Building Foundation Models - Potentials, Challenges and Research Directions for using LLM and LVM in AEC


Joern Ploennigs
University of Rostock


Markus Berger
University of Rostock


Thomas Wortmann
University of Stuttgart

2025 European Conference on Computing in Construction
CIB W78 Conference on IT in Construction
Porto, Portugal
July 14-17, 2025

BUILDING FOUNDATION MODELS - POTENTIALS, CHALLENGES AND RESEARCH DIRECTIONS FOR USING LLM AND LVM IN AEC

Joern Ploennigs[1], Markus Berger[1], Thomas Wortmann[2], Jakob Kirchner[3], Jakob Beetz[4], Alina Roitberg[2], Karsten Menzel[5], and Björn Ommer[6]
[1]University of Rostock, Germany
[2]University of Stuttgart, Germany
[3]University of Kassel, Germany
[4]RWTH Aachen, Germany
[5]Technical University Dresden, Germany
[6]Ludwig-Maximilians-University Munich, Germany

Abstract

Foundation Models like Large Language Models (LLM) and Large Vision Models (LVM) show huge potential in automating processes in building design, construction, and operation. However, there are concerns about their capability to create coherent, usable, sustainable, and safe architecture. This paper analyzes the challenges and potentials of LLMs and LVMs in the architecture, engineering, and construction (AEC) sector by executing a systematic literature review of research papers. We discuss a collection of relevant AEC modalities like Computer Aided Design (CAD) models, semantic graphs, and time series and how they could integrate into the current landscape of models. From the analysis we derive research directions toward the development of domain-specific "Building Foundation Models" which are in the fields of reliable & consistent foundation models, explainable tools & collaborative workflows, and benchmark methods and datasets.

Introduction

The architecture, engineering, and construction (AEC) industry was an early driver of the development of modern computer technologies. The sketchpad by Sutherland in 1964 (Sutherland, 1964) pioneered the development of software for *Computer Aided Design (CAD)* as well as hardware such as graphical computer screens, digital pens and the computer mouse by D. Engelbert a few years later. With the introduction of personal computers and 3D modeling capabilities in the 1970s, CAD systems became more sophisticated in the 1980s (Requicha and Voelcker, 1982) and established themselves as the main tool for designing and planning constructions.

The establishment of CAD lead to new challenges in the AEC industry, in which many companies and individuals tend to work together on a single project. To improve collaboration and information exchange the concept of *Building Information Modeling* (BIM) emerged in the early 1990s (van Nederveen and Tolman, 1992) as a common workflow and exchange format. However, while these technological developments have had a positive impact on collaboration and information exchange, productivity has not improved as much as first hoped. The fundamental problem with engineering buildings is that each building is to some degree unique. As such, technologies like BIM do not immediately create large productivity gains, because they still do not automate many of the individual processes.

New technologies like artificial intelligence (AI) can build on and assist existing BIM technologies. The question of *Computational Design* has been an obvious use case for years (Caetano et al., 2020). Various approaches developed over time, from Parametric (Jabi, 2013) to Generative (Krish, 2011; Gan, 2022), and Algorithmic Design (Oxman, 2017; Sun et al., 2022; Mandow et al., 2020). The challenge in these approaches is that they are constrained in their generation capability by the underlying generation rules and representations and cannot recombine them in creative ways (Eloy et al., 2018) without introducing randomness and non-determinism, which bring their own issues (Caetano et al., 2020).

Recent AI advancements simplify rule creation by using a black-box approach. With proper training, AI enables creative recombination within implicit limits, avoiding the need for strict rules or pure randomness. The most recent models, called *Foundation Models (FM)*, like ChatGPT or StableDiffusion, are trained on vast datasets and promise massive productivity gains. Many studies experiment with these models and show wide applicability across use cases. But they also uncover limitations that are intrinsic to these models and create unreliable results in multiple ways (Manduchi et al., 2024). This unreliability is especially problematic in engineering, where incorrect information can have dramatic real-world consequences. To truly increase productivity within the AEC industry, FMs need to be able to support our modalities in engineering as input and output, and do so reliably within long-term workflows involving people of various knowledge levels.

The paper analyzes the current state of the art in form of a *systematic literature review (SLR)* on the types of Foundation Models and their application in the AEC domain. Therefore, we searched for recent publications about LLMs and LVMs in AEC on Google Scholar, Scopus and Web of Science. With a systematic data extraction and analysis process, we identified the common strengths and challenges of these models to derive core research topics for our scientific community. Similar studies exist for urban planning (Zhang et al., 2024b), but, not for AEC in general. Finally, we propose the development of special **Building Foundation Models (BFMs)**, which we consider a family of multi-modal foundation models specialized for domain-specific scenarios and datasets. They need to be capable of assisting users in design, build, and operation, with deep domain and standardization knowledge beyond the generic text and image prompts of current models. To focus the development of these BFMs, we will lay out design requirements and a research roadmap.

Foundation Models for the AEC Sector

Large Language Models

There are three important benefits of LLMs: (i) First, their *capability to generate* elaborated responses to questions of common knowledge that is well represented in their training data (Raiaan et al., 2024). This also extends to knowledge on architecture (Ploennigs and Berger, 2024), worker safety education (Uddin et al., 2023), or education (Abedi et al., 2023). (ii) Their *capability to summarize* texts from standards (Zentgraf and König, 2024), work orders and incidents (Smetana et al., 2024), reporting (Pu et al., 2024), or even performing energy analysis (Forth et al., 2023). (iii) Their *capability to adapt* to new knowledge by tuning (few-shot-learning) to train problem specific models. This is particularly useful for training complex information like regulations (Fuchs et al., 2024), advising in BIM development (Du et al., 2024), performing compliance checking (Iversen, 2024), and executing simple planning tasks like scheduling in project management (Prieto et al., 2023).

While these early results demonstrate a wide applicability of LLMs, they also uncover similar limitations. A large challenge of LLMs is their tendency to hallucinate, which describes the case that LLMs generate factually wrong output that sounds reasonable when they are actually uncertain. This risk of hallucination has been identified as main issue in engineering across all the papers cited above. Due to its probabilistic nature, LLMs cannot differentiate this case from the correct one and thus have no easy way to prevent it (Xu et al., 2024). Recent research has confirmed (Compton et al., 2023) that this cannot be solved by adding more training data, as it increases spurious correlations.

Large Vision Models

The benefits of LVMs are similar to those of LLMs: (i) Their *capability to generate* simplifies the generation of images for art (Kotovenko et al., 2019), concept design (Cheng et al., 2023), urban planning (Seneviratne et al., 2022), or floor plan design (Shim et al., 2024). (ii) Their *capability to summarize* allows the interpretation of existing images for inspection (Prajapati et al., 2024) or historical reconstruction (Moral-Andrés et al., 2023). (iii) Their *capability to adapt* with few-shot learning allows them to address new tasks. In particular, StableDiffusion stands out as a freely available model (Rombach et al., 2022). It allows for quick fitting of the model to specific tasks like computational design (Kotovenko et al., 2019).

On the other hand, these image generators lack contextual, spatial, and semantic understanding of the generated content. While images initially look properly structured, they contain many errors in detail (Shim et al., 2024; Seneviratne et al., 2022; Kotovenko et al., 2019; Moral-Andrés et al., 2023), because the models generate pixels based on probability rather than on a semantic understanding. In consequence, the images lack conceptual representation of the 3D space, constructability, material properties, building sequences and other fundamental context present in engineering. In AEC this is akin to issues with early CAD formats, which led to the development of BIM. Beyond being *incapable of creating correct details*, the generators also *cannot create consistent multi-view images*, like consistent views on an object from different angles (Rombach et al., 2022; Blattmann et al., 2022). This limits their applicability in the ideation phase and does not bring productivity increases in latter steps. In summary, the understanding of even simple technical, domain-specific drawings is lackluster, and little training data pertaining to multi-modal information carriers in the built environment exists to date.

Foundation Models

Many recent LLMs and LVMs are multi-modal models that operate on both images and text. They are generalized under the term *Foundation Model* (Zhou et al., 2023). Recent work in FMs focuses on developing specific domain models such as medicine (Moor et al., 2023), climate (Nguyen et al., 2023), mechanics (Göpfert et al., 2024) or robotics (Firoozi et al., 2023). Here, modalities range beyond text and images into more structured forms like e. g. code (Kwak et al., 2023), video (Wang et al., 2024), 3D models (Mohiuddin et al., 2024; Bauscher et al., 2024), point clouds (Zhang et al., 2024a), speech and voice (Li et al., 2023), time series (Steinberg et al., 2021) or graphs (You et al., 2020). The benefits can again be classified into: (i) the *capability to generate* multi-modal content (Bauscher et al., 2024; Zhang et al., 2024a); (ii) *capability to summarize* content (Wang et al., 2024; Mohiuddin et al., 2024); and (iii) *capability to adapt* domain models by few-shot-learning (Li et al., 2023; Mohiuddin et al., 2024).

Challenges of Foundation Models in AEC

FMs face the same challenges as LLMs and LVMs individually. Because of this, there has been growing interest in the AEC research community in using semantic BIM models as a basis for new AI approaches, utilizing their graph structure to simplify model architecture and increase scalability. The semantic models can be used to automate the training of individual AI models at large scale, utilizing symbolic reasoning or combinations of reasoning and ML approaches in the form of *neuro-symbolic AI approaches*. These methods combine symbolic reasoning with neural networks to leverage the strengths of both paradigms: (i) deterministic semantic knowledge representation and reasoning and (ii) flexible learning capabilities to create more adaptive and context-aware solutions. Of particular importance in this development are Graph Neural Networks (GNNs) (Scarselli et al., 2008), that can exploit the graph structure of semantic knowledge to learn and classify those structures or even the underlying data (Haurilet et al., 2019). This approach shows promise for many tasks like classification (Wang et al., 2022), model enrichment (Kaltenegger et al., 2024), automated code checking and structural design optimization (Bloch et al., 2023; Nakhaee et al., 2024), data querying (Stéphane et al., 2023), causal understanding (Haurilet et al., 2019), and performance predictions (Grauer and Ploennigs, 2025).

Construction projects also collect a lot of unstructured information beside BIM in the form of standards, documents, tables, equations, drawings, pictures, and time-series data that are collected during design, construction and operation (Sobhkhiz and El-Diraby, 2023). Attempts to add this information to BIM primarily resulted in more complicated formats and lack of adaptation in practice (Rexhaj and Střelec, 2024). FMs provide new ways to extract this unstructured information from text and images (Zentgraf and König, 2024; Smetana et al., 2024). This capability needs to be extended to extract information from other modalities (see Fig. 1). Many historical CAD plans exist only as line drawings (Sobhkhiz and El-Diraby, 2023). Point cloud scans require separation of spatial, geometric and image information (Werbrouck et al., 2020; Mohiuddin et al., 2024). Documents and standards require understanding of tables and equations (Lynch et al., 2024). FMs can improve these approaches by researching *multi-modal few-shot-learning approaches that can analyze unstructured information*, to increase productivity by learning historical domain and project knowledge. This requires different approaches for each modality.

To enable the previously described approaches, it is essential *to create good datasets to train and benchmark the trained foundation models*. Creating those datasets is non-trivial, as their quality can significantly influence model quality and they need to represent the body of knowledge in its multi-modality in a fair and robust manner (Chang et al., 2024; Manduchi et al., 2024). Furthermore, we need ways to evaluate the correctness of the generated results according to established engineering standards. Although likelihood-based metrics are extensively used to evaluate FMs (e. g. Mean Absolute Error, Accuracy, Precision) they do not provide a good assessment of the engineering quality of the generated outputs (Theis et al., 2015), as it remains unclear whether an increase in accuracy translates to safe usage. Another challenge is the subjective nature of attributes like realism and style. Human inspection is the gold standard but can be subjective, especially in AEC (Zhou et al., 2019). Consequently, learning a reward function from human preferences has gained importance, and public benchmarks of these could provide valuable evaluation tools for generative models (Ouyang et al., 2022). We have to create *validated benchmark datasets and evaluation metrics* to ensure reliable FMs in the AEC industry. Some IFC datasets exist (Emunds et al., 2021), but for AI they are often generated (Bloch et al., 2023). For meaningful FM training we need labeled, realistic datasets.

Current FMs also face significant computational challenges related to the generative process design. The iterative multi-stage denoising process in diffusion models notably slows down inference, often requiring hundreds to thousands of network function evaluations (NFE) to generate high-quality samples (Song et al., 2020). Similarly, the autoregressive nature of LLMs results in slow inference due to sequential token generation. In contrast, alternative generative models such as VAEs and GANs require only a single NFE, but struggle with issues like blurry sample generation and mode collapse (Dosovitskiy and Brox, 2016; Arjovsky et al., 2017). Thus, speeding up inference in diffusion models is a crucial research problem that spans multiple, potentially complementary, directions. Approaches for pruning large models (Ma et al., 2023) are commonly used, but, increase the risk of hallucination. *We need FMs that are less computationally complex to train and compute in order to reduce the environmental impact*. Therefore, alternative lossy model approaches, such as VQ-GAN (Esser et al., 2021) for images, promise to significantly reduce data dimensionality while preserving relevant details.

Last but not least, FMs need to be accessible to AEC practitioners. The domain is strongly fragmented and dominated by SMEs (Small and Medium Enterprises), where individuals cannot afford to build those FMs on their own. We need a *community effort to build those models together and to release them as open source*. Accessibility also implies that the models can be seamlessly integrated into common AEC workflows. Current established workflows for FMs are prompt-based and require expertise and multiple iterations to retrieve the desired outcome (Zechmeister et al., 2023). Unfortunately, many design, planning, and construction processes do not involve only iterative refinement towards a final choice. They are inherently multi-stage and multi-participant workflows. Current research towards such workflow-driven development focuses on multi-tool workflows for LLMs like LangChain/LangGraph (Topsakal and Akinci, 2023). In AEC, these processes should ideally be managed through intuitive interfaces, such as deep CAD integration or voice-controlled assistants. *We need new tools and processes for integrating FMs into AEC workflows that allow multi-modal interaction in design, construction, and operation.* This also includes AI-based understanding of user behavior (Roitberg et al., 2015) to enable intuitive and efficient interaction with professionals.

The preceding literature analysis demonstrates the *broad applicability of FMs to increase productivity and sustain-*

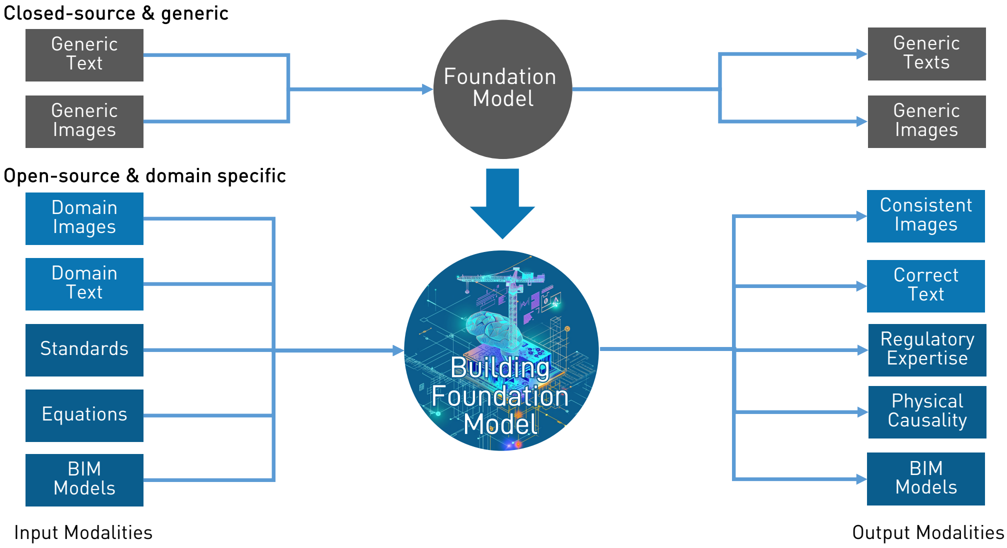

Figure 1: Foundation models need to be extended with new modalities and knowledge representations for AEC

ability in the AEC industry across a multitude of use cases in the lifecycle. It also identified common challenges, discussed in the last section, across all applications that reside deeply in the models themselves. The research results also show that we cannot solve those with general purpose FM approaches, as AEC offers special challenges:

- *Incorrect:* Hallucinations are a major risk for AEC
- *Incompatible:* Well established models like BIM are not accessible to current FM.
- *Inconsistent:* FMs have little spatial and causal understanding to create consistent content.
- *Non-contextual:* Complex constraints such as brownfield scenarios are not considered.
- *Incomplete:* A large body of knowledge exists in standards, regulations, etc. that remains unused.
- *Disconnected:* FMs are not well integrated into common AEC workflows to unlock productivity gains.
- *Untrusted:* Good benchmark and training datasets to validate FM's correctness are missing.

Research Directions for BFM

The diversity and large multi-modality required across the AEC lifecycle indicate that there cannot be a single large FM that serves them all. We must focus on *researching Building Foundation Models (BFMs) as a family of models and technologies, that enable us to build reliable domain-specific models*. They may have generalized versions (like a BFM for ISO standards), but also need to be tuneable to specific projects to account for their uniqueness. Possible solutions here are quick-shot learning approaches, chain-of-thought explainable reasoning, knowledge graph-based verification, and neuro-symbolic modelling to *enable a new generation of BFMs for engineering that reasons cautiously to provide reliable and consistent assistance*. Fig. 2 shows the different proposed research topics and how they connect with each other. The topics are:

Consistent novel multi-modal BFM approaches focus on creating advanced FMs that seamlessly integrate multiple data modalities such as text, images, and geometric data, as both inputs and outputs. These models should *leverage existing knowledge from BIM systems* to ensure they meet the complex constraints of the AEC domain and to produce consistent and accurate engineering designs.

Causal neuro-Symbolic BFMs are needed to unlock the potential of neuro-symbolic, graph-based foundation models that are capable of representing and reasoning about causal relationships within engineering contexts. These models should not only be able to understand and model cause-and-effect relationships, but also apply this reasoning to support decision-making processes in engineering tasks. The objective is to develop models that can incorporate domain-specific constraints and knowledge, enabling them to provide more accurate, context-aware outputs that are essential for critical engineering decisions.

Novel knowledge extraction approaches access knowledge from non-textual sources such as tables, equations, diagrams, and CAD drawings. The goal is to enhance the foundational understanding of multi-modal models by incorporating information that is traditionally difficult to process, such as complex engineering calculations or design specifications. By developing tools and methods that can accurately interpret and integrate this diverse information, BFMs can be made more robust and versatile, capable of addressing a wider range of engineering challenges.

Utilizing few-shot learning enables FMs to quickly adapt to specific engineering problems. This objective focuses on developing models that can be trained with minimal data to perform effectively in niche areas of the AEC domain. By leveraging few-shot learning, these models can be fine-tuned to address specific tasks, such as structural analysis or materials optimization, without the need for extensive retraining. This approach aims to make foundation models more flexible and adaptable, allowing them to be deployed rapidly in various engineering scenarios.

Development of interfaces and tools for new input

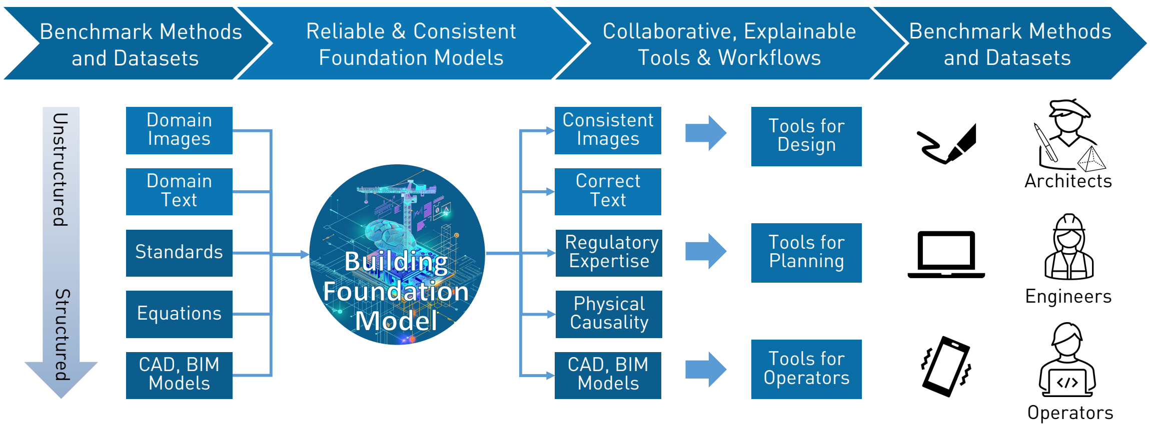

Figure 2: Topic areas to research Foundation Models in Architecture and Civil Engineering

modalities creates interfaces that seamlessly integrate new input modalities into the BFM workflow. This includes designing tools that can interpret and process sketches and diagrams, converting them into actionable data that BFMs can use to generate accurate design outputs. It can also involve developing robust image processing and machine learning techniques that can accurately capture and interpret design intent from various types of visual input. The goal is to bridge the gap between traditional design methods and modern AI-driven processes, allowing for a more fluid transition between manual and digital design stages.

Integration into human design workflows needs interfaces and tools that enable the smooth integration of generative BFMs into design workflows by human-centered-design, ensuring that these models can be effectively utilized in construction design and other engineering processes. This includes developing solutions that facilitate easy integration with project management systems, design documentation, and collaborative CAD platforms. The goal is to streamline the design process and make it easier for teams to collaborate and manage projects efficiently.

Construction and operational integration tools focus on developing interfaces and tools that facilitate the integration of BFMs into construction and operational processes throughout the building lifecycle. This includes creating mixed reality and mobile tools that enable the use of BFMs in the field, ensuring their practical application in real-world construction and operational scenarios.

Data labeling and MLOps workflows requires research into tools and processes tailored to different stages of BFM development. Starting with workflows to efficiently manage and label unstructured datasets for BFMs training, to automated model testing and monitoring of model quality.

Novel benchmarking and evaluation methods are needed to evaluate the performance of BFMs, particularly in terms of their correctness, fairness, and robustness when applied to engineering tasks. This includes developing benchmarks that reflect the diverse and multi-modal nature of the AEC domain, ensuring that models are tested against realistic and challenging scenarios. The objective is to establish a comprehensive evaluation framework that can identify and mitigate potential issues, such as bias or hallucination, ensuring that BFMs meet the stringent quality standards of engineering.

The development of large-scale reference datasets is immanent to train and benchmarking BFMs on real-world conditions. These datasets should be designed to capture the multi-modal and diverse nature of engineering knowledge, enabling the extraction of generalizable domain knowledge that can enhance the robustness of BFMs.

Usability studies across application scenarios need to assess the effectiveness of BFMs in practical, real-world settings. These studies should evaluate the impact, usability, and productivity of BFMs across various stages of the building lifecycle, ensuring that these models can be effectively integrated into everyday engineering workflows.

Fairness, bias, and ethical considerations require methods to detect and address gender, racial, or cultural bias in datasets. This also includes approaches to evaluate the ethical implications of privacy violations, misinformation propagation, and harmful outputs, as well as datasets and metrics that test a model's ability to explain its predictions and decision-making process.

Enabled Use Cases throughout the Lifecycle

Addressing the research questions will unlock the safe use of BFMs in AEC industries across the lifecycle, from architecture, planning, construction, operation, to asset and facility management, and recycling. They will significantly increase productivity and sustainability by assisting architects and engineers with AI solutions that analyse, summarize, advise, and automate tasks in their daily work to reduce complexity of information intake, support understanding of complex knowledge, and increase the quality of output produced under constraints.

The design phase embeds BFMs in a number of ways. This begins with *summarizing* the requirements of a construction by using BFMs to collect requirements from initial sketches, assisted with text or speech dialogues, and complementing them with boundary constraints like shape, style, and regulatory requirements by analysing existing documentation about a site. Following those constraints, images and 3D models for design variants are *generated*,

Table 1: Capabilities enabled by Building Foundation Models in different lifecycle phases

Capability	Design	Build	Operate
Generate	- 3D BIM-Models - Consistent Multi-View Images - Descriptive Texts and Reqirements	- Schedules - Action lists - Translations	- Work orders - Reports - Maintenance
Summarize	- User Input (e. g., sketches, voice) - Site constraints from documents - Standards/Regulations	- 360° progress reports - Construction standards - Project communication	- Historical data - Maintenance History - Spatio-temporal causality
Adapt	- Site constraints - Local regulation & governance - Multi-modal user interaction	- Communication context - Project context - Supply-chain & logistics	- Green Digital Twins - Change documentation - Circular Economy

which provide consistent views from different angles and can fuse with existing images of the surrounding environment to deliver a convincing vision. Those variants are evaluated to meet engineering standards like structural and energetic requirements. These designs are then refined by BIM planners, supported by AI assistants that provide contextually correct regulatory advice through BFMs that are *adapted* to these specific use cases. Ultimately, the designs are documented with generated text that reflects the design, requirements and technical details.

The construction phase utilizes BFMs to assist construction managers and builders in the field. They can use BFMs to *summarize* construction progress by analysing pictures with references to the original designs and linking relevant communications and documents. The installation of complex systems is supported by BFMs that condense handbooks and standards to the current context. This includes *generating* schedules of causally correct sequences of actions to provide action plans for workers and construction robots alike. These BFMs also overcome language barriers on construction sites and allow multi-lingual communication at different knowledge levels. This presumes that the BFMs *adapt* to the contextual knowledge of a project and site to provide the right input at the right time.

The operation phase extends BFMs capabilities of earlier phases in order to enable predictive and preventive maintenance. BFMs are used to *summarize* historical documentation like work orders, BIM models or handbooks to quickly assess the state of the construction and identify recurring issues for predictive maintenance. They can be used to *generate* and verify incomplete work orders with more detailed problem descriptions and solutions, improving documentation quality while reducing documentation effort. By continuously *adapting* the BFMs to the specific behaviour of the building, its lifecycle documents, and operational data, building-specific Digital Twins can be retrieved. They can support operators by answering questions about the construction and by linking information from various sources across the lifecycle in consistent, understandable views. This can also be a key driver for sustainable renovations and recycling.

In all of these lifecycle phases BFMs support architects and engineers in their tasks as shown in Table 1 by *summarizing* multi-modal documents contextually to their project, by automating workflow steps with content that is *generated* according to engineering standards, and by *adapting* to domain- and project-specific knowledge. Particularly in our world, where the data that needs to be processed is constantly increasing, these assisted technologies are crucial to stay productive and focused.

Conclusions

This paper discusses the development of FMs in the context of AEC use cases. The current generation of FMs demonstrates wide applicability across use cases and remarkable first results. They also reveal intrinsic issues that prevent FMs from unlocking their full productivity potential for engineering generally and AEC specifically. We discussed those challenges like halluzination, interpretability and domain applicability and showed examples of recent research approaches that may levigate them. Based on this analysis, we formulated a number of research directions toward the development of Building Foundation Models, which are domain-specific FMs that adapt to engineering and project-specific knowledge from ISO standards to BIM models and seamlessly assist existing processes. Those models should be open-sourced to be adoptable by the SME-dominated AEC community. Our community is best positioned for this development with our long investment in standardization and BIM. To leverage this, **we need a joint effort in our community to fully embrace the potential of Building Foundation Models** to transform the productivity of workflows across the lifecycle and increasing reliability of increase the reliability of future engineering.

References

Abedi, M., Alshybani, I., Shahadat, M. R. B., and Murillo, M. (2023). Beyond traditional teaching: The potential of large language models and chatbots in graduate engineering education. Qeios.

Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein generative adversarial networks. In ICML.

Bauscher, E., Dai, A., Elshani, D., and Wortmann, T. (2024). Learning and generating spatial concepts of modernist architecture via graph machine learning. In CAADRIA.

Blattmann, A., Rombach, R., Oktay, K., Müller, J., and Ommer, B. (2022). Retrieval-augmented diffusion models. In NeurIPS.

Bloch, T., Borrmann, A., and Pauwels, P. (2023). Graph-based learning for automated code checking–exploring the application of graph neural networks for design review. Adv. Eng. Inform., 58.

Caetano, I., Santos, L., and Leitão, A. (2020). Computational design in architecture: Defining parametric, generative, and algorithmic design. Front. Archit. Res., 9.

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., et al. (2024). A survey on evaluation of large language models. ACM Trans. Intell. Syst. Technol., 15(3).

Cheng, K., Neisch, P., and Cui, T. (2023). From concept to space: a new perspective on AIGC-involved attribute translation. Digit. Creat.

Compton, R., Zhang, L., Puli, A., and Ranganath, R. (2023). When more is less: Incorporating additional datasets can hurt performance by introducing spurious correlations. In Machine Learning for Healthcare Conf.

Dosovitskiy, A. and Brox, T. (2016). Generating images with perceptual similarity metrics based on deep networks. In NeurIPS.

Du, C., Nousias, S., and Borrmann, A. (2024). Towards a copilot in BIM authoring tool using large language model based agent for intelligent human-machine interaction. In EG-ICE.

Eloy, S., Pauwels, P., Economou, A., Wortmann, T., and Stouffs, R. (2018). Algorithmic complexity of shape grammar implementation. AI EDAM, 32(2).

Emunds, C., Pauen, N., Richter, V., Frisch, J., and van Treeck, C. (2021). IFCNet: A benchmark dataset for IFC entity classification. In EG-ICE.

Esser, P., Rombach, R., and Ommer, B. (2021). Taming transformers for high-resolution image synthesis. In IEEE/CVF CVPR.

Firoozi, R., Tucker, J., Tian, S., Majumdar, A., Sun, J., Liu, W., et al. (2023). Foundation models in robotics: Applications, challenges, and the future. arXiv:2312.07843.

Forth, K., Abualdenien, J., and Borrmann, A. (2023). Calculation of embodied GHG emissions in early building design stages using BIM and NLP-based semantic model healing. Energy Build., 284.

Fuchs, S., Witbrock, M., Dimyadi, J., and Amor, R. (2024). Using large language models for the interpretation of building regulations. arXiv:2407.21060.

Gan, V. J. (2022). BIM-based graph data model for automatic generative design of modular buildings. Automat. Constr., 134.

Göpfert, J., Weinand, J. M., Kuckertz, P., and Stolten, D. (2024). Opportunities for large language models and discourse in engineering design. Energy and AI.

Grauer, D. and Ploennigs, J. (2025). Does ChatGPT know building physics? exploiting foundation models for building performance prediction with GNNs. In EC3 Conf.

Haurilet, M., Roitberg, A., and Stiefelhagen, R. (2019). It's not about the journey; it's about the destination: Following soft paths under question-guidance for visual reasoning. In IEEE/CVF CVPR.

Iversen, O. (2024). Leveraging large language models for BIM-based automated compliance checking of building regulations. Master's thesis, NTNU.

Jabi, W. (2013). Parametric design for architecture.

Kaltenegger, J., Petrova, E., Borrmann, A., and Pauwels, P. (2024). A conceptual system architecture for enriching digital twins with material performance data using symbolic and sub-symbolic artificial intelligence. In LDAC.

Kotovenko, D., Sanakoyeu, A., Lang, S., and Ommer, B. (2019). Content and style disentanglement for artistic style transfer. In IEEE/CVF ICCV.

Krish, S. (2011). A practical generative design method. Comput.-Aid. Des., 43(1).

Kwak, C., Jung, P., and Lee, S. (2023). A multimodal deep learning model using text, image, and code data for improving issue classification tasks. Appl. Sci., 13(16).

Li, B., Hwang, D., Huo, Z., Bai, J., Prakash, G., Sainath, T. N., et al. (2023). Efficient domain adaptation for speech foundation models. In IEEE ICASSP.

Lynch, K., Eck, B., and Ploennigs, J. (2024). Symbol description reading. In AAAI, volume 38.

Ma, X., Fang, G., and Wang, X. (2023). LLM-pruner: On the structural pruning of large language models. NeurIPS.

Mandow, L., Pérez-de-la Cruz, J.-L., Rodríguez-Gavilán, A. B., and Ruiz-Montiel, M. (2020). Architectural planning with shape grammars and reinforcement learning. Eng. Appl. Artif. Intell., 96.

Manduchi, L., Pandey, K., Bamler, R., Cotterell, R., Däubener, S., Fellenz, S., et al. (2024). On the challenges and opportunities in generative AI. arXiv:2403.00025.

Mohiuddin, R., Prakhya, S. M., Collins, F., Liu, Z., and Borrmann, A. (2024). OpenSU3D: Open world 3D scene understanding using foundation models. arXiv:2407.14279.

Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., and Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. Nature, 616(7956).

Moral-Andrés, F., Merino-Gómez, E., Reviriego, P., and Lombardi, F. (2023). Can artificial intelligence reconstruct ancient mosaics? Stud. Conserv.

Nakhaee, A., Elshani, D., and Wortmann, T. (2024). A vision for automated building code compliance checking by unifying hybrid knowledge graphs and large language models. In Scalable Disruptors.

Nguyen, T., Brandstetter, J., Kapoor, A., Gupta, J. K., and Grover, A. (2023). ClimaX: A foundation model for weather and climate. arXiv:2301.10343.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al. (2022). Training language models to follow instructions with human feedback. In NeurIPS.

Oxman, R. (2017). Thinking difference: Theories and models of parametric design thinking. Des. Stud., 52.

Ploennigs, J. and Berger, M. (2024). Generative AI and the History of Architecture. Springer.

Prajapati, S., Singh, T., Hegde, C., and Chakraborty, P. (2024). Evaluation and comparison of visual language models for transportation engineering problems. arXiv:2409.02278.

Prieto, S. A., Mengiste, E. T., and García de Soto, B. (2023). Investigating the use of ChatGPT for the scheduling of construction projects. Buildings, 13(4).

Pu, H., Yang, X., Li, J., and Guo, R. (2024). Autorepo: A general framework for multimodal LLM-based automated construction reporting. Expert Syst. Appl., 255.

Raiaan, M. A. K., Mukta, M. S. H., Fatema, K., Fahad, N. M., Sakib, S., Mim, M. M. J., et al. (2024). A review on large language models: Architectures, applications, taxonomies, open issues and challenges. IEEE Access.

Requicha, A. A. and Voelcker, H. B. (1982). Solid modeling: a historical summary and contemporary assessment. IEEE CG&A, 2(02).

Rexhaj, G. and Střelec, L. (2024). Digitalization in engineering firms: The role and impact of BIM on productivity. Eu. J. Bus. Sci. Technol.

Roitberg, A., Somani, N., Perzylo, A., Rickert, M., and Knoll, A. (2015). Multimodal human activity recognition for industrial manufacturing processes in robotic workcells. In ACM ICMI.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In IEEE/CVF CVPR.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. (2008). The graph neural network model. IEEE Trans. Neural Netw., 20(1).

Seneviratne, S., Senanayake, D., Rasnayaka, S., Vidanaarachchi, R., and Thompson, J. (2022). DALLE-URBAN: Capturing the urban design expertise of large text to image transformers. In DICTA.

Shim, J., Moon, J., Kim, H., and Hwang, E. (2024). Floordiffusion: Diffusion model-based conditional floorplan image generation method using parameter-efficient fine-tuning and image inpainting. J. of Building Eng.

Smetana, M., Salles de Salles, L., Sukharev, I., and Khazanovich, L. (2024). Highway construction safety analysis using large language models. Appl. Sci., 14(4).

Sobhkhiz, S. and El-Diraby, T. (2023). Dynamic integration of unstructured data with BIM using a no-model approach based on machine learning and concept networks. Automat. Constr., 150.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. (2020). Score-based generative modeling through stochastic differential equations. arXiv:2011.13456.

Steinberg, E., Jung, K., Fries, J. A., Corbin, C. K., Pfohl, S. R., and Shah, N. H. (2021). Language models are an effective representation learning technique for electronic health record data. J. Biomed. Inform., 113.

Stéphane, R., Anthony, D., and Ana, R. (2023). Neuro-symbolic approach for querying BIM models. In IEEE Int. Conf. on Signal-Image Tech. & Internet-Based Syst.

Sun, J., Wu, W., Liu, L., Min, W., Zhang, G., and Zheng, L. (2022). Wallplan: synthesizing floorplans by learning to generate wall graphs. ACM Trans. Graph., 41(4).

Sutherland, I. E. (1964). Sketch pad a man-machine graphical communication system. In SHARE design automat.

Theis, L., Oord, A. v. d., and Bethge, M. (2015). A note on the evaluation of generative models. arXiv:1511.01844.

Topsakal, O. and Akinci, T. C. (2023). Creating large language model applications utilizing langchain: A primer on developing LLM apps fast. In ICAENS.

Uddin, S. J., Albert, A., Ovid, A., and Alsharef, A. (2023). Leveraging ChatGPT to aid construction hazard recognition and support safety education and training. Sustain., 15(9).

van Nederveen, G. A. and Tolman, F. P. (1992). Modelling multiple views on buildings. Automat. Constr., 1(3).

Wang, Y., Li, K., Li, X., Yu, J., et al. (2024). Internvideo2: Scaling video foundation models for multimodal video understanding. arXiv:2403.15377.

Wang, Z., Sacks, R., and Yeung, T. (2022). Exploring graph neural networks for semantic enrichment: Room type classification. Automat. Constr., 134.

Werbrouck, J., Pauwels, P., Bonduel, M., Beetz, J., and Bekers, W. (2020). Scan-to-graph: Semantic enrichment of existing building geometry. Automat. Constr.

Xu, Z., Jain, S., and Kankanhalli, M. (2024). Hallucination is inevitable: An innate limitation of large language models. arXiv:2401.11817.

You, Y., Chen, T., Wang, Z., and Shen, Y. (2020). When does self-supervision help graph convolutional networks? In ICML.

Zechmeister, C., Pérez, M. G., Knippers, J., and Menges, A. (2023). Concurrent, computational design and modelling of structural, coreless-wound building components. Automat. Constr., 151.

Zentgraf, S. and König, M. (2024). Anhancing information extraction from building standards using kosmos-2: A multimodal approach. In ICCCBE.

Zhang, G., Junnan, C., Gao, G., Li, J., and Hu, X. (2024a). Hednet: A hierarchical encoder-decoder network for 3D object detection in point clouds. In NeurIPS.

Zhang, W., Han, J., Xu, Z., Ni, H., Liu, H., and Xiong, H. (2024b). Towards urban general intelligence: A review and outlook of urban foundation models. arXiv preprint arXiv:2402.01749.

Zhou, C., Li, Q., Li, C., Yu, J., Wang, G., Zhang, K., et al. (2023). A comprehensive survey on pretrained foundation models: A history from Bert to ChatGPT. arXiv preprint arXiv.2302.09419.

Zhou, S., Gordon, M., Krishna, R., Narcomey, A., Fei-Fei, L. F., and Bernstein, M. (2019). Hype: A benchmark for human eye perceptual evaluation of generative models. In NeurIPS.