parties, seek advice of your Faculty

***Contract Manager* when answering this question**

I will be the owner of the data and the data will be restricted by me during the research. Data that will be used in my report will be shared anonymously.
The intellectual property rights are framed by a graduation agreement between Delft University of Technology, myself and Count & Cooper.

11. Which personal data or data from human participants do you work with? (Select all that apply.)

- Other types of personal data or other data from human participants – please provide details below
- Gender
- Proof of consent (such as signed consent materials which contain name and signature)
- Audio recordings
- Telephone number, email addresses and/or other addresses as contact details for administrative purposes
- Date of birth and/or age
- Names as contact details for administrative purposes

For getting information of ages, I will use age ranges instead of exact ages with gaps of 5 years to make it more anonym.
Other information about the experiences in tendering, how long and what they have been working on in tendering, and in the use of AI will be gathered.
Also opinions on the judgement on a tender will be collected and participants will be asked to give their reasoning to their answers on how 'good' the tender is done according to the criteria points the participants got.

12. Please list the categories of data subjects and their geographical location.

I will select participants that are professionals working Dutch construction with experience in tenders, project management.

13. Will you be receiving personal data from or transferring personal data to third parties (groups of individuals or organisations)?

- No

16. What are the legal grounds for personal data processing?

- Informed consent

17. Please describe the informed consent procedure you will follow below.

The participant will be asked to fill in the digital informed consent form before we are starting the experiment/interview. I will ask them to sign it. After the experiment/interview I will store the digital form in OneDrive of TU Delft till the end of the thesis. After, I will transfer these forms to my supervisor Tong Wang, she will keep this information stored safely, using Project Data Storage.

18. Where will you store the physical/digital signed consent forms or other types of proof of consent (such as recording of verbal consent)?

The storage option indicated earlier in the plan will be used, OneDrive of TU Delft. After the project is done the responsibility will be of my TU Delft supervisor Tong Wang, using Project Data Storage.

19. Does the processing of the personal data result in a high risk to the data subjects? (Select all that apply.)

If the processing of the personal data results in a high risk to the data subjects, it is required to perform a Data Protection Impact Assessment (DPIA). In order to determine if there is a high risk for the data subjects, please check if

any of the options below that are applicable to the processing of the personal data in your research project.

If any category applies, please provide additional information in the box below. Likewise, if you collect other type of potentially sensitive data, or if you have any additional comments, include these in the box below.

If one or more options listed below apply, your project might need a DPIA. Please get in touch with the Privacy team (privacy-tud@tudelft.nl) to get advice as to whether DPIA is necessary.

- None of the above apply

23. What will happen with the personal data used in the research after the end of the research project?

- Anonymised or aggregated data will be shared with others

24. For how long will personal research data (including pseudonymised data) be stored?

- Personal data will be deleted at the end of the research project

25. How will your study participants be asked for their consent for data sharing?

- In the informed consent form: participants are informed that their personal data will be anonymised and that the anonymised dataset is shared publicly

V. Data sharing and long term preservation

27. Apart from personal data mentioned in question 23, will any other data be publicly shared?

Please provide a list of data/code you are going to share under 'Additional Information'.

- No other data/code can be publicly shared – please explain below why data/code cannot be publicly shared

I will only share the anonymised interview transcripts to the appendix of the report of my master thesis.

VI. Data management responsibilities and resources

33. If you leave TU Delft (or are unavailable), who is going to be responsible for the data/code resulting from this project?

My supervisor Tong Wang, assistant professor at the TU Delft, with email address t.wang-12@tudelft.nl

34. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

I will do the data management myself and therefore no other resources will be required

F.2. Consent Form

Delft University of Technology
HUMAN RESEARCH ETHICS
PARTICIPANT INFORMATION

Beste lezer,

U wordt uitgenodigd om deel te nemen aan een onderzoek genaamd Generative AI in Infrastructure Tendering. Dit onderzoek wordt uitgevoerd door Roos van Duuren van de TU Delft, in samenwerking met Count & Cooper.

Het doel van dit onderzoek is de interactie tussen mens en Generative AI te onderzoeken en zal maximaal 120 minuten in beslag nemen. U wordt gevraagd om mee te doen aan een experiment om een tender te reviewen. Dit zal in samenwerking gaan met Generative AI. Om de onderzoeksresultaten zo min mogelijk te beïnvloeden, wordt geen extra informatie blootgegeven omtrent het experiment.

Afsluitend wordt u gevraagd deel te nemen aan een discussiegroep om uw ervaringen van het experiment te delen.

De data zal gebruikt worden voor een master thesis en deze zal worden gepubliceerd op TU Delft Repository.

Zoals bij elke online activiteit is het risico van een databreuk aanwezig. Wij doen ons best om uw antwoorden vertrouwelijk te houden. We minimaliseren de risico's door alle data vertrouwelijk te waarborgen en alle data die worden gebruikt in het rapport zullen anoniem worden vermeld.

Uw deelname aan dit onderzoek is volledig vrijwillig, en u kunt zich elk moment terugtrekken zonder reden op te geven. U bent vrij om vragen niet te beantwoorden. Na uiterlijk een week na de einddatum van de master thesis, zal de data, verkregen uit het experiment, worden verwijderd.

Naam onderzoeker: Roos van Duuren

Technische Universiteit Delft
Toestemmingsformulier

GELIEVE DE JUISTE VAKJES AAN TE KRUISEN	Ja	Nee
A: ALGEMENE OVEREENSTEMMING - ONDERZOEKSDOELEN, TAKEN VAN DEELNEMERS EN VRIJWILLIGE DEELNAME		
1. Ik heb de informatie over het onderzoek gelezen en begrepen. Ik heb de mogelijkheid gehad om vragen te stellen over het onderzoek en mijn vragen zijn naar tevredenheid beantwoord.	☐	☐
2. Ik doe vrijwillig mee aan dit onderzoek en ik begrijp dat ik kan weigeren vragen te beantwoorden en mij op elk moment kan terugtrekken uit de studie, zonder dat ik hiervoor een reden moet hebben.	☐	☐
3. Ik begrijp dat mijn deelname aan het onderzoek de volgende punten bevat: o Digitale vragenlijsten zullen worden voorgelegd, informatie hieruit zal anoniem gebruikt worden in het verslag van de master thesis. o De interviews worden opgenomen en deze zullen tijdelijk worden opgeslagen in de TU Delft OneDrive. o De opgenomen interviews zullen worden getranscribeerd en geanonimiseerd, waarna de audiobestanden verwijderd zullen worden.	☐	☐
4. Ik begrijp dat de studie op uiterlijk 1 juli 2025 eindigt en dat geanonimiseerde onderdelen van het interview gepubliceerd kunnen worden als onderdeel van het onderzoek.	☐	☐
B: POTENTIAL RISKS OF PARTICIPATING (INCLUDING DATA PROTECTION)		
5. Ik begrijp dat mijn deelname de volgende risico's met zich meebrengt: databreuk en verlies van opgenomen bestanden/transcripties. Ik begrijp dat deze risico's worden geminimaliseerd door gebruik te maken van TU Delft officiële OneDrive en de verwerkte bestanden direct te verwijderen.	☐	☐
6. Ik begrijp dat mijn deelname betekent dat er persoonlijke identificeerbare informatie en onderzoeksdata worden verzameld, met het risico dat ik hieruit geïdentificeerd kan worden.	☐	☐
7. Ik begrijp dat binnen de Algemene Verordering Gegevensbescherming (AVG) een deel van deze persoonlijk identificeerbare onderzoeksdata als gevoelig wordt beschouwd.	☐	☐
8. Ik begrijp dat de volgende stappen worden ondernomen om het risico van een databreuk te minimaliseren en dat mijn identiteit op de volgende manieren wordt beschermd in het geval van een databreuk: o Data wordt alleen opgeslagen op de TU Delft OneDrive. o Opgenomen interviews worden getranscribeerd en geanonimiseerd. Alleen begeleiding van de TU Delft en Count & Cooper heeft toegang tot de geanonimiseerde data. o Nadat de opnames zijn getranscribeerd en anoniem zijn gemaakt worden deze opnames verwijderd.	☐	☐
9. Ik begrijp dat de persoonlijke informatie die over mij verzameld wordt en mij kan identificeren, zoals naam, werkplaats en contactgevens, niet gedeeld worden buiten het onderzoek.	☐	☐
10. Ik begrijp dat de persoonlijke data die over mij verzameld wordt, uiterlijk 1 juli 2025 vernietigd wordt.	☐	☐
C: PUBLICATIE, VERSPREIDING EN TOEPASSING VAN ONDERZOEK		

GELIEVE DE JUISTE VAKJES AAN TE KRUISEN	Ja	Nee
11. Ik begrijp dat na het onderzoek de geanonimiseerde informatie gebruikt zal worden voor het verslag van de master thesis die door de onderzoeker zal worden geschreven en dat dit verslag openbaar beschikbaar zal zijn in de TU Delft Repository.	☐	☐
12. Ik geef toestemming om mijn antwoorden, ideeën of andere bijdrages anoniem te quoten in resulterende producten. Hierbij zal ik voorafgaand aan de publicatie het rapport in kunnen zien en kan ik tegengaan dat informatie dat volgens mijn weten niet klopt, niet gepubliceerd wordt.	☐	☐
D: (LANGDURIGE) OPSLAG EN TOEGANG VAN GEGEVENS		
13. Ik geef toestemming om de geanonimiseerde data, verwerkte transcripties die over mij verzameld worden, gebruikt mogen worden in dit het verslag van de masterthesis dat deze vervolgens gepubliceerd mag worden in de TU Delft Repository.	☐	☐

Handtekeningen

_____	_____	_____
Naam deelnemer	Handtekening	Datum

Ik, de onderzoeker, verklaar dat ik de informatie en het toestemmingsformulier correct aan de potentiële deelnemer heb voorgelegd en, naar mijn beste kunnen, heb verzekerd dat de deelnemer begrijpt waar hij/zij vrijwillig mee instemt.

Roos van Duuren

_____	_____	_____
Naam deelnemer	Handtekening	Datum

Contactgegevens van de onderzoeker voor verdere informatie:

Roos van Duuren

+31 6 3742 5724

R.S.vanDuuren@student.tudelft.nl

F.3. Ethics Approval Application

Date 23-Feb-2025
Correspondence hrec@tudelft.nl

Human Research Ethics
Committee TU Delft
(http://hrec.tudelft.nl)

Visiting address
Jaffalaan 5 (building 31)
2628 BX Delft

Postal address
P.O. Box 5015 2600 GA Delft
The Netherlands

Ethics Approval Application: Generative AI in infrastructure tendering
Applicant: Duuren, Roos van

Dear Roos van Duuren,

It is a pleasure to inform you that your application mentioned above has been approved.

Thanks very much for your submission to the HREC which has been approved.

In addition to any specific conditions or notes, the HREC provides the following standard advice to all applicants:
• In light of recent tax changes, we advise that you confirm any proposed remuneration of research subjects with your faculty contract manager before going ahead.
• Please make sure when you carry out your research that you confirm contemporary covid protocols with your faculty HSE advisor, and that ongoing covid risks and precautions are flagged in the informed consent - with particular attention to this where there are physically vulnerable (eg: elderly or with underlying conditions) participants involved.
• Our default advice is not to publish transcripts or transcript summaries, but to retain these privately for specific purposes/checking; and if they are to be made public then only if fully anonymised and the transcript/summary itself approved by participants for specific purpose.
• Where there are collaborating (including funding) partners, appropriate formal agreements including clarity on responsibilities, including data ownership, responsibilities and access, should be in place and that relevant aspects of such agreements (such as access to raw or other data) are clear in the Informed Consent.

Good luck with your research!

Sincerely,

Dr. Ir. U. Pesch
Chair HREC
Faculty of Technology, Policy and Management