Bus Inf Syst Eng 68(1):1–5 (2026)
https://doi.org/10.1007/s12599-025-00978-8

EDITORIAL

Generative AI and Conceptual Modeling

Hans-Georg Fill · Jennifer Horkoff · Peter Fettke · Julius Köpke

Published online: 9 February 2026

1 Introduction

The release of ChatGPT in November 2022 has caused generative artificial intelligence to become a global phenomenon that is being explored in many areas of science and practice including conceptual modeling. Once known mainly to specialists in deep learning and text generation, machine learning models that are trained on vast amounts of data and usable across many different downstream tasks have now become commodity tools. Today, a large number of high-quality, pre-trained large language models (LLMs) are available. These models allow users to issue natural language prompts in order to create and analyze text, conceptual models, code, images, videos, and audio data.

H.-G. Fill (✉)
Research Group Digitalization and Information Systems, University of Fribourg, 1700 Fribourg, Switzerland
e-mail: hans-georg.fill@unifr.ch

J. Horkoff
The University of Gothenburg and Chalmers University of Technology, Gothenburg, Sweden
e-mail: jennifer.horkoff@gu.se

P. Fettke
German Research Center for Artificial Intelligence (DFKI), 66123 Saarbrücken, Germany
e-mail: Peter.Fettke@dfki.de

P. Fettke
Saarland University, 66123 Saarbrücken, Germany

J. Köpke (✉)
University of Klagenfurt, Klagenfurt, Austria
e-mail: julius.koepke@aau.at

LLMs can be categorized based on how readily available information is about their training data and the weights that determine a model's capabilities. Proprietary models such as OpenAI's advanced GPT models,[1] Google's Gemini,[2] or Anthropic's Claude[3] models are typically only accessible through some service endpoint and neither disclose their training data nor their weights. Access to them requires the provision of paid or unpaid access through a provider. In contrast, the weights of open-weight models such as OpenAI's GPT-OSS models,[4] Meta's Llama,[5] Google's Gemma,[6] or Apple's Ferret[7] can be downloaded and run on one's own machine. Although the training data is not available in this case, this allows for independent experimentation. Most recently, open models such as the Swiss Apertus model (Hernández-Cano et al. 2025) disclose their training data and procedure as well their weights and are thus fully transparent. From the viewpoint of open science, open models are preferable in order to inspect and validate all components, ensure the reproducibility of research, and contribute to the quality of scientific work. However, as of today, the proprietary models are still far ahead in terms of generation quality, which makes it necessary for researchers to work with these models when conducting experiments. This applies also to the field of conceptual modeling and generative AI.

[1] https://platform.openai.com/docs/models.
[2] https://ai.google.dev/gemini-api/docs/models.
[3] https://platform.claude.com/docs/en/about-claude/models/overview.
[4] https://platform.openai.com/docs/models/gpt-oss-120b.
[5] https://www.llama.com/.
[6] https://ai.google.dev/gemma/docs.
[7] https://github.com/apple/ml-ferret.

Conceptual modeling is a pivotal academic field not only in business and information systems engineering, but also in (management) information systems, (business) informatics, software engineering, process science and other disciplines (Frank et al. 2014; Michael et al. 2024; Mayr and Thalheim 2021). It supports human understanding and communication in general (Mylopoulos 1992), aligns business and IT aspects (Sandkuhl et al. 2018; Fill 2020), formalizes requirements for software-based systems (Horkoff & Yu 2016), and analyzes domain concepts and terminologies (Van Gils et al. 2022). Conceptual models are based on schemata in the form of modeling languages including syntax, semantics, and a visual or textual notation (Harel and Rumpe 2004). In addition, they are also used in model-driven engineering (Brambilla et al. 2017), for code generation (Sebastián et al. 2020) and simulation (Rosenthal et al. 2021). Further, conceptual models may be processed semantically, thereby acting as an interface to knowledge graphs and reasoning (Smajevic and Bork 2021; Fill 2017). Conceptual models are the result of higher-order cognitive processes such as abstraction, relational reasoning based on the integration, and maintenance of information. They are thus subject to individual differences between modelers and their interpretations (Wilmont et al. 2013).

Soon after the introduction of ChatGPT, it became apparent that large language models are highly capable of creating and analyzing conceptual models in a variety of modeling languages (Fill et al. 2024). Based on some textual input, LLMs can create conceptual models by using existing syntax formats such as PlantUML or BPMN-XML as well as newly specified ones. At first this was surprising since the models were not specifically trained for such tasks. However, among the large amounts of training data, there was obviously sufficient information about conceptual models to allow LLMs to reason about requests for creating models which use established languages as well as ones in newly designed languages, which are explained in a prompt at runtime using fewshot learning (Fill et al. 2024). Since then, many experiments have been conducted for determining the best approaches for generating conceptual models (e.g., Calamo et al. 2025; Reinhartz-Berger et al. 2025; Safan and Köpke 2025; Köpke and Safan 2024; Muff and Fill 2024, 2025; Kolev et al. 2025; Klievtsova et al. 2025). The scope of investigated modeling languages until now includes well-established languages such as the Unified Modeling Language (UML), Entity-Relationship (ER) diagrams, or business process models in BPMN notation, as well as newly developed and domain-specific languages such as Heraklit or ARWFML (Reinhartz-Berger et al. 2025; Fill et al. 2023; Muff and Fill 2025; Baumann et al. 2024).

From the viewpoint of conceptual modeling, the phenomenon of generative AI has several implications, which touch both upon the way conceptual models are technically created as well as in terms of how humans conceptualize the world and interact with AI. The mere generation of conceptual models from textual input has been explored in the field of natural language processing (NLP) for quite some time (e.g., Bellan et al. 2023; van der Aa et al. 2018). In this regard, the use of LLMs for model creation may be viewed as another tool that surpasses the performance and quality of previous approaches and does not require specific training or configurations. However, LLMs' ease of use and foundational nature allow them to be applied to many different formal and semi-formal languages and tasks, contributing to their disruptive potential (Buchmann et al. 2024; Storey et al. 2025). This has implications for the practice of information systems engineering. Requirements for systems do not need to be manually elicited from textual descriptions and then translated into conceptual models. Rather, LLMs can generate a preliminary version of potential requirements and their formalization, which is subsequently refined by human actors as needed (Ronanki et al. 2024). Further, the advancement of LLMs has enabled them to perform well on agentic tasks. Thereby, their capabilities in terms of generating code allow for the derivation of actions, e.g., via calls to APIs, thus realizing agentic workflows (Fettke et al. 2025). From the viewpoint of teaching, the use of LLMs has many implications that are not yet fully understood. Some in the conceptual modeling community claim that education on conceptual modeling has to shift from model creation to the critical evaluation of models (Snoeck and Pastor 2025). Others, however, put the focus on conceptual models as one potential interface to GenAI systems where both humans and AI systems are forced to formalize their thoughts in the form of conceptual models in order to better understand each other and more easily detect errors in their conceptualizations (Fill et al. 2024).

Lastly, a major open issue is the evaluation of the output of LLMs when generating or interpreting conceptual models. Today, this is typically checked against some baseline in the form of ideal models created by human modelers (e.g., Calamo et al. 2025). Given, however, the fact that conceptual modeling is a cognitive effort to structure the world as some subject or a group of subjects perceive it (Wilmont et al. 2013), the question remains whether such baselines are indeed valid for evaluations as this would implicitly assume the existence of a single "ground truth" representation of a domain. The value of generated conceptual models may thus need to be evaluated in the future from multiple dimensions other than strict conformance metrics. A step in this direction may be the use of quality frameworks for models, as for example proposed by Krogstie (2012), that take a holistic view.

2 Overview of the Special Issue

For the special issue at hand, three papers were selected in a highly competitive review process with 11 initial submissions. All papers were evaluated by at least three experts in the field and have undergone at least two revisions.

The paper by Schinckus, Simonofski, and Rosselló on "Large Language Models for Process Knowledge Acquisition: Designing and Evaluating a Multi-Agent System" focuses on improving the acquisition of process knowledge using large language models (LLMs). Using a design science research approach, they investigate the requirements for using LLMs during process discovery. They then translate these requirements into a prototype and empirically evaluate it. Their prototype uses a multi-agent approach, in which process analysts and domain experts interact with LLM-based agents, such as preparator, interviewer, elicitor, and modeler agents. Through experiments with domain experts and process analysts, they found that using LLM-generated conceptual models enhances the exploratory process discovery phase. This is because it empowers domain experts to independently create initial formalized sketches of their process knowledge. Furthermore, they discovered that the iterative refinement applied in their approach improves the semantic and pragmatic quality of the elicited information. The second paper by Elham Motamedi, Inna Novalija, and Luis Rei on "Semi-automatic Hierarchical Taxonomy Creation From Existing Taxonomies With Large Language Models" proposes a method for the semi-automatic refinement of taxonomies using large language models. Thereby, LLMs are used for supporting tasks such as the definition of metaheuristics, proposing merges of concepts or the generation of representative labels. They apply their approach in a case study for an innovations taxonomy for assisting users in classifying patent documents that is based on the International Patent Classification. They could show that the LLM-based approach reduces the need for expert involvement in the taxonomy creation process by shifting the role of experts to overseeing and refining an automated process. They conclude that LLMs can be successfully used for assisting experts in taxonomy refinement, which scales more efficiently than traditional methods.

Finally, the paper by Luca Franziska Hörner, Maximilian Josef Möller, Michael Winter, and Manfred Reichert on "Automatically Generating BPMN 2.0 Process Models from Natural Language Process Descriptions: Challenges, Framework, Quality Assessment" studies the semantic quality and comprehensibility of models in BPMN generated by LLMs from textual descriptions. It compares LLM generated models against expert-created models. For this purpose they present an LLM-based conversational software application based on a client–server architecture. In an empirical study they subsequently investigated the semantic accuracy of process models, the cognitive load during the comprehension of process models, the acceptability of process models, and the comprehension performance of process models, either generated by an LLM or by human experts. Their findings indicate that the comprehension of LLM-generated process models was mentally not more demanding than expert-created models and that both types of models are comparable in terms of accessibility and ease of understanding.

The special issue is concluded with an interview with Frank van Harmelen, full professor at the Vrije Universiteit Amsterdam. In this interview we asked him to reflect on the future of generative AI and the role of conceptual modeling, given his background in knowledge representation and symbolic AI.

3 Directions for Future Research on GenAI in Conceptual Modeling

The papers in this special issue have made innovative contributions to advancing generative AI in conceptual modeling. However, the papers point out several limitations of current approaches that must be addressed in the future. We categorize these limitations as *structural*, *content-related*, and *technological*.

In terms of structure, some approaches have so far reverted to models of only limited size and complexity. Given the fact that real world conceptual models, e.g., in business process management or enterprise architecture management, exhibit much larger sizes and greater complexity, this will need to be considered by future approaches.

As pointed out above, the evaluation of the content of GenAI generated conceptual models is an ongoing challenge. Given the nature of conceptual models as artifacts created through cognitive processes by individuals, the value of models cannot fully be evaluated using traditional conformance checking methods. Rather, the evaluation approaches will need to be expanded to multiple dimensions, also considering the nature of LLMs as large repositories of knowledge that may actually point to new interpretations in inputs that may have been overlooked by humans.

On the technological level, first approaches have appeared that not only consider textual prompts as inputs for generating conceptual models but that also revert to existing model repositories, modeling guidelines, or other forms of documentation for enhancing the generation results. These can be provided through RAG- (retrieval augmented generation) based approaches, few shot learning, or the fine tuning of LLMs.

References

Baumann N, Diaz JS, Michael J, Netz L, Nqiri H, Reimer J, Rumpe B (2024) Combining retrieval-augmented generation and few-shot learning for model synthesis of uncommon DSLs. https://dl.gi.de/handle/20.500.12116/43781

Bellan P, Dragoni M, Ghidini C, van de Aa H, Ponzetto SP (2023) Process extraction from text: Benchmarking the state of the art and paving the way for future challenges. https://doi.org/10.48550/arXiv.2110.03754

Brambilla M, Cabot J, Wimmer M (2017) Model-driven software engineering in practice 2nd edition. Synth Lect Softw Eng. 3(1):1–207

Buchmann R, Eder J, Fill HG, Frank U, Karagiannis D, Laurenzi E, Mylopoulos J, Plexousakis D, Santos MY (2024) Large language models: expectations for semantics-driven systems engineering. Data Knowl Eng 152:102324. https://doi.org/10.1016/j.datak.2024.102324

Calamo M, Mecella M, Snoeck M (2025) Assessing the suitability of large language models in generating UML class diagrams as conceptual models. In: Guizzardi R et al (eds) Enterprise, business-process and information systems modeling. Springer, Cham, pp 211–226. https://doi.org/10.1007/978-3-031-95397-2_13

Fettke P, Fill HG, Köpke J (2025) LLM, LAM LxM Agent: from talking to acting machines: Insights from the perspective of conceptual modeling. Enterp Model Inf Syst Arch (EMISAJ) Int J Concept Model. https://doi.org/10.18417/emisa.20.3

Fill HG (2017) SeMFIS: a platform for semantic annotations of conceptual models. Semantic Web 8(5):747–763. https://doi.org/10.3233/SW-160235

Fill HG, Fettke P, Köpke J (2023) Conceptual modeling and large language models: Impressions from first experiments with ChatGPT. Enterp Model Inf Syst Archit Int J Concept Model. https://doi.org/10.18417/EMISA.18.3

Fill HG, Härer F, Vasic I, Borcard D, Reitemeyer B, Muff F, Curty S, Bühlmann M (2024) CMAG: A framework for conceptual model augmented generative artificial intelligence. In: Gallinucci E, et al (eds) Companion proceedings of the 43rd international conference on conceptual modeling, Pittsburgh. https://ceur-ws.org/Vol-3849/forum5.pdf

Fill HG (2020) Enterprise modeling: From digital transformation to digital ubiquity. In: Ganzha M, et al (eds) Proceedings of the 2020 federated conference on computer science and information systems, Sofia. https://doi.org/10.15439/2020F001

Frank U, Strecker S, Fettke P, vom Brocke J, Becker J, Sinz E (2014) The research field "Modeling business information systems": current challenges and elements of a future research agenda. Bus Inf Syst Eng 6:39–43. https://doi.org/10.1007/s12599-013-0301-5

Harel D, Rumpe B (2004) Meaningful modeling: what's the semantics of "semantics"? Computer 37(10):64–72. https://doi.org/10.1109/MC.2004.172

Hernández-Cano A, Hägele A, et al (2025) Apertus: Democratizing open and compliant LLMs for global language environments. https://doi.org/10.48550/arXiv.2509.14233

Horkoff J, Yu E (2016) Interactive goal model analysis for early requirements engineering. Requir Eng 21(1):29–61. https://doi.org/10.1007/s00766-014-0209-8

Klievtsova N, Benzin JV, Mangler J, Kampik T, Rinderle-Ma S (2025) Process modeler vs Chatbot: Is generative AI taking over process modeling? In: Delgado A, Slaats T (eds) Process mining workshops. Springer Nature, Cham, pp 637–649. https://doi.org/10.1007/978-3-031-82225-4_47

Kolev PA, Pruss HH, Wilken JR, Sandkuhl K (2025) Grass-root enterprise modelling: How large language models can help. In: Paja E et al (eds) The practice of enterprise modeling. Springer Nature, Cham, pp 123–139. https://doi.org/10.1007/978-3-031-77908-4_8

Köpke J, Safan A (2024) Efficient LLM-based conversational process modeling. In: Gdowska K et al (eds) Business process management workshops—BPM 2024 international workshops. Springer, Krakow, Revised Selected Papers, pp 259–270. https://doi.org/10.1007/978-3-031-78666-2_20

Krogstie J (2012) Quality of business process models. In: Sandkuhl K et al (eds) The practice of enterprise modeling, vol 134. Springer, Heidelberg, pp 76–90. https://doi.org/10.1007/978-3-642-34549-4_6

Mayr HC, Thalheim B (2021) The triptych of conceptual modeling. Softw Syst Model 20(1):7–24. https://doi.org/10.1007/s10270-020-00836-z

Michael J, Bork D, Wimmer M, Mayr HC (2024) Quo vadis modeling? Softw Syst Model 23(1):7–28. https://doi.org/10.1007/s10270-023-01128-y

Muff F, Fill HG (2024) Limitations of ChatGPT in conceptual modeling: Insights from experiments in metamodeling. Workshop Modeling in the Age of Large Language Models (LLM4Modeling). GI. https://doi.org/10.18420/MODELLIERUNG2024-WS-008

Muff F, Fill HG (2025) Creating augmented reality applications using large language models: experiments with the CMAG framework and ARWFMM. Proc AAAI Symp Ser 5(1):374–378. https://doi.org/10.1609/aaaiss.v5i1.35615

Mylopoulos J (1992) Conceptual modelling and Telos. Conceptual modelling, databases, and CASE: an integrated view of information system development. Wiley, Chichester, pp 49–68

Reinhartz-Berger I, Ali SJ, Bork D (2025) Leveraging LLMs for domain modeling: the impact of granularity and strategy on quality. In: Krogstie J et al (eds) Advanced information systems engineering. Springer, Cham, pp 3–19. https://doi.org/10.1007/978-3-031-94569-4_1

Ronanki K, Cabrero-Daniel B, Horkoff J, Berger C (2024) Requirements engineering using generative AI: Prompts and prompting patterns. In: Nguyen-Duc A et al (eds) Generative AI for effective software development. Springer nature, Cham, pp 109–127. https://doi.org/10.1007/978-3-031-55642-5_5

Rosenthal K, Ternes B, Strecker S (2021) Business process simulation on procedural graphical process models. Bus Inf Syst Eng 63(5):569–602. https://doi.org/10.1007/s12599-021-00690-3

Safan A, Köpke J (2025) A framework for LLM-based conceptual modeling: Application to BPMN collaboration diagrams. In: ER2025: companion proceedings of the 44th international conference on conceptual modeling: industrial track, ER Forum, 8th SCME, Poitiers. https://ceur-ws.org/Vol-4099/forum_paper7.pdf

Sandkuhl K, Fill HG, Hoppenbrouwers S, Krogstie J, Matthes F, Opdahl A, Schwabe G, Uludag O, Winter R (2018) From expert discipline to common practice: a vision and research agenda for extending the reach of enterprise modeling. Bus Inf Syst Eng 60(1):69–80. https://doi.org/10.1007/s12599-017-0516-y

Sebastián G, Gallud JA, Tesoriero R (2020) Code generation using model driven architecture: a systematic mapping study. J Comput Lang 56:100935. https://doi.org/10.1016/j.cola.2019.100935

Smajevic M, Bork D (2021) From conceptual models to knowledge graphs: A generic model transformation platform. In: 2021 ACM/IEEE international conference on model driven engineering languages and systems companion (MODELS-C). pp. 610–614. https://doi.org/10.1109/MODELS-C53483.2021.00093

Snoeck M, Pastor O (2025) Teaching conceptual modelling in the age of LLMs: shifting from model creation to model evaluation skills. Softw Syst Model. https://doi.org/10.1007/s10270-025-01307-z

Storey VC, Pastor O, Guizzardi G, Liddle SW, Maaß W, Parsons J, Ralyté J, Santos MY (2025) Large language models for conceptual modeling: assessment and application potential. Data Knowl Eng 160:102480. https://doi.org/10.1016/j.datak.2025.102480

van der Aa H, Carmona J, Leopold H, Mendling J, Padró L (2018) Challenges and opportunities of applying natural language processing in business process management. In: Bender EM, et al (eds.) Proceedings of the 27th international conference on computational linguistics. Santa Fe, pp. 2791–2801. https://aclanthology.org/C18-1236/

van Gils B, Hoppenbrouwers S, Proper HA (2022) Conceptual modeling in digital transformations — Enabling enterprise design dialogues. In: Clark T, et al (eds.) Proceedings of the forum at practice of enterprise modeling 2022, London, pp. 41–51. https://ceur-ws.org/Vol-3327/paper05.pdf

Wilmont I, Hengeveld S, Barendsen E, Hoppenbrouwers S (2013) Cognitive mechanisms of conceptual modelling. In: Ng W, et al (eds) Conceptual Modeling. pp. 74–87. Springer. https://doi.org/10.1007/978-3-642-41924-9_7