The Generative AI Paradigm: Architectural Praxis Reshaped?

Christoph Geiger, Clemens Lindner

In recent years we have all witnessed the slow but undeniably transformative introduction of Generative AI in many professions, including the field of architectural production, where it has received substantial attention.

While this recent hype has fired many in the profession to imagine a future for design enhanced by increasingly capable emergent AI, there is an undeniable gap between the perceived potential and the current status of the industry.

Whereas computational architecture during previous decades primarily focused on procedural and linear logic-based methodologies, AI demonstrates the potential to introduce an entirely new paradigm to the design process—and has already initiated disruption of established practices in this domain.

While traditional algorithmic and logic-based definitions also strive towards increasingly general and applicable solutions, they still necessitate relatively rigid problem formulation, wherein computational designers create definitions to produce solutions with varying degrees of specificity. Although the scope and applicability of these solutions are substantially determined by the computational designers themselves, the majority of resulting computational tools remain quite constrained to specific use cases.

The novel paradigm that AI introduces to computational architecture discourse—and consequently to the architectural industry holistically—resides predominantly in its generative nature: the capacity to produce diverse solution sets modifiable through mere prompt reformulation rather than extensive manual reworking.

The speed of iteration production could be massively increased through these methodologies. Optioneering transcends manual labor processes, variable manipulation of parametric systems, and tedious programmatic redefinition. New solutions could be nearly instantaneous and accessible through simple natural language descriptions.

However, as this analysis will demonstrate, the generic nature of current AI systems simultaneously constitutes their greatest potential and their most significant limitation, but nonetheless offers undeniably massive opportunities for the industry as a whole to become significantly more data-driven and eventually more automated.

Current State

Research into deep learning (DL) methodologies for addressing specific architectural planning challenges has existed for numerous years, but the emergence of large language models (LLMs) and large-scale general text-to-image systems has catalyzed unprecedented attention throughout the broader industry toward these technologies.

In contrast to preceding predominantly academic research focused on specialized architectural applications at smaller scales and for rather specific problems, the high-quality outputs generated by general generative models—developed by major AI research laboratories with substantial financial and intellectual resources—provided initial indications of transformative technological potential, thereby attracting attention widely beyond academic spheres.[1]

Predominantly younger, technologically proficient architects rapidly began applying these large-scale general models to architectural applications. Implementation of diffusion-based text-to-image models initially involved prompting systems to generate architectural visualizations and imagery, while LLMs were tested regarding their architecture–engineering–construction domain knowledge through queries about architectural codes or building regulations.

1 Oliver Wainwright, "'It's already way beyond what humans can do': will AI wipe out architects?," *Guardian*, August 7, 2023, https://www.theguardian.com/artanddesign/2023/aug/07/ai-architects-revolutionising-corbusier-architecture; N. Hitti, "AI will 'completely change the way we design buildings' says Zaha Hadid Architects," *Dezeen*, April 26, 2023, https://www.dezeen.com/2023/04/26/zaha-hadid-architects-patrik-schumacher-ai-dalle-midjourney/.

Primarily through the proliferation of high-quality architectural imagery across social media and professional platforms, interest in these tools increased substantially. Additionally, the exceptional simplicity and capacity to produce high-quality images with unprecedented speed and volume attracted considerable attention toward Generative AI within architectural circles, particularly among firms and practitioners engaged in early design phases.

Subsequently, challenges emerged regarding the integration of generative visual models into established architectural workflows beyond conceptual design optioneering. Although available solutions—both open-source and commercial—have enhanced user interfaces and incorporated methodologies to address previous limitations such as visual consistency across multiple images or selective editing while maintaining overall visualization integrity, other fundamental constraints remain less amenable to resolution.

For instance, general image generation with stylistic similarity fails to maintain spatial consistency. Consequently, while carefully curated image sets evoking stylistic coherence can be readily generated, they do not actually represent consistent spatial configurations. These images effectively convey atmospheric intentions; however, even with intensive prompting and human curation, achieving solutions that authentically reflect designer intentions while satisfying spatial and functional requirements proves challenging.

Initial architectural applications of LLMS encountered comparable limitations. While model outputs frequently appeared valid and likely benefited from exposure to architectural texts within training datasets, results lacked reliability. These models' tendency toward hallucination and limited knowledge bases for specialized queries resulted in erroneous outputs and incorrect assumptions.

Progressing beyond initial experimentation, certain architects adopted techniques to mitigate described limitations. Ongoing research into text-to-image models has produced various methodologies to enhance visual output guidance, with researchers developing novel model architectures that substantially improved prompt conformity for image generation, exemplified by ControlNet for image-to-image guidance and rectified flow models for enhanced prompt adherence.[2]

2 Zhang Lvmin et al., "Adding Conditional Control to Text-to-Image Diffusion Models," *arXiv*, last modified November 26, 2023, https://arxiv.org/pdf/2302.05543; Patrick Esser et al., "Scaling Rectified Flow Transformers for High-Resolution Image Synthesis," in

Methodologies employing image-guided generation have demonstrated particular utility for architectural applications. By utilizing guidance images extracted from 3D models, spatial conformity can be significantly enhanced, with image generation informed by architect-designed massings or interior spaces that comply with established spatial and functional requirements and constraints.

This approach potentially positions guided text-to-image models between conceptual imagery and definitive visualization, where outputs are informed by established parameters while retaining generative randomness that can inform subsequent design refinement processes. Alternative approaches to directing text-to-image models toward specific outputs include fine-tuning techniques. Methods such as low-resolution adaptation—involving training small parallel neural networks—enable substantial influence on style using limited image sets with similar characteristics and minimal computational resources.[3]

Various fine-tuning methodologies applicable to larger datasets can enhance text-to-image models to conform with specific styles or improve architectural imagery generally through training on comprehensive architectural image collections.

Similar guidance potential exists for LLMs, with substantial research concerning content generation and knowledge retrieval through LLMs for specialized domains. These include low-resolution adaptation and more straightforward implementation methodologies that eliminate model training requirements, such as retrieval-augmented generation—which identifies relevant text segments from provided natural language data and incorporates this information within the language model's context window.[4]

As generative DL research organizations have expanded into substantial commercial entities with continued massive investment streams, more general generative models with architectural adjacency have emerged, including

Forty-first International Conference on Machine Learning (2024), https://arxiv.org/abs/2403.03206.

3 Pareesa Ameneh Golnari, "LoRA-Enhanced Distillation on Guided Diffusion Models," *arXiv*, December 12, 2023, https://arxiv.org/pdf/2312.06899.

4 Edward J. Hu et al., "LoRA: Low-Rank Adaptation of Large Language Models," *arXiv*, last modified October 16, 2021, https://arxiv.org/abs/2106.09685; Gao Yunfan et al., "Retrieval-Augmented Generation for Large Language Models: A Survey," *arXiv*, last modified March 27, 2024, https://arxiv.org/pdf/2312.10997.

text-to-video and text-to-3D systems that have attracted architectural experimentation.

Fig. 37: Zaha Hadid Architects, Image generation based on stylistic finetuning

Text-to-video models attract interest through their capacity to generate temporally and spatially consistent outputs, enabling flythrough videos that describe coherent spatial arrangements suitable for digital 3D reconstruction or direct utilization as conceptual visualization.

While architectural interest in text-to-3D models appears more immediately apparent, contemporary state-of-the-art systems predominantly reflect training on extensive video game asset datasets, with architectural outputs primarily applicable to conceptualization, optioneering, and visualization rather than actual production processes.[5] This contrasts with domains such as game design or visual effects, where digital 3D models constitute final products without physical translation requirements—allowing generated textured meshes to integrate directly into the design pipelines.

These DL model types theoretically support similar fine-tuning and control methodologies as text-to-image systems; however, architectural applications have demonstrated comparatively limited implementation or success—potentially due to challenges in acquiring sufficient architectural video or 3D model datasets, or technical complexities in successfully fine-tuning models competitive with emerging general generative systems.

Limitations

All previously described techniques were not originally designed or specifically trained for architectural applications; however, as in numerous industries, specification and enhanced guidance toward visual outputs or specific concepts and topics are crucial for practical utility.

The essential limitation of current models resides in their generic nature, which inherently introduces randomness and hallucination tendencies—producing outcomes that deviate from specific requirements given by the user. Numerous startups focus on addressing these limitations by offering tailored solutions and specialized interfaces, primarily integrations based on foundational research and guidance models from DL research laboratories.

5 Ren Xuanchi et al., "XCube: Large-Scale 3D Generative Modeling using Sparse Voxel Hierarchies," *Nvidia*, 2024, https://research.nvidia.com/labs/toronto-ai/xcube/; Zibo Zhao et al. "Hunyuan3D 2.0: Scaling Diffusion Models for High Resolution Textured 3D Assets Generation," *arXiv*, January 21, 2025, https://arxiv.org/abs/2501.12202.

While this approach represents a promising direction likely to integrate increasingly into architectural workflows, general generative base models from the AI industry continue to advance rapidly. This creates the potential for general models to surpass specialized solutions in capability. While industry-specific integrations and user interfaces provide substantial value, the trend from open-source toward closed-source models may increasingly challenge these projects in maintaining competitive capabilities without backend integration of general solutions, particularly as major technology corporations increasingly capture revenue through usage-based pricing models.

Beyond individual architectural experimentation with generative models, software companies providing CAD and 3D modeling solutions have actively integrated text-to-image, text-to-3D, and similar functionalities, or enhanced search and automation capabilities with deep search and generative methodologies based on LLMs—a trend expected to continue.

In summary, utilizing diffusion and rectified flow models exclusively for visualization and transformer-based LLMs primarily for information retrieval fails to fully leverage the demonstrated capabilities in both domains, and the industry would profit from a more focused research effort in domain-specific topics.

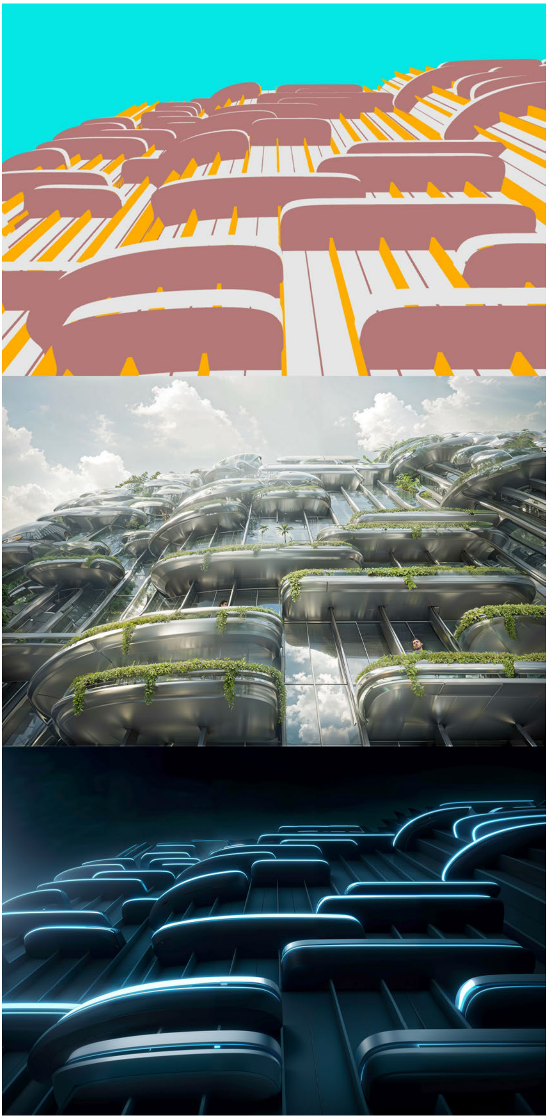
Architecture-specific DL research at fundamental levels exists, including for example topics like floor plan generation through diffusion models[6] or surrogate models for environmental or structural simulations.[7] This research domain merits expansion, as general models appear unlikely to adequately address highly specialized architectural requirements anytime soon.

But substantial potential remains for further integration, particularly regarding LLMs as context-aware agents—either through general-purpose systems or domain-specific training or fine-tuning. This approach could leverage the emergent intelligence of AI systems in ways the architecture–engineering–construction industry and academic research have yet to explore and effectively utilize.

6 Amin Shabani Mohammad, Hosseini Sepidehsadat and Furukawa Yasutaka, "HouseDiffusion: Vector Floorplan Generation via a Diffusion Model with Discrete and Continuous Denoising," *arXiv*, November 23, 2022, https://arxiv.org/pdf/2211.13287.

7 T. Wortmann, A. Costa, G. Nannicini, and T. Schroepfer, "Advantages of surrogate models for architectural design optimization," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 29 (2015): 471–81.

Fig. 38: Zaha Hadid Architects, Image generation controlled by a segmented guidance image



Towards Architectural Intelligence?

The following considerations are predominantly speculative; however, architectural discourse regarding AI-enhanced planning potential appears essential. As automation within the planning industry potentially accelerates—whether beneficially or detrimentally—architects must neither overestimate the technologies' capabilities nor dismiss their possibilities. Developing influential perspectives on tool development could help to ensure benefits extend beyond real estate developers and clients to creative practitioners.

Recent implementations of LLMs as context-aware agents for decision-making and content generation suggest potential integration as planning and design software assistants, comparable to the programming environment's "copilots." Given the impressive capabilities—and substantial market valuation—of code-generating models, numerous tools have emerged embedded directly within integrated development environments as programmer assistants.

With comprehensive project context, language models receive substantial code segments and provide informed development suggestions. While architectural software integration as design, drawing, and modeling copilots offering continuous suggestions and natural language interaction appears conceptually straightforward, implementation presents significant challenges.

Primary obstacles include insufficient structured and labeled datasets and the absence of a unified architectural framework. Each challenge describes different approaches to architectural data utilization for generative DL methods, with distinct problems, limitations, and potentials.

Comparing architectural data to software engineering easily reveals the fundamental disparities. While the Internet provides abundant code for LLM training—a volume incomprehensible to human cognition—architectural data exists under different conditions. Unlike software engineering's structured, collaborative, annotated, and cloud-stored codebases, architectural plans and 3D models lack standardized formats or consistent data structures across projects or practices, with no centralized repository for sharing and potential training utilization.

Despite efforts toward data interoperability and format standardization—particularly through Building Information Modeling (BIM) with Industry Foundation Classes (IFC)—and dataset generation for machine learning research, effective utilization for generative design assistance remains uncertain.

Whether existing architectural data could at all support emergent intelligence comparable to code-generating language models represents an open research question—potentially answerable only by major CAD software developers with access to substantial cloud datasets.

Beyond challenges in data collection lies the absence of a unified computational framework capturing essential design information. Current project data files often omit critical contextual knowledge that humans implicitly incorporate. While BIMs provide extensive detail on the elements and geometry of a building, they frequently lack more fundamental information for informed decision-making, including basic things such as interior–exterior differentiation, spatial connectivity, area utilization metrics, or environmental factors.

Generative models would consequently lack crucial context information for intelligent design progression—information architects typically acquire through site research, constant visual interpretation during the design, and drawing from their education-based intuition. Even assuming inherent emergent architectural intelligence, DL models would require comprehensive contextual data access for informed generation and decision-making and therefore frameworks providing this as partial snapshots for effectively guided generation.

This approach could demonstrate particular value for complex or large-scale design decisions, often found in urban planning or complex public buildings, where requirement, regulation, and constraint volumes challenge even human cognitive capacity.

The idea and proposal of a unified framework concept aligns partly with a recent research strain in computational architecture utilizing gamified systems.[8] Architectural configurator games are constantly tracking key information such as spatial layout and connectivity, area sizes, simplified environmental considerations, and basic structural calculations. These projects can be seen as representing simplified approaches to such frameworks, potentially serving as experimental platforms for testing generative or agent-based approaches.

Well-thought-out, comprehensive, and structured frameworks for architectural data, ideally complemented with standardized formats, could be a big

8 José Luis Sánchez. "Block'hood—Developing an Architectural Simulation Video Game," eCAADe 33 (2015): 89–97; A. Peacock, "Zaha Hadid Architects creates parametric London for Fortnite game," *Dezeen*, September 23, 2024, https://www.dezeen.com/2024/09/23/parametric-london-zaha-hadid-architects-fornite/.

step towards establishing extendable and common foundations for automation-focused architectural research.

Fig. 39: Zaha Hadid Architects, Gamification as design exploration

Final Thoughts

As Generative AI continues its rapid evolution, architectural DL applications in the end may manifest differently than anticipated in this article. But given potential impacts on core architectural competencies, the discipline should actively participate in technological development alongside data and computer scientists.

While this text has assumed progression toward increased automation and co-piloting systems, this trajectory is neither inevitable nor likely to manifest in the immediate future.

Nevertheless, the profession's engagement with these technologies seems crucial, particularly considering global urbanization trends and migration patterns necessitating substantial built environment expansion. Research into automation tools that could enhance qualitative mass production of buildings and urban quarters warrants serious consideration.

Generative DL presents significant opportunities for architectural research, requiring a technically grounded exploration. Ideally, a robust academic research community will investigate these possibilities in collaboration with data scientists, engineers, and programmers—similar to developments following parametric modeling tools' emergence. Establishing an independent architectural research ecosystem with common frameworks is essential for maintaining both academic rigor and professional autonomy from computer-aided design (CAD) software developers and full-fledged AI corporations.

Despite the delayed engagement of the profession on a more foundational level with these technologies, meaningful participation remains possible and undoubtedly necessary—also to ensure a future implementation that ultimately benefits the occupants of buildings and urban environments and not only their financiers.