The Opportunities and Challenges of Multimodal GenAI in the Construction Industry: A Brief Review

Meng Sun[1]*, Rui Zhao[2], Fan Xue[3]

This is the authors' version of the paper:

Sun, M., Zhao, R., & Xue, F. (2024). The opportunities and challenges of multimodal GenAI in the construction industry: A brief review. *Proceedings of the 29th International Symposium on Advancement of Construction Management and Real Estate (CRIOCM2024)*, Springer, in press.

Abstract: In the recent decade, generative artificial intelligence (GenAI) has revolutionized image and text generation. Novel large language models (LLMs) promise handling of construction's complexity in unimodal text, and the application of multimodal GenAI is also promising for the construction processes for productivity, innovations, and sustainability in construction management. This study first revisits the latest research of multimodal GenAI in other fields, then summarizes potential application scenarios within the construction industry, involving multimodal elements like contracts and documents, drawings and images, schedules, the network of stakeholder communication, 3D geometry, videos, time-dynamic 4D point clouds, and building information modeling (BIM). New challenges, such as low-cost performance and lack of evaluation systems, of multimodal GenAI applications in construction are also pinpointed. Finally, future research directions are categorized into three groups, i.e., strategy and policy, technology, and scenario-guided best practice communications, to promote the further impact of multimodal GenAI in construction.

Keywords: Multimodal GenAI, LLM, construction, lifecycle

[1]* Meng Sun
Corresponding author, Department of Real Estate and Construction, The University of Hong Kong, Pokfulam, Hong Kong, SAR, China
E-mail: sunmhku01@connect.hku.hk

[2] Rui Zhao
Department of Real Estate and Construction, The University of Hong Kong, Pokfulam, Hong Kong, SAR, China

[3] Fan Xue
Department of Real Estate and Construction, The University of Hong Kong, Pokfulam, Hong Kong, SAR, China

1. Introduction

Artificial intelligence (AI) has experienced surging development in the past decade and is seen as a promising and potential research field. Beyond typically trained discriminative models based on traditional machine learning and deep learning, generative artificial intelligence (GenAI) has become mainstream research since 2022[1–3]. However, the construction industry has been less affected due to the resistance to change and unsatisfactory error rates of cutting-edge technologies in handling the complex nature of construction. Novel large language models (LLMs) promise handling of complexity in unimodal text; it thus attracted researchers in the construction industry. In practice, construction activities and interactions are multimodal by involving contracts and documents, drawings and images, schedules, the network of stakeholder communication, 3D geometry, videos, time-dynamic 4D point clouds, and building information modeling (BIM).

GenAI is a subset of deep learning using a generative model to process both labeled and unlabeled data to synthesize novel content[4,5]. Unlike prior AI methods, GenAI focuses on automatically understanding or learning the natural features. In this way, the goals are more than solving particular tasks such as numerical forecasts or internal rules, but acquiring general performance in the constructed world of training data[1,5,6]. Nowadays, many modalities are brought into the GenAI methods including text, image, audio, and video [5,7]. The processing of generative models allows either modality to be used as input or output without any other restrictions, and the convenience and intelligence of this interaction soon gained the attention of professionals and the whole society. Miikkulainen believes that the rise of GenAI represents a paradigm of AI in the making[6].

Some of the most known GenAI are as follows. Text-Text -- ChatGPT based on LLM; Text-Image -- DALL-E/Midjourney based on diffusion; Text-Video -- Sora based on diffusion. Among those models, ChatGPT's release in November 2022 marked a turning point and also led to the rapid development of LLMs and field applications[8,9]. LLMs evolved from Language Models(LMs) utilizing the transformer architecture and are generally considered to begin with the release of the GPT series[8,10]. In the early stage of the development of large models, the self-coding model represented by Bert has made rapid progress. Recently, the autoregressive branch represented by the GPT series has taken the upper hand and become the mainstream branch. In addition, the LLaMA series, Claude and Bard in this branch are also in continuous iterative development[7,11].

Researchers also identified limitations while being inspired by text-based GenAI like ChatGPT. The main issue is the limited unimodality functions of these GenAI's Natural Language Processing (NLP). Thus, researchers are motivated to start working on integrated multimodal GenAI, which is a subset of GenAI and often related closely to LLMs (**Fig. 1**)[12–15]. State-of-the-art research identified some directions for multimodal GenAI to expand and strengthen the existing LLMs[16,17]. For example, GPT-4o realized seamless human-computer interaction[18,19], with emphasis on complex unimodal data types, including video, 3D, and point clouds(PCD). As a result, GPT-4o conducts high-quality fusion based on textual information[20–23]. Some researchers are trying to construct an empowered agent combined with digital twins to realize Artificial General Intelligence(AGI)[24,25].

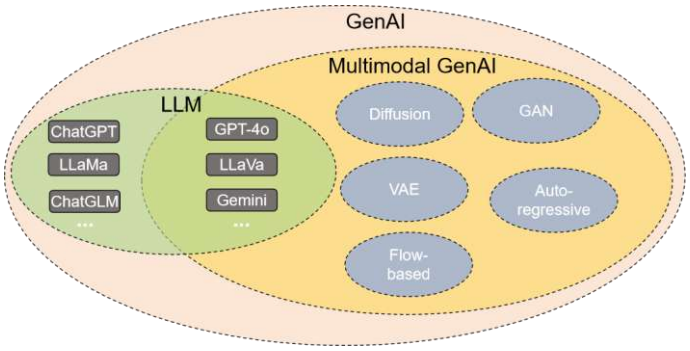

Fig. 1 The relationship among GenAI, LLM, and multimodal GenAI

GenAI, so far, represented by LLM has shown amazing capability and has been integrated into many industries such as finance, consulting, medicine, and other fields[1]. Relevant personnel in the construction industry also continuously launch research results and landing products related to GenAI[4,9,26–28]. Applications of LLMs and Diffusion models in architectural design demonstrate the huge potential in design creativity, enhancing efficiency and safety, regulatory compliance, and uniforming project requirements [29–38]. Digital transformation in the construction industry also provides an available data base for construction site risk assessment[39,40], text-based BIM retrieval[28,41–44], schedule optimization[45,46], construction robotics[47], and so on. The emergence of LLMs has also further advanced research that was difficult to automate, including scene reconstruction[48], object and space understanding[26], material and waste management[27], and communication-aided service[49,50]. However, it should be noted that the above studies are mostly based on unimodal data, while there are few studies on the application of the state-of-the-art multimodal GenAI in the construction industry, and there is a lack of relevant summary and induction[51]. Consequently, this review aims to summarize application scenarios in existing studies, reveal potential challenges, and look forward to promising opportunities for the latest multimodal GenAI in the construction industry.

2. Application Scenario of Multimodal GenAI in Construction

2.1. Overview

GenAI-based studies, especially those related to LLMs, have recently sparked a wave of research in the construction industry. As shown in **Table 1**, some reviews have summarized important application scenarios and related modalities of Input-Output of GenAI in all phases of the construction industry.

Table 1 GenAI applications in the construction lifecycle[4,9,27]

Phase	Application	Modality of Input-Output	Ability
Feasibility	Expert guidance for design and construction	text-text	B
	Procurement decision support	text-text	B
	Project brief/PID	text-text	A
	Project planning	text-text, text-task	E
	Feasibility report	text-text	A
	Visual representations of data& prototyping	text-X	D
	Creation of contracts and agreements	text-text	B

Design	Design concept	text-text, text-image, text-task	B, F
	Design specification (design requirements)	3D-text, image-text, PCD-text	B, F
	Regulatory compliance	3D-text, image-text, PCD-text	B
	Optimizing material selection	3D-text, image-text, PCD-text	F
	Quantity take-off and costing	text-text, text-task	E
	Energy efficient analysis	text-text, text-task	B
Procurement	The material delivery schedule	text-text, text-task, text-3D	E
	A request generation for a quotation	text-text	B
	Identification of optimal suppliers	text-text	B
	Streamlining subcontractor bidding&selection	text-text, text-task	E
	Automated inventory management	text-text	B
Construction	Scheduling and logistics	text-text, text-task	E
	Documentation	text-text	A, B
	Regulatory compliance	text-text	B
	Risk management	text-text, text-task	B
	Monitoring and reporting	image-text, text-text, text-task	A, B
	Resources management	text-text, text-task	E
	Change order and quality management	text-text, text-task	E
	Claim and dispute resolution	text-text	A, B
	Safety management	image-text, text-text, 3D-task, PCD-task	B
	Budgeting(cost estimation)	text-text, text-task	E
	Training	text-text, text-image, text-video	B
	Human-robot interaction	text-task, text-audio, audio-task	B, E
Operation & Maintenance	Occupation communication	text-text	B
	Creating a work order from logs	text-text	A
	Incident resolution	text-text, image-text	B
	Emergency support	text-text	B
	Predictive maintenance	text-text, text-task	C
	Energy management	text-text, text-task	E
	Life cycle management of asset	text-text, text-task	B, C
	Regulatory compliance	text-text	B
	Waste management	text-text, image-text	B
	Space optimization	text-task, X-text	B, E
	Project marketing	text-text, text-image, text-video, image-text	A, D
	Sustainability	text-text	A, B
Demolition	Demolition protocol	text-text	B
	Waste management	text-text, image-text	B
	Redevelopment plan	X-text	B, E
	Regulatory compliance	text-text	B
	Environmental impact assessment	text-text	B
	Structural issues	text-text, X-text, image-text, 3D-text	B, E

Risk assessment	text-text	B
Material recovery	text-text, image-text	B

A: Summary; B: Analysis C: Prediction; D: Visualization; E: Calculation-Optimization; F: Aided-design

Six phases are divided from the entire construction lifecycle, including feasibility, design, procurement, construction, operation & maintenance, and demolition. In these phases, potential applications of multimodal GenAI are summarized as well as required abilities, as shown in **Table 1**. These required abilities are used to meet the needs of different tasks, and there are various degrees of difficulty in implementing them.

(1) *Summary*: LLMs extract and summarize the key point from complex textual and image data, which is the basic NLU & NLG task. Based on the summary ability, complex construction data, lengthy reports, and unstructured textual information can be efficiently turned into required structured templates and concise conclusions, which are applied in the whole lifecycle. This ability is highly text-dependent, so due to the emergence of large models, it can be well applied to relevant scenarios.

(2) *Analysis*: GenAI aims to output analyzed results under typical constraint conditions for pointed tasks. In the project, GenAI is expected to output some text content (such as protocols, and contracts) according to the input information, or provide decision support for specific processes (material classification, risk analysis, etc.), while complying with the requirements of norms and regulations. The analysis is text-based, but since the input information comes from the whole process stage of the project and has the characteristics of multimodal, the ability is based on both text and multimodal fusion.

(3) *Prediction:* Make predictions and judgments about the future by learning historical data. The typical applications include building cycle prediction, maintenance, and risk forecast. These predictions have been studied for a long time by using machine learning methods and statistics, but lack general and uniform models. The fusion of the previous approach and GenAI allows for better use of time series data for general-purpose forecasting.

(4) *Visualization*: Different from the above text-dependent abilities highly based on LLMs, visualization focuses on the generation of modalities including image, video, and 3D, based on other generative models[20,27]. It is efficient to present architectural data, design concepts, and analysis results in a visual form. Considering the appearance of Sora and GPT-4o, the application of visualization ability is promising though some challenges like hallucination exist.

(5) *Calculation-Optimization*: Perform complex logical calculations and optimization to meet specific performance criteria and requirements while following basic physical and mathematical laws. The requirements of engineering itself, as well as ESG evaluation, are so time-consuming and necessary to allow complicated calculations to exist at all stages of the life cycle in the construction industry. GenAI is expected to present potentials in engineering calculation, cost calculation, energy efficiency optimization, and so on. Although Calculation-Optimization is based on text representation, it is based on the complicated logic of physics and mathematics, which goes beyond the traditional realm of natural language processing (NLP). However, due to the black box characteristics of the large model, the Calculation-Optimization still has obvious shortcomings in the logical structure, and there is still a long way from practical application.

(6) *Aided-design:* It is designed to generate multiple modalities to meet design needs. The ability is mainly applied in the feasibility and design phase to generate floorplans, construction details, structural designs, and so on. It is apparent that the aided-design is multimodal-dependent and that target design results can be achieved by inputting different kinds of data. Though the ability can efficiently generate various designs, promote the generation of creative designs, and enhance design and decision-making efficiency, it is constrained by the technical limitations of GenAI models[30].

The construction industry currently demands multimodal abilities, yet often relies on text. Depending on the degree of reliance on text, multimodal GenAI can be categorized into two types: text-based LLMs and general multimodal fusion (**Fig. 2**). Text-based LLMs strongly depend on dealing with textual data as an information original and require it to be either an input or output for professionals. On the other hand, general multimodal fusion only uses text as the interaction modality. For instance, the summary represents a text-based fusion with tasks, as it relies on textual input and produces textual output. In contrast, the visualization exemplifies general multimodal fusion, as it incorporates text alongside other forms of interaction, such as graphics and images, without solely relying on textual information.

Text-based LLMs are currently rapidly evolving, with direct applications in the construction industry due to their convenience and efficiency. On the other hand, general multimodal fusion GenAI, despite being more complex, represents the future direction of development[14]. It even aims directly towards the concept of embodied agents. While this approach may present more challenges, it holds immense potential and could revolutionize AI applications in various industries.

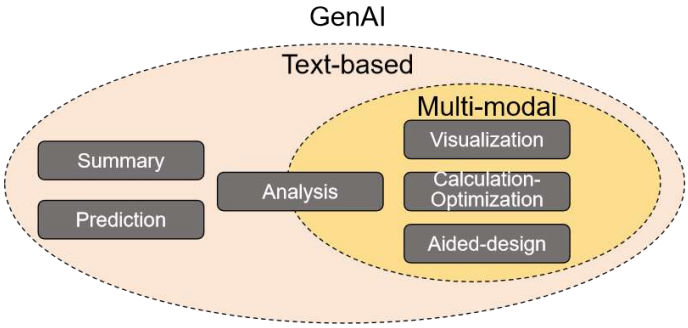

Fig. 2 The ability classification of multimodal GenAI

2.2. Text-based LLMs Application

Textual data is one of the most fundamental information originals in the construction industry, making the implementation of text-based GenAI essential across various stages of the construction lifecycle. The rise of GenAI, as represented by LLMs, is propelling the field of NLP from Natural Language Understanding (NLU) towards Natural Language Generation (NLG). This progression has made text interactions more streamlined and efficient. Consequently, professionals in the construction industry can capitalize on the advantages of GenAI throughout the entire project lifecycle.

Different from past text-text conversational AI, text-based LLMs, thanks to the iterations of models, allow other modalities as the direct input or output. As shown in **Fig. 2**, only the summary, prediction, and part of analysis are classified as text-based LLMs. In **Table 1**, 40% of the total 50

applications in the construction lifecycle are text-text, which highly depend on LLMs. Due to the existing LLMs having a good performance on NLP, they can be well qualified for these text tasks in the construction industry.

2.3. General Multimodal Fusion GenAI Application

Multimodality allows AI to bypass the intermediate representation of humans and interact directly with the world. Human natural language text has information extraction, loss, redundancy, and even errors. In a real multimodal GenAI, the information original can be transferred directly by these multiple modalities including image, video, audio, point cloud, 3D, and task instead of secondary expression through text. The general multimodal fusion GenAI tries to realize the seamless transitions between any modalities of information (X-X)[14]. For multimodal GenAIs, encoding and decoding of modalities are directly conducted on the input side and output side, and the existence of textual content is only for understanding and interaction. This means that multimodal GenAI can be evolved from text-based LLM packaging, but due to circumventing the necessity for textual mediation, information loss is reduced, thereby enhancing precision and efficiency. Except for the LLM, there are five major types of GenAI models, named for GAN, VAE, Autoregressive, Diffusion, and Flow-based[4,52].

In the construction industry, multiple modalities exist in all aspects and have the potential for GenAI applications, and 60% of application scenarios rely on multimodal (**Table 1**). Also, according to Accenture's research report, only about 20% of task time in construction engineering can be reduced by using text-based LLMs, while about 30% of the time spent on non-text tasks can be enhanced by the multimodal GenAI[53]. Multimodal GenAI serves as an interactive medium that allows for more intuitive and dynamic communication between designers, engineers, and customers. During the construction life cycle, it can help from the feasibility to the demolition stage, including decision support (text-task) and design drawings (text-image/3D). In addition, multimodal GenAI has a promising prospect for the monitoring of sites (image/3D/PCD-text), as it serves as an end-to-end tool to deal with complicated “on-site” affairs[14]. However, due to the limitation of the layout of sensors and the current frontier GenAI technology, the existing research is not mature. Cost issues and industry acceptance are also reasons why multimodal GenAI has not yet been fully implemented in the construction industry, which will be elaborated in the next section.

3. Challenges

It also brings new challenges in applying new multimodal GenAI to the construction industry. It is inevitable due to the complex nature of work in the industry and the limitations of the GenAI technologies. The construction industry's entrenched procedural standards, despite facilitating stability, pose substantial challenges to the integration of new technologies due to its historically low industrialization. While general drawbacks and challenges associated with multimodal GenAI have been extensively documented in prior research, it is crucial to also consider the industry-specific issues that arise within the construction context. The unique challenges in the construction industry should be addressed to clear the way for future applications[4,15]:

- GenAI cannot be the subject of responsibility, as the architect responsibility system is an internationally accepted management mode in the field of construction. Therefore, due to the

perspective of regulations and ethics, GenAI intelligence currently acts as an auxiliary tool for professionals but cannot replace the work link, which undoubtedly limits the application of GenAI. In addition, the black-box nature of GenAI has also led experts to question whether it can handle the corresponding job responsibilities.

- Although the construction industry contributes a lot to the economy, the industry remains under-penetrated by digitalization. This means that the cost-performance of developing multimodal GenAI models is very low, considering training a model is extremely expensive. Current industry-specific models thereby have not emerged on a large scale. It is significant to find a multimodal GenAI development method that is suitable for the industry and related enterprises.

- At present, there are many kinds of evaluation for GenAI, especially LLM, but there is a lack of evaluation systems and evaluation sets in the field of construction[10,54]. Evaluation should be treated as an essential discipline to better assist the industry application.

- The efficacy of GenAI in construction hinges on high-quality, annotated data. However, it is difficult to assess the quality and quantity of construction-related data due to the unavailability of cutting-edge GenAI training data. In addition, dealing with specialized construction data requires a highly technical profession, which makes general-purpose multimodal GenAI models less effective in construction applications.

- The construction industry, unlike others, is heavily reliant on multiple modalities like 3D models and point clouds for spatial understanding. However, the question remains whether GenAI possesses the spatial imagination capability necessary to interpret and generate complex architectural designs effectively.

- Construction management often requires strong planning and decision-making ability (text-task), such as project planning, scheduling, and logistics. However, state-of-the-art LLMs can hardly complete any planning tasks under multi-constraint conditions[55].

4. Promising Opportunities of Multimodal GenAI in Construction

Applications of multimodal GenAI in the construction industry highlight promising research opportunities. Such opportunities could be structured into 3 groups: 1) strategy and policy at the government/regulator side, 2) technology development & adoption for both university & industry and 3) best practice communications for industry & client.

(1) Strategy and policy at the government/regulator side:

- The application of multimodal GenAI is bound to change the traditional construction pattern, thus new industry standards and guidelines are needed to ensure the consistency and reliability of technology use, especially considering the resistance to change and unsatisfactory error rates of cutting-edge technologies in handling the complex nature of construction.

- Regulatory authorities, endowed with extensive construction data and case studies, can leverage multimodal GenAI to organize and analyze engineering projects, thereby assisting in risk identification, fraud detection, and also the recognition of exemplary projects.

(2) Technology development & adoption for both university & industry:

- Understanding and semantic generation of point clouds is a promising direction due to few GenAI studies on point clouds in the construction field. Huang et al. encapsulated the LLM, implementing a multimodal encoding and decoding operation, thereby enabling the capability to interpret and comprehend point cloud data[26]. Moreover, the sparsity of point clouds leads to the difficulty of segmentation and semantic learning, thus the path of PCD-voxel-image may be practicable.

- Nature creates a virtuous cycle through "spatial intelligence", so being able to understand and describe 3D space would be a big step forward for the multimodal GenAI[17,48,56]. The breakthrough in understanding spatial-related modalities (i.e., 3D, PCD) will provide a new direction for the development of architectural space.

- Some new studies are trying to improve the text-task capability based on the rational method and gameplay-style value iteration. Integrating the knowledge of construction management will enhance the efficiency of calculation-optimization tasks throughout the entire construction lifecycle.

- The current multimodal GenAI is model-driven to simulate "observation" and "perception" of the environment. Integrating sensor data from construction sites to enhance the perceptual abilities of these AI systems represents a promising new direction for research and application, potentially leading to more context-aware in the construction industry.

(3) Best practice communications for industry & client:

- Due to the complexity of construction links and sites, various multimodal data are complicated and difficult to deal with. Multimodal GenAI, serving as middleware, can unify the processing of diverse data types, store them in databases, and retrieve them as needed[22]. This capability has the potential to significantly enhance engineering and communication efficiency.

- Though BIM plays a crucial role in the construction industry, BIM information exchange still has many problems due to complicated structures and identifiers of the IFC format[57,58]. IFC format can added to GenAI models as a BIM modality to make it understand and better serve BIM information exchange.

5. Conclusions

The rise of multimodal GenAI is rapidly impacting all walks of life, and it is necessary to think about how to use this technology in the construction industry. This paper summarizes 50 application scenarios in all construction phases, including feasibility, design, procurement, construction, operation & maintenance, and demolition. These applications need to adopt different abilities of multimodal GenAI models, such as Summary, Analysis, Prediction, Visualization, Calculation-Optimization, and Aided-design. The application of these abilities to the construction industry faces different challenges. On the one hand, the challenges reflect the technical limitations of multimodal GenAI, and on the other hand, challenges are well attributed to the complex nature of the tasks in the construction industry. Once the challenges are resolved, partially or wholly, significant improvements and impact are expected.

In summary, there exist some promising opportunities for multimodal GenAI in the

construction industry, spanning the groups of strategy and policy, technology, and scenario-guided best practice communications. The key is to meet or combine domain knowledge and the breakthrough of technology in its own right. We are optimistic that multimodal GenAI is a convincing technology and should have far-reaching significance in the digital transformation of the construction industry, thus better promoting the development of large-scale digital twins and smart cities.

Acknowledgment

The work presented in this paper was supported by the Hong Kong Research Grants Council (No. C7080-22GF) and The University of Hong Kong (CCMC-CCST6003).

References

[1] García-Peñalvo, F., & Vázquez-Ingelmo, A. (2023). *What Do We Mean by GenAI? A Systematic Mapping of The Evolution, Trends, and Techniques Involved in Generative AI*. https://doi.org/10.9781/ijimai.2023.07.006

[2] Jebara, T. (2004). Generative Versus Discriminative Learning. In T. Jebara (Ed.), *Machine Learning: Discriminative and Generative* (pp. 17–60). Springer US. https://doi.org/10.1007/978-1-4419-9011-2_2

[3] van der Zant, T., Kouw, M., & Schomaker, L. (2013). Generative Artificial Intelligence: Philosophy and Theory of Artificial Intelligence. In V. C. Mueller (Ed.), *Philosophy and Theory of Artificial Intelligence* (Vol. 5, pp. 107–120). Springer. https://doi.org/10.1007/978-3-642-31674-6_8

[4] Ghimire, P., Kim, K., & Acharya, M. (2024). Opportunities and Challenges of Generative AI in Construction Industry: Focusing on Adoption of Text-Based Models. *Buildings*, *14*(1), 220. https://doi.org/10.3390/buildings14010220

[5] Karpathy, A., Abbeel, P., Brockman, G., Chen, P., Cheung, V., Duan, Y., Goodfellow, I., Kingma, D., Ho, J., Houthooft, R., Salimans, T., Schulman, J., Sutskever, I., & Zaremba, W. (2016, June). *Generative models*. Generative Models. https://openai.com/index/generative-models/

[6] Miikkulainen, R. (2024). Generative AI: An AI paradigm shift in the making? *AI Magazine*, *45*(1), 165–167. https://doi.org/10.1002/aaai.12155

[7] Gozalo-Brizuela, R., & Garrido-Merchan, E. C. (2023). *ChatGPT is not all you need. A State of the Art Review of large Generative AI models* (arXiv:2301.04655). arXiv. https://doi.org/10.48550/arXiv.2301.04655

[8] van Dis, E. A. M., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. *Nature*, *614*(7947), 224—226. https://doi.org/10.1038/d41586-023-00288-7

[9] Yu, Z., & Gong, Y. (2024). ChatGPT, AI-generated content, and engineering management. *Frontiers of Engineering Management*, *11*(1), 159–166. https://doi.org/10.1007/s42524-023-0289-6

[10] Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., & Xie, X. (2024). A Survey on Evaluation of Large Language Models. *ACM Transactions on Intelligent Systems and Technology*, *15*(3), 39:1-39:45. https://doi.org/10.1145/3641289

[11] Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., Zhong, S., Yin, B., & Hu, X. (2024). Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond.

ACM Trans. Knowl. Discov. Data, *18*(6). https://doi.org/10.1145/3649506
[12] Akkus, C., Chu, L., Djakovic, V., Jauch-Walser, S., Koch, P., Loss, G., Marquardt, C., Moldovan, M., Sauter, N., Schneider, M., Schulte, R., Urbanczyk, K., Goschenhofer, J., Heumann, C., Hvingelby, R., Schalk, D., & Aßenmacher, M. (2023). *Multimodal Deep Learning* (arXiv:2301.04856). arXiv. https://doi.org/10.48550/arXiv.2301.04856
[13] MIT Technology Review Insights. (2024, May 8). *Multimodal: AI's new frontier*. MIT Technology Review. https://www.technologyreview.com/2024/05/08/1092009/multimodal-ais-new-frontier/
[14] Wu, S., Fei, H., Qu, L., Ji, W., & Chua, T.-S. (2023). *NExT-GPT: Any-to-Any Multimodal LLM* (arXiv:2309.05519). arXiv. https://doi.org/10.48550/arXiv.2309.05519
[15] Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T., & Chen, E. (2024). *A Survey on Multimodal Large Language Models* (arXiv:2306.13549). arXiv. https://doi.org/10.48550/arXiv.2306.13549
[16] Wang, Z., Zhang, Z., Zhang, H., Liu, L., Huang, R., Cheng, X., Zhao, H., & Zhao, Z. (2024). *OmniBind: Large-scale Omni Multimodal Representation via Binding Spaces* (arXiv:2407.11895). arXiv. https://doi.org/10.48550/arXiv.2407.11895
[17] Shang, Y., Chen, J., Fan, H., Ding, J., Feng, J., & Li, Y. (2024). *UrbanWorld: An Urban World Model for 3D City Generation* (arXiv:2407.11965). arXiv. https://doi.org/10.48550/arXiv.2407.11965
[18] OpenAI. (2024, May 13). *Hello GPT-4o*. https://openai.com/index/hello-gpt-4o/
[19] Fei, N., Lu, Z., Gao, Y., Yang, G., Huo, Y., Wen, J., Lu, H., Song, R., Gao, X., Xiang, T., Sun, H., & Wen, J.-R. (2022). Towards artificial general intelligence via a multimodal foundation model. *Nature Communications*, *13*(1), 3094. https://doi.org/10.1038/s41467-022-30761-2
[20] Chan, E. R., Nagano, K., Chan, M. A., Bergman, A. W., Park, J. J., Levy, A., Aittala, M., De Mello, S., Karras, T., & Wetzstein, G. (2023). *Generative Novel View Synthesis with 3D-Aware Diffusion Models*. arXiv. http://arxiv.org/abs/2304.02602
[21] OpenAI. (2024, February 15). *Sora: Creating video from text*. https://openai.com/index/sora/
[22] Shao, R., Yang, C., Li, Q., Zhu, Q., Zhang, Y., Li, Y., Liu, Y., Tang, Y., Liu, D., Yang, S., & Li, H. (2024). *AllSpark: A Multimodal Spatio-Temporal General Intelligence Model with Thirteen Modalities* (arXiv:2401.00546). arXiv. https://doi.org/10.48550/arXiv.2401.00546
[23] Wu, Y., Shi, L., Cai, J., Yuan, W., Qiu, L., Dong, Z., Bo, L., Cui, S., & Han, X. (2024). *IPoD: Implicit Field Learning with Point Diffusion for Generalizable 3D Object Reconstruction from Single RGB-D Images* (arXiv:2404.00269). arXiv. https://doi.org/10.48550/arXiv.2404.00269
[24] Liu, Y., Yang, H., Si, X., Liu, L., Li, Z., Zhang, Y., Liu, Y., & Yi, L. (2024). *TACO: Benchmarking Generalizable Bimanual Tool-ACtion-Object Understanding* (arXiv:2401.08399). arXiv. https://doi.org/10.48550/arXiv.2401.08399
[25] Wang, T., Mao, X., Zhu, C., Xu, R., Lyu, R., Li, P., Chen, X., Zhang, W., Chen, K., Xue, T., Liu, X., Lu, C., Lin, D., & Pang, J. (2023). *EmbodiedScan: A Holistic Multi-Modal 3D Perception Suite Towards Embodied AI* (arXiv:2312.16170). arXiv. https://doi.org/10.48550/arXiv.2312.16170
[26] Huang, K.-C., Li, X., Qi, L., Yan, S., & Yang, M.-H. (2024). *Reason3D: Searching and Reasoning 3D Segmentation via Large Language Model* (arXiv:2405.17427). arXiv. https://doi.org/10.48550/arXiv.2405.17427
[27] Saka, A., Taiwo, R., Saka, N., Salami, B. A., Ajayi, S., Akande, K., & Kazemi, H. (2024). GPT models in construction industry: Opportunities, limitations, and a use case

validation. *Developments in the Built Environment*, *17*, 100300. https://doi.org/10.1016/j.dibe.2023.100300

[28] Zheng, J., & Fischer, M. (2023). *BIM-GPT: A Prompt-Based Virtual Assistant Framework for BIM Information Retrieval* (arXiv:2304.09333). arXiv. https://doi.org/10.48550/arXiv.2304.09333

[29] Rane, N., Choudhary, S., & Rane, J. (2023). Integrating ChatGPT, Bard, and leading-edge generative artificial intelligence in architectural design and engineering: Applications, framework, and challenges. *International Journal of Architecture and Planning*. https://www.semanticscholar.org/paper/Integrating-ChatGPT%2C-Bard%2C-and-leading-edge-in-and-Rane-Choudhary/e97aef909cdfa5ab968d43ca4f952070c25cf2cc

[30] Chen, J., Shao, Z., & Hu, B. (2023). Generating Interior Design from Text: A New Diffusion Model-Based Method for Efficient Creative Design. *Buildings*, *13*(7), Article 7. https://doi.org/10.3390/buildings13071861

[31] Tan, L., & Luhrs, M. (2024). Using Generative AI Midjourney to Enhance Divergent and Convergent Thinking in an Architect's Creative Design Process. *DESIGN JOURNAL*, *27*(4), 677–699. https://doi.org/10.1080/14606925.2024.2353479

[32] Wu, A. N., Stouffs, R., & Biljecki, F. (2022). Generative Adversarial Networks in the built environment: A comprehensive review of the application of GANs across data types and scales. *Building and Environment*, *223*, 109477. https://doi.org/10.1016/j.buildenv.2022.109477

[33] Pouliou, P., Horvath, A.-S., & Palamas, G. (2023). Speculative hybrids: Investigating the generation of conceptual architectural forms through the use of 3D generative adversarial networks. *INTERNATIONAL JOURNAL OF ARCHITECTURAL COMPUTING*, *21*(2), 315–336. https://doi.org/10.1177/14780771231168229

[34] Bucher, M. J. J., Kraus, M. A., Rust, R., & Tang, S. (2023). Performance-Based Generative Design for Parametric Modeling of Engineering Structures Using Deep Conditional Generative Models. *AUTOMATION IN CONSTRUCTION*, *156*, 105128. https://doi.org/10.1016/j.autcon.2023.105128

[35] Han, D., Zhao, W., Yin, H., Qu, M., Zhu, J., Ma, F., Ying, Y., & Pan, A. (2024). Large language models driven BIM-based DfMA method for free-form prefabricated buildings: Framework and a usefulness case study. *Journal of Asian Architecture and Building Engineering*, *0*(0), 1–18. https://doi.org/10.1080/13467581.2024.2329351

[36] Forth, K., & Borrmann, A. (2024). Semantic enrichment for BIM-based building energy performance simulations using semantic textual similarity and fine-tuning multilingual LLM. *Journal of Building Engineering*, *95*, 110312. https://doi.org/10.1016/j.jobe.2024.110312

[37] Jang, S., Lee, G., Oh, J., Lee, J., & Koo, B. (2024). Automated detailing of exterior walls using NADIA: Natural-language-based architectural detailing through interaction with AI. *Advanced Engineering Informatics*, *61*, 102532. https://doi.org/10.1016/j.aei.2024.102532

[38] Du, C., Esser, S., Nousias, S., & Borrmann, A. (2024). *Text2BIM: Generating Building Models Using a Large Language Model-based Multi-Agent Framework* (arXiv:2408.08054). arXiv. https://doi.org/10.48550/arXiv.2408.08054

[39] Kamari, M., & Ham, Y. (2022). AI-based risk assessment for construction site disaster preparedness through deep learning-based digital twinning. *AUTOMATION IN CONSTRUCTION*, *134*, 104091. https://doi.org/10.1016/j.autcon.2021.104091

[40] Uddin, S. J., Albert, A., Ovid, A., & Alsharef, A. (2023). Leveraging ChatGPT to Aid Construction Hazard Recognition and Support Safety Education and Training.

Sustainability, *null*, null. https://doi.org/10.3390/su15097121

[41] Chung, S., Moon, S., Kim, J., Kim, J., Lim, S.-H., & Chi, S. (2023). Comparing natural language processing (NLP) applications in construction and computer science using preferred reporting items for systematic reviews (PRISMA). *Automation in Construction*, *null*, null. https://doi.org/10.1016/j.autcon.2023.105020

[42] Wang, H., Meng, X., & Zhu, X. (2022). Improving knowledge capture and retrieval in the BIM environment: Combining case-based reasoning and natural language processing. *Automation in Construction*, *139*, 104317. https://doi.org/10.1016/j.autcon.2022.104317

[43] Wang, N., Issa, R. R. A., & Anumba, C. J. (2022). *Query Answering System for Building Information Modeling Using BERT NN Algorithm and NLG*. 425–432. https://doi.org/10.1061/9780784483893.053

[44] Lin, T.-H., Huang, Y.-H., & Putranto, A. (2022). Intelligent question and answer system for building information modeling and artificial intelligence of things based on the bidirectional encoder representations from transformers model. *Automation in Construction*, *142*, 104483. https://doi.org/10.1016/j.autcon.2022.104483

[45] Prieto, S. A., Mengiste, E. T., & García de Soto, B. (2023). Investigating the Use of ChatGPT for the Scheduling of Construction Projects. *Buildings*, *13*(4), Article 4. https://doi.org/10.3390/buildings13040857

[46] Amer, F., Jung, Y., & Golparvar-Fard, M. (2021). Transformer machine learning language model for auto-alignment of long-term and short-term plans in construction. *Automation in Construction*, *132*, 103929. https://doi.org/10.1016/j.autcon.2021.103929

[47] You, H., Ye, Y., Zhou, T., Zhu, Q., & Du, J. (2023). *Robot-Enabled Construction Assembly with Automated Sequence Planning based on ChatGPT: RoboGPT*. https://doi.org/10.48550/arXiv.2304.11018

[48] Avetisyan, A., Xie, C., Howard-Jenkins, H., Yang, T.-Y., Aroudj, S., Patra, S., Zhang, F., Frost, D., Holland, L., Orme, C., Engel, J., Miller, E., Newcombe, R., & Balntas, V. (2024). *SceneScript: Reconstructing Scenes With An Autoregressive Structured Language Model* (arXiv:2403.13064). arXiv. https://doi.org/10.48550/arXiv.2403.13064

[49] Saka, A., Oyedele, L. O., Àkànbí, L., Ganiyu, S., Chan, D. W. M., & Bello, S. (2023). Conversational artificial intelligence in the AEC industry: A review of present status, challenges and opportunities. *Adv. Eng. Informatics*, *55*, 101869. https://doi.org/10.1016/j.aei.2022.101869

[50] Abioye, S. O., Oyedele, L. O., Akanbi, L., Ajayi, A., Davila Delgado, J. M., Bilal, M., Akinade, O. O., & Ahmed, A. (2021). Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges. *Journal of Building Engineering*, *44*, 103299. https://doi.org/10.1016/j.jobe.2021.103299

[51] Miret, S., & Krishnan, N. M. A. (2024). *Are LLMs Ready for Real-World Materials Discovery?* (arXiv:2402.05200). arXiv. https://doi.org/10.48550/arXiv.2402.05200

[52] Liu, Y., Yang, Z., Yu, Z., Liu, Z., Liu, D., Lin, H., Li, M., Ma, S., Avdeev, M., & Shi, S. (2023). Generative artificial intelligence and its applications in materials science: Current situation and future perspectives. *Journal of Materiomics*, *9*(4), 798–816. https://doi.org/10.1016/j.jmat.2023.05.001

[53] Daugherty, P., Ghosh, B., Narain, K., Guan, L., & Wilson, J. (2023). *Gen AI LLM - A new era of generative AI for everyone*. Accenture. https://www.accenture.com/content/dam/accenture/final/accenture-com/document/Accenture-A-New-Era-of-Generative-AI-for-Everyone.pdf

[54] Guo, Z., Jin, R., Liu, C., Huang, Y., Shi, D., Supryadi, Yu, L., Liu, Y., Li, J., Xiong, B., & Xiong, D. (2023). *Evaluating Large Language Models: A Comprehensive Survey* (arXiv:2310.19736). arXiv. https://doi.org/10.48550/arXiv.2310.19736

[55] Xie, J., Zhang, K., Chen, J., Zhu, T., Lou, R., Tian, Y., Xiao, Y., & Su, Y. (2024). *TravelPlanner: A Benchmark for Real-World Planning with Language Agents* (arXiv:2402.01622). arXiv. http://arxiv.org/abs/2402.01622

[56] Li, F.-F. (2024, April). *Fei-Fei Li: With spatial intelligence, AI will understand the real world | TED Talk*. https://www.ted.com/talks/fei_fei_li_with_spatial_intelligence_ai_will_understand_the_real_world

[57] Lou, J., Lu, W., & Xue, F. (2021). A Review of BIM Data Exchange Method in BIM Collaboration. In X. Lu, Z. Zhang, W. Lu, & Y. Peng (Eds.), *Proceedings of the 25th International Symposium on Advancement of Construction Management and Real Estate* (pp. 1329–1338). Springer. https://doi.org/10.1007/978-981-16-3587-8_90

[58] Jang, S., & Lee, G. (2024). *Interactive Design by Integrating a Large Pre-Trained Language Model and Building Information Modeling*. 291–299. https://doi.org/10.1061/9780784485231.035