How Does UML Look and Sound? Using AI to Interpret UML Diagrams through Multimodal Evidence

Aleksandar Gavric[0009−0005−1243−7722], Dominik Bork[0000−0001−8259−2297], and Hend*erik* A. Proper[0000−0002−7318−2496]

Business Informatics, TU Wien, Vienna, Austria
{aleksandar.gavric, dominik.bork, henderik.proper}@tuwien.ac.at

Abstract. Despite the availability of abundant digital data, there is a research gap in effectively utilizing this empirical evidence to inform the process of constructing or learning conceptual models. This gap highlights the need to address empirically-informed conceptual modeling practice. This paper underscores the potential of Artificial Intelligence (AI) for improving evidence-based conceptual modeling practice, specifically with Unified Modeling Language (UML) diagrams. We use generative AI to provide a look (visual) and sound (audio) perspective of UML diagrams. Our solution, named **mod**Sense, can link diagram elements to existing image and audio data or generate new image and audio content to improve understanding through real-world examples. Furthermore, our solution incorporates human preferences and feedback to dynamically adjust the generated or retrieved content to the user's comprehension level, providing a tailored human-model interaction experience. Through increased engagement and real-world connections in UML diagrams, we aim to make models more aligned with business logic, resulting in better conceptual models and, subsequently, more effective computer programs. We report on the initial results of **mod**Sense and the **mod**Sense**4**All empirical study that focuses on assessing the educational impact of these multimodal resources in the domain of programming assistant software for surgery applications.

Keywords: UML · Empirical Evidence · Generative AI · Multimodal Data · Conceptual Modeling.

1 Introduction

> *"If you cannot explain a program to yourself, the chance of the computer getting it right is pretty small."* – Bob Frankston

Understanding and clearly communicating the structure and behavior of a system is fundamental in software development, enabling practitioners to specify, visualize, construct, and document the artifacts of software systems. The Unified Modeling Language (UML) provides a standardized way to visualize

and document these systems, but how does UML translate beyond static diagrams? This paper explores innovative approaches to interpreting UML diagrams through generated images and audio, offering a multi-sensory experience to enhance comprehension and communication among model readers/creators and stakeholders. Although UML is fundamental in technical education and professional practice, its complexity can be daunting, particularly for newcomers [3]. This complexity necessitates innovative educational tools that enhance comprehension and facilitate a more intuitive learning process. This research contributes to the evidence-based conceptual modeling practice through pedagogical methods in software engineering education and aligns with AI advancements. We take the application of AI a step further—aiming to foster the reading and creation of UML diagrams by providing real-world examples of diagram concepts and structures. This raises our research question (**RQ**): *How can generative AI enhance the comprehension and creation of UML diagrams by providing contextual real-world examples?*

To respond to those research questions, we utilize generative AI to offer both visual and auditory perspectives on UML diagrams. Our approach can associate diagram components with specifications to generate new content and, therefore, facilitate understanding through practical, real-world examples. Moreover, our approach adapts to individual user preferences and feedback, dynamically modifying the generated or retrieved content to match the user's level of comprehension, thus providing a personalized human-model interaction experience. By increasing engagement and making real-world connections within UML diagrams, we aim to ensure that models are more closely aligned with business logic and related to the domain represented by the models. This leads to evidence-based rather than confidence- or assumption-based conceptual models and, ultimately, more effective computer programs. Our solution is particularly beneficial for simulating domains where evidence data collection is hindered by constraints such as limited time, spatial opportunities, or prohibitive costs. Our approach aims to represent and manage these constraints within our simulations. Furthermore, our approach is invaluable in domains where the content is complex and challenging for beginners, such as surgical procedures.

In the remainder of this paper, Section 2 provides related work. Section 3 describes our solution, termed ***mod****Sense*, which is an umbrella term for the method and the tool used for interpreting UML diagrams with multimodal AI [7]. We report on our current evaluation results and the on-going empirical study in Section 4 and conclude in Section 5.

2 Related Work

Understanding conceptual models of business domains is a key skill for practitioners tasked with systems analysis and design [12].

2.1 Generative AI and UML

Our work is positioned in the intersection of Generative AI, such as Generative Adversarial Networks [8], that offered a generative approach to visualizing empirical knowledge and interpreting UML diagrams, such as [16]. When it comes to composing UML diagrams, natural language models have been developed to transform textual requirements into structured UML class diagrams, enhancing the comprehension and visualization of software design [1]. Conversational AI further enhances the learning experience by providing interactive and adaptive feedback to learners. Intelligent tutoring systems like *COLLECT-UML* have demonstrated the effectiveness of using constraint-based feedback to teach object-oriented analysis and design using UML. Automated tools for evaluating and correcting UML diagrams also play a fundamental role in the learning process. For example, tools that automatically classify and correct UML diagrams based on images have been developed, allowing learners to receive instant feedback on their diagrams. These tools utilize machine learning models to detect and correct errors, making the learning process more efficient and effective [10].

2.2 Education in the Context of UML

When it comes to the education of UML context, the potential of generative AI in computing education was explored by Zastudil et al. (2023) [17], who conducted interviews with students and instructors to gauge their experiences and preferences. Their findings suggest a growing acceptance of AI tools in classrooms, albeit with concerns regarding their optimal use. Farrelly et al. (2023) [5] expanded on these implications, discussing the impact of AI on higher education and the challenges of maintaining academic integrity.

These studies collectively demonstrate the potential of AI in education while also highlighting the ethical and practical challenges that must be addressed to ensure its effective and responsible use. Our empirical study is in compliance with the mentioned guidelines and takes the application of AI a step further while also taking a step toward a more specific application to foster UML diagram reading and creation by providing real-world examples of diagram concepts and structures. When it comes to organizing the empirical study, research in this field predominantly uses experiments with specific user proxy cohorts to examine factors that explain how well different types of conceptual models can be comprehended by model viewers [12]. A recent article by Abrahão et al. (2024) [2] provides significant insights into this area, highlighting key findings and methodologies that shape current practices. By synthesizing data from various sources, the authors present a nuanced view of how evidence-based approaches can drive innovation and efficiency in software development processes [2].

3 *The modSense* Method

In conceptual modeling education, the comprehension of UML diagrams stands as a fundamental skill [9]. Recognizing the diverse learning styles and preferences of students, a learning-centric approach becomes imperative for effective

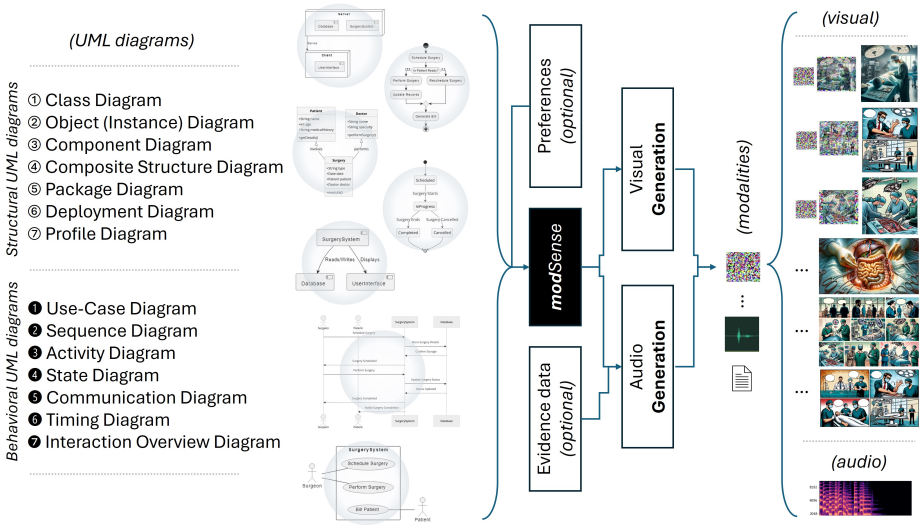

Fig. 1. ***mod****Sense* platform overview.

pedagogy. In this paper, we present the ***mod****Sense* platform that is designed to transform the interpretation of UML diagrams by placing learning at the forefront of its design philosophy. At its core, ***mod****Sense* embodies a commitment to enhancing the educational experience through innovative AI-driven technologies, catering to the individual needs and preferences of learners. By prioritizing learning outcomes and adopting a multifaceted approach that integrates image and audio generation, and multimodal data retrieval, ***mod****Sense* redefines the landscape of UML diagram interpretation, offering a dynamic and engaging learning environment for students in the field of conceptual modeling. Through the integration of advanced generative AI technologies, ***mod****Sense* provides a multifaceted approach (see Fig. 1) to enhance the understanding of UML diagrams, catering to diverse learning styles and preferences. ***mod****Sense* employs state-of-the-art Generative AI techniques to dynamically generate visual representations and audio descriptions of UML diagrams. This multimodal approach aims to enrich the learning experience by providing learners with multiple sensory inputs. Visual representations offer a clear representation of the diagram's structures, while audio descriptions provide additional context and explanation, catering to auditory learners and those with visual impairments. By presenting information in multiple modalities, ***mod****Sense* accommodates different learning preferences and enhances overall comprehension and accessibility [15,11].

3.1 Understanding UML

UML diagrams [6] are a set of graphical representations that help in visualizing, specifying, constructing, and documenting the artifacts of a software system. UML diagrams are broadly categorized into structural and behavioral diagrams. Structural diagrams focus on the static aspects of the system, portraying the physical and conceptual elements and their relationships. Common structural

diagrams include class diagrams, object diagrams, component diagrams, composite structure diagrams, deployment diagrams, and package diagrams. These diagrams help in detailing the system's architecture, such as classes and their relationships, system components, and the arrangement of software deployments. Our solution can interpret UML Structural diagrams as illustrated in Fig. 2.

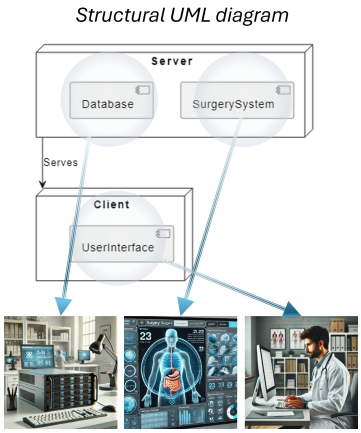

Fig. 2. Structural UML diagram mapped to generated content.

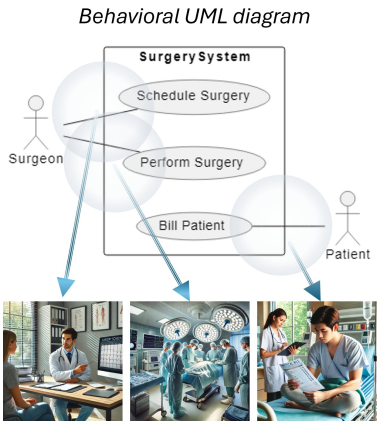

Fig. 3. Behavioral UML diagram mapped to generated content.

Behavioral diagrams, in contrast, represent the dynamic aspects, detailing how the system behaves over time and in response to events. This category includes use case diagrams, sequence diagrams, activity diagrams, state machine diagrams, communication diagrams, interaction overview diagrams, and timing diagrams. These diagrams map out the processes, workflows, and operational sequences within a software system, providing insights into the flow of data and control among different system components. Our solution can interpret UML Behavioral diagrams as illustrated in Fig. 3.

Both types of diagrams are constructed from a variety of elements that provide the building blocks for modeling software systems. We have selected 59 such elements across all 14 types of UML diagrams following the UML specification [14] (see Table 1). The elements include but are not limited to, classes, interfaces, components, use cases, actors, and various types of interaction units such as messages and transitions.

3.2 Visual Generative Capabilities

To connect theoretical knowledge and practical application in the field of conceptual modeling education, this paper introduces a novel (to the best of our knowledge) application of Generative AI to enhance the learning and understanding of UML diagrams. By taking the advanced capabilities of discrete Variational Autoencoders (dVAE), autoregressive transformers, and diffusion models [13],

Table 1. UML Key Elements for Generating ***mod****Sense* interpretations

No.	UML Diagram	Key Elements for Generating *modSense* interpretations
1	Class	Classes, Interfaces, Associations, Generalizations, Dependencies
2	Object	Objects, Links, Fields
3	Use Case	Actors, Use Cases, Relationships, System Boundary
4	Sequence	Objects, Lifelines, Messages, Activation Bars, Time Constraints
5	Collaboration	Objects, Links, Messages, Sequence Numbers
6	Activity	Activities, Transitions, Swimlanes, Forks, Joins, Start/End
7	State Machine	States, Transitions, Events, Actions, Initial and Final States
8	Component	Components, Interfaces, Ports, Connections
9	Deployment	Nodes, Artifacts, Links
10	Timing	Objects, Time Rulers, State or Value Lifelines, Events
11	Interaction	Interaction Frames, Activity Nodes, Control Flows, Decisions
12	Communication	Objects, Messages, Links, Sequence Numbers
13	Composite Struct.	Parts, Connectors, Ports, Collaborations
14	Package	Packages, Dependencies, Mergings, Access

we employ a system capable of generating dynamic visual and auditory content that contextualizes UML diagrams.

The conditioned sample generation starts with the encoding of a textual description y, which specifies the desired image attributes. This text is converted into a sequence of Byte Pair Encoding (BPE) tokens and then embedded into the model. The autoregressive transformer models the generation process where each image token x_i is sequentially predicted based on the previously generated tokens $x_{<i}$ and the input text y.

The conditional probability of generating an image given a text description is given by:

$$p(x \mid y) = \prod_{i=1}^{N} p(x_i \mid x_{<i}, y),$$

where N is the total number of image tokens to be generated. This formulation illustrates the sequential nature of generating image tokens, where the probability of each token x_i depends on its preceding tokens $x_{<i}$ and the text y.

3.3 Auditory Generative Capabilities

For generating audio, we incorporate the solution by [13] into our platform's pipeline. This is achieved through conditioned audio generation using MUSICGEN [4]. MUSICGEN is based on autoregressive modeling, where the sequence of tokens U from the quantized representation Q is modeled such that each token U_t at time t is predicted based on all previous tokens, $p_t(U_t|U_{t-1}, \ldots, U_0) = P(U_t|U_{t-1}, \ldots, U_0)$. Here, U_0 is often a special token indicating the start of the sequence, and p_t is the autoregressive probability model at time t.

Then, the Inexact Autoregressive Decomposition approach is used for the simultaneous prediction of multiple codebooks to reduce complexity and increase efficiency. If the sequence V represents tokens with K codebooks at time t, then

the probability of observing $V_{t,k}$, the token from the k-th codebook at time t, given all previous tokens is $p_{t,k}(V_{t,k}|V_{t-1},\ldots,V_0) = P(V_{t,k}|V_{t-1},\ldots,V_0)$ where $V_{t,k}$ is the token at time t from codebook k, and $p_{t,k}$ is the autoregressive model specific to codebook k.

3.4 Focus Masks

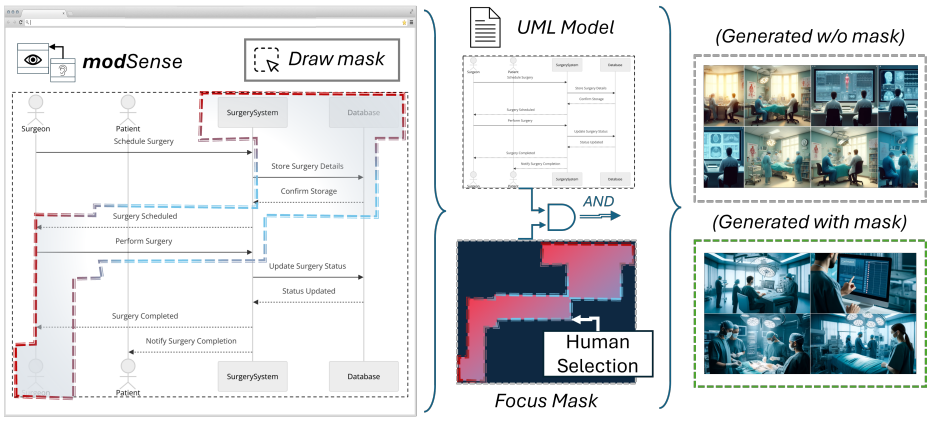

Fig. 4. Focus mask concept for human-guided selective model interpretation.

To make the generation of the content even more customized, we implemented a concept of Focus Masks (see Fig. 4). Implementing a focus mask in the context of UML diagram generation offers a highly personalized and interactive approach to learning and understanding complex systems. A focus mask is essentially a user-defined selection within a UML diagram that highlights specific parts or elements of the diagram for deeper exploration. This functionality enhances the educational experience by allowing users to concentrate on particular components of a system, simplifying the learning process by isolating complex interactions or structures. Users can apply a focus mask to any part of a UML diagram—be it a cluster of classes in a class diagram, specific interactions in a sequence diagram, or particular states in a state machine diagram.

Upon defining a focus mask, the system then generates the specific part of the UML diagram highlighted by the mask. This generation process utilizes generative AI technologies that can dynamically create or adjust the visual representation of the selected diagram portion based on user inputs and predefined modeling rules. For example, if a user selects a series of interactions in a sequence diagram, the AI generates an enhanced view of these interactions, possibly expanding on details like message sequencing, timing, and participant lifelines. This tailored generation helps users see only the relevant parts of a diagram in greater detail, making it easier to understand specific functionalities or relationships without the distraction of unrelated diagram components.

4 Empirical Study in Conceptual Modeling: Multimodal UML Education

Our current evaluation of **mod**Sense involved generating interpretations for various test domains and comparing the results. We selected test domains with intensive manual work, specifically pottery making, organic farming, and bookbinding. So far, we report positive results on diagrams where the temporal dimension is more visible. The generated visual content for these test domains is presented in Figures 5, 6, and 7, respectively, while the generated audio content is available on our GitHub repository[1]. Through manual observation, we confirmed that the visual generations with basic UML models we created are satisfactory. However, we believe that the reasons for not achieving better results in covering all UML types and better audio generation are (1) the simplicity of the UML models chosen for the test, and (2) the lack of engagement from modelers who did not invest sufficient thought and time to customize their requirements. Consequently, we are addressing these issues with an ongoing study that tracks UML creation and learning in a control group.

The on-going empirical study, **mod**Sense**4**All, aims to investigate the efficacy of using Generative AI to synthesize images for educational purposes in the context of UML conceptual modeling. Specifically, the study focuses on the educational impact of multimodal learning resources generated by AI across 14 different types of UML diagrams. Participants are recruited from undergraduate and graduate computer science courses, tasked to program a surgery-assistant system, with a balanced mix of those who have prior exposure to UML and those who do not. Participants are randomly divided into two groups: a control group that receives traditional UML instruction and an experimental group that utilizes AI-generated multimodal resources. These resources include AI-synthesized images illustrating key concepts and elements of each UML diagram type, aimed at enhancing comprehension through visual learning.

The primary domain for **mod**Sense**4**All evaluation are surgery applications, given its complex and highly technical nature. The study is conducted over a period of one university semester. The first phase involves a pre-test to assess the baseline understanding of UML concepts among participants. This is followed by a four-week instructional phase where each group receives their respective learning materials. The final phase includes a post-test identical to the pre-test to measure any changes in comprehension. Throughout the study, participants' interactions with the learning materials are documented using learning management system (LMS) analytics, capturing metrics such as time spent on each resource and frequency of access. Additionally, participants complete weekly surveys to provide qualitative feedback on their learning experiences. The assessment of the study involves a comparative analysis of pre- and post-test results between the control and experimental groups, supplemented by qualitative insights from the surveys. This mixed-methods approach aims to provide a com-

[1] `https://github.com/aleksandargavric/modsense`

prehensive understanding of the potential benefits and challenges of integrating AI-generated visual aids in UML education.

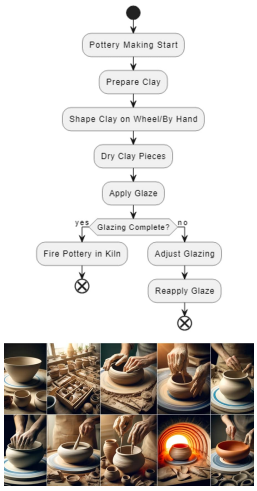

Fig. 5. Example case study in the domain of pottery making.

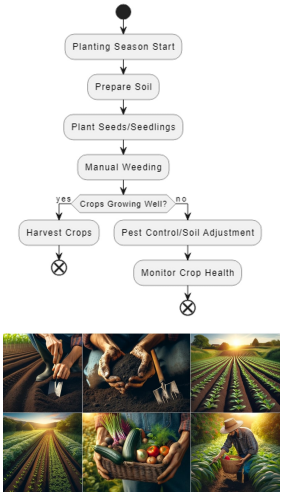

Fig. 6. Example case study in the domain of organic farming.

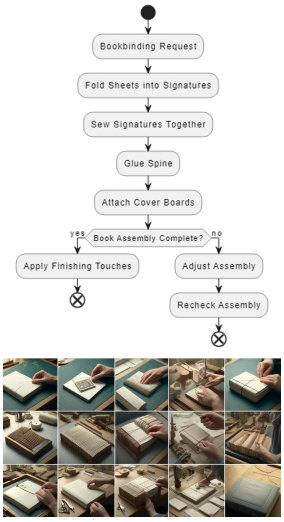

Fig. 7. Example case study in the domain of bookbinding.

5 Conclusion

This paper explored the application of AI technologies—specifically generative and conversational AI—in enhancing the teaching and understanding of Unified Modeling Language (UML) diagrams across domains. By integrating these AI capabilities, we have demonstrated a significant advancement in the educational use of UML diagrams, making them more interactive, accessible, and tailored to individual learning needs. The focus mask feature and the dynamic generation of content based on user preferences provide a customized learning experience, allowing users to focus on specific areas of a diagram relevant to their interests or challenges. Looking forward, the potential of this AI-driven approach extends beyond the initial domains tested, promising broader applications in various educational and professional training settings. By incorporating user feedback and refining AI interactions, the tool can evolve to meet educational demands and adapt to new learning environments.

References

1. Abdelnabi, E.A., Maatuk, A.M., Hagal, M.: Generating uml class diagram from natural language requirements: A survey of approaches and techniques. In: IEEE 1st International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering MI-STA. pp. 288–293 (2021)

2. Abrahão, S., Staron, M., Baldassarre, M., Horkoff, J., Penzenstadler, B., Ralph, P., Serebrenik, A.: Research highlights in evidence-based software engineering. IEEE Software **41**, 133–136 (01 2024). https://doi.org/10.1109/MS.2023.3321418
3. Baghaei, N., Mitrovic, A., Irwin, W.: Supporting collaborative learning and problem-solving in a constraint-based CSCL environment for UML class diagrams. International Journal of Computer-Supported Collaborative Learning **2**, 159–190 (2007). https://doi.org/10.1007/s11412-007-9018-0
4. Copet, J., Kreuk, F., Gat, I., Remez, T., Kant, D., Synnaeve, G., Adi, Y., Défossez, A.: Simple and controllable music generation. In: Thirty-seventh Conference on Neural Information Processing Systems (2023)
5. Farrelly, T., Baker, N.: Generative artificial intelligence: Implications and considerations for higher education practice. Education Sciences (2023)
6. Fowler, M.: UML Distilled: A Brief Guide to the Standard Object Modeling Language. Addison-Wesley Professional, 3 edn. (2003)
7. Gavric, A., Bork, D., Proper, H.: Multimodal process mining. In: 26th International Conference on Business Informatics. IEEE (2024), https://model-engineering.info/publications/papers/CBI-MultiModalProcessMining.pdf, in press
8. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative Adversarial Networks (Jun 2014), https://arxiv.org/abs/1406.2661
9. Hitz, M., Kappel, G.: UML@ Work: von der Analyse zur Realisierung (2002)
10. Lokonon, M.S., Houndji, V.R.: Automatic uml defects detection based on image of diagram. In: DeLTA. pp. 193–198 (2022)
11. Lukyanenko, R., Bork, D., Storey, V.C., Parsons, J., Pastor, O.: Inclusive conceptual modeling: Diversity, equity, involvement, and belonging in conceptual modeling (short paper). In: Companion Proceedings of the 42nd International Conference on Conceptual Modeling: ER Forum, 2023. CEUR Workshop Proceedings, vol. 3618 (2023)
12. Mendling, J., Recker, J., Reijers, H., Leopold, H.: An empirical review of the connection between model viewer characteristics and the comprehension of conceptual process models. Information Systems Frontiers pp. 1–25 (2019). https://doi.org/10.1007/S10796-017-9823-6
13. Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., Sutskever, I.: Zero-shot text-to-image generation (2021)
14. Rumbaugh, J., Jacobson, I., Booch, G.: Unified Modeling Language Reference Manual, The (2nd Edition). Pearson Higher Education (2004)
15. Sarioglu, A., Metin, H., Bork, D.: How inclusive is conceptual modeling? A systematic review of literature and tools for disability-aware conceptual modeling. In: Conceptual Modeling - 42nd International Conference, ER 2023. vol. 14320, pp. 65–83. Springer (2023). https://doi.org/10.1007/978-3-031-47262-6_4
16. Whittle, J., Jayaraman, P., Elkhodary, A., Moreira, A., Araújo, J.: MATA: A Unified Approach for Composing UML Aspect Models Based on Graph Transformation, pp. 191–237. Springer Berlin Heidelberg, Berlin, Heidelberg (2009). https://doi.org/10.1007/978-3-642-03764-1_6
17. Zastudil, C., Rogalska, M., Kapp, C., Vaughn, J.L., Macneil, S.: Generative ai in computing education: Perspectives of students and instructors. ArXiv **abs/2308.04309** (2023)