Large Language Models in Enterprise Modeling: Case Study and Experiences

Leon Görgen[1], Eric Müller[1], Marcus Triller[1][a], Benjamin Nast[1][b] and Kurt Sandkuhl[1,2][c]
[1]*Institute of Computer Science, Rostock University, Albert-Einstein-Str. 22, 18059 Rostock, Germany*
[2]*School of Engineering, Jönköping University, Gjuterigatan 5, 55111 Jönköping, Sweden*

Keywords: Enterprise Modeling, Large Language Model, ChatGPT, Artificial Intelligence, Proxy Domain Expert, Process Modeling.

Abstract: In many engineering disciplines, modeling is considered an essential part of the development process. Examples are model-based development in software engineering, enterprise engineering in industrial organization, or digital twin engineering in manufacturing. In these engineering disciplines, the application of modeling usually includes different phases such as target setting, requirements elicitation, architecture specification, system design, or test case development. The focus of the work presented in this paper is on the early phases of systems development, specifically on requirements engineering (RE). More specifically, we address the question of whether domain experts can be substituted by artificial intelligence (AI) usage. The aim of our work is to contribute to a more detailed understanding of the limits of large language models (LLMs). In this work, we widen the investigation to include not only processes but also required roles, legal frame conditions, and resources. Furthermore, we aim to develop not only a rough process overview but also a detailed process description. For this purpose, we use a process from hospitality management and compare the output of ChatGPT, one of the most popular LLMs currently, with the view of a domain expert.

1 INTRODUCTION

In many engineering disciplines, modeling is considered an essential part of the development process. Examples are model-based development in software engineering, enterprise engineering in industrial organization, or digital twin engineering in manufacturing. In these engineering disciplines, the application of modeling usually includes different phases such as target setting, requirements elicitation, architecture specification, system design, or test case development. The focus of the work presented in this paper is on the early phases of systems development, specifically on requirements engineering (RE). More specifically, we address the question of whether domain experts can be substituted by artificial intelligence (AI) usage.

The aim of our work is to contribute to a more detailed understanding of the limits of large language models (LLMs).

In previous work (Sandkuhl et al., 2023), we focused on retrieving domain knowledge from ChatGPT regarding processes common in an application domain or general tasks to be performed. The core result of this previous study was that the more specific the domain knowledge required, the less suitable LLMs seem to be.

In this work, we widen the investigation to include not only processes but also required roles, legal frame conditions, and resources. Furthermore, we aim to develop not only a rough process overview but also a detailed process description. For this purpose, we use a process from hospitality management and compare the output of ChatGPT, one of the most popular LLMs currently, with the view of a domain expert.

The paper is structured as follows: section 2 describes the background for our work from enterprise modeling (EM), LLMs, and the application potential of LLMs in EM. Section 3 introduces the research method applied in our work, followed by a systematic literature review (SLR) in section 4. Section 5 describes the experiment and discusses the results. Section 6 gives a conclusion and implications for future work.

[a] https://orcid.org/0000-0002-7709-0353
[b] https://orcid.org/0000-0003-4659-9840
[c] https://orcid.org/0000-0002-7431-8412

Görgen, L., Müller, E., Triller, M., Nast, B. and Sandkuhl, K.
Large Language Models in Enterprise Modeling: Case Study and Experiences.
DOI: 10.5220/0012387000003645

In *Proceedings of the 12th International Conference on Model-Based Software and Systems Engineering (MODELSWARD 2024)*, pages 74-85
ISBN: 978-989-758-682-8; ISSN: 2184-4348

2 BACKGROUND AND RELATED WORK

2.1 Large Language Models

LLMs belong to the broader category of deep learning models, address the area of natural language processing, and are designed to interpret and generate human-like text. Essential concepts of LLMs and their evolution have been widely documented, for example, in the publication by Brown et al. (Brown et al., 2020).

The most influential architecture in recent times for building LLMs is the Transformer. It uses attention mechanisms (Vaswani et al., 2017) to weigh the importance of different words or tokens in a sequence when producing an output. LLMs are trained on vast amounts of text data to be able to generate coherent and contextually relevant text across a wide range of topics. For instance, models like OpenAI's GPT (Generative Pre-trained Transformer) series have been trained on books, articles, and web pages. The pre-training of LLMs is supposed to be task-agnostic (Huang et al., 2022).

Capabilities of LLMs include tasks such as translation, question-answering, summarization, and text generation without needing task-specific training data. One of the most important features of models like GPT is their ability to generate coherent, diverse, and contextually relevant text over long passages. One of the currently most popular LLMs, OpenAI's GPT-4 with its Chatbot frontend - ChatGPT[1] can also be used for translation, grammar correction, or email composition (Floridi and Chiriatti, 2020). While LLMs are powerful, they can sometimes produce incorrect or nonsensical answers, which often are termed "hallucinations".

The use of LLMs starts from inputs (called prompts) stating the task to be completed by the LLM. LLMs are sensitive to the input phrasing. Thus, prompt engineering and prompting methods (Liu et al., 2023) have developed into a critical topic of study for LLMs as they investigate the techniques by which end-users can use LLMs to perform tasks.

2.2 Enterprise Modeling

EM is addressing the "systematic analysis and modeling of processes, organization structures, products structures, IT-systems or any other perspective relevant for the modeling purpose" (Vernadat, 2003). The role of EM is usually to provide methods, tools, and practices for capturing and visualizing the current ("as-is") situation and developing the future ("to-be") situation. In particular, a model of the current situation forms one of the fundamentals for supporting the future development of organizations. Without knowledge of the "as-is", a systematic design and development of future capabilities, products, or services are usually difficult.

The variety and dynamics of methods, languages, and tools supporting EM are visible in work on research roadmaps and future directions, originating both from the information systems community (see, e.g., (Sandkuhl et al., 2018)) and from scholars in industrial organizations (e.g., (Vernadat, 2020)).

Given the complexity of enterprises, in the course of modeling an enterprise, there is the need to understand, analyze, capture, and represent what is relevant for different stakeholders and/or modeling purposes. In this context, there seems to be an agreement in the academic literature related to enterprise modeling that a key feature of an enterprise model is that it includes various perspectives. (Frank, 2014), e.g., states that "a perspective as a psychological construct constitutes a conception of reality, comparable to a particular viewpoint in spatial perception [...], which helps to reduce complexity by constituting sense [...]." EM projects can have different purposes.

In some cases, making an EM activity is helpful when capturing, delimiting, and analyzing the initial problem situation and deciding on a course of action. In such cases, EM is mostly used as a problem-solving and communication tool. The enterprise model created during this type of modeling is used for documenting the discussion and the decisions made. The main characteristics of this purpose are that the company does not intend to use the models for further development work and that the modeling activity has been planned to be only a single iteration.

2.3 Application Potential of LLMs in EM for Requirements Engineering

The potential of LLMs as a proxy for domain experts has been investigated by (Sandkuhl et al., 2023) starting from the expected contribution of domain experts to EM. We consider this perspective as also suitable for our investigation of EM use in RE. (Stirna and Persson, 2018) describe the role of domain experts in EM in general as "supplying domain knowledge, knowledge about organization units involved [...]; examining and evaluating the results of enterprise modeling, and integration of modeling results of different teams into a consistent whole." In RE, these contributions are required for different aspects of the system to be developed, for example, the required

[1] https://chat.openai.com

Table 1: Potential application areas of LLMs in EM for Requirements Engineering.

	Task of Domain Experts in RE		
EM support for RE	*Supply of Domain Knowledge*	*Integrate Modeling Results*	*Evaluate Results*
Model of current situation	perspectives relevant for the scope (e.g., goals, organisation, structure, process, products, IT, resources)	models developed for the perspectives	individual models and inter-model integration
Required changes and alternatives	potential changes; how realistic and accepted are they?		
Model of future situation	all perspectives relevant for the change	models developed for specifying the change	individual models and inter-model integration

functionality or organizational integration. For organizational integration, the different perspectives of EM (see section 2.2) are useful. An analysis by (Vernadat, 2020) showed that frequently used perspectives are goals, organization structure, process, products, and IT and resources.

In addition to different contributions expected from domain experts and various perspectives, the different modeling tasks in the course of RE require different ways of participation from the domain experts. (Krogstie, 2016) concludes that the most relevant modeling phases to be distinguished in this context are scoping of the project, modeling of the current situation, analysis of required changes and potential alternatives, and modeling of the future situation. As scoping usually has to be finished before starting RE, we exclude this phase from our investigation.

The results of modeling the current situation have to be examined by the domain expert for accuracy and completeness. In the process of analysis and finding alternatives, creativity in designing feasible and acceptable changes is most important. In modeling the future situation, the domain experts have to make sure that the different perspectives add to a consistent whole.

The above considerations result in a variety of tasks that could potentially be supported by LLM. Table 1 summarizes these tasks by showing the phases of EM support for RE as rows and the different contributions of domain experts as columns.

3 RESEARCH METHOD

The starting point of our work is the question presented in the introduction and the decision to focus on supporting domain experts in the task of modeling the current situation. Based on this, two research questions (RQ) were defined for the paper:

- RQ 1: How consistent and complete is the output of ChatGPT compared to the knowledge of a domain expert in the context of an EM project?
- RQ 2: How can prompt pattern improve the output of ChatGPT?

The overall research strategy for work presented in this paper is of an explorative nature, i.e., we aim to gather new knowledge by exploring the potential of ChatGPT use in EM. More concretely, the work combines literature studies with quasi-experiments and argumentative-deductive work.

The literature review objective was to identify relevant studies and findings from other researchers to consider when exploring the potential of LLMs for use in EM. Kitchenhams's SLR approach (Kitchenham, 2004) was utilized for this purpose. Six steps are suggested, which we briefly introduce below and document in detail in section 4.

The first step is to develop the research questions (RQ) to be answered by the SLR. The process of paper identification begins with the definition of the overall search space (step 2), which basically consists of determining the literature sources to be considered in light of the research questions. Paper identification continues with the population phase (step 3). In this step, the search string is developed and applied by searching the literature sources. This is followed by the paper selection step, in which inclusion and exclusion criteria are defined, and relevant papers found in the population phase (step 4) are manually selected. The data collection phase (step 5) focuses on extracting the information relevant to answering the research question from the set of identified relevant papers. The final step is data analysis and interpretation, i.e., answering the research question defined in step 1 using the collected data from relevant papers.

We structured the field of EM along with the tasks to be performed during a project (see section 2.2). This is the argumentative-deductive part of our work.

In our work, we conduct a quasi-experiment using ChatGPT and domain experts as the study objects.

The treatment is the task of eliciting required roles, legal frame conditions, and resources for a process from hospitality management. A quasi-experiment is "an experiment in which units are not assigned to conditions randomly" (Cook et al., 2002). The purpose of the experiment is to conduct exploratory research to answer the defined research questions rather than to test a specific hypothesis. The experiment design is described in detail in section 5.1.

4 SYSTEMATIC LITERATURE REVIEW

Related work was identified through an SLR following the six-step method proposed by Kitchenham (see section 3). The research questions (step 1) were already introduced in section 3. Scopus, IEEE Xplore, and AISeL databases constituted the search space (step 2). The search string used in these databases combines the term "Enterprise Modeling" with "large language model" and its synonyms, such as "Process Modeling", "LLM", "neural text", and "ChatGPT". The final search string used was *("Enterprise Modeling" OR "Process Modeling") AND ("Large Language Model" OR "LLM" OR "Neural Text" OR "ChatGPT")*. The search in title, abstract, and keywords yielded 18 papers. The inclusion criterion (step 4) required that the papers discuss LLM use in the context of EM.

Of the 6 results found in Scopus, 2 were excluded due to their status as conference proceedings that contained papers on either EM or LLM, but no papers covered both topics in the same work. 2 papers used synonyms in the title or abstract but did not address the topic of the use of LLM in EM, and 1 paper appeared but meant something else by the abbreviation "LLM". 1 paper was relevant to our work: (Simon et al., 2023) describe systematic experiments on using ChatGPT-3 to interpret a textual process description and to convert it into a formal representation. This work is not intended to substitute an expert or assist in creating a new process description or model. However, an LLM is also used to support certain modeling phases.

No results were found in IEEE Xplore.

In AISeL, the query interface only allowed for search within all metadata. 11 of the 12 results did not mention EM and LLM or synonyms for them together in title, abstract, or keywords. One paper was found that is relevant to our work: In our previous work (Sandkuhl et al., 2023), we investigated the potential of LLMs as a proxy for domain experts, starting from the role of domain experts in EM and their expected contribution. The focus was on the preparation of RE in EM and the identification of alternatives for change. While the scope was limited to supporting the role of the domain expert, the research in this paper is expanded to include legal frame conditions and resources of the process. The previous results show that ChatGPT can work with domain experts to improve productivity, completeness, and accuracy. ChatGPT can assist during the preparation phase by gathering comprehensive information on the application domain as well as on general business processes and their flow of information. However, the results should not be considered complete, and an expert is always needed for the specifics of a company.

Table 2 summarizes the number of papers found in the different databases and the relevant ones. In conclusion, the SLR returned 1 paper (Sandkuhl et al., 2023) addressing LLM use in EM, focusing on the same phases of EM projects as our work. However, we expanded the scope to include more aspects (legal frame conditions and resources). (Simon et al., 2023) are focusing on creating models based on existing knowledge. For this reason, we decided not to consider the identified work to improve the output of ChatGPT with prompt pattern (RQ2). However, parts of our previous work (Sandkuhl et al., 2023) aim to answer similar questions as we want to answer in this work (RQ1). It was also identified that future work should consider investigating tasks that cover a range of specificity, from general to specific. This work includes an experiment with such tasks, so a comparison of the results could be interesting.

Table 2: Results of the SLR.

Database	Results	Relevant Papers
Scopus	6	(Simon et al., 2023)
IEEE Xplore	none	none
AISeL	12	(Sandkuhl et al., 2023)

5 EXPERIMENT

5.1 Experiment Design

This section presents the experiment design of this work. First, an overview of the modeling task to be solved by ChatGPT is given. In this context, how the results of ChatGPT are compared to those of the domain expert is also outlined. Then, the sequence of prompts identified in this work to obtain a business process model of ChatGPT is shown.

5.1.1 Modeling Task

The task described below is intended to explore the potential of using ChatGPT in the context of EM. Specifically, the task involves the planning of a hypothetical corporate event. Targeted and precise questions will be posed to ChatGPT in order to gain an in-depth understanding of the diversity and complexity of existing processes, roles, resources, and legal frame conditions within an enterprise. The hypothetical event used as a testing ground is a corporate meeting. The meeting is scheduled to start at noon and end with a celebration in the evening. This experiment focuses on supporting the domain experts' role in modeling the current situation (see section 2.3).

It starts with the survey of ChatGPT. The answers obtained serve as the basis for developing an initial business process model. In a subsequent step, a domain expert is consulted and confronted with the same questions. Together with the expert, a new business process model is then developed for the same event.

To conclude the investigation and to fully round out the experiment, the model generated by ChatGPT is subjected to a thorough analysis. Under the guidance and with the technical support of the domain expert, the model will be examined for possible errors or weaknesses. Here, not only is a comprehensive error analysis to be performed but also the uncovering and highlighting of potential improvement opportunities is of central importance. This phase of the experiment thus forms a comprehensive evaluation of the suitability of ChatGPT as a proxy for a domain expert in EM.

5.1.2 Comparison of ChatGPT and Domain Expert

The comparison of the business process models is performed in a systematic and controlled manner. This ensures that a direct and meaningful analysis is possible between the model created using ChatGPT and the model developed in collaboration with the domain expert.

In the first step, the domain expert is interviewed. The same questions that were previously asked in ChatGPT are used here. The domain expert's answers serve as the basis for developing a business process model that addresses the same context as the model created with ChatGPT. After both models have been created, the domain expert evaluates them. Here, the two models (ChatGPT vs. domain expert) are juxtaposed and evaluated with the help of various metrics. This systematic approach enables an objective and comprehensive evaluation and provides valuable insights into the strengths and weaknesses of model development using ChatGPT compared to traditional model development of a domain expert.

The following section presents the methodology for evaluating and comparing the business process model developed using ChatGPT with a business process model created by a domain expert. Here, four key metrics are applied to evaluate different aspects of the models:

- **Accuracy** aims to determine the correctness of the information provided by ChatGPT and the domain expert with respect to several issues (e.g., correct, out of scope, or hallucination). It is an important indicator of the reliability of the generated business process models.
- **Completeness** is intended to determine the extent to which the responses provided comprehensively cover all necessary information. If any missing information is identified, it is checked to see if it is required or optional. A high level of completeness ensures that all relevant aspects of the business process are included in the model.
- **Comprehensibility** assesses the ease with which the responses can be interpreted and understood by those conducting the experiment. This is crucial to ensure that the models are clear and comprehensible to all participants.
- **Time** captures the duration that both the domain expert and ChatGPT need to provide the required information. This is an important indicator of the efficiency of model building and may have an impact on the practicality of the approach.

5.1.3 Prompt Engineering

Achieving meaningful and optimal results in working with ChatGPT requires careful design and formulation of appropriate input requirements (prompts). An input prompt represents a sequence of instructions or directives that are used to guide and control the LLM. By specifically formulating these input prompts, the model can be programmed to generate certain responses or to improve and refine its response capabilities in specific ways (Liu et al., 2023).

The focus of the following section is on the area of "prompt engineering", a process by which LLMs can be programmed and controlled by providing carefully designed prompts (White et al., 2023). This is followed by a presentation of input prompts, which includes the context of the situation as well as the issues involved in developing a business process model. The prompts were entered in German.

The prompt engineering in this work can be divided into three phases:

1. **Input Refinement:** Using ChatGPT-4, we enriched our formulated questions to improve the output.
2. **Output of the Process Description:** We used the enriched questions in ChatGPT-4 to get the information about the process.
3. **Textual Description of the Model:** Finally, we asked ChatGPT-4 to provide a textual description of the model.

The foundation of the discussed prompt stems from the work outlined in (White et al., 2023). To enhance the understanding of the subject matter and emphasize the significance of the research findings, this summary highlights the patterns utilized here to enhance communication with an LLM:

- **Question Refinement:** ChatGPT is actively involved in the prompt engineering process. The pattern enables ChatGPT to optimize the questions asked by the user in order to obtain additional information or to fill any gaps in understanding.
- **Cognitive Verifier:** Helps ChatGPT to better understand the intention of the question. The goal is to encourage ChatGPT to decompose the current question into additional questions to provide a more precise answer.
- **Persona:** Allows the LLM to be given a specific point of view or perspective (in this case, an expert in EM).
- **Reflection:** This pattern is used to provide an automatic reasoning for the given answers. This allows for a better assessment of the validity of the output and provides insight into how ChatGPT arrived at a particular answer.

Applying the patterns has equipped ChatGPT with the requisite communication methods necessary for more nuanced and precise interactions with users. This includes the ability to refine inquiries, take varying perspectives, reflect on response processes, and manage contextual reference points within ongoing conversations.

After integrating these patterns, it becomes critical to specify the most relevant context possible. This context allows the system to develop a clear understanding of the current situation and lays the foundation for the subsequent conversation. In this context, providing a precise and comprehensive description of the situation plays a vital role in enabling ChatGPT to concentrate on relevant aspects and generate a suitable response.

During the *1. Input Refinement* phase, our aim is to gain precise comprehension of ChatGPTs evaluations and perspectives to develop a sophisticated business process model. It is critical to use carefully worded questions to obtain an accurate assessment of ChatGPT's capabilities and responses. With the intent of conducting such an in-depth assessment, a series of specific questions were formulated and directed to ChatGPT. These questions were designed to explore various aspects of ChatGPT's performance and behavior, highlighting the nuances and complexities of the interaction. The questions were extensively edited and expanded using two of the above patterns (marked with square brackets in the prompt) within an earlier dialog with ChatGPT. This methodological adaptation served to ensure a better formulation and selection of questions, thus generating better results in the subsequent research:

"[Question Refinement:] If I ask a question and you find a better wording that could avoid possible misunderstandings, suggest this improved version of the question. Also, think of additional questions that could help me design a more accurate business process model. [Cognitive Verifier:] Think of an additional one to three questions that will help you provide a more accurate answer. After answering the additional questions, combine the answers to provide a final answer to my original main question. [Questions:] (...)"

We then created the main prompt for the *1. Input Refinement* phase, which only required one interaction in the dialog with ChatGPT:

"[Persona:] You are now acting as a domain expert in the field of enterprise modeling. I would like to design a business process model for an event. Use your expert knowledge as an enterprise modeler to improve the questions I ask you. [Question Refinement:] If I ask a question and you find a better wording that could avoid possible misunderstandings, suggest this improved version of the question. Also, think of additional questions that could help me design a more accurate business process model. [Context:] Here is the context for the event: A company wants to plan a conference in an exhibition hall that starts at 12 noon. The conference should last until 6 pm and be rounded off with an appropriate evening event after 6 pm until 11 pm at the latest. Answer the following questions from the perspective of the most responsible person. Around 200 guests are expected. During the conference and the celebration, a catering company will be hired to take care of all the food and drinks. An average budget is available for the event. [Questions:]

1. *What are the specific roles and responsibilities in the planning and execution of such an event? Who is typically responsible for what?*

2. *What resources are typically required to successfully execute such an event?*
3. *What are the processes involved in such a meeting?*
4. *What is a typical schedule for the day of the event? What happens at what time, and who is responsible?*
5. *How will the number and type of personnel required be determined? How are the work schedules and tasks of the staff organized?*
6. *How will the process of procuring ingredients and materials for the event be controlled?*
7. *What special regulations and safety requirements must be observed in the planning and execution of such an event?*
8. *What follow-up activities are necessary after the event (e.g., cleaning, evaluation, accounting with suppliers and partners)?"*

The results of this query can be seen in the prompt for the creation of the business process model. Some questions were rephrased (e.g., questions 1 and 8) or made more specific (e.g., questions 3 and 4) in this process. In addition, two new questions were added (questions 2 and 10). We used an additional dialog with ChatGPT for the phase 2. *Output of the Process Description*, which required four interactions:

User (1): *"[Persona:] You are now acting as a domain expert in the field of enterprise modeling. Use your expert knowledge as an enterprise modeler to model a company based on the following situation and answer my questions to the best of your ability. [Reflection:] When answering a main question, please explain the reasoning and assumptions behind your answer. You should also address any ambiguities or limitations in your answer. Answer the questions in as much detail as possible. Do not answer in general terms, but be specific to the given context for the planned event. In the next message, I will give you the context of the event that you are supposed to model. Did you understand that?"*

ChatGPT: *"Yes, I have understood that. I look forward to modeling the proposed company based on the situation provided. Please provide the context for the company you would like modeled."*

User (2): *"[Context:] A company would like to plan a conference in an exhibition hall that starts at 12 noon. The conference is to last until 6 p.m. and be rounded off with an appropriate evening event after 6 p.m. until 11 p.m. at the latest. Answer the following questions from the perspective of the most responsible person. 200 participants are expected. During the conference and the celebration, a catering company will be hired to take care of all the food and drinks. An average budget is available for the event. Your task is to find out which players, resources, and processes are needed for such an event. Your task is to answer my questions about these players, resources, and processes. In the next message, I will ask you questions about the event, and we will try to create a business process model together. Did you understand that?"*

ChatGPT: *"Yes, I have understood that. It sounds like an interesting task, and I am ready to answer your questions and create a business process model. Please ask your questions, and we can start the modeling process."*

User (3): *Here are my questions:*

1. *What specific responsibilities must be taken into account when planning and implementing such an event, and what tasks usually fall within their areas of responsibility?*
2. *Which actors are still present in addition to the main responsible roles? Which processes fall within their areas of responsibility?*
3. *What resources are needed to successfully plan and execute an event of this size and nature?*
4. *What specific processes and procedures should be considered during the planning, execution, and follow-up of the event?*
5. *What might a detailed schedule for the day of the event look like, including responsibilities and timing for the various activities?*
6. *On the basis of which criteria is the required number and type of personnel determined, and how are the work plans and tasks of the personnel organized and coordinated?*
7. *How is the process of procuring resources for the event efficiently managed and coordinated?*
8. *What special legal regulations and safety requirements must be observed and complied with when planning and holding such an event?*
9. *Which post-event activities are necessary, and how can they be carried out effectively and efficiently (e.g., cleaning, evaluation, billing suppliers and partners)?*
10. *How is communication and coordination between the various stakeholders (e.g., staff, catering companies, guests) ensured?"*

Using the Persona and Reflection Pattern in conjunction with the extended questions provides a systematic approach to building and reflecting on the model to effectively align it with the requirements of the business process.

5.2 Results

This section presents the results of the experiment. First, the output of ChatGPT is given in section 5.2.1, followed by the comparison of the business process models in section 5.2.2. Section 5.2.3 is a discussion on the use of ChatGPT in the area of EM.

5.2.1 ChatGPT Output

This section describes the output of ChatGPT using the developed prompt described in section 5.1.3 (*2. Output of the Process Description*). The output is displayed and divided according to the questions regarding the process, roles, resources, and legal frame conditions. We then asked for the textual description of the process model.

1. Specific Responsibilities and Tasks

- Event Manager: Is responsible for the overall organization and planning of the event. This includes setting the budget, booking the venue, hiring service providers (such as catering, security, and cleaning), creating the schedule, and coordinating communication between all parties involved.
- Technical Director: Is responsible for the technical aspects of the event, such as the sound system, lighting, video technology, and stage construction.
- Security Director: Responsible for creating and implementing safety plans, including evacuation plans, fire code compliance, and general participant safety.
- Marketing: Responsible for promoting the event, creating marketing materials, maintaining social media and website, and communicating with participants.

2. Other Players

- Catering Company: Responsible for providing food and drinks during the meeting and evening event.
- Cleaning Company: Responsible for cleaning before, during, and after the event.
- Suppliers: Supply necessary equipment, such as furniture, technology, decoration, etc.
- Participants: Are the target audience of the event.

3. Required Resources

- Venue: An exhibition hall large enough to accommodate 200 participants.
- Technical Equipment: Audio-visual equipment, microphones, lighting, stage, etc.
- Staff: Technicians, security staff, cleaning staff, event managers, PR/marketing staff.
- Food and Drinks: Provided by the catering company.
- Decoration: On the design of the venue.

4. Specific Processes and Procedures

- Planning Phase: Selecting and booking the venue, setting the budget, hiring staff and service providers, and creating the schedule.
- Implementation Phase: Setting up the equipment and decorations, running the event, monitoring the safety and well-being of participants, providing food and drinks, and ongoing cleaning.
- Follow-up Phase: Dismantling and cleaning, evaluation of the event, accounting with service providers and partners.

5. Proposal for a Detailed Schedule

- 07:00 - 10:00: Setup Technology and Decoration
- 10:00 - 12:00: Last Checks and Safety Briefings
- 12:00 - 18:00: Implementation of the Event, Catering Provides Food and Drinks
- 18:00 - 23:00: Implementation of the Evening Event, Catering Continues to Provide Food and Drinks
- 23:00 - 01:00: Dismantling and Cleaning

6. Number and Type of Staff

- Staffing levels and types are determined based on the scope of the event, the number of participants, and the specific requirements of the event. Staff work schedules and tasks are organized and coordinated by the event manager.

7. Resource Procurement

- The procurement process is managed by the event manager and/or a dedicated procurement manager. This includes selecting and negotiating with suppliers, tracking deliveries, and managing stock and inventory.

8. Legal Frame Conditions

- When planning and staging such an event, various legal regulations must be observed, including fire protection regulations, noise protection regulations, hygiene regulations (especially with regard to catering services), data protection regulations, and possibly regulations on serving and consuming alcohol.

9. Post-Event Activities

- Post-event activities include dismantling and cleaning the venue, evaluating the event (e.g., through surveys of participants), settling accounts with suppliers and partners, and documenting and analyzing the event for future events.

10. Communication

- Communication and coordination between the various stakeholders are mainly ensured by the event manager, often supported by communication tools and platforms. This may include regular meetings and updates, clear communication guidelines and channels, and dedicated communication or project management software where appropriate.

Afterward, we asked for a textual description of a business process model (*3. Textual Description of the Model*) using the business process model and notation (BPMN):

User (4): "*Create a textual description for a business process model for the entire event (before the event, during the event, after the event) using BPMN. Include all processes and actors in the event. There are several actors in the notation. Each process in the event must be associated with an actor. Represent each process of the event in this model. If some processes also require resources, point them out in the respective processes. The processes must also be linked to other processes. Try to break down the process as detailed as possible. In addition, show me the connection to the further processes for each process mentioned. Assign IDs to the processes so that your assignment is unique. Each process must be linked to at least one other process.*"

The answers were divided into before, during, and after the event, and the processes were assigned to the different roles. Based on this, we were able to develop the business process model (see Figure 1). The following is an example output for the role of Event Manager before and after the event:

Before Event:

- Process 1.1: Event starts
- Process 1.2: Selection and booking of the venue → Connected with Process 1.3
- Process 1.3: Budget setting → Connected with Process 1.4
- Process 1.4: Hiring staff and service providers → Connected with Process 1.5, Process 2.1, Process 3.1
- Process 1.5: Creation of a schedule → Connected with Process 1.6, Process 3.2
- Process 1.6: Coordination of communication with service providers → Connected with Process 1.7
- Process 1.7: Organisation of procurement (Resource: Inventory) → Connected with Process 2.2
- Process 1.8: Event planning completed → Connected with Process 4.1, Process 5.1

After Event:

- Process 7.1: Event ends → Connected with Process 4.4, Process 5.3, Process 6.3, Process 7.2
- Process 7.2: Selection and booking of the venue → Connected with Process 1.3
- Process 7.3: Event evaluation → Connected with Process 7.1, Process 7.3
- Process 7.4: Documentation and analysis of the event → Connected with Process 7.3, Process 6.4
- Process 7.5: Event follow-up completed → Connected with Process 7.4

5.2.2 Comparison of the Models

As described above, the assessment is based on four essential parameters: Accuracy, Completeness, Comprehensibility, and Time.

Regarding the Accuracy of ChatGPT, significant errors in the created connections are revealed. Processes that, upon closer inspection, were considered to have been initiated for no reason came to light (e.g., Process 6.1). In addition, it was recognized that ChatGPT neglected the temporal sequencing of processes, which is a fundamental violation of BPMN. Despite the identified flaws, the results of ChatGPT and the model created were overall comprehensible.

In terms of Completeness, ChatGPT showed a stronger focus on the execution of the event, while the domain expert placed more weight on the planning phase, which usually makes up the bulk of such an event. Further, the domain expert placed considerable focus on communication between roles, particularly between customers and responsible parties, while ChatGPT omitted customers or participants from the model entirely. Additionally, the domain expert incorporated an additional role for support staff that was left out of ChatGPT's modeling. In general, the model created by ChatGPT was less detailed, which was reflected in the integration of the few resources associated with the processes. In addition, ChatGPT was strongly oriented toward the supporting questions posed in the prompt and integrated fewer branches for different alternatives and end states.

Both subjects showed high Comprehensibility, although ChatGPT occasionally caused confusion due to illogical links (e.g., Process 6.3 → Process 7.1),

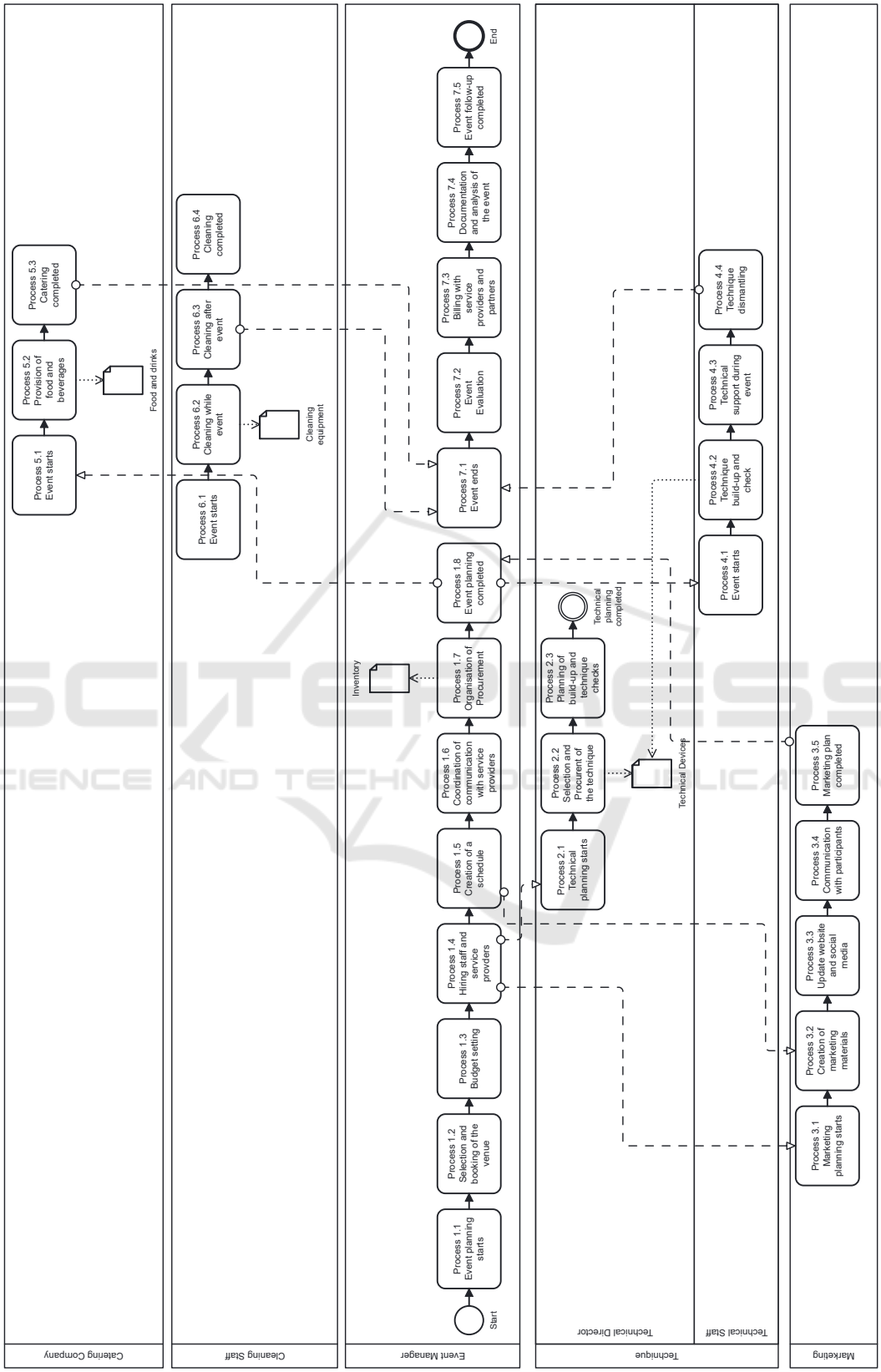

Figure 1: Business process model created with ChatGPT.

elements, and contradictions. Nevertheless, modeling with ChatGPT was considered somewhat easier because all processes, roles, and resources were accurately described and linked. In contrast, the domain expert left more room for interpretation. Due to the domain expert's extensive knowledge, he became aware of additional aspects of a previous topic at a later stage, which made it slightly more difficult to understand.

With regard to Time, the study showed that the actual modeling with the domain expert could be completed much more quickly while waiting for an appointment took more time. In contrast, ChatGPT is available immediately. In addition, a significant time advantage could be achieved through the use of prebuilt templates and targeted prompt engineering.

5.2.3 On the Use of ChatGPT in Enterprise Modeling

The results of the comparison (see section 5.2.2) reflect many different facets. It was recognized that the actual added value of ChatGPT depends on the complexity of the modeling task and the prompt. Although the generated model of ChatGPT is not as expressive, it still proves to be useful for simpler comprehension questions for those without expertise.

The use of a variety of patterns and prompts has shown that this can increase the effectiveness of the results. At the same time, however, difficulties in understanding the specific requirements became apparent. In addition, it was found that ChatGPT is able to generate process models even though they are not yet available in graphical form and are not completely semantically correct.

The experiment showed that the limitations of ChatGPT are mainly in the following areas:

- There are limitations in verifying the accuracy of information: ChatGPT is based on trained data and can potentially provide false or misleading information. It is not able to verify facts or check the accuracy of information like a knowledgeable human. However, using the Input Refinement pattern is a starting point to improve the output.
- There is limited understanding of context: ChatGPT may have difficulty grasping the full context of a question or conversation. This can lead to inconsistent or inaccurate responses, especially if the context is complex or ambiguous.
- Sensitivity to input variations: The smallest changes in the wording of a question can lead to different answers. ChatGPT is sensitive to nuances and word choice, potentially yielding inconsistent results.
- It is not possible to express uncertainty: ChatGPT tends to present answers with some conviction, even if it is uncertain. It cannot express uncertainty or lack of knowledge, which can lead to misleading or inaccurate information.

It is obvious that the results of this study indicate that ChatGPT can already be used in EM as a supporting tool. In summary, and with respect to RQ1, it proves our assumption (see section 2.3 and (Sandkuhl et al., 2023)) that it helps more with general questions than with specific ones. ChatGPT, for example, can provide novices with a rudimentary understanding when familiarizing themselves with a subject area. However, it is important to emphasize that ChatGPT can by no means completely replace a domain expert, as the results of this study reveal.

The part *1. Input Refinement* led to an overall improvement of ChatGPT's outputs and thus provided good insight into how to use prompt patterns effectively (RQ 2). Nevertheless, more and more errors or deficiencies occurred in detailed questions in the *2. Output of the Process Description* phase. In addition, problems occurred in *3. Textual Description of the Model*, especially when it came to relations or resources. Thus, it can be concluded that general aspects are better supported in the modeling of the current situation than specific ones regarding the EM support for RE.

To use ChatGPT more effectively, targeted work on prompts and providing even better and more significant context is needed. This can be done by developing a specific guide in collaboration with domain experts, which can then be used repeatedly to integrate understanding and knowledge into ChatGPT. Only then should consideration be given to which interactions could be automated.

6 CONCLUSIONS AND FUTURE WORK

This work has shown that ChatGPT can be a useful tool in enterprise modeling. Especially for beginners, ChatGPT offers the possibility to develop a basic understanding of different topics and help aspiring domain experts identify missing aspects. However, it is important to note that ChatGPT cannot be seen as a proxy for a domain expert in our specific case.

ChatGPT can assist in compensating for human errors through a synthetic approach, working alongside domain experts to complement their explanations. The benefit is also a more efficient utilization of the domain expert's time. However, it is important to

critically scrutinize all statements provided by ChatGPT to ensure accuracy and objectivity. Although ChatGPT provides valuable support, the expertise and knowledge of human domain experts are still crucial.

While our research has led to a number of results, it also has many limitations that identify some aspects for future work. The development of the prompts is mainly based on the application of existing patterns rather than on systematic development. It is possible that the prompts could be improved to provide a more relevant and complete output. As ChatGPT was intentionally used without prior knowledge of the domain in this work, it would be interesting to investigate to what extent the expert's knowledge (e.g., the model) can be emulated by subsequent prompt engineering.

Since our results are based on only one experiment, further research is needed to make them generalizable. Future work should consider additional patterns and focus more on evaluating the resulting changes in responses. The ChatGPT response-based process model is founded in textual descriptions. It is recommended to try to generate the model in an appropriate visual modeling language. In addition to evaluating how responses vary based on different prompts, future studies should also aim to investigate their potential usefulness for other LLMs.

It is crucial to verify if other domain experts provide identical evaluations on this topic. Additionally, there is a need to explore whether the quality of ChatGPT's output changes when considering other phases of an EM project or targeting other application areas or model types.

Improving the accuracy and quality of ChatGPT results is of great importance. This can be achieved by developing methods for verifying correctness and a better understanding of the context. In addition, collaboration with domain experts and optimization of the interaction between humans and AI models offer promising approaches for further improving ChatGPT and enhancing its performance.

Overall, the use of ChatGPT in enterprise modeling opens promising opportunities but also presents challenges and limitations. With further research and consideration of the identified limitations, ChatGPT can be better integrated into the enterprise context in the future to provide valuable support. This paper's contribution highlights the significance of further research in this area.

REFERENCES

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., and Askell, A. et al. (2020). Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Cook, T. D., Campbell, D. T., and Shadish, W. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Houghton Mifflin Boston, MA.

Floridi, L. and Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30:681–694.

Frank, U. (2014). Multi-perspective enterprise modeling: Foundational concepts, prospects and future research challenges. *Software & Systems Modeling*, 13(3):941–962.

Huang, W., Abbeel, P., Pathak, D., and Mordatch, I. (2022). Language models as zero-shot planners: Extracting actionable knowledge for embodied agents. In *International Conference on Machine Learning*, pages 9118–9147. PMLR.

Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004):1–26.

Krogstie, J. (2016). *Quality of Business Process Models*. Springer.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., and Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9):1–35.

Sandkuhl, K., Barn, B., and Barat, S. (2023). Neural text generators in enterprise modeling: Can ChatGPT be used as proxy domain expert? In *31st International Conference on Information Systems Development*.

Sandkuhl, K., Fill, H.-G., Hoppenbrouwers, S., Krogstie, J., Matthes, F., Opdahl, A., Schwabe, G., Uludag, Ö., and Winter, R. (2018). From expert discipline to common practice: A vision and research agenda for extending the reach of enterprise modeling. *Business & Information Systems Engineering*, 60:69–80.

Simon, C., Haag, S., and Zakfeld, L. (2023). Experiments on GPT-3 assisted process model development. In *37th ECMS International Conference on Modelling and Simulation (ECMS 2023)*, page 270 – 276.

Stirna, J. and Persson, A. (2018). Enterprise modeling. *Cham: Springer*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Vernadat, F. (2003). Enterprise modelling and integration: From fact modelling to enterprise interoperability. *Enterprise inter- and intra-organizational integration: Building International Consensus*, pages 25–33.

Vernadat, F. (2020). Enterprise modelling: Research review and outlook. *Computers in Industry*, 122:103265.

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., and Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT. *arXiv preprint arXiv:2302.11382*.