Experiments on GPT-3 Assisted Process Model Development

Carlo Simon and Stefan Haag and Lara Zakfeld
Department of Computer Science
Worms University of Applied Sciences
Erenburgerstr. 19, 67549 Worms
Email: {simon,haag,zakfeld}@hs-worms.de

KEYWORDS

GPT-3, Process Modeling, Artifical Intelligence

ABSTRACT

Computer assisted process model development from textual descriptions is still an open research question. Advantages of such a technology lie in shorter development times and possibly a more concise interpretation of the narrative input. A solution to this problem necessarily relies on methods from formal modeling and linguistics. In the latter field, the new GPT-3 model is recognized as a breakthrough that outperforms previous technologies whose limitations hindered success of earlier research in this context. But are GPT-3's capabilities to summarize text, detect cause-and-effect, or to classify terms sufficient to succeed? The presented research describes the results of systematic experiments to use GPT-3 to interpret a textual process description and transform it into a formal representation. The different settings demonstrate how to exploit the capabilities of large language models and how to avoid pitfalls. Although the observations made are promising, further work is needed. The outcome of this paper identifies the direction in which this future research should proceed.

THE NEXT DISRUPTION?

ChatGPT, developed by OpenAI, is a tangible example of an emerging technology that has brought machine learning and natural language processing to the attention of a broad audience. Many authors see the potential for this technology to disrupt existing ones. One example of the many publications on this new technology is `https://www.theguardian.com/commentisfree/2023/jan/07/chatgpt-bot-excel-ai-chatbot-tech`.

ChatGPT uses a chat-like interface to communicate with its users. Its breakthrough strength lies in the capability to answer questions, having been trained on an enormous amount of input texts. More than 235.000.000 text documents in English and more than 10.000.000 in other languages have been taken from the internet (Kublik and Saboo, 2022, p. 6). In addition, sources and documentations taken from GitHub and discussions concerning development projects hosted at Stack Overflow have been used. These materials enable GPT-3 to answer questions on how to develop software and to generate formal specifications.

First and foremost, the modeling and simulation community should ask which impact this technology does have on the community. The authors assume that the initial transformation might be in the way people develop models. Especially the capability of GPT-3, the Large Language Model (LLM) that drives ChatGPT, to hold conversations and handle narrative input opens up new fields of application.

In this paper, the textual descriptions and models considered are process models. Three systematic experiments performed by the authors demonstrate the practical capabilities the environment offers. Some difficulties needed to be overcome:

- Training of LLM needs input. Despite the above mentioned amount of text used to train GPT-3, the amount of informal process descriptions together with their formal specifications in some modeling notation is quite rare. (The authors assume that the proprietary storage of graphical models drawn in some modeling tools hinder a free access.) Consequently, some training of formal methods must be included into the dialog.
- Process models base on current business facts. However, the training material of GPT-3 is general text input taken from the internet, with the most recent data being from 2021. Currently, there exists no straight forward approach to enrich GPT-3 with current facts.
- There is a difference in the way GPT-3 interprets a process description and the way this is done by humans. So, how do we have to adapt these two processes to give modelers a new modeling experience?

Nonetheless, the possibilities of this new technology cannot been overseen as it offers an easily accessible API to perform linguistic analysis of input text for further processing. In the field of business process modeling, this might lead to faster model implementation and earlier use of the models.

The idea of the presented approach is to use process descriptions as input for GPT-3 and ask the system questions that help to develop an appropriate formal specification step-by-step. The research question is:

Which steps in the process of (process-)model development have the potential to be assisted by GPT-3 today?

This paper has limitations. The experimental design is intended to provide initial insights, and since OpenAI updates its model regularly, it might be difficult to repeat the experiments and to achieve the same results.

The remainder of the paper is organized as follows: Next, the section Related Work embeds the paper in a broader scope and provides basic information about LLMs. In the section Laboratory, the used systems and an overview of the experiment phases is given. The following three sections cover these Experiments and their Results. The paper closes with a Discussion and Conclusion.

Communications of the ECMS, Volume 37, Issue 1,
Proceedings, ©ECMS Enrico Vicario, Romeo Bandinelli,
Virginia Fani, Michele Mastroianni (Editors) 2023
ISBN: 978-3-937436-80-7/978-3-937436-79-1 (CD) ISSN 2522-2414

RELATED WORK

Even for humans who are creative opposed to computer algorithms the transformation of process descriptions into formal process specifications is a difficult undertaking. Figure 1 shows typical steps humans conduct to create a process model from a given process description. After the process elements (roles and actors, events and triggers, activities, and business and information objects) are identified, the process flow is composed in sequences, alternatives, iterations, and in concurrent structures. Quantitative information enrich the model.

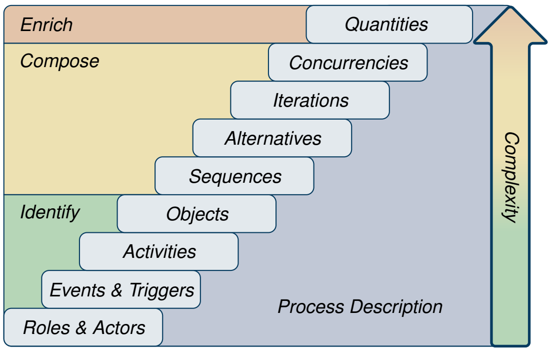


Figure 1: Procedure Model for Formal (Process) Modeling

Also GPT-3 must conduct these steps and therefore has to "understand" a process description to transform it into a formal structure using previously learned process patterns. Users might correct wrong transformations using the ChatGPT user interface.

(Kecht et al., 2023) evaluate the capability of chatbots to learn business processes from a large set of customer service conversations on Twitter. Chatbots, however, have a very restricted way to produce answers. Since the transformer model of GPT-3 is more powerful and less determinated, the approach explained in this contribution is entirely different. To understand this, the terms conversational agent, transformer language model, and especially GPT-3 are explained. Finally, the approach presented here is embedded in a broader research context.

At the time of submission, GPT-4 was not yet available and there was no final release schedule. Due to the short timespan after its release, no revision was made. Moreover, as access to GPT-4 is limited and paywalled, it is not readily available to the general public. Thus, the experiments retain their overall validity.

Conversational Agents

Conversational agents simulate human-like text- or voice-based conversations with users (Adamopoulou and Moussiades, 2022). While the complexity of conversational agents is relatively low, the variants can be classified somewhere in the following spectrum:

- Rule-based chatbots respond to specific inputs (cf. (Weizenbaum, 1966)).
- Voice assistants use artificial intelligence and machine learning to understand natural language inputs and respond in a human-like manner (cf. (Lopez et al., 2017)).

Transformers and The GPT Language Models

Transformer language models use deep learning to transform input into output sequences. Such a model is trained probabilistically to relate words and their contexts to other words and their respective contexts. The model can then decide which of these contexts are more likely to be related to a specific topic than others. Given a sequence of input tokens, a transformer takes its context into account to find the most likely word to follow (Vaswani et al., 2017).

Each token - user input or system output - is stored in a context vector which keeps the model in context during a dialog. This vector represents the working memory of the model and can include new information provided by the users (Vaswani et al., 2017).

The probabilistic nature of the language model makes it a natural enemy of deterministic process descriptions. Even if users enter process facts to nudge the system into a specific direction, other already learned patterns might prevent this. This problem needs to be solved.

GPT-3

GPT-3 is a state-of-the-art transformer language model to generate human-like natural text developed by OpenAI. Trained on a vast amount of text data, GPT-3 is capable of producing coherent and fluent text. Because of its large training corpus, task-specific finetuning through few-shot learning can produce surprisingly good results.

One of the key strengths of GPT-3 is its versatility in a wide range of applications. It can be used for tasks such as language translation, text completion, writing summaries, and to create chatbots and virtual assistants. GPT-3 is particularly notable for its capability to generate context-aware text that adapts to user preferences and language styles (Brown et al., 2020).

Speech to Model

The ability to *Transform Text to Model* is challenge 13 out of 25 challenges of semantic process modeling (Mendling et al., 2014). Figure 2 puts challenge 13 into a context and explains the different kinds of artifacts that must be produced so solve it and how its results may be provided as input to other tasks. These other challenges are to *Verify Model Correctness* and to *Validate Model Completeness* (Haag and Simon, 2022).

Challenge 13 is neither new nor unaddressed. Most of the contributions, however, use preprocessed data from texts (e.g. (Fliedl et al., 2005; Nolte, 2020)). Current NLP technologies such as GPT-3 or other transformers have not yet been investigated.

Challenge 13 is still unsolved for two reasons: 1) Rule-based automatisms can't reliably handle the complex nature of formal modeling, so a more flexible approach is needed. 2) Methods based on machine learning require large amounts of correctly annotated examples in their training phase, which simply aren't available. GPT-3 seems to break these limitations which motivates the following experiments.

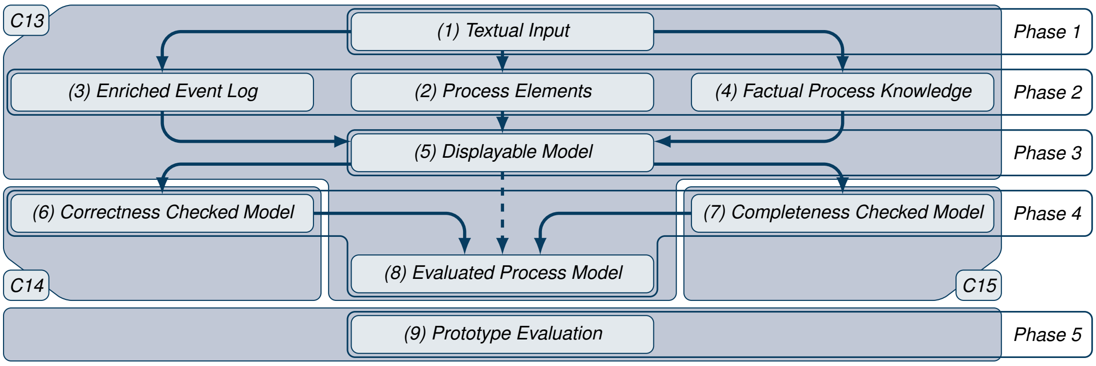


Figure 2: Phases and Artifacts of a Research Plan to Extract Models from Natural Language Text, cf. (Haag and Simon, 2022)

LABORATORY FOR THE EXPERIMENTS

The experiments have been conducted with the tools provided by OpenAI as a "laboratory". The execution of each experiment is documented in a laboratory logbook which is published on the working group's website. It provides all settings, describes the experiments, and summarizes the full prompts (Haag, 2023).

GPT-3 and ChatGPT

ChatGPT (`https://chat.openai.com/chat`) is a web-based tool that offers a conversational user interface to GPT-3. It generates relevant and human-like responses based on the large amount of training texts used to build the language model (OpenAI, 2022). ChatGPT is able to conduct a dialog which uses earlier interactions to improve the next output in a running conversation.
Text send to and received from GPT-3 is called a prompt. It is a kind of programming language in plain English, but also the system's output. Three types of prompts are distinguished (Tingiris, 2021, p. 6-10):
Zero-shot prompts provide a description of a task, or some text for GPT-3 to get started with.
One-shot prompts provide one example that GPT-3 can use to learn how to best complete a task.
Few-shot prompts provide multiple examples showing a pattern GPT-3 should continue.

OpenAI Playground

The OpenAI Playground (`https://platform.openai.com/playground`) is a more elaborated version of ChatGPT that still can be used without writing a single line of code (Tingiris, 2021, p. 20). It provides access to various GPT-based language models where *text-davinci-003* is the current default. Several parameters control the system's output (Brown et al., 2020), e. g.:
Temperature is a value between 0 and 1 that controls the randomness. The lower the value, the more deterministic and repetitive the system behaves.
Maximum length limits the magnitude of output text which is broken down into tokens - a numeric representation of parts of words. For most language models the maximum length is 2.048 tokens which corresponds to around 1.500 words (Tingiris, 2021, p. 12).
Token highlighting indicates how likely a token was to be generated.
The experiments have mainly been conducted with the default settings except for the following changes:
- The default temperature of 0.7 was partially changed to 0 to gain insight into a baseline.
- The maximum length was increased from 256 to higher values for a better user experience without an effect on the actually generated answers.
- Token highlighting was used.

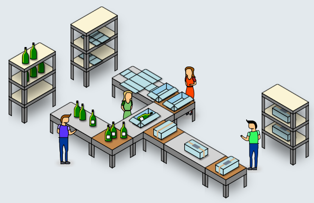
Phases of the Experiment

The experiments have been conducted in three phases:
1. Extending the description: GPT-3 was used to enrich an initial process description by further details.
2. Information Extraction: Afterwards, GPT-3 was tasked to extract process relevant information from this extended description.
3. Formal Transformation: Finally, GPT-3 was persuaded to present its answers in a formal way.
After this short introduction to the technical environment, the following sections explain the experimental setup and the observed results.

EXTENDING THE DESCRIPTION

For the experiments, a process description was taken from literature (Simon et al., 2022a). The process and its description are illustrated in Figure 3.

The process is the handling of personalized wine gifts in a winery. Unlabeled bottles of wine are taken from the storage and delivered to the input inventory of a workbench where they receive a personalized label according to customer wishes.



In parallel, cartons are dispatched in form of sheets to another workplace to be assembled there. Both items are taken to the packing station where bottle and newly created supporting material like greeting cards are put into the carton. Afterwards, the packed gift box (for short: box) is taken to another workplace where the shipping label is attached, and the box is sealed. The completed box is deposited in the outgoing goods area.

Figure 3: Example Process and its Visualization

Table 1: Probabilities for the Occurrence of Tokens in 10 OpenAI Playground Prompts Calculated at Temperature 0

token	run 1	run 2	run 3	run 4	run 5	run 6	run 7	run 8	run 9	run 10
corresponding	27.41	28.46	28.18	27.83	27.83	28.15	28.06	27.89	27.96	28.73
personalized	23.31	22.18	22.67	22.80	22.80	22.43	22.69	22.96	23.19	21.90
appropriate	19.55	19.74	19.47	19.54	19.54	19.67	19.57	19.44	19.36	19.70
label	5.62	5.62	5.79	5.71	5.71	5.81	5.73	5.78	5.74	5.64
necessary	5.34	5.22	5.25	5.30	5.30	5.21	5.22	5.18		5.26
respective									5.20	
cumulated	81.23	81.22	81.36	81.18	81.18	81.27	81.27	81.25	81.45	81.23

In a first step, this simple process description was used to explore the capability of GPT-3 to detect activities, events and process structures. Since this worked pretty well, it was the goal to increase the complexity with the aid of GPT-3. For this, the OpenAI Playground was prompted with the original process description and a list of process modifications shown in Figure 4 together with a request to produce an extended process description. ten runs were conducted with a temperature at 0.7 and another ten runs with a temperature at 0.

1. *The process starts with a customer order*
2. *From this order, amount and type of wine bottles are used to determine the bottles to be packed*
3. *The process not only handles gifts, but also standard orders. If an order is a gift, the personalized label for the bottle needs to be printed at an own workplace. If it is not a gift, this step isn't necessary as a stock label is used*
4. *The wine bottles are not necessarily packaged in size one, but there are different package sizes: 1, 3, 6, 12 bottles per package*
5. *Before sealing the box, a quality check is conducted for breakage, leakages, package contents, and overall looks*
6. *If during quality check an issue is discovered, the process should continue with the corresponding process step. The possible issues are: damaged bottle, damaged label on bottle, missing or damaged gift card, insufficient stuffing material, damaged cartons, missing or damaged shipping label*
7. *The process doesn't start immediately after receiving an order, but all orders are processed once a day*

Figure 4: Additional Information to Extend the Description

Figure 5 shows the result. It is a condensed version of the answers given by GPT-3. Although it has been lightly edited, it still demonstrates the impressive possibilities. The reasons for the edit are explained below.

Results of the First Phase

The goal of the first phase of experiments was to establish a process description for the next phases using the GPT-3 language generation capabilities. GPT-3 was able to add finer details and modify process elements. When asked about the quality of the produced results, GPT-3 correctly complained about a lack of business process management concepts like process goals, performance measures, or interfaces to a wider organizational context.

All ten runs at temperature 0.7 yielded different results concerning phrasing and process details, while all runs at temperature 0 resulted in the same phrasing. Thus, the latter should be used as a standard for repeatability.

The daily process of handling customer orders starts with the determination of the amount and type of wine bottles required for each order. Customers can order packages of one, three, six or twelve bottles each. The bottles are taken from storage to the input inventory at the workbench. If the order is a gift, the corresponding bottles receive a personalized label according to customer specifications that is printed at a small workstation. If it is a standard order, the bottles receive a stock label instead.
In parallel, cartons are dispatched to the assembly workplace, where they are assembled. At the packing station, the bottles and materials are put into the carton. This includes printing of these materials, such as accompanying documents and, for gift orders, greeting cards.
Next, a quality check is conducted to ensure that there are no breakages, leakages, or discrepancies in the package contents and, if it is a gift package, to verify that the overall appearance is suitable. If any issues are found during the quality check, the process continues with the corresponding step to rectify the issue.
The shipping label is attached to the box at another workplace, where it is also sealed. Lastly, the package is deposited in the outgoing goods area.

Figure 5: Automatically Extended Description

But this is not guaranteed which may be explained at the example of the term "corresponding" in the fourth sentence of the output shown in Figure 5.

Table 1 shows the probabilities calculated by GPT-3 for the occurrence of this token. The value varies by about 4.6% over the different results although the temperature was set to 0. Alternative terms always have lower values for all ten runs and for all of them the next 3 most probabilistic terms are always the same. But the lowest distance between the two uppermost probabilities is 4.1% which is quite close. In the ninth run, the fifth ranked tokens switch without having an impact on the actual output. Based on this observations, results may deviate.

The description of Figure 5 is not a copy of a single prompt as none of them was entirely convincing. For example, the last added information (No. 7: process starts once a day for all orders) was always placed as a separate sentence at the very end of the process description and not as supposed at it's beginning. This led to a tentative manual combination of the different prompts by the authors.

Furthermore, the description still leaves out some information: neither actors nor roles are defined, no information objects are given, and the control flow is described without explicitly mentioning control structures. This motivated the decision to investigate GPT-3's capability to reveal implicit information and to infer missing information.

INFORMATION EXTRACTION

The second phase of the experiment is about the capability of GPT-3 to extract process relevant information out of a given process description. The respective requests are formulated in 13 questions that are prompted to the system in a structured way. This step and its results are explained in the following.

Setting of the Experiment

The process description of Figure 5 was prompted to the Davinci model with maximum length settings of 2.048 and temperatures of 0 and 0.7. Afterwards, the tasks of Figure 6 were applied to this process description in four different settings:

1. In the first setting, all prompts were conducted with temperature 0. The process together with one of the questions each was send to the system in a separate prompt. This hindered the system to build a context vector linking the different questions.
2. The temperature for the second setting was also set to 0. Now the questions have been sent to the system one after another in a single prompt so that GPT-3 built a context vector.
3. In the third setting, the temperature was set to 0.7. Apart from that, the questions were send to GPT-3 like in the second setting. This procedure was repeated for five times.
4. Lastly, the temperature was set to 0.7. But in this setting the order of questions was randomized. This procedure was repeated for 20 times.

A full exploration of all 13! possible permutations which would gain deep insights into the behavior of the system fails with respect to its complexity.

1. Provide an ordered list of the workplaces that are used in this process.

2. Provide an ordered list of business roles or actors in this process.

3. Provide an ordered list of events that occur in this process.

4. Provide an ordered list of activities that need to be conducted in this process.

5. Provide an ordered list of business objects in this process.

6. Provide an ordered list of information objects in this process.

7. Is there a sequence in this process? If so, which?

8. Is there a decision in this process? If so, which?

9. Is there an alternative in this process? If so, which?

10. Is there an exclusivity in this process? If so, which?

11. Is there a concurrency in this process? If so, which?

12. Is there an iteration in this process? If so, which?

13. Is there a loop in this process? If so, which?

Figure 6: Tasks to Analyze the Process Description

The answers given by GPT-3 were evaluated for completeness and correctness from a human perspective as discussed next.

Results of the Second Phase

In addition to the results of the first phase, the following observations could be made:

- Explicit information concerning workplaces, activities, and some of the control flow information was extracted successfully.
- Implicit information was guessed:
  - Answers concerning roles and actors ranged from two roles (customers, workers) to ten (one employee per activity identified). In three runs, though, no roles were detected at all but workplaces were specified instead.
  - Business and information objects were also guessed and ranged from material to workplaces to control structures.
- Activities were identified almost correctly.
- Extraction of events almost failed. These attempts rather produced a list of activities some of which did not even fit the description.
- Control structures show an ambiguous picture:
  - Sequences were identified in all runs. They are, however, ordered in the way the activities occur in the description which is not the intended sequence. Also, some activities that are intended to occur concurrently were serialized.
  - When asked about concurrency, 22 of 27 runs gave a positive answer, even though they only had detected a single sequence before.
  - 23 runs detected at least (the intended) alternative.
  - 15 runs identified the intended iteration after the quality check. 3 misinterpreted information concerning the bottle quantities, the carton assembly, or the label printing as iterations. In 9 runs, no iteration was identified at all.

From the authors' perspective, this is a remarkable result when keeping in mind that these are early experiments on this new technology. Further improvements of the results seem to be possible if the context vector is systematically improved by correctly phrased questions asked in the right order.

FORMAL TRANSFORMATION

The previously discussed transformations produced text. But this paper is about GPT-3 assisted Process Model Development. Process models must be specified in a formal specification language which follows dedicated syntactic rules like in flow-diagrams or BPMN diagrams, probably even models using a semantic like Petri nets. The extracted results must then be merged in order to establish a representation of the entire process consisting of activities, events, process structures and other typical process model components. These elements must then be represented in a form processable by a process modeling environment.

Setting of the Experiment

The transformation of the narrative answers to the questions of Figure 6 into a formal representation requires a modeling environment with an open API or a plain-text file-format.

The authors chose to use the specification language of the Process-Simulation.Center (P-S.C), an Integrated Management System that allows for modeling, simulation, and documentation of processes with the aid of Petri nets. The tool includes specification languages for organigrams and swimlanes. Also, the connection of processes among each other in a process map can be specified in a textual form (Simon et al., 2022b). These experiments were also executed in two parts.

The first part was organized as follows:

1. Prompt of the process description at temperature 0.
2. Provide the workplaces first, because based on the previous experiments, a better extraction of other process elements was expected.
3. Extract the activities in a verb-noun phrasing and format the output as needed for the P-S.C.
4. Extract the events. However, as the word "event" did not work well the term "trigger" was used instead. This output was also formatted as needed.
5. Extract structures of "sequence", "branching", "merging", "iteration", and "concurrency" in the mentioned order and format the output as needed.

This experiment was repeated once at a temperature of 0.7, and five times in ChatGPT with varying orders of the requests for the control structure.

In the second part, the tasks prompted to the system were rephrased and enriched with examples. The first questions were posed in the same order, however the one for the control structures was changed to "branching", "merging", "concurrency", "iteration", and "sequence". Again, these tasks were conducted once at temperature 0 and five times at temperature 0.7.

Results of the Third Phase

The first steps delivered valuable results when GPT-3 was provided with precise instructions and examples of correct formatting. But the system deviated more and more from the established process elements and structures as the chat progressed.

- Although the prompted style required named Petri net elements, in several runs at temperature 0.7 ChatGPT numbered them instead. Also closing semicolons were omitted. The system performed better at temperature 0.
- The extraction of triggers worked worse compared with activities, but rephrasing and providing examples improved the results.
- The extraction of control structures failed, but changing the order of structures asked for in the prompts improved the results slightly. The system, however, still missed intended structures or built structures not described. Furthermore, activities were serialized as they occur in the description ignoring deviating statements.

In summary, GPT-3 was able to produce rudimentary process models concerning activities and events, but the results for control structures are far from being usable. From this point of view, the problems of GPT-3 to format a correct formal syntax can be put aside for the time being.

DISCUSSION AND CONCLUSION

The experiments described in this contribution give a first answer to the question:
Which steps in the process of (process-)model development have the potential to be assisted by GPT-3 today?
The experiments have been divided into the phases *Extending the Description*, *Information Extraction*, and *Formal Transformation* that address different linguistic capabilities of GPT-3: text generation, text summary, and translation (into a formal language).

Keeping in mind that this technology is quite new, the achieved results are impressive. They are a promising development compared to former work on chatbots for process patterns. The transformer technology and the massive amount of training data opens new horizons. But the current capabilities are still far away from being conducted unsupervised. Assisted process model development is possible, but an automatism is still out of reach.

The reasons for this can be observed in each phase of the experiments:

Extending the Description: Although GPT language models are trained on a very large dataset, they don't have world knowledge or common sense. They are not capable of inferring new knowledge, but are limited to what they have been exposed to during their training. This hinders an extension of existing descriptions in a reasonable way fine-tuning may be difficult, because specific domain is less available than general training material.

Another limitation is important for this task: although it may appear so, GPT is incapable of reasoning and calculation. What looks like reasoning is just the output of the most likely word and number combinations given the context vector.

Information Extraction: The descriptions need to be as precise and clear as possible to achieve accurate results. Then, knowledge extraction is possible if the system is asked the right questions in the right order. In the experiments described here, however, it was not always possible to gather all relevant information.

Formal Transformation: GPT-3 can also help to create formal models. Process elements and (at least partial) structures can be extracted and serialized. But this must be carefully prepared in order to use the result for modeling software.

Future work on the use of GPT-3 for assisting business process model development is needed in various fields. Detailed work will be on the various language models and the different process models they generate. A large number of well documented and evaluated experiments will have to be conducted.

Another field will be systematic prompt design to solve the numerous tasks related to process modeling.

However, the authors expect that the main area of research will be to add factual knowledge to the inherently probabilistic knowledge base.

For the improvements in GPT-4 over GPT-3 and their differing capabilities, please refer to (Bubeck et al., 2023), which was recommended by a reviewer.

ACKNOWLEDGEMENT

This research was supported by "ProFIL - Programm zur Förderung von Forschungspersonal, Infrastruktur und forschendem Lernen der Hochschule Worms".

REFERENCES

Adamopoulou, E. and L. Moussiades (2022): An Overview of Chatbot Technology. In: *Artificial Intelligence Applications and Innovations*, eds. I. Maglogiannis; L. Iliadis; and E. Pimenidis. Springer, Cham, pp. 373–383.

Brown, T.; B. Mann; N. Ryder; M. Subbiah; J. D. Kaplan; P. Dhariwal; A. Neelakantan; P. Shyam; G. Sastry; A. Askell; S. Agarwal; A. Herbert-Voss; G. Krueger; T. Henighan; R. Child; A. Ramesh; D. Ziegler; J. Wu; C. Winter; C. Hesse; M. Chen; E. Sigler; M. Litwin; S. Gray; B. Chess; J. Clark; C. Berner; S. McCandlish; A. Radford; I. Sutskever; and D. Amodei (2020): Language Models are Few-Shot Learners. In: *Advances in Neural Information Processing Systems*, eds. H. Larochelle; M. Ranzato; R. Hadsell; M.F. Balcan; and H. Lin. Curran Associates, Inc., vol. 33, pp. 1877–1901.

Bubeck, S.; V. Chandrasekaran; R. Eldan; J. Gehrke; E. Horvitz; E. Kamar; P. Lee; Y. T. Lee; Y. Li; S. Lundberg; H. Nori; H. Palangi; M. T. Ribeiro; and Y. Zhang (2023): Sparks of Artificial General Intelligence: Early experiments with GPT-4. `https://arxiv.org/abs/2303.12712/` (last accessed: 10.04.2023).

Fliedl, G.; C. Kop; and H. C. Mayr (2005): From textual scenarios to a conceptual schema. *Data and Knowledge Engineering*, 55(1):20–37.

Haag, S. (2023): *Protocol of Experiments on GPT-3 for Process Model Generation*. Tech. rep., HS Worms.

Haag, S. and C. Simon (2022): Extracting Models from Spoken Text - A Research Proposal. In: *EMISA 2022*. p. in Veröffentlichung.

Kecht, C.; A. Egger; W. Kratsch; and M. Röglinger (2023): Quantifying chatbots' ability to learn business processes. *Information Systems*, 113.

Kublik, S. and S. Saboo (2022): *GPT-3 - Building Innovative NLP Products Using Large Language Models*. O'Reilly, Sebastopol, CA.

Lopez, G.; L. Quesada; and L. A. Guerrero (2017): Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-Based Natural User Interfaces. In: *Advances in Human Factors and Systems Interaction*, ed. I. L. Nunes. Springer, Cham, pp. 241–250.

Mendling, J.; H. Leopold; and F. Pittke (2014): 25 Challenges of Semantic Process Modeling. *International Journal of Information Systems and Software Engineering for Big Companies*, 1(1):78–94.

Nolte, F. R. (2020): *Text to Process Model: Automating Process Model Creation form Text*. Ph.D. thesis, WWU Münster.

OpenAI (2022): ChatGPT: Optimizing Language Models for Dialogue. `https://openai.com/blog/chatgpt/` (last accessed: 10.04.2023).

Simon, C.; S. Haag; and L. Zakfeld (2022a): Stratification of Timed Petri Nets at the Example of a Production Process. In: *ECMS 2022: 36th International ECMS Conference on Modelling and Simulation*. pp. 128–134.

Simon, C.; S. Haag; and L. Zakfeld (2022b): The Process-Simulation.Center. In: *SIM-SC: Special Track at SIMUL 2022: The Fourteenth International Conference on Advances in System Simulation*, ed. F. Herrmann. Lissabon (Portugal), pp. 74–77.

Tingiris, S. (2021): *GPT-3 - An unofficial first look at the general-purpose language processing API from OpenAI*. Packt, Birmingham, UK.

Vaswani, A.; N. Shazeer; N. Parmar; J. Uszkoreit; L. Jones; A. N. Gomez; L. Kaiser; and I. Polosukhin (2017): Attention Is All You Need. In: *31st Conference on Neural Information Processing System*. Curran Associates.

Weizenbaum, J. (1966): ELIZA - A Computer Program For the Study of Natural Language Communication Between Man And Machine. *Communications of the ACM*, 9(1):36–45.

AUTHOR BIOGRAPHIES

CARLO SIMON studied Informatics and Information Systems at the University of Koblenz-Landau. For his PhD, he applied process thinking to automation technology in the chemical industry. For his state doctorate, he considered electronic negotiations from a process perspective. Since 2007, he is a Professor for Information Systems, first at the Provadis School of Technology and Management Frankfurt and since 2015 at the Hochschule Worms. His e-mail address is: simon@hs-worms.de.

STEFAN HAAG holds degrees in Business Administration and Engineering as well as Economics. His main interests are related to modeling and simulation in graphs and practical applications of computational linguistics. After working at the Fraunhofer Institute for Systems and Innovation Research ISI Karlsruhe for several years, he has been a Research Fellow at the Hochschule Worms since 2017. His e-mail address is: haag@hs-worms.de.

LARA ZAKFELD graduated in International Logistics Management (B.A.) after completing an apprenticeship as a management assistant in freight forwarding and logistics services. She is currently pursuing a master's degree in Entrepreneurship and works as a Research Assistant at the Hochschule Worms. Her e-mail address is: zakfeld@hs-worms.de.