Utilizing ChatGPT as a Virtual Team Member in a Digital Transformation Consultancy Team

Tim de Wolff and Sietse Overbeek[a]

Department of Information and Computing Sciences, Faculty of Science, Princetonplein 5, 3584 CC, Utrecht University, The Netherlands

Keywords: Business Architecture, ChatGPT, Consultancy, Digital Transformation, GenAI.

Abstract: This study aims to design and evaluate a method that leverages ChatGPT for efficiency improvement in digital transformation projects, specifically while designing target business architecture products. The main research question is stated as follows: 'How can a large language model tool be utilized to support the development of target business architecture products?' The resulting method, GenArch, enables utilization of ChatGPT throughout business architecture design processes. This method is validated by means of expert interviews and an experiment. The perceived ease of use, perceived usefulness, and intention to use of the method are analyzed to assess the perceived efficacy, which serves as an indicator for efficiency. The results show that GenArch possesses at least a moderately high level of perceived efficacy.

1 INTRODUCTION

Recent developments in the field of Generative Artificial Intelligence (GenAI) and ChatGPT have affected both academia and practice in a wide variety of domains. GenAI can be defined as the automated construction of intelligence Zant et al. (2013). ChatGPT is a Large Language Model (LLM)-based chatbot that allows users to conduct, refine, and steer meaningful discussions about topics of their preference. The models behind ChatGPT are trained on hundreds of terabytes of textual data and are therefore highly effective Shanahan (2024). LLMs are generative mathematical models of statistical distribution that generate statistically likely continuations of words based on the input prompt. Because of context dependency, the inclusion of many activities, and the large amount of digital possibilities Westerman et al. (2014), Digital Transformation (DT) is a suitable domain for different types of support and guidance, like text generation and advice that ChatGPT can deliver Liu et al. (2023). DT is defined as the use of new digital technologies, like for example social media, big data or AI, to enable major business improvements, such as enhancing customer experience, improving operations or creation of new business models Fitzgerald et al. (2013). A specific part of DT is Business Architecture (BA). This is defined as an enterprise map that provides a common understanding and is used to align strategic objectives and tactical demands Simon and Schmidt (2015). The target BA is the to-be state that is desired by an organization and the goal of the DT. This map can be divided in a variety of BA products which can most effectively be designed and deployed by the use of feedback cycles Hohpe (2017). Examples are a capability map, which is a visualization of an organization based on distinct abilities to perform unique business activities Wißotzki and Sandkuhl (2015), and a value stream map, which is a map of sequences of activities that are required to deliver products or services LeanIX (2024).

This study aims to contribute to the domain of GenAI by the design and validation of a method that describes how ChatGPT can be leveraged in DT projects to improve efficiency when designing target BA products. The professional application, the associated risks, and the influence of prompts are investigated. ChatGPT has the potential to transform businesses by automating and executing language-based tasks with unprecedented speed and efficiency KPMG (2023). Making BA decisions in advance is suboptimal, as additional information becomes available over time, allowing for more informed decisions Hohpe (2017). We therefore focus on the design of target BA products rather than entire BAs to ensure that not all decisions need to be made in advance. Thereby, tar-

[a] https://orcid.org/0000-0003-3975-200X

Wolff, T. and Overbeek, S.
Utilizing ChatGPT as a Virtual Team Member in a Digital Transformation Consultancy Team.
DOI: 10.5220/0013286200003929

In *Proceedings of the 27th International Conference on Enterprise Information Systems (ICEIS 2025) - Volume 2*, pages 707-721
ISBN: 978-989-758-749-8; ISSN: 2184-4992

get BA product design processes have a specific output, namely the product itself, and could likely be executed more efficiently while utiziling ChatGPT.[1 2 3] This is expected as BA involves lots of stakeholders, activities and uncertainties, and requires proper alignment over all aspects.[1 2 3] Inefficiencies in practice result from insufficient input, the need to start from scratch, time investment in modeling tasks, and deliberations with stakeholders.[2 3] This study aims to provide a clear overview of how target BA product design processes could incorporate the use of ChatGPT in a structural and efficient manner. Within this research, the focus specifically lies on ChatGPT because of the available (grey) literature, the isolated interface, and more advanced storage and sharing options compared to alternatives. The design problem Wieringa (2014) is stated as follows: Improve *the design process of target Business Architecture products* by *employing a method that guides the proper use of an LLM tool (taking the example of ChatGPT)* that satisfies *requirements of the target Business Architecture products* in order to *increase efficiency of the design processes for the designer*.

This research is performed in collaboration with accounting firm KPMG to gain access to experts and company documentation. A case study is performed at KPMG Netherlands within the Digital Transformation team, which is part of the overarching Digital team in their Advisory business unit. This team focuses on transformation projects in which they guide transformation processes. The aim of this study is to design and validate a method. A method is defined as an approach to perform a systems development project, based on a specific way of thinking, consisting of directions and rules, structured in a systematic way in development activities and deliverables Brinkkemper (1996). The designed method is aimed at DT projects, though, and does not solely involve development activities. The proposed method will serve as guidance for its stakeholders, prescribing the appropriate timing and utilization of ChatGPT when designing a target BA product. Furthermore, an overview of risks associated with ChatGPT in a professional context are provided as part of our study and we contribute to the fields of prompt engineering and BA. BA is often viewed as merely one of the elements making up an Enterprise Architecture (EA), while it helps establish the pivotal connection between the business and IT sides of companies Bouwman et al. (2011). This research aims at improving the efficiency of the target BA product design process. Both scientific and grey literature focus on the quality of the BA, tools or frameworks, and the content of the BA. However, to the best of our knowledge, no literature is existent that aims at improving efficiency of such a process.

The remainder of this paper is as follows. First, the research approach is presented in section 2. Second, the results of a Multivocal Literature Review (MLR) are discussed in section 3. Third, the research process, including two iterations of the design cycle, is described in section 4. Fourth, the results of the experiment and analysis of these results are discussed in section 5. Finally, the paper ends with a discussion in section 6 and conclusions and future research in section 7.

2 RESEARCH APPROACH

This section presents the research approach that is adopted throughout this study. This includes the research questions, the conducted method and the literature review and case study protocols. The main aim of this research is to design and evaluate a method that describes how ChatGPT can be leveraged in DT projects to improve efficiency when designing target BA products. The proposed method will serve as guidance for stakeholders, prescribing the appropriate timing and utilization of ChatGPT within a BA context.

2.1 Research Questions

To be able to fulfill this aim, the following Main Research Question is formulated:

Using the example of ChatGPT, how can a Large Language Model tool be effectively integrated into the design process of target Business Architecture products to mitigate risks, optimize prompt impact, and improve efficiency?

To be able to answer this MRQ, five Sub Research Questions are constructed. **SRQ1.** *In what ways does ChatGPT pose risks within a professional context of target Business Architecture design?* This SRQ is focused on risks that occur when using ChatGPT in a professional context. A MLR is performed, aiming to identify risks that could occur during professional use. These risks are to be mitigated during the design of the method in SRQ4. Section 3.1 presents the findings of this review. **SRQ2.** *What is the impact of prompts on the output generated by ChatGPT*

[1] KPMG Senior Manager, personal communication, May 2024

[2] KPMG Manager1, personal communication, March 2024

[3] KPMG Manager2, personal communication, May 2024

within a professional context of target Business Architecture design? This SRQ also requires an extensive literature review. Section 3.2 presents the findings of this MLR. **SRQ3.** *How are design processes of target Business Architecture products shaped?* The exact activities and responsibilities are retrieved during the last part of the MLR in section 3.3. This review aims to identify the BA products and the steps that are part of the design process. Together with additional input of KPMG experts, this SRQ is answered in iteration Alpha. **SRQ4.** *How can a method be introduced to effectively leverage ChatGPT in the design process of target Business Architecture products?* With the input of the previous SRQs, a method can be constructed. The technical version of this method is designed using a Process Deliverable Diagram (PDD). This is a meta-modeling technique, based on UML activity and class diagrams Weerd van de and Brinkkemper (2008). The used method engineering protocol can be found in Wolff de (2024). **SRQ5.** *To what extent does the method incorporating ChatGPT improve the efficiency of the target Business Architecture design process?* To be able to find out how the constructed method performs, a case study at KPMG is performed. The focus lies on efficiency, for which the perceived ease of use, perceived usefulness and intention to use are considered. Section 2.2.2 elaborates on these variables. The case study consists of two separate parts: expert interviews and an experiment. These are executed at separate stages in the research process. Section 2.2 outlines the research method and covers this case study.

2.2 Design Science Method

Design science can be seen as the design and investigation of artifacts in context Wieringa (2014). We design a method to interact with the problem context of DT consultancy processes to improve the efficiency when designing target BA products. This study applies the design cycle that is shown in Figure 1. In this cycle, design is decomposed in three tasks: problem investigation, treatment design and treatment validation Wieringa (2014). It is presented as a cycle as researchers iterate over these activities multiple times. During this research, we iterate over the design cycle twice. These iterations involve dissimilar validation methods that lead to distinct insights. During both cycles, a large amount of the data will be generated at KPMG by performing a case study Heale and Twycross (2018). The case study consists of analyzing company documentation, conducting two expert interviews and executing an experiment. Explanations of both iterations, as well as their accompanying steps and deliverables, are presented in the remaining part of this section.

2.2.1 Iteration Alpha

We refer to the first iteration as iteration Alpha. This iteration is focused on designing an initial version of the method and validating this through expert interviews with scholars and practitioners. This iteration consists of the steps mentioned hereafter. **Problem investigation**. First, the problem and its context are analyzed. This is done by reviewing related works and conducting an MLR. The MLR focuses on the risks of ChatGPT in a professional context, prompt engineering and the BA process. This type of literature review is chosen to be able to explore recent insights in the relatively new research topic GenAI. At the completion of this stage, SRQs 1, 2 and 3 should be answered. **Treatment design**. Second, the first version of the treatment is designed based on insights of the problem investigation. The treatment is the interaction between the artifact and the problem context Wieringa (2014). The artifact in this research is a method, as this fits the goal of specifying when and how ChatGPT should be used throughout a process to improve efficiency. The name of this method is GenArch, a merger of the terms GenAI and architecture. The GenArch method aims to serve as guidance for stakeholders, prescribing the appropriate timing and utilization of ChatGPT within the context of BA. The inclusion of ChatGPT in the method is based on insights of the experts from KPMG and literature about GenAI. A method engineering approach is adopted for the creation of the GenArch method. The Process Deliverable Diagram is selected as the meta-modeling technique. This technique is especially developed for method engineering purposes and can be used for analyzing and assembling method fragments Weerd van de and Brinkkemper (2008). A separate version, which does not follow the PDD notation and shows a ballpark view of the method, is designed as well to increase understandability of GenArch. **Treatment validation**. Third, the initial version of the method is validated using expert interviews. In total, six semi-structured interviews are conducted with KPMG experts, external experts and scholars to generate qualitative data about potential improvements. Semi-structured interviews start with an interview protocol comprised of open-ended questions and allow for follow-up questions of the interviewer Magaldi and Berler (2020). The interviewees are selected based on their knowledge of and experience with DT consultancy, EA and GenAI. The focus of the validation lies on validating the design process

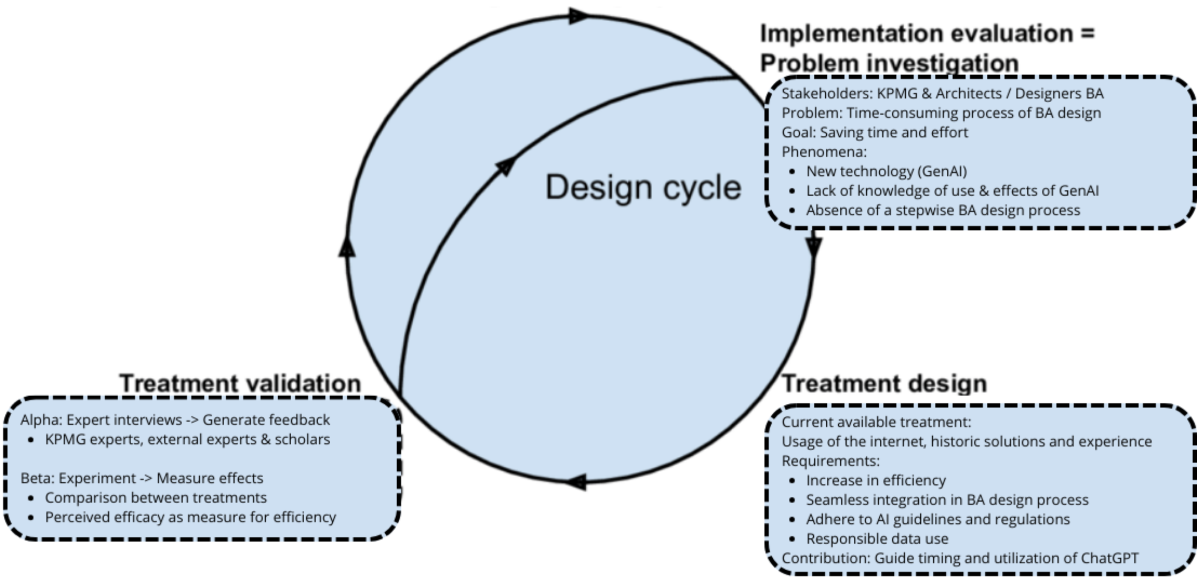

Figure 1: Design cycle.

of BA products and the categorization of the activities within this process.

2.2.2 Iteration Beta

We refer to the second iteration as iteration Beta. This iteration is focused on analyzing the gathered qualitative data from the expert interviews to identify improvement possibilities. Subsequently, the method is modified based on these insights and validated by an experiment, performed at KPMG. This iteration consists of the following steps: **Case evaluation**. The qualitative interview data is analyzed to identify improvement possibilities. **Treatment redesign**. The method is improved based on the insights gathered during the case evaluation. **Treatment validation**. The improved, final version of the method is validated using an experiment, the last part of the case study. The focus of this validation lies on validating if participants perceive (a segment of) the method as more efficient. The experiment is performed in an off-line setting with professionals from the Digital Transformation team of KPMG Netherlands. Figure 2 illustrates the design of the experiment. The goal of this experiment is to measure the efficiency via the variables introduced hereafter. Both treatment validation steps are aimed at finding potential improvements for the method and validating the model based on criteria. We retrieve these criteria from the method evaluation model Moody (2003). This model provides mechanisms for evaluating both the likelihood of acceptance and the actual impact of a method in practice Abrahão et al. (2009). The advantage over other comparable models is the incorporation of actual efficacy and actual usage Abrahão et al. (2009). To be able to validate if the GenArch method is an improvement over the current approaches, the perceived efficacy is used as the measure and taken as indicator for the efficiency. This measure consists of three sub-measures: The perceived ease of use, which refers to the expected required effort to learn and use the method; the perceived usefulness, which refers to the expected degree to which the method will achieve intended objectives; the intention to use, which refers to the extent to which a person intends to use a particular method Abrahão et al. (2009). The ethical considerations of this research are found in Wolff de (2024).

3 LITERATURE STUDY RESULTS

The results of the MLR are divided into four parts: the risks associated with professional use of ChatGPT, the state-of-the-art developments in the field of prompt engineering, the steps and products needed to create a target BA, and a discussion of four works related to this study.

3.1 Risks ChatGPT

As mentioned in section 1, ChatGPT is trained on lots of input data Shanahan (2024). This data partially consists of copyrighted texts, leading to concerns about legal complications Piñeiro-Martín et al. (2023). Thereby, it is not known how much data is

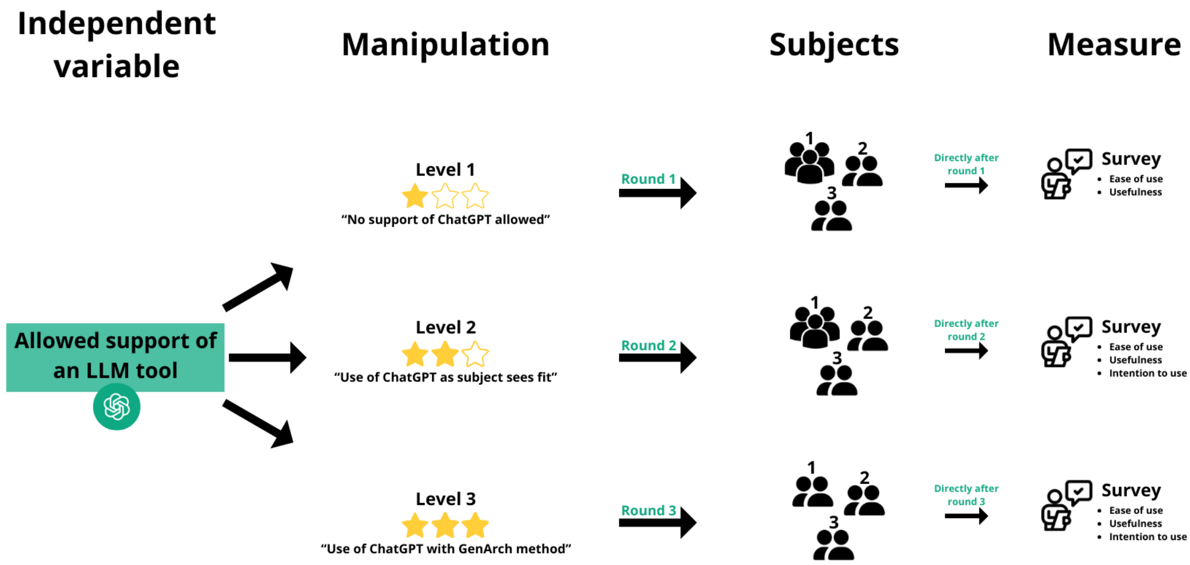

Figure 2: Experiment design.

used for the training of the LLM. LLMs often operate as black boxes, which makes it challenging to understand the decision-making processes behind the provided results and expert knowledge is then required to verify produced content Piñeiro-Martín et al. (2023). User instructions are part of the GenArch method and include reminders to fact-check important results. Furthermore, there are risks associated with the current inability of GenAI to distinguish between concepts acquired through data compression and singular memorized concepts. This inability leads to hallucination, i.e., the fabrication of credible but factually incorrect content and to talk about this with confidence Piñeiro-Martín et al. (2023). Therefore, LLMs are still considered far from reliable Qiu et al. (2023). ChatGPT also does not memorize mathematical concepts and the corresponding rules but aims to achieve pattern recognition, sometimes resulting in wrongly performed mathematics Ferrini (2023). It could occur that a GenAI algorithm leaks an example of text or visual art that it has memorized, which in turn leads to plagiarism Ferrini (2023). This is likely one of the reasons that organizations are reluctant to use sensitive data in input prompts. Therefore, expert verification is an important recurring step in GenArch. Another risk is that LLMs learn from vast amounts of training data, which may inadvertently contain biases Piñeiro-Martín et al. (2023). Thereby, they could partially embody the biases of their creators Tacheva and Ramasubramanian (2023). A consequence then is the potential to intensify discrimination and inequality Tacheva and Ramasubramanian (2023). LLMs present language bias as well, as they perform well in English compared to other languages as most training data was English Qiu et al. (2023). Therefore, in GenArch the results are used as input or suggestions, but decisions are made by humans to increase the chance biases are avoided. The last risk relates to privacy. As LLMs memorize input data it is possible to extract sensitive information from this data using prompts Qiu et al. (2023). Examining the privacy issues that are associated with sensitive information before putting it in a prompt is therefore crucial and it is advised to use GenArch in private GenAI tools to mitigate these risks.

3.2 Prompt Engineering

In essence, prompt engineering entails optimizing textual input to effectively communicate with Large Language Models Bains (2023). Prompt engineering is the process of formulating a prompt in a way that a GenAI system produces an output that closely matches the expectations Bains (2023). It involves considering the inner working of an AI model to be able to construct inputs that work well with the model Bains (2023). Prompt engineering skills are vital for fully leveraging LLMs, but they do not come naturally and need to be learned Wang et al. (2024). The designed method includes a prompt template that increases the likelihood of useful results. Prompt patterns can be defined as summaries of effective prompt-tuning techniques that provide an approach to crafting the input and interaction to achieve desired output Wang et al. (2024). Proven patterns will be included in the prompt templates of GenArch. The exact patterns included in the method are chosen during the treatment design phase. In the persona adoption pattern, a user asks the LLM to play a particular role without providing further details Wang et al. (2024); Bains (2023). Closely related to this is the appliance in reverse by instructing to complete a task with a specific audience in mind Bains (2023). Both appliances of this pattern will be included in the prompt templates and suggestions. Chain-of-thought prompting is appropriate for problem-solving Bains (2023) and generates a sequence of steps with explanations before inferring the output. This manner of prompting will be advised when ChatGPT does not answer as expected during use of GenArch.

Another prompting technique is few-shot prompting, where examples are included within a prompt to assure the output meets expectations Willey et al. (2023). Target-your-response prompting is focused on the output of the system. If the GenAI tool is not explicitly told about the appearance of an answer, it can give results in many forms. This prompt contains two elements, the question or problem and an explanation of what the response should be like Eliot (2023). The inclusion of the desired response in the prompt should cause the use of GenAI to be sufficiently efficient, and if users pay per transaction, to possibly be less costly Eliot (2023). It will therefore be included in the prompt templates. Multi-turn or continuous prompting involves a dialogue or conversation between the model and the user to iteratively get to the desired output rather than optimizing a single prompt Bains (2023). Multi-turn prompting will not be a standard component of the method but rather an advise to further explore the results that ChatGPT provides. Meta prompting denotes using higher-order prompts to let the LLM generate its own natural language prompts for certain tasks Korzynski et al. (2023). This approach aims to leverage the inherent capabilities and understanding of natural language of such models to create more effective prompts Korzynski et al. (2023). This prompting strategy will be used as input during prompt template design.

3.3 Business Architecture Steps

This section is focused on identifying required products and steps to create a target BA. These are capabilities, value streams, principles, business processes, roles, business functions, a gap analysis, policies and change management. All of these concepts are included in the initial version of GenArch as steps or products. A capability is defined as the capacity of an organization to successfully perform a unique business activity to create a specific outcome Wißotzki and Sandkuhl (2015). Examples are human resource management (HRM) and customer relationship management (CRM). Capabilities can be mapped in a business capability map Group (2022), which is a visualization of an organization based on distinct business capabilities Smith (2024). Value streams are sequences of activities that are required to deliver a product or service to a customer LeanIX (2024); Smith (2024). It is a collection of all activities, value adding as well as non-value adding, that are required to go from raw material to the end customers Singh et al. (2011). Value streams show how the capabilities enable value, as the value flows through the capabilities and gives context to why capabilities are needed Group (2022). An example is to first receive requirements, then verify these and subsequently develop software based on these requirements to be able to ultimately deploy a fitting software solution. Value stream mapping is defined as the outlining of activities that an organization performs to create value that is being delivered to stakeholders Group (2022).

Principles are general rules and guidelines that inform and support the way in which an organization fulfills its mission Group (2022). In this research, we specifically focus on architecture principles that bridge the gap between high-level intentions and concrete design decisions, and document fundamental choices in an accessible form Greefhorst and Proper (2011). BA is principle-driven and the principles are preferably understandable, robust, complete, consistent and stable Group (2022). Models of business processes are one of the most established elements of BA and describe components in business processes with additional information like ownership or type of activity Smith (2024). An example is a sales process of generating leads, qualifying them as potential customers, and closing deals. In the context of the broader approach that is BA, value streams and business processes are integrated to help provide a comprehensive understanding of the vital processes Smith (2024). The different roles and responsibilities that should be allocated for capabilities and activities are closely related to the organizational structure. The structure displays the formal hierarchy of the organization including departments, teams, roles and responsibilities LeanIX (2024). The decisions related to the allocation of the responsibility of capabilities and business processes are key in BA. Other elements closely related to the organizational structure are business functions and units and BA tries to achieve integration between these. The technique known as gap analysis is widely used in TOGAF to validate a developed target BA Group (2022). The basic idea is to highlight the differences between the baseline and target Group (2022). Identified gaps are also input for the roadmap towards achieving the target BA. Policies and standards define the rules and guidelines that have to be adhered to in various aspects of the organization LeanIX (2024). Having such rules causes consistency and compliance with regulations. Standards are often derived from best practices. Policies come from regulations, these could be sector-specific or enforced by law LeanIX (2024). The last element discussed relates to how to manage the execution of the actual DT. This is referred to as change management LeanIX (2024) which is an approach of dealing with change in terms of strategy, control and helping stakeholders adapt.

3.4 Related Work

Four related works were identified. First, we discuss an evaluation of ChatGPT as a tool in common business decision-making cases Chuma and Oliveira (2023). In this study, ChatGPT is provided with three simple questions, where ChatGPT showed simple text in all cases but appears useful to present topical overviews. Another paper is concerned with the idea of having ChatGPT as a virtual member of a software development team to inform, coach and execute a share of the development work Bera et al. (2023). An experiment is included to assess the performance of ChatGPT with tasks that would be performed by scrum masters. Both of these studies are executed in fields related to the DT field. Although both studies are exploratory, they both reveal the potential of the professional use of ChatGPT. The subsequent paper can be described as an exploration of the use of GenAI in the software industry Ebert and Louridas (2023). It is concluded that GenAI has the potential to significantly improve software production by automation, enhancing creativity, improving accuracy, and streamlining development processes. Lastly, a paper related to HRM and prompt engineering is identified. It is shown that GenAI can be a helpful assistant for strategic and operational tasks that HRM specialists perform Aguinis et al. (2024). Guidelines are provided regarding effective prompt design and a verification process is implemented to check outputs. All of these papers relate to this research as they contain case studies, risk assessments and prompt engineering. However, no papers have been found that are concentrated on the field of DT. Therefore, this research addresses the gap of the application of GenAI in BA design and also aims to extend knowledge on prompt engineering in this field by the performed MLR. Furthermore, no literature has been found that aims at improving efficiency of the BA design process.

The case study in this research is executed over multiple iterations in the context of method validation. The aim is to generate results on efficiency when compared to the exploratory case studies in the mentioned papers. The findings of the MLR have implications for method design. First, the identified risks (Section 3.1) are taken into account throughout the entire design. Elements that are included are: limiting the recommended use to activities that involve common concepts such that a decent amount of training data can be expected, providing guidelines regarding proper use of the method, and urging users to fact-check results. Moreover, decisions should be made by humans and, preferably, users apply the GenArch method in a paid version of ChatGPT where the GenAI model cannot use input prompts to learn OpenAI (2023). Second, the findings on prompt engineering (Section 3.2) are leveraged by including a prompt template in the method itself. The full prompt template is presented in Wolff de (2024). Third, the identified BA products and steps (Section 3.3) are included in the method.

4 THE GenArch METHOD

This section describes the GenArch method that is designed during this research. First, the method and all of its components are explained. Subsequently, the application of the design cycle as explained in section 2 is outlined. The design cycle of this research includes two iterations: Iteration Alpha is focused on designing and validating an initial version of the GenArch method. Iteration Beta is focused on analyzing the insights gathered during the validation in Alpha, using these to redesign the initial version of the GenArch method and validating this through an experiment with KPMG experts. Extensive explanations of the process of both these iterations can be found in Wolff de (2024). The experiment is explained in more detail in section 4.2. The treatment in this research is a method that serves as guidance for stakeholders, prescribing the appropriate timing and utilization of ChatGPT in the design of BA products. The model uses colors of the visual identity of KPMG KPMG (2024b), recognizable illustrations, and short descriptions. Figure 3 presents the ballpark view of the GenArch method.

The BA products and main activities are based on the insights of the problem investigation of iteration Alpha. Each main activity, from now on referred to as phase, is divided into smaller activities. The key stakeholders, policies, baseline assessment, design product / decision, gap analysis and change management sub-activities are identified in the MLR as presented in section 3.3. The remaining sub-activities are identified in collaboration with KPMG experts.[1234] BA products are often redesigned and updated more than once over time[1] KPMG (2024a). Therefore, the method is iteratively designed so users can loop through the process multiple times. Activities that are marked with OpenAI logos are classified as 'suited for ChatGPT'. This does not mean that ChatGPT completely takes care of this activity. It means that ChatGPT can be utilized as a reference, as an input source or a conversation partner. Human input is still always needed to refine, update, judge or provide context to the results.

If the capability map is selected as the example BA product for applying the GenArch method, it is plugged in the process in the middle of figure 3. The method starts with creating a vision for this capability map by answering the following questions: Who are the stakeholders and what are their needs?; What are the objectives?; What is the scope?; finally: How do we measure progress towards our goals? After this phase is complete, architecture principles need to be designed based on best practices and relevant policies. Architecture principles can be reused from earlier projects.[234] Subsequently, a baseline is analyzed by identifying strengths, weaknesses and root causes for these weaknesses. After this phase, the target capability map needs to be designed. This could be an update of a current version or a completely new map. After finishing this capability map, the method continues with conducting a gap analysis on the baseline and target capability map while also taking risk mitigation into account. This process is finished by designing a roadmap that includes an implementation strategy, a change management strategy and a timeline with important milestones. For each activity that is marked with an OpenAI logo, the tool ChatGPT is used. The recommended use of ChatGPT can be found at the bottom of figure 3. The 'recommended use of ChatGPT' section consists of six sub sections. The 'intended role of ChatGPT' and 'use cases' sections are derived from the categorization in iteration Alpha where the ways of using ChatGPT are explained. The 'potential risks' section is directly derived from the sub review of the MLR about the risks of ChatGPT (see: Section 3.1). The 'risk mitigation strategy' section is also partly based on this sub review but additionally includes parts of the guidelines of KPMG regarding the use of GenAI. The 'prompting strategy' and 'refine output' sub sections describe the patterns and elements used in the prompt template of the GenArch method (see: Wolff de (2024) for more details). The iterative nature and suitability for smaller decisions within BA products is assumed to make GenArch well-suited for an agile architecture. The method takes the following principles of the agile manifesto into account: prioritizing continuous delivery and welcoming changing requirements Alliance (2001). The other principles are out of scope of the method due to their specific nature or focus on software. Recall that besides the ballpark view of the GenArch method (see: figure 3), a technical version is designed to give more detail about the functioning of the GenArch method. Appendix A presents the technical version of the GenArch method. Wolff de (2024) exhibits corresponding activity and concept tables in which the activities and concepts as part of the technical version are explained in detail.

[4]KPMG Manager3, personal communication, April 2024

4.1 Iteration Alpha and Beta

The first step of iteration Alpha is the problem investigation of which the literature review of section 3 is the primary part. In addition to the literature review, a baseline BA product design process is created. The second step is the design of the first version of a method that serves as guidance for stakeholders, prescribing the appropriate timing and utilization of ChatGPT in the design of BA products. To be able to decide how ChatGPT can be best leveraged in the BA product design process, each of the activities are classified as either 'suited for ChatGPT' or 'not suited'. The following indicators below are used for this classification: Needed data to perform the activity is (expected to be) publicly available; Execution of the activity does not depend on highly confidential company data; Successful execution is not highly dependent on understanding the context; The activity does not mainly consist of a decision; and: The activity does not require very recent data or information. A prompt template is created to help users of the GenArch method utilize the potential of GenAI. As a last step of iteration Alpha, this initial version of the GenArch method is validated by means of interviewing experts.

The first step of iteration Beta is to analyze the qualitative interview data that is produced as part of iteration Alpha to identify improvement possibilities. The outcomes of this analysis are meant to provide reasoning for the implemented changes. Based on the results of this analysis, the GenArch method was redesigned. The inclusion of a section that outlines the recommended use of ChatGPT is the most prominent modification. A validation of the redesigned GenArch method concludes the Beta iteration.

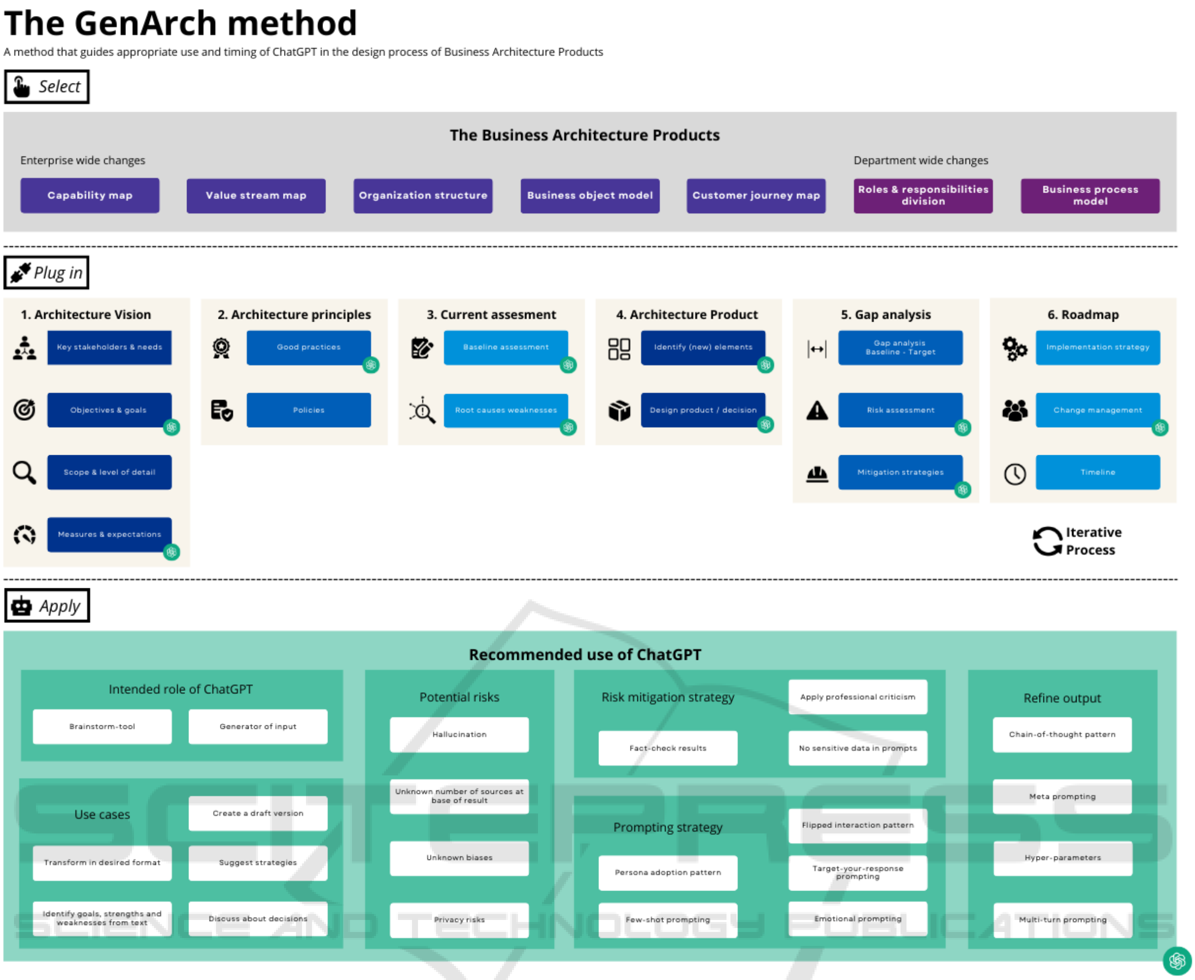

Figure 3: Ballpark view of the GenArch method version 2.

4.2 Validation Experiment

As the last step of iteration Beta, the final version of the GenArch method is validated using an experiment. Recall that the experiment design is illustrated in figure 2. The focus of this validation lies on validating if participants perceive a segment of the method to be more efficient. Wolff de (2024) includes a detailed explanation of the experiment, the workshop it is incorporated into and reasoning for its construction. The case of a hospital is taken as the basis throughout the entire workshop. In round one, subjects design the capabilities of the current state of the hospital. In round two, subjects designed a set of architecture principles for the target state of the hospital and in round three, subjects designed new capabilities for this target state. The choices for this capability and activity are based on the input of the interviewees from the Alpha validation stage.

Seven subjects participated in the workshop that was held on 7 June 2024 at the KPMG office in Amstelveen, the Netherlands. They used the free version of ChatGPT at that point in time, ChatGPT 3.5, in an empty conversation. One participant had to leave early and only participated in rounds one and two. Wolff de (2024) includes a template of the informed consent form that all subjects were required to sign before the experiment. The answers to the surveys are taken as the results of this validation to be able to draw conclusions about the perceived efficacy. Wolff de (2024) includes the full set of answers to the surveys. Section 5 discusses and analyzes these results.

5 ANALYSIS OF THE EXPERIMENT RESULTS

This section concentrates on analyzing the results of the performed experiment. Wolff de (2024) includes full details of the results for each of the three rounds of the workshop. This section focuses on analyzing these results to be able to answer SRQ 5: To what extent does the method incorporating ChatGPT improve the efficiency of the target Business Architecture design process? The experiment evaluated a segment of the GenArch method to be able to draw conclusions about the efficiency of the overall method. The capabilities segment was evaluated as the interviewees from the validation stage in iteration Alpha expected this segment to be very suitable for the use of the GenArch method. For the measurement of efficiency, we adopt the perceived efficacy of the method evaluation model as our measure (Section 2.2.2). This measure consists of three sub measures which are taken as the dependent variables in the experiment. These are the perceived ease of use, the perceived usefulness and the perceived intention to use. These are adopted from the method evaluation model as well Moody (2003). In the remainder of this chapter, each of these measures are discussed individually to be able to draw conclusions about the perceived efficacy.

5.1 Perceived Ease of Use

The first variable that is discussed is the perceived ease of use. This refers to the expected required effort to learn and use the method Abrahão et al. (2009). In round one of the experiment, subjects use the current approach, referred to as common sense, to fulfill the task. They are already familiar with this approach meaning that a comparison for the ease of use seems unnecessary. A majority of the comments of subjects were positive towards the ease of use of the GenArch method. The most relevant comments that were made are categorized and listed below. These comments were made by multiple subjects. The following are **positive** comments on **ease of use**: "*It was very helpful with a clear explanation*"; "*The method had some good questions and checks*"; and: "*I had guardrails on how to do it, and how to use it*". There was also a **negative** comment on the **ease of use**: "*Helpful but not fully convenient*". All other comments sound positive and indicate a high level of perceived ease of use. The last remark was made because of the longer startup time that was experienced by some subjects. To quantify the results, subjects of the experiment were asked to specify how much they agreed with certain statements in round three. We assign range 1 - 5 to the range 'strongly disagree' to 'strongly agree'. We take the average values for the statements as an indication of how true that statement is. This means that the higher the value, the more true the statement. Two statements relate to the perceived ease of use. The average values for these statements are as follows: **4.00** - *It was clear how to utilize ChatGPT with the GenArch method for this task*; and **4.00** - *It was easy to use ChatGPT with the GenArch method for this task*. Both of the values are high, meaning that based on these values and the comments, it can be assumed that the segment of the GenArch method has a good perceived ease of use. Due to multiple remarks about the longer startup time and large amount of information needed for the prompts, it is expected that the GenArch method is suited for larger and more complex cases as such cases need to be analyzed in detail. Thereby, appropriate training in the use of the GenArch method is assumed to further increase the perceived ease of use.

5.2 Perceived Usefulness

Second, the perceived usefulness is discussed. This corresponds to the expected degree to which the method will achieve intended objectives Abrahão et al. (2009). Subjects mostly perceived the first round as achievable in the given time. However, most did not seem completely satisfied with the results they produced as they described them as quite high level with somewhat short descriptions. As this was the round in which they applied the current approaches, other tools like search machines and reusing old KPMG material were allowed. One subject for example mentioned to have used old deliverables of KPMG for a similar case. According to him, having such a reference made the assignment achievable. Another subject mentioned the assignment to be doable but with moderate results. He appears to be of the opinion that it is possible to finish the exercise in the given time if moderate results are accepted. To summarize, the collective opinion about the usefulness of the application of their current own approaches appears to be that producing a high-level answer to the assignment was achievable in the given time. However, most subjects appeared not satisfied with their results in terms of level of detail. This seems logical, as participants were given a short time frame and restricted to current approaches. Creating capabilities tailored to the specific context in a detailed manner was anticipated to be challenging.

The results include certain positive comments about the perceived usefulness of the GenArch method. The subjects perceive the method as useful and stipulate the high quality output that GenArch manages to produce. On first glance, these results seem more positive compared to the results of round one. The most relevant comments that were made about the perceived usefulness are categorized and listed below. These comments were made by multiple subjects. **Quality output**: "*Our experience was, in short, very good. The answers we needed to give did not give room to work through it quickly which resulted in more quality input*"; "*It was very helpful. This resulted in clearer results*"; "*GenArch enhanced quality*"; "*The answers were clear and quite detailed*"; "*I prefer the GenArch method over the current approach. The answers were more detailed because you are more specific in what you want*". **Complete output:** "*GenArch resulted in more complete answers*"; "*It gives more creative answers and makes them more complete*". **Useful GenArch elements**: "*The method had some good questions and checks. Asking ChatGPT to ask the user questions before giving an answer was helpful*".

All comments seem positive and indicate a high level of perceived usefulness. Subjects seem to agree on the fact that GenArch provokes clear, detailed and complete answers which are useful for the given task. To again quantify the results, a similar range as for the perceived ease of use (Section 5.1) is given to the statements relating to the perceived usefulness. Three statements relate to the perceived usefulness and their average values are as follows: **4.17** - *ChatGPT was helpful for this task*; **4.17** - *The results that ChatGPT provided were useful*; **3.83** - *ChatGPT with the GenArch method made my answer for the exercise more complete*. The first two values are very high. The third is well above neutrality but a bit lower. This means that we can assume, based on these values and the comments, that the segment of the GenArch method has a very good perceived usefulness. However, designers cannot fully depend on the completeness of the results of ChatGPT. It is expected that this slightly lower completeness of ChatGPT is caused by the specific domain or situational knowledge that is required. Some subjects explained that designing a complete prompt that included all information was a challenge in the given time. It is anticipated that this completeness increases when users have more time to design their prompt. However, adding specific domain and situational knowledge to the result of ChatGPT by a human will always be necessary to deliver a complete answer. It is expected that this is the cause for the other values not scoring even higher as well. ChatGPT is not able to deliver results that are 100% correct due to its inner workings. Consequently, results always need to be refined, updated, judged or be put in context by a human user.

5.3 Intention to Use

As the last of the three sub variables, the intention to use is discussed. The intention to use refers to the extent to which a person intends to use a particular method Abrahão et al. (2009). Talking about the application of the current approach, which subjects use in round one, is unnecessary as this approach is currently already in use. Hence, this section does not make a comparison between round one and round three. In the results of round three, parts of the intention to use of the subjects are already discussed. Although some remarks are already discussed in this particular section, the most relevant comments that were made about the intention to use are categorized and mentioned hereafter. These comments were made by multiple subjects. There is one comment on **more training needed to prefer the GenArch method**: "*For now I prefer my own way, however I do see that the GenArch method results in higher quality output, with more experience on it I would use GenArch*". There is also one comment on the **preference of free ChatGPT use**: "*GenArch is a nice reference but I do not see it as something I would go through every time*". Two comments are made on the **preference of the GenArch method**: "*I prefer the third round, because it helped in a specific scenario, with a given prompt. So there was not much thinking required from me*"; and: "*I prefer the GenArch method. The answers were more detailed because you are more specific about your wishes*".

The comments seem a bit mixed. This is the first time the subjects use the GenArch method, which requires them to think more extensively rather than allowing them to start immediately. Hence, these results seem rational. As mentioned, the first remark could be solved by providing users with a more extensive explanation with examples or a small practice round, and the second remark could be catered for by creating different scenarios for cases that differ in size. This calls for further investigation in future research (Section 7). Subjects do seem to agree on the potential of GenArch. As was done in previous sections (Sections 5.1 and 5.2), a range is assigned to the statement relating to the intention to use. This statement has an average value of **3.83** - *I intend to use ChatGPT with the GenArch method for similar tasks in my work*. This value gravitates towards having the intention to use GenArch. However, this value does not convincingly give an indication that the subjects will immediately start using GenArch. To achieve this, additional training regarding the use of the GenArch method and more information about its corresponding benefits is expected to be needed. Accordingly, the adoption of GenArch is a subject for future research (see: Section 7).

5.4 Perceived Efficacy

The aforementioned sub variables all come together in the perceived efficacy (Section 2.2.2), which is taken as the main indicator for efficiency. The previous sections of show that the segment of the GenArch method has a moderately high to high level for all of the three sub variables. Accordingly, we derive that the GenArch method has at least a moderately high level of perceived efficacy. This is based on the comments in the surveys as well as the assigned values to the rated statements. In the survey of round three, subjects rate an additional statement regarding the efficiency of the GenArch method. If we apply the same range as in the previous rounds, the average value is **3.67**. Similarly to the sub variables, this value can be described as moderately high to high. Accordingly, we conclude that the GenArch method has at least a moderately high level of perceived efficacy. This is explained in greater detail in section 7.

5.5 GenArch versus Free Use of ChatGPT

The main goal of this analysis was to compare the current approach with the segment of the GenArch method (Section 2.2). As an additional effort, the segment of the GenArch method is also compared with the free use of ChatGPT. The aim of this comparison is to assess if practitioners could already use this method as-is. We reuse the range that was applied in the previous sections to be able to compare the quantified results. Figure 4 uses these assigned values to visualize the results. Figure 4 shows that currently, based on the statements in the survey, subjects view their own manner as slightly better in terms of usefulness, ease of use and intention to use. Certain subjects do indicate that they prefer the GenArch method (Sections 5.2 and 5.3), but on average this seems not to be the case. Two subjects did mention that they preferred their own way of working because of the extensiveness of GenArch and their lack of experience with it.

It is believed that these results are partly caused by the following three external elements. First, the workshop consisted of short tasks in which using a known approach is more efficient. Therefore, this could have boosted the results of round two as subjects could use their own approach of using ChatGPT. Second, the case was kept simple and common to make sure that subjects could quickly understand the task. This also implies that ChatGPT has a lot of relevant data, potentially causing that simple GenAI interaction was enough to provoke useful answers. Multiple subjects mentioned the extensiveness of the input prompt in the third round, implying that they used simpler prompts in round two. This again could have boosted the variables in round two. Third, subjects used GenArch for the first time while the majority indicated to have used ChatGPT before. It is plausible that this familiarity caused a small bias towards the free use of ChatGPT. Besides these partial possible causes, the results show that the GenArch method is valuable but could still improve to further increase the perceived efficacy of its users. Extended explanations including examples or small practice rounds and different scenarios depending on the case at hand are identified as areas in which the GenArch method could possibly improve.

6 DISCUSSION

The design and validation of GenArch in an iterative manner to ensure that the method is rigorously tested and refined is considered to be one of the strengths of this research. Moreover, the execution of the validation experiment in collaboration with KPMG allows for practical insights and applicability in the real world. This increases relevance and impact of the research findings. During both validation phases, the GenArch method was evaluated based on criteria as they were experienced by participants, with the aim to enhance acceptance and adoption of the method. It is noteworthy to mention that before conducting the actual interviews and experimental procedures we practiced these to be as prepared as possible for the actual interviews and the experiment. Beside strengths, this research also contains limitations. First, the sole focus is on ChatGPT which may have limited the generalizability. Other LLM tools could offer different performance characteristics that are not explored. For example, the indicators used to classify activities (see: Section 4.1) could differ for alternative LLM tools. The indicator relating to the fact that ChatGPT is unable to take recent events into account would differ as alternative LLM tools are able to do this Lau (2023). Furthermore, recommended use of such an LLM tool needs to be investigated and testing the difference in performance is essential. Performing a similar study with a different LLM tool could be a subject for future research (Section 7). Second, the expert interviews relied on subjective opinions of the participants. Their perspectives might not have encompassed the wide variety of opinions which might have affected the validation. To minimize the effects

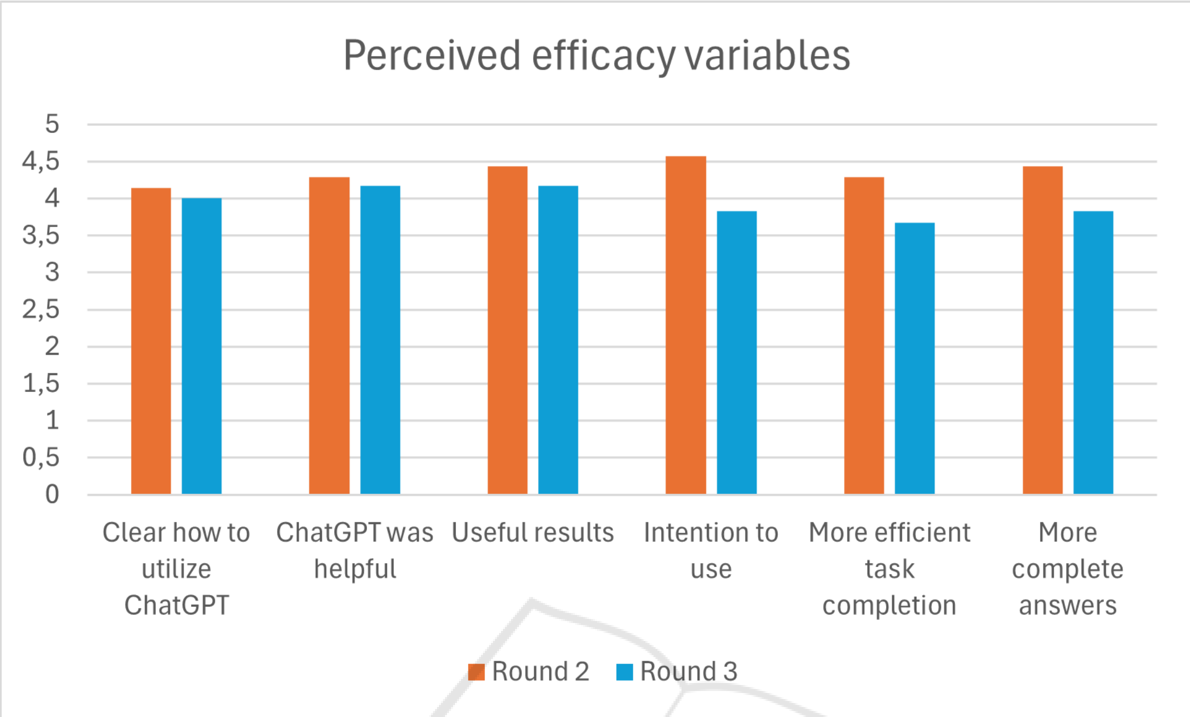

Figure 4: Comparison of experiment rounds two and three.

of this limitation, the interviews were conducted with different types of experts. Also, the execution of the validation experiment within KPMG might not have captured challenges in other organizations or industries. To mitigate this, a toy problem was used and research goals were not communicated during the experiment. Two iterations were performed, using six interviewees, seven experiment-subjects and three experiment rounds. Larger sample sizes would have potentially made the results more representative. For example, conducting additional experiments that focus on other activities and BA products could have produced additional results. The validation experiment, which corresponds to the second validation, produced positive results concerning perceived efficacy. Nevertheless, it also highlighted areas for potential improvements. These could have consisted of the inclusion of additional validation cycles or experiments, tests with alternative LLM tools to assess the generalizability of the results, the development of different scenarios tailored to differing cases, and an exploration of strategies surrounding the adoption of the GenArch method.

7 CONCLUSIONS AND FUTURE RESEARCH

This paper is centered around exploring the potential of utilizing an LLM tool to support the development of target BA products within the broader context of DT projects. The Main Research Question is as follows: 'Using the example of ChatGPT, how can a Large Language Model tool be effectively integrated into the design process of target Business Architecture products to mitigate risks, optimize prompt impact, and improve efficiency?' The resulting GenArch method aims at guiding stakeholders, prescribing the appropriate timing and utilization of ChatGPT throughout the target BA product design process. The GenArch method was developed through an iterative process comprising two design cycles including expert interviews and an experiment as validation (see: Section 2.2). In the experiment, a segment of the GenArch method was evaluated to assess the perceived ease of use, perceived usefulness, and intention to use of the GenArch method in its entirety. These factors were utilized to determine the perceived efficacy which was taken as the indicator for the efficiency of the GenArch method (see: Section 2.2.2). The results of the experiment indicate that the GenArch method scores moderately high to high on the perceived ease of use, perceived usefulness, and intention to use (see: Section 5). More specifically, perceived usefulness was rated the highest (see: Section 5.2), while the perceived ease of use and intention to use, though slightly lower, still produced positive results (see: Sections 5.1 and 5.3). From these findings, we derive that the GenArch method has at least a moderately high level of perceived efficacy (see: Section 5.4). This suggests that the GenArch method effectively guides the integration of ChatGPT into the target BA product design process, while enhancing the efficiency of this process. To answer the Main Research Question, we conclude that the GenArch method provides a structured approach for the timely and effective utilization of ChatGPT in the design process of target BA products. The GenArch method positively influences the efficiency of this design process. In summary, this research underscores the potential of LLM tools to enhance BA design. The findings suggest that the GenArch method can contribute to the success of DT projects by providing an approach that integrates ChatGPT.

Future research could include additional experiments focusing on other activities or BA products that are part of the GenArch method. This would provide a more comprehensive understanding of how GenArch can be utilized and optimized. Some of these additional experiments could also be executed with the use of a pre-trained language model to observe if results improve. Exploring strategies for adoption of the GenArch method within organizations could be another subject for future research. This could positively influence the practical applicability of the method and support a more effective integration of GenArch in different organizational contexts. Comparative studies could be conducted to determine if and how the GenArch method needs to be adapted when utilized with an alternative LLM tool. This could also help identify the most effective LLM tools for specific BA design tasks or contexts. Lastly, other potential application areas for LLM tools within DT could be explored.

ACKNOWLEDGEMENTS

We would like to sincerely thank the Digital Transformation team of KPMG Netherlands for their support during the conduction of this research, including the provisioning of access to company documentation, domain experts, and the support provided during the experiment.

REFERENCES

Abrahão, S., Insfran, E., Carsí, J. A., Genero, M., and Piattini, M. (2009). Evaluating the ability of novice analysts to understand requirements models. In *International conference on quality software*, pages 290–295. IEEE.

Aguinis, H., Beltran, J. R., and Cope, A. (2024). How to use generative AI as a HRM assistant. *Organizational Dynamics*, 53(1):101029.

Alliance, T. A. (2001). Principles behind the Agile Manifesto. https://agilemanifesto.org/principles.html. Accessed: 13-05-2024.

Bains, C. (2023). AI prompt engineering: learn how not to ask a chatbot a silly question. https://tinyurl.com/4jtax73a. Accessed: 14-02-2024.

Bera, P., Wautelet, Y., and Poels, G. (2023). On the use of ChatGPT to support agile software development. In *International Conference on Advanced Information Systems Engineering*, pages 1–9. CEUR.

Bouwman, H., Houtum van, H., Janssen, M., and Versteeg, G. (2011). Business architectures in the public sector: experiences from practice. *Communications of the Association for Information Systems*, 29:411–426.

Brinkkemper, S. (1996). Method engineering: engineering of information systems development methods and tools. *Information and Software Technology*, 38:275–280.

Chuma, E. L. and Oliveira, G. G. D. (2023). Generative AI for business decision-making: a case of ChatGPT. *Management Science and Business Decisions*, 3(1):5–11.

Ebert, C. and Louridas, P. (2023). Generative AI for software practitioners. *IEEE Software*, 40(4):30–38.

Eliot, L. (2023). Upping your prompt engineering superpowers via target-your-response techniques when using generative AI. https://tinyurl.com/466p4h37. Accessed: 14-02-2024.

Ferrini, M. (2023). Generative AI: just a zip file of the web? https://assets.kpmg.com/content/dam/kpmg/ch/pdf/generative-ai.pdf. Accessed: 26-02-2024.

Fitzgerald, M., Kruschwitz, N., Bonnet, D., and Welch, M. (2013). Embracing digital technology: A new strategic imperative. *MIT Sloan Management Review*, 55(2):1–12.

Greefhorst, D. and Proper, H. A. (2011). *Architecture Principles*, volume 4. Springer.

Group, T. O. (2022). Phase B: Business Architecture. https://pubs.opengroup.org/architecture/togaf92-doc/arch/index.html. Accessed: 01-02-2024.

Heale, R. and Twycross, A. (2018). What is a case study? *Evidence-Based Nursing*, 21(1):7–8.

Hohpe, G. (2017). The Architect Elevator — Visiting the upper floors. https://martinfowler.com/articles/architect-elevator.html. Accessed: 16-04-2024.

Korzynski, P., Mazurek, G., Krzypkowska, P., and Kurasinski, A. (2023). Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as Chat-

GPT. *Entrepreneurial Business and Economics Review*, 11(3):25–37.

KPMG (2023). Generative AI models — the risks and potential rewards in business. https://tinyurl.com/3vv4ap68. Accessed: 01-05-2024.

KPMG (2024a). Digital Architecture Talkbook.

KPMG (2024b). Visual identity overview. https://assets.kpmg.com/content/dam/kpmg/dk/pdf/dk-2020/12/alexander.pdf. Accessed: 08-05-2024.

Lau, J. (2023). Bing Chat vs. ChatGPT: Which AI chatbot should you use? https://zapier.com/blog/chatgpt-vs-bing-chat. Accessed: 15-01-2024.

LeanIX (2024). Business Architecture. https://www.leanix.net/en/wiki/ea/business-architecture. Accessed: 26-03-2024.

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., and Ge, B. (2023). Summary of ChatGPT-related research and perspective towards the future of large language models. *Meta-Radiology*, 1(2):100017.

Magaldi, D. and Berler, M. (2020). *Semi-structured interviews*, pages 293–298. Springer.

Moody, D. L. (2003). The Method Evaluation Model: A Theoretical Model for Validating Information Systems Design Methods. In *European Conference on Information Systems (ECIS)*.

OpenAI (2023). Introducing ChatGPT Enterprise. https://openai.com/index/introducing-chatgpt-enterprise. Accessed: 11-07-2024.

Piñeiro-Martín, A., García-Mateo, C., Docío-Fernández, L., and del Carmen López-Pérez, M. (2023). Ethical challenges in the development of virtual assistants powered by large language models. *Electronics*, 12(14):3170.

Qiu, J., Li, L., Sun, J., Peng, J., Shi, P., Zhang, R., Dong, Y., Lam, K., Lo, F. P. W., Xiao, B., Yuan, W., Wang, N., Xu, D., and Lo, B. (2023). Large AI models in health informatics: Applications, challenges, and the future. *IEEE Journal of Biomedical and Health Informatics*, 27(12):6074–6087.

Shanahan, M. (2024). Talking about large language models. *Communications of the ACM*, 67(2):68–79.

Simon, D. and Schmidt, C. (2015). *Introduction: Demystifying Business Architecture*, pages 1–17. Springer.

Singh, B., Garg, S. K., and Sharma, S. K. (2011). Value stream mapping: Literature review and implications for Indian industry. *International Journal of Advanced Manufacturing Technology*, 53(5):799–809.

Smith, V. (2024). Business architecture: How to build a successful model. https://tinyurl.com/3xmn4yha. Accessed: 26-03-2024.

Tacheva, J. and Ramasubramanian, S. (2023). AI Empire: Unraveling the interlocking systems of oppression in generative AI's global order. *Big Data and Society*, 10(2).

Wang, X., Attal, M. I., Rafiq, U., and Hubner-Benz, S. (2024). *Turning Large Language Models into AI Assistants for Startups Using Prompt Patterns*, pages 192–200. Springer.

Weerd van de, I. and Brinkkemper, S. (2008). *Metamodeling for situational analysis and design methods*, pages 35–54. IGI Global.

Westerman, G., Bonnet, D., and McAfee, A. (2014). *Leading digital : turning technology into business transformation*. Harvard Business Press.

Wieringa, R. J. (2014). *Design Science Methodology for Information Systems and Software Engineering*. Springer, 1st edition.

Willey, L., White, B. J., and Deale, C. S. (2023). Teaching AI in the college course: introducing the AI prompt development life cycle (PDLC). *Issues in Information Systems*, 24(2):123–138.

Wißotzki, M. and Sandkuhl, K. (2015). Elements and characteristics of enterprise architecture capabilities. In *International Conferences on Perspectives in Business Informatics Research*, volume 229 of *LNBIP*, pages 82–96. Springer Verlag.

Wolff de, T. (2024). The use of ChatGPT as a virtual team-member in a digital transformation consultancy team. Master's thesis, Utrecht University, Utrecht, the Netherlands.

Zant, T. V. D., Kouw, M., and Schomaker, L. (2013). *Generative AI*, volume 5, pages 107–120. Springer, 1st edition.

APPENDIX A

Click on this link to open Appendix A.