

Review

Shaping Architecture with Generative Artificial Intelligence: Deep Learning Models in Architectural Design Workflow

Socrates Yiannoudes 

Department of Interior Architecture, School of Applied Arts and Culture, University of West Attica, 12243 Athens, Greece; syiannoudes@uniwa.gr

Abstract

Deep-learning generative AI promises to transform architectural design, yet its potential employment and ready-to-use capacity for professional workflows are unclear. This study presents a systematic review conducted in accordance with PRISMA 2020 guidelines, synthesizing peer-reviewed work from 2015 to 2025 to assess how GenAI methods align with architectural practice. A total of 1566 records were initially retrieved across databases, of which 42 studies met eligibility criteria after structured screening and selection. Each was evaluated using five indicators with a three-tier rubric: Output Representation Type, Pipeline Integration, Workflow Standardization, Tool Readiness, and Technical Skillset. Results show that most outputs are raster images or non-editable objects, with only a minority producing CAD/BIM-ready geometry. Workflow pipelines are often fragmented with manual hand-offs and most GenAI methods map only onto the early conceptual design stage. Prototypes frequently require bespoke coding and advanced expertise. These findings indicate a persistent gap between experimentation with ideation-oriented GenAI and the pragmatism of CAD/BIM-centered delivery. By framing the proposed rubric as a workflow maturity model, this review contributes a replicable benchmark for assessing practice readiness and identifying pathways toward mainstream adoption. For GenAI to move from prototypes to mainstream architectural design practice, it is essential to address not only technical barriers, but also cultural issues such as professional skepticism and reliability concerns, as well as ecosystem challenges of data sharing, authorship, and liability.

Keywords: generative artificial intelligence; deep learning; architectural design; workflow integration



Academic Editors: Changfeng Fu, Yusuf Arayici, Efcharis Balodimou and Mohammad Saki

Received: 27 August 2025

Revised: 2 October 2025

Accepted: 3 October 2025

Published: 10 October 2025

Citation: Yiannoudes, S. Shaping Architecture with Generative Artificial Intelligence: Deep Learning Models in Architectural Design Workflow. *Architecture* **2025**, *5*, 94. <https://doi.org/10.3390/architecture5040094>

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Generative AI Models in Architectural Design

Generative design methodologies have a long history in architectural practice. Since the 1960s, architects have been employing rule-based generative systems such as pattern language [1], shape grammars [2], expert systems [3], and optimization techniques [4] to automate design processes and generate architectural form. Although these methods were initially met with skepticism [5,6], the generative design paradigm re-emerged in the 1990s, when architects used techniques and models like evolutionary algorithms [7–10], cellular automata [11,12], and multi-agent systems [13,14], to generate and optimize architectural form, shape, and floor-plans in relation to functional, structural, or environmental parameters. However, despite their promising endeavors, an early study by Grobman et al. [6],

showed that up till 2009, generative design methods were rare in mainstream architectural practice. Yet, recently, academic interest in generative design has resurfaced. Castro Pena et al. (2021) report an 85% increase in publications related to generative methods since 2015 [15], reflecting a broader shift toward data-driven and approximation-based approaches enabled by recent advances in deep learning (DL) and Generative Artificial Intelligence (GenAI). Recent systematic reviews on GenAI in architectural design indicate a significant surge, particularly since 2020, in studies about the application of data-driven deep learning models. These models have emerged as central to a rapidly evolving research agenda for architectural design, which is primarily oriented toward the development of novel computational methods for early-stage concept imagery, three-dimensional massing, floor-plan generation, and urban-scale design.

1.2. Deep Learning-Based Generative AI

The rise in deep learning has brought renewed attention to GenAI systems—which diverge from earlier rule-based models by learning patterns from large datasets [16]. Deep learning (DL) is a subsymbolic associative machine learning method that encompasses artificial neural networks trained on large datasets, most commonly consisting of images and texts. Deep learning generative systems employ Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), Diffusion models, and Transformers. GANs comprise two neural networks, a generator which creates fake data and the discriminator, which learns to distinguish fake from real data, thus making the generator produce increasingly valid outputs. Conditional GANs are variants of GANs in which both the generator and discriminator are conditioned on additional information (such as class labels or other attributes), enabling the model to generate data that meet specific criteria or contexts. VAEs are generative models that consist of an encoder–decoder pair. VAEs generate new data by sampling from a learned latent space of input data representations, enforcing a probabilistic structure for smooth interpolation within the latent space. Diffusion Models are a class of generative models that gradually add noise to training data and learn to reverse this process to generate new data. They produce high-quality synthetic samples by iteratively denoising an initial random noise distribution seed. Transformers are neural network architectures based on self-attention mechanisms that can parallel process sequential data and thus achieve state-of-the-art performance in tasks such as Natural Language Processing [17].

While the use of DL models as generative design techniques signals a growing enthusiasm for AI-driven design experimentation, their practical utility for architectural design is an open question. This study addresses this issue by systematically reviewing recent literature and documented case studies, following the PRISMA 2020 guidelines, to evaluate the relevance of data-driven generative techniques for architectural design. Specifically, it examines the degree to which DL-based GenAI methods can be meaningfully integrated into professional design workflows for everyday architectural practice. To this end, the study introduces a system of five assessment indicators: (1) the relevance of the type of generated output for architectural design, (2) the extent of seamless integration of design tools along the pipeline, (3) the extent to which the method used maps onto standard design workflow from schematic to construction documents stages, (4) the requirement for custom-made tools and scripting skills, and (5) the need for technical support/expertise. Such a structured system is necessary because existing reviews often remain descriptive, whereas our contribution provides a systematic, replicable, and comparable assessment of how these methods map onto professional architectural design workflows.

1.3. Review Objectives and Questions

The objective of this systematic review is to evaluate the extent to which DL-based GenAI methods can be effectively integrated into architectural design workflows. While recent studies have highlighted the rapid growth of GenAI applications in architecture, their practice-readiness and alignment with standard CAD/BIM-centered processes remain unclear. To address this gap, the study is guided by the following research questions:

- What forms of output (e.g., raster images, meshes/voxels/graphs, CAD/BIM-native models) are produced by DL-based GenAI methods, and how suitable are they for architectural practice?
- To what extent are GenAI toolchains integrated or fragmented, and how does this affect their applicability in design workflows?
- How do GenAI methods align with conventional architectural workflow stages (schematic design, design development, construction documentation)?
- What level of customization or coding effort is required to employ these methods, and are they accessible through off-the-shelf interfaces?
- What degree of specialized expertise beyond typical architectural training is necessary to operate these tools effectively?

By systematically addressing these questions through the five-indicator framework (Output Representation Type, Pipeline Integration, Workflow Standardization, Tool Readiness, and Technical Skillset), the study seeks to provide a structured, replicable, and transparent assessment of the capacity of DL-based GenAI to contribute to professional design practice.

1.4. Previous Reviews

Reviews about GenAI methods in architectural design have significantly increased since 2020. Some reviews see data-driven GenAI models as part of a wider field of generative design research that encompasses additional generative techniques. These include search-based and rule-based systems—most often genetic algorithms, shape grammars, and cellular automata. Other, more recent reviews, focus exclusively on the use and impact of data-driven DL-based GenAI models on architectural practice. These reviews discuss the problems architects face with GenAI models in the design process, sometimes proposing strategies to overcome them.

Castro Pena et al. (2021) [15] reviewed 75 scientific publications from 1995 onward to rank the frequency of occurrence of various types of generative design methods in the literature. The authors observe that DL-based AI generators for conceptual stage shape or floor plan optimization are a rising trend, although outweighed—in terms of study count—by interactive evolutionary computation and cellular automata. The authors implicitly suggest that DL methods are mostly research prototypes, with limited integration in conventional CAD/BIM architectural design workflows.

Bölek et al. (2023) [18] sorted generative methods into six classes in order to subsequently examine their application domains in architectural design. They argue that GenAI holds transformative potential for architectural practice—enabling new modes of design thinking. They observe that in 242 reviewed papers, data-driven DL models are an emerging and increasing popular trend, but not as prevalent as other generative methods (like evolutionary computation). Citing uses of GANs for floor-plan generation or façade design, they point out that workflow integration of DL-based GenAI methods is generally prototype level. Realizing the potential of GenAI requires advances in accessibility, technical innovation, and interdisciplinary collaboration.

Vissers-Similon et al. (2024) [16] found that DL generators in architectural design have been sharply increasing since 2020. They point to Transformer models, along with

graph machine learning and evolutionary computation, as those methods with the greatest potential to impact the early stages of architectural design. Despite the fact that Transformer models are easy to slot into current ideation workflows for fast sketch and image generation there is a growing research trend in employing GANs, VAEs, and Diffusion models for 3D output and direct integration into CAD/BIM environments [19–24].

Zhuang et al.'s (2025) [25] research highlighted the fact that generative design techniques, which were traditionally based on rule-driven algorithms, have lately shifted towards machine learning models. These include both classic machine learning systems as well as data-driven DL models, like GANs and VAEs. Outcomes of this latter paradigm may range from raster images, topology graphs and vector shapes, to mesh models and, more rarely, NURBS surfaces and solids—depending on the toolchain of the design workflow. This approach, as the authors argue, can eventually transform the practice of architecture by enabling efficient conceptual design exploration, complex performance evaluation, and simulation-informed creativity. Tighter workflow integration can be achieved by tactics such as pairing smaller conditional GAN/VAE models with scripted constraints. However, the authors implicitly suggest that realizing the full potential of DL-based GenAI requires overcoming several hurdles: heterogeneous AEC data formats, insufficient data availability, and lack of controllability and interpretability.

Li et al. (2025) [26] observed that architects are rather hesitant in using AI models in architectural design, with only occasional deployment, due to algorithmic complexity and the demand for specialized expertise. Despite architects' awareness of generative AI's potential, integration into everyday design practice remains narrow and uneven and when employed, it is mostly for image generation. The authors argue that research efforts should focus on mapping different generative tools across the entire workflow, as well as building unified, modular platforms where tools can be chained or swapped without rebuilding the pipeline. In addition, they propose embedding domain-specific requirements and regulation-aware, real-time, user-friendly features so that advanced GenAI can transform into an everyday, end-to-end design partner.

Lystbæk (2025) [17] suggests that, despite technical limitations, GenAI models can substantively enhance creativity and productivity across design stages, while respecting the discipline's standards and constraints. He observes that many architectural firms are experimenting with general-purpose AI platforms rather than industry-specific or custom-trained systems, which could have allowed for a more reliable and domain-specific deployment in architectural design workflow. Professional architects recognize the transformative potential of generative models. Yet, concerns over AI "hallucinations", unresolved issues of intellectual property, and even the environmental footprint of large models, have mitigated their enthusiasm for broad deployment. The author suggests that AI models in architectural practice could be used for managing secondary and mundane tasks. These could include automated documentation, data analysis, or early-stage massing studies, that hardly impact critical design decisions. Then, large pre-trained models could be fine-tuned on domain-specific data, including validation metrics and output controls. This could improve contextual awareness and potential incorporation into existing CAD/BIM workflows.

These reviews show that DL-based GenAI methods are mostly used at the early stages of architectural design, for brainstorming ideas and exploring formal configurations. These could be images, 3D massing models, or floor plan layouts. Yet, the integration of GenAI methods in later phases of design development is still limited, due to technical and practical barriers. There is also a lack of specialized expertise and computational resources, scarcity of high-quality architectural datasets, and concerns about outcome reliability and low interpretability in generated outputs. When DL-based GenAI models are used beyond the conceptual design stage, they remain stand-alone prototypes, with limited CAD/BIM

workflow integration. Moreover, they are largely contingent upon bespoke scripting, which demands programming expertise, custom bridges between heterogeneous platforms and file formats, as well as ad hoc pipelines. This results in fragmented toolchains that depend on manual hand-offs between design environments and design phases. However, although the above reviews suggest a chasm between research prototyping and design workflow integration, they only describe the problem. They do not provide a systematic, transparent, and replicable assessment method, i.e., a model that could evaluate, in quantitative terms, workflow integration maturity and practice readiness of these methods.

2. Methods

This study was conducted as a systematic review and reported in accordance with the PRISMA 2020 guidelines. PRISMA principles [27] were used throughout the study to ensure the validity and transparency of the methodology. The objective was to identify, select, and synthesize peer-reviewed studies on the use of deep learning-based generative AI (GenAI) methods in architectural design workflows. A diagram of the overall review method and a PRISMA flowchart are depicted in Figures 1 and 2, respectively.

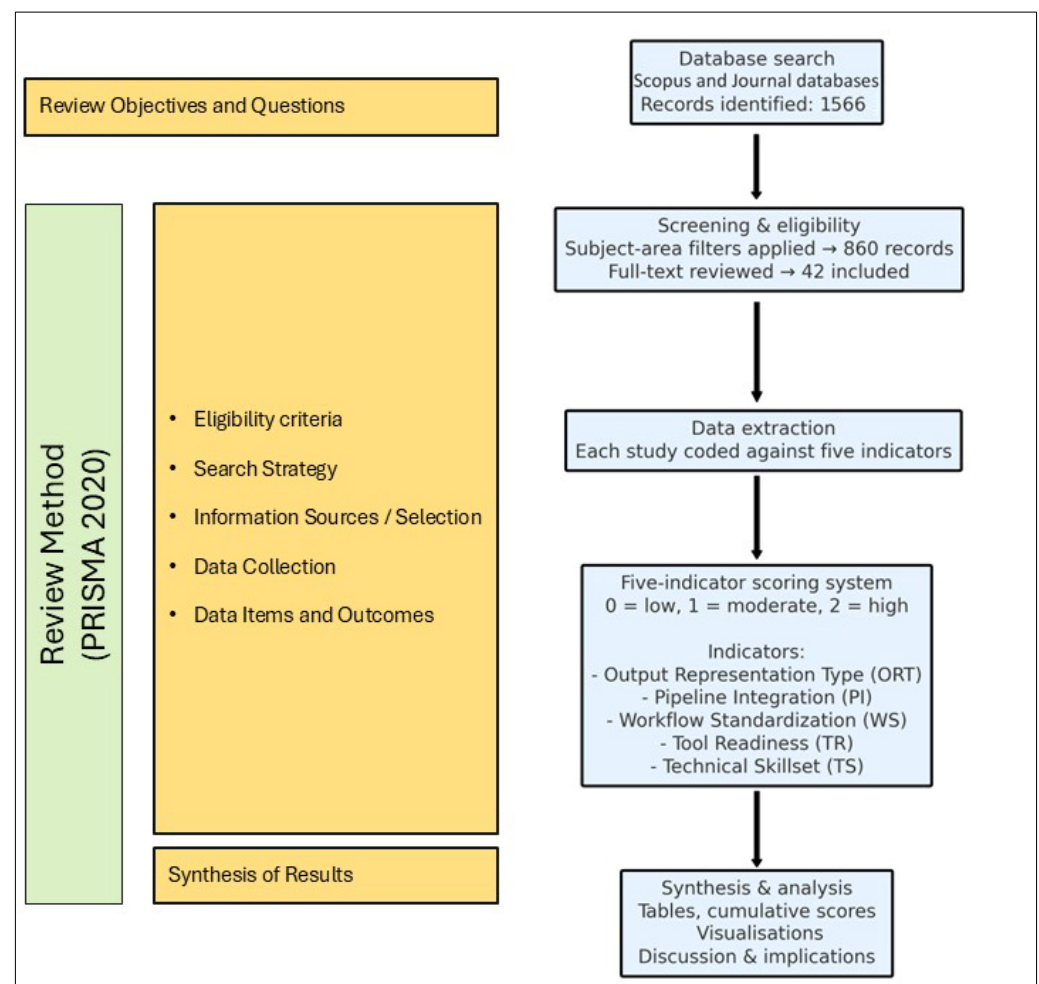


Figure 1. Review method diagram.

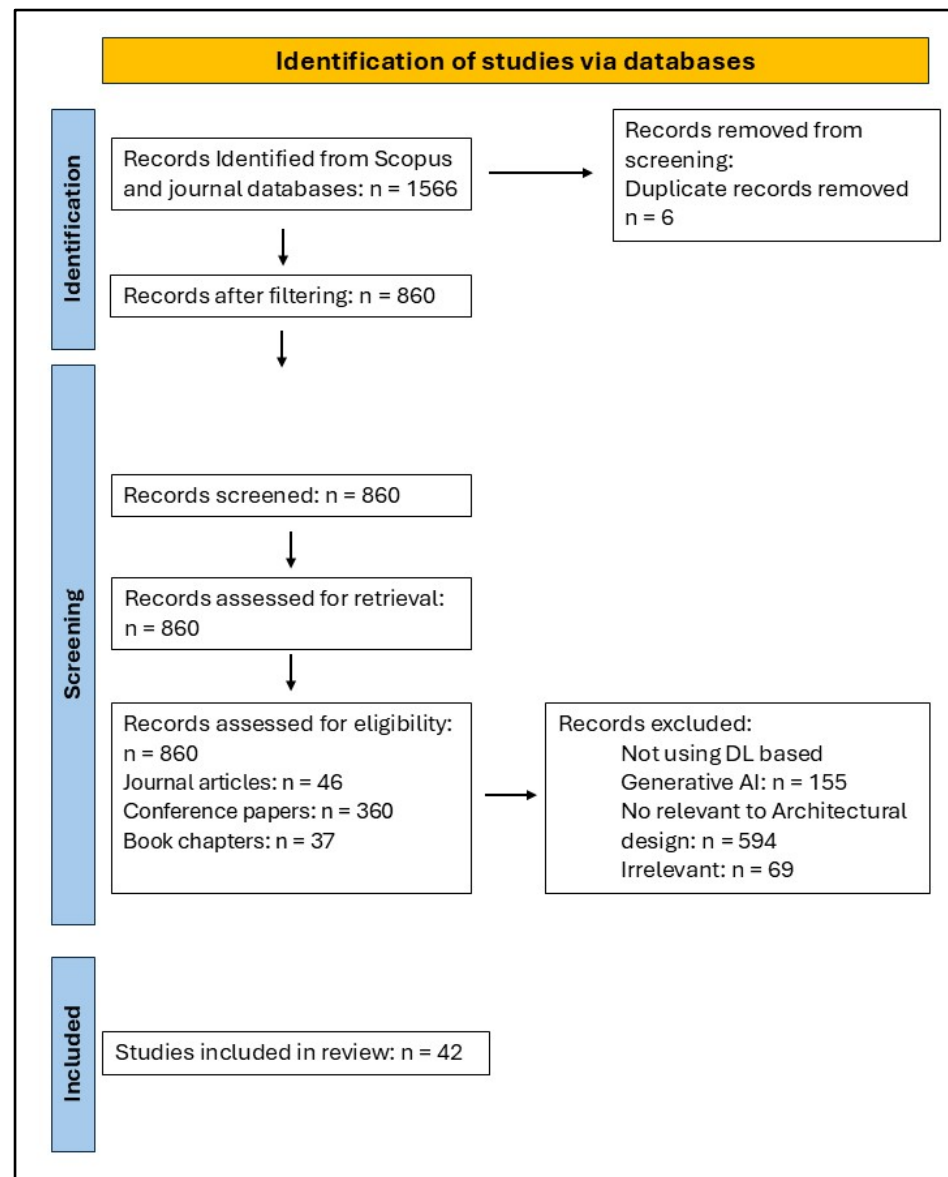


Figure 2. PRISMA flowchart.

Eligibility Criteria. The studies included fulfilled the following criteria:

- (1) They reported on the application of DL-based GenAI methods to architectural design regardless of the degree of implementation. DL models that were combined with other generative systems (such as agent-based models, genetic algorithms, or shape grammars) that cooperatively produced the design outcome were also included.
- (2) They focused on the generation of architectural form (including three-dimensional configurations, spatial layouts, massing, and façades). Although form, in the context of architecture, is not easy to define, in the context of this study, form would involve conceptual representations of three-dimensional spatial configurations and shapes, volumes, spatial layouts, and massing of buildings or structures. Techniques that generated building facades or floorplans were also selected as long as they implied or led to some three-dimensional configuration.
- (3) When optimization techniques were employed (e.g., structural, environmental, or comfort-related) the generation of form was significantly affected.

Search Strategy. Only peer-reviewed journal articles, edited book chapters, and conference proceedings published between 2015 and 2025 were considered, as these reflect

the state-of-the-art and ensure methodological rigor. Non-academic sources (e.g., white papers, proprietary reports) were excluded. The search was limited to academic journals, edited volumes and well-known conference proceedings because (1) they represent the state-of-the-art in deep learning generative AI in architectural design and routinely report on prototypes with potential professional uptake; (2) they allow for quality control—leading journals and conferences use rigorous peer review, so methods are well-documented; and (3) they involve a comparable scope because they target design-stage generative workflows rather than purely construction-engineering papers. Industry case studies, which might be documented elsewhere, may capture more highly scored methods and projects, but were outside our present scope.

Protocol and Registration. No review protocol was registered for this study.

Information Sources and Selection. A comprehensive search was carried out in Scopus and relevant architectural and computer science journal databases within the 2015–2025 timeframe. The search was last updated in early 2025. This review window was chosen because of the radical increase in publications on generative design since 2015, as well as the mainstream employment of DL models in architectural generative design practice—especially after 2020. A Boolean search was conducted with the following keywords: “deep learning”, AND “architectural design”, AND “artificial intelligence”. The full search string and database filters are available upon request. Subject-area filtering was applied to exclude irrelevant domains (e.g., unrelated engineering or non-design computer science studies). The initial search yielded 1566 records. After applying subject-area filters, 860 records remained. These comprised 463 journal articles, 360 conference papers, and 37 book chapters. These were screened for relevance by reviewing titles and abstracts. A total of 42 studies met all eligibility criteria and were included in the review.

Data Collection Process. The data collection method combined qualitative indicators with quantifiable proxies to assess workflow integration for each study. For each study included, data were extracted according to five (5) workflow-integration indicators which were then evaluated on a three-tier scale. These indicators assess the output format representation; the level of integration of the generative workflow pipeline; the alignment of techniques used with standard architectural design practice workflows; the need for scripting and custom-made tools; and the requirement for specialized knowledge and technical support from non-architects. These indicators are described more analytically below.

1. *Output Representation Type (ORT)*. This indicator determines the final output format of the workflow. Workflow output may vary, from raster imagery (Tier-0) to voxels, topology graphs, or mesh models (Tier-1) and vector-based or CAD/BIM-native geometries (Tier-2). Output is considered to be the final outcome of the process, even when pipelines are customized for indirect conversion of images or meshes into CAD/BIM-ready formats.
2. *Pipeline Integration (PI)*. The PI indicator assesses the extent to which the tools used along the pipeline of the workflow are integrated. When the workflow combines more than four loosely coupled tools with usually manual hand-offs, studies score Tier-0. When two to three tools are linked via scripts or plug-ins, studies score Tier-1. When a single platform or fully embedded plug-in is used, with no exports or imports, then studies score Tier-2.
3. *Workflow Standardization (WS)*. Standard design workflows usually follow the Schematic Design/Design Development/Construction Documents (SD/DD/CDs) pipeline, which indicates the typical phases of a design and construction project commonly used in architecture and engineering. This is a structured approach that takes a project from initial concept to detailed construction plans, ensuring a systematic and organized process. In the context of our research, for papers to align with this

standard scheme, they must output CAD/BIM native models coupled with site and code constraints, structural and environmental performance, comfort metrics documentation, etc. These should be Tier-2 case studies, which almost always output parametric geometry, NURBS editable models, or IFC. A mesh-based, raster-to-vector or raster-to-3D-massing pipeline should be Tier-1. Studies in which only stylistic, conceptual, and mood board generation are present stay at Tier-0.

4. *Tool Readiness (TR)*. This indicator determines the requirement for heavy or light custom or off-the-shelf tools in the design workflow. Sometimes there is a demand for custom-made tools and heavy bespoke programming essential for dataset training and further design development. These are Tier-0 studies. Tier-1 studies introduce occasional short scripts, visual code, or macros, while Tier-2 studies do not require heavy programming or coding, often because they use off-the-shelf user interfaces.
5. *Technical Skillset (TS)*. This indicator assesses the requirement for technical expertise beyond typical architectural skillsets, including technical support from programmers and computer scientists. Standard architectural skillsets do not go beyond mainstream digital drafting or visual coding software and parametric modeling plugins. Tier-0 studies should be those that require the competent skills of data scientists and engineers, such as heavy deep ML/RL expertise, as well as Python/C#/API scripting and GPU management. Moderate use of algorithmic node-graph editors (like Grasshopper and Dynamo) and off-the-shelf API bridges, light scripting, and plug-in configuration skills, should indicate Tier-1 studies. Studies that operate with familiar CAD/BIM or prompt-based web user interfaces, not requiring scripting or any kind of model training, should be Tier-2.

Data Items and Outcomes. These five indicators were scored on a three-tier scale (0 = low integration, 1 = moderate integration, 2 = high integration). This rubric was designed to assess the degree to which GenAI methods map onto professional design workflows, from early schematic design through to construction documentation. Table 1 presents the three-tier scale for each indicator.

Table 1. Three-tier indicator system with descriptions.

	Indicator	0—Low Integration	1—Moderate Integration	2—High Integration
1	Output Representation Type (ORT)	Pure raster imagery (no geometry)	Discrete geometry but not industry-native: topology graphs, voxel grids, meshes	BIM/CAD-ready geometry: NURBS surfaces and solids, vector-based geometries, parametric families, IFC
2	Pipeline Integration (PI)	≥four loosely coupled tools/manual hand-offs	Two to three tools linked via scripts or plug-ins	Single platform or fully embedded plug-in—no exports/imports
3	Workflow Standardization (WS)	Experimental pipeline, diverges from typical design phases	Partially maps onto conventional concept/DD/CD flow	Seamless fit with standard BIM/CAD + project-delivery processes
4	Tool Readiness (TR)	Heavy bespoke programming essential	Occasional short scripts or macros	No coding required—off-the-shelf UI
5	Technical Skillset (TS)	Advanced ML/DL expertise.	Some grasp of model training, dataset preparation, and moderate scripting.	Typical architect skillset suffices

Risk of Bias and Quality Assessment. In line with PRISMA 2020, we considered potential sources of bias in the included studies. Because architectural design research does not employ standardized trial or experimental protocols, conventional bias assessment tools (e.g., ROBIS) were not applicable. Instead, we evaluated studies using our five-indicator

rubric (Output Representation Type, Pipeline Integration, Workflow Standardization, Tool Readiness, and Technical Skillset), which indirectly reflects the maturity and reproducibility of each workflow. We recognize that this approach does not assess all possible biases (such as selective reporting or dataset availability), but it provides a structured, transparent framework tailored to the architectural design domain.

3. Results

A total of 42 peer-reviewed studies on DL-based GenAI methods in architectural design were evaluated in terms of five indicators: ORT, PI, WS, TR, and TS. Three discrete representation tiers were used for each indicator. The titles, publication details, and reference list numbers of 42 selected studies are listed chronologically in Table 2, along with their scores on a three-tier scale for each indicator.

Table 2. Three-tier scale scores for each of five indicators for 42 studies.

Title of Paper	Name of First Author	Ref No.	Name of Publication	Year	ORT	PI	WS	TR	TS
Artificial intelligence in architecture: Generating conceptual design via deep learning	As	[28]	International Journal of Architectural Computing	2018	1	0	1	0	0
Generative Deep Learning in Architectural Design	Newton	[19]	Technology Architecture + Design	2019	1	0	0	0	0
Training deep convolution network with synthetic data for architectural morphological prototype classification	Cai	[29]	Frontiers of Architectural Research	2020	0	0	0	0	0
An Academy of Spatial Agents: Generating spatial configurations with deep reinforcement learning	Veloso	[30]	eCAADe	2020	1	0	1	0	0
On GANs, NLP and Architecture: Combining Human and Machine Intelligences for the Generation and Evaluation of Meaningful Designs	Huang	[31]	Technology Architecture + Design	2021	0	0	0	0	0
A generative architectural and urban design method through artificial neural networks	Zheng	[32]	Building and Environment	2021	2	1	1	0	0
Self-learning Agents for Spatial Synthesis	Veloso	[33]	Formal Methods in Architecture	2021	1	1	1	0	0
The AI-teration Method and the Role of AI in Architectural Design	Danchenko	[34]	Proceedings of the Future Technologies Conference	2021	0	0	1	0	0
Intuitive Behavior: The Operation of Reinforcement Learning in Generative Design Processes	Wang	[35]	CAADRIA	2021	1	0	1	0	0
Automatic generation of architecture façade for historical urban renovation using generative adversarial network	Sun	[36]	Building and Environment	2022	0	0	0	0	0
Architectural Form Explorations through Generative Adversarial Networks	Eroglu	[37]	eCAADe	2022	0	0	1	1	1
Design across multi-scale datasets by developing a novel approach to 3DGANs.	Ennemoser	[21]	International Journal of Architectural Computing	2023	1	1	0	0	0
Speculative hybrids: Investigating the generation of Conceptual architectural forms through the use of 3D generative adversarial networks	Pouliou	[22]	International Journal of Architectural Computing	2023	1	1	1	0	0

Table 2. Cont.

Title of Paper	Name of First Author	Ref No.	Name of Publication	Year	ORT	PI	WS	TR	TS
Synthesis and generation for 3D architecture volume with generative modeling.	Zhuang	[38]	International Journal of Architectural Computing	2023	1	1	0	0	0
Spatial synthesis for architectural design as an interactive simulation with multiple agents	Veloso	[39]	Automation in Construction	2023	1	1	1	0	0
Using text-to-image generation for architectural design ideation	Paananen	[40]	International Journal of Architectural Computing	2023	0	2	1	2	2
Using Artificial Intelligence to Generate Master-Quality Architectural Designs from Text Descriptions	Chen	[41]	Buildings	2023	0	2	1	2	2
Research on Architectural Generation Design of Specific Architect's Sketch Based on Image-To-Image Translation	Li	[42]	Hybrid Intelligence, Computational Design and Robotic Fabrication	2023	0	2	0	0	0
The Role of Artificial Intelligence for The Architectural Plan Design: Automation in Decision-making	Çelik	[43]	Proceedings of the 8th International Conference on Machine Learning Technologies	2023	0	0	1	2	2
Generating Conceptual Architectural 3D Geometries with Denoising Diffusion Models Showcasing a deep learning based 3D generative prototype.	Sebestyen	[20]	eCAADe	2023	1	0	1	0	0
AI for conceptual architecture: Reflections on designing with text-to-text, text-to-image, and image-to-image generators	Horvath	[44]	Frontiers of Architectural Research	2024	0	0	0	0	0
Data-driven generative contextual design model for building morphology in dense metropolitan areas	Peng	[45]	Automation in Construction	2024	1	1	2	1	0
Vitruvio: Conditional variational autoencoder to generate building meshes via single perspective sketches	Tono	[46]	Automation in Construction	2024	2	1	1	0	0
Automated layout generation from sites to flats using GAN and transfer learning	Wang	[47]	Automation in Construction	2024	2	1	2	1	0
Generative artificial intelligence and building design: early photorealistic render visualization of façades using local identity-trained models	Jo	[48]	Journal of Computational Design and Engineering	2024	0	2	1	1	1
Generative early architectural visualizations: incorporating architect's style-trained models	Lee	[49]	Journal of Computational Design and Engineering	2024	0	2	1	1	1
Generative AI-powered architectural exterior conceptual design based on the design intent	Shi	[50]	Journal of Computational Design and Engineering	2024	0	1	1	1	1
Generative design experiments with artificial intelligence: reinterpretation of shape grammar	Çelik	[51]	Open House International	2024	0	0	1	2	2
SketchPLAN Recognition and Vectorization of Floor Plan Sketches for Building Information Modelling Design Environment	Abdelmoula	[23]	Advancements in Architectural, Engineering, and Construction Research and Practice	2024	2	0	2	1	0

Table 2. Cont.

Title of Paper	Name of First Author	Ref No.	Name of Publication	Year	ORT	PI	WS	TR	TS
Autocompletion of Architectural Spatial Configurations Using Case-Based Reasoning, Graph Clustering, and Deep Learning	Eisenstadt	[52]	Case-Based Reasoning Research and Development	2024	1	0	1	0	0
Using Generative Adversarial Networks to Create 3D Building Geometries	Mueller	[22]	eCAADe	2024	1	1	1	0	0
Research on Machine Learning-assisted Floor Plan Generation in Old-style Residential Buildings: Taking Tong Lau in Macau as an Example	Tam	[53]	Proceedings of the 3rd International Conference on Computer, Artificial Intelligence and Control Engineering	2024	0	0	1	0	0
Research on Interior Intelligent Design System Based on Image Generation Technology	Zhang	[54]	The 4th International Conference on Machine Learning and Big Data Analytics for IoT Security and Privacy	2024	0	2	1	0	1
A performance-based generative design framework based on a design grammar for high-rise office towers during early design stage	Chen	[55]	Frontiers of Architectural Research	2025	2	1	2	1	0
A diffusion-based machine learning method for 3D architectural form-finding	Zheng	[56]	Frontiers of Architectural Research	2025	1	1	1	1	0
Automated residential layout generation and editing using natural language and images	Zeng	[57]	Automation in Construction	2025	1	1	1	0	1
Generative Architectural Design from Textual Prompts: Enhancing High-Rise Building Concepts for Assisting Architects	Yang	[58]	Applied Sciences	2025	1	1	1	1	1
A deep learning-based framework for intelligent modeling: From architectural sketch to 3D model	Li	[59]	Frontiers of Architectural Research	2025	1	0	1	1	0
Enhancing architectural space layout design by pretraining deep reinforcement learning agents	Kakooee	[60]	Journal of Computational Design and Engineering	2025	1	1	1	0	0
An Intelligent Natural Language Processing (NLP) Workflow for Automated Smart Building Design	Okonta	[61]	Buildings	2025	2	2	2	0	0
A structured prompt framework for AI generated biophilic architectural spaces	Lee	[62]	Journal of Building Engineering	2025	0	1	1	0	0
A hybrid deep learning approach to investigating architectural morphology: A workflow combining graph and image data to classify high-rise residential building floorplans	Wang	[63]	Journal of Building Engineering	2025	1	0	1	0	0

Output Representation Type (ORT): Table S1 (see Supplementary Materials) shows the three-tier scale score and rationale for each study for Output Representation Type indicator. Table 3 shows the number and % distribution of studies for each tier score for the ORT. Seventeen (17) papers scored Tier-0 (40%). Their respective generative methods produced raster outputs, through text-to-image diffusion models and GANs or conceptual diagram generators. Typical examples include [34,61], where text-to-image diffusion or GAN models are leveraged for rapid visual ideation, mood board creation, or conceptual stage exploration and dialog with clients. Although these are useful for practitioners, they

cannot be readily scaled, dimensioned, or adapted within CAD/BIM environments, forcing designers to redraw or remodel outputs before schematic design can commence.

Table 3. No and % distribution of studies for each tier score for the ORT indicator.

Tier	Representation Type	Papers (No)	% Distribution
0	Raster images	17	40%
1	Mesh/voxel/graph	19	45%
2	CAD/BIM-native	6	15%

Tier-1 studies account for nineteen (19) papers, almost as many as the raster output methods. Studies such as [20,52] export density voxels, triangular meshes, or layout topology graphs, either via three-dimensional format dataset GAN training (3DGANs) or by converting 2D image heat maps and signed distance fields via modelers like Rhino, SketchUp, or Blender. These processes can be used for diagrammatic form conceptualization, syntactic exploration, early-stage massing, or environmental performance analyses without complete reconstruction. These studies move beyond pure imagery, but they lack editable parametric geometry, layers, or metadata for CAD/BIM-ready modeling. Also, they demand further manual or scripted conversion for downstream design development.

Tier-2 studies (15%), accounted for only six (6) papers, as exemplified, for example, in [24,61]. Their methods manage to export AI-derived data into platforms like Grasshopper and Revit to ultimately produce CAD/BIM-native geometries (such as editable NURBS surfaces/B-Reps/solids, etc.), preserving semantic object attributes and material parameters. However, achieving this level of integration requires custom APIs and Python plug-ins, CAD-to-BIM bridges like Rhino.Inside, or advanced middleware, like Autodesk Forge and Dynamo. Nonetheless, Tier-2 outputs demonstrably support multi-phase workflows, allowing for concept, analysis, and documentation to proceed without data loss. For example, workflows generating parametric floor plans in Grasshopper or Revit allow architects to integrate generative results directly into documentation and coordination. The relative scarcity of CAD/BIM-native outputs demonstrates that GenAI tools—although promising for creative exploration—remain largely disconnected from professional production environments.

Pipeline Integration (PI): Table S2 (see Supplementary Materials) shows the three-tier scale score and rationale for each study for the PI indicator. Table 4 shows the number and % distribution of studies for each Tier score for the Pipeline Integration indicator. Tier-0 and Tier-1 studies dominate the corpus. For 43% of the studies (Tier-0), workflow is fragmented into four or more distinct bits with usually manual hand-offs, such as separate AI training code (e.g., CycleGAN), and image-to-mesh post-processing methods. A typical sequence starts with Python script and a custom ML model; then, Photoshop is used for image masking, a module for image vectorization, Illustrator for vector drawing refinement, and Revit for redrawing the imported vector file. This manual conversion of raster output into CAD/BIM-ready drawing means that the workflow is profoundly interrupted along the process, with potential file-format loss, version mismatch, and human error.

Table 4. No and % distribution of studies for each tier score for the PI indicator.

Tier	Representation Type	Papers (No)	% Distribution
0	≥four loosely coupled tools/manual hand-offs	18	43%
1	Two to three tools linked via scripts or plug-ins	17	40%
2	Single platform or fully embedded plug-in—no exports/imports	7	17%

Tier-1, representing 17 studies, reflects workflows with a moderate degree of tool integration. In these cases, two or three applications are linked together with partial automation, allowing for smoother hand-offs than Tier-0 but still requiring movement across distinct platforms. Typical examples include pipelines where a diffusion model produces a height-map that is then translated into a Grasshopper mesh, or where a variational autoencoder is coupled with an artificial neural network to drive EnergyPlus simulations through Rhino/Grasshopper. These systems often rely on API calls, Rhino.Inside connections to Revit, or Grasshopper plug-ins to bridge between environments, creating a quasi-continuous pipeline. One illustrative case is Abdelmoula et al.'s SketchPLAN, which connects Python-based GAN segmentation with Rhino and Revit, enabling raster outputs to be transformed into editable BIM elements. Such strategies act as the “glue” between research code and CAD/BIM software, preserving metadata and leveraging specialized engines not natively available in design platforms. However, even at Tier-1, the workflow remains fragmented: custom scripts and bridging modules are necessary, and the toolchain does not yet collapse into a seamless, single-platform experience accessible to practitioners.

Only 17% of the corpus—seven (7) studies—attain Tier-2. In this case, studies employ a single design environment or plug-in, meaning that the designer works entirely inside a single user interface, or through a bespoke pipeline. A single interface would deliver only 2D raster images, such as Zhang et al.'s ComfyUI (a node-graph graphical user interface), which means they acquire Tier-0 on the ORT indicator. A bespoke development on the other hand, like Okonta et al.'s system that links Natural Language Processing with Revit, demonstrates that deep integration is technically feasible but usually demands custom python scripts and add-ins, API wrappers, or extended visual code components. Nevertheless, Tier-2 workflows deliver the greatest downstream value: reduced translation effort, consistent parameter sets, and immediate compatibility with practice standards.

Workflow Standardization (WS): Table S3 (see Supplementary Materials) shows the three-tier scale score and rationale for each study for WS indicator. Table 5 shows the number and % distribution of studies for each tier score for the Workflow Standardization indicator. Eight (8) papers scored Tier-0 because they follow an experimental track where the generated output is interesting research material but cannot be mapped onto any typical architectural design workflow without wholesale rework. Tier-0 studies showcase novel AI engines—GAN-based images and reinforcement learning sandboxes—yet explicitly state that outputs are “for inspiration only” or require manual redrawing before practical use.

Table 5. No and % distribution of studies for each tier score for the WS indicator.

Tier	Representation Type	Papers (No)	% Distribution
0	Experimental pipeline, diverges from typical design phases	8	19%
1	Partially maps onto conventional concept → DD → CD flow	29	69%
2	Seamless fit with standard BIM/CAD + project-delivery processes	5	12%

Twenty-nine (29) papers scored Tier-1 because their respective workflow supports a single stage integration—typically concept sketching, façade mood-boarding, or massing studies. Yet, hand-off to schematic design or BIM phase is manual. Tier-1 studies embed GenAI into one single task. Examples include early massing exploration through Sebestyén et al.'s density voxels, space-planning through Eisenstadt et al.'s graph-based floor-plan topologies, and Kakooee & Dillenburger's reinforcement learning layout, or interior mood-boarding through Lee et al.'s prompt framework.

Only five (5) papers (12%) reach multi-stage integration in the workflow. These Tier-2 studies include, for example: Abdelmoula et al.'s SketchPLAN and Okonta et al.'s bridge

between natural language processing and Revit, both of which deposit native BIM objects that remain editable through documentation. Veloso's agent academy is paired with a Grasshopper pipeline that can carry bubble diagrams into performance analysis without metadata loss. These methods' outputs feed directly into schematic design and carry editable drawing data (such as NURBS and Grasshopper definitions) forward to design development in standard BIM-friendly pipelines.

Tool Readiness (TR): Table S4 (see Supplementary Materials) shows the three-tier scale score and rationale for each study for the TR indicator. Table 6 shows the number and % distribution of studies for each Tier score for the Tool Readiness indicator. The highest proportion of studies (No = 27, 65%) scored Tier-0. These papers introduce custom machine learning training algorithms and datasets [20]), reinforcement learning spatial agents [35,39], and graph-neural networks [52]. Even when the generative engine itself is open-source, authors routinely append preprocessing, post-processing, and evaluation code that lies outside commercial CAD/BIM ecosystems. The result is a patchwork of notebooks, scripts, and API calls—powerful for experimentation but unusual in day-to-day practice.

Table 6. Number and % distribution of studies for each tier score for the TR indicator.

Tier	Representation Type	Papers (No)	% Distribution
0	Heavy bespoke coding essential	27	65%
1	Helper scripts or light visual code definitions and macros	11	26%
2	No custom code; commercial, off-the-shelf GUI	4	9%

Tier-1 papers, which account for one fourth of the corpus (No = 11, 26%), show a transitional pattern. Researchers use ready-made plug-ins (Karamba for Finite Element Methods, LunchBox-ML for regression, ComfyUI for diffusion images), adding only data flow guiding modules, like Grasshopper canvases or CSV import macros. Although flexible, these methods still assume some fluency in scripting and parametric modeling.

Only four studies achieved Tier-2. These methods [40,41,51] usually employ web-based text-to-image off-the-shelf services (Midjourney, Dall-E2) or open-source platforms (Stable Diffusion) with no need for coding, just prompt literacy. Although a no-code generative system might be gradually possible, currently off-the-shelf generative techniques that do not demand bespoke scripting are usually lightweight ideation tools that stop at raster imagery.

Technical Skillset (TS): Table S5 (see Supplementary Materials) shows the three-tier scale score and rationale for each study for the TS indicator. Table 7 shows the number and % distribution of studies for each tier score for the Technical Skillset indicator. The dominance of Tier-0 studies (No = 31, 74%) reflects a field still driven by research prototypes. Reinforcement learning graph-based space planning, diffusion/3DGAN pipelines, and optimizers coupled with performance modules commonly require custom scripting, data curation, GPU setup, and API bridges that typically exceed the skillsets of architects. In practice, these methods imply interdisciplinary teams and outsourcing to machine learning specialists.

Table 7. No and % distribution of studies for each tier score for the TS indicator.

Tier	Definition	Papers (n)	% Distribution
0	High specialist demand	31	74%
1	Moderate scripting literacy	7	17%
2	Ordinary design skills	4	9%

Tier-1 papers make up 17% of the corpus. Workflow complexity is mitigated by parametric design environments and custom modules like Grasshopper with Karamba and machine learning components, or node-graph user interfaces such as ComfyUI. End-users still need to understand data flow, parameters, and plug-in interactions, but do not train models or write substantial code. This level is increasingly attainable for computationally literate practices.

Tier-2 remains rare and bifurcates into two types of studies. One is prompt-only ideation (text-to-image), which is technically accessible but delivers raster outputs with limited downstream design utility. The other involves compiled or tightly embedded add-ins (e.g., Revit-native tools or Rhino.Inside bridges packaged for end-users). These achieve genuine “no-code” operation inside common design software, but they demand significant engineering skills up front—hence their scarcity in the corpus.

4. Discussion

In terms of *Output Representation Type* (ORT), 85% of the studies in the corpus attain Tier-0 (raster imagery) and Tier-1 (voxel grids, graphs, and meshes). This implies that, at the moment, GenAI design tools operate mainly as visual ideation and conceptual or massing exploration aids. For more studies to achieve Tier-2, this will require diffusion or GAN models capable of generating structured geometry for CAD/BIM downstream integration, or plug-and-play exporters that embed semantic metadata automatically. Tier-1 cases are usually conducted within web-based raster output diffusion services that naturally score highly in Tool Readiness (TR). However, both Tier-1 and some Tier-0 cases demonstrate considerable efforts to convert outputs into usable file formats; this includes mesh clean-up, retopology, API bridges, mesh-to-NURBS and raster-to-vector strategies, highlighting a persistent chasm between research prototypes and mainstream practice. Although these methods achieve low scores in Workflow Standardization (WS), they showcase an emerging pattern, that of hybrid pipelines, that start with raster images and finish with parametric geometries without data loss.

Pipelines are often fragmented, particularly when researchers need to manually import raster output into standard drafting or post-processing tools, risking file-format compatibility, version inconsistencies, and potential human error. For the *Pipeline Integration* indicator (PI), Tier-2 studies usually involve single web-based platforms like Stable Diffusion and Midjourney. However, these studies score low on workflow standardization because they do not proceed beyond the conceptual and schematic design stage. Studies that integrate or directly embed GenAI models in CAD/BIM platforms are rare. For pipelines to achieve higher integration with less tool hand-offs, they would require advanced software-engineering resources and skills, overcoming licensing restrictions and open APIs. In this case, common machine learning tasks should be packaged into CAD/BIM environments, pushing more workflows into Tier-2.

The dominance of Tier-1 papers (No = 29, 40%) in *Workflow Standardization* (WS) demonstrates that, with modest engineering effort, workflows can be easily mapped onto the schematic stage of design development. This commonly includes images, meshes, voxel grids, and topology diagrams, for concept and mood board sketching, site massing exploration, room adjacency diagramming, etc. For Tier-1 studies to be able to deliver recognizable hand-offs to the next design stage, it would require geometric re-modeling of the outcomes and possible restructuring of the workflow to account for code compliance, site metrics, and construction limitations. However, some Tier-2 workflows (No = 5) demonstrate a sort of standard end-to-end design thinking that seems to bridge the gap between creative novelty and professional pragmatism. Using native geometry exporters, metadata API bridges, custom scripting modules, and intricate visual code definitions,

these workflows achieve some level of seamless multi-stage alignment. This indicates their dependence on custom tools and highlights the need for more technical expertise and computer engineering resources.

Indeed, for the *Tool Readiness* indicator (TR), two-thirds (No = 27, 65%) of the studies rely on substantial custom programming, ad hoc model training, and dataset curation efforts. This reflects both the novelty of the domain and the scarcity of ready-to-use domain-specific datasets. It also shows the rarity of open-source modules for advanced machine learning integration in CAD/BIM environments. Only four (No = 4) studies use off-the-shelf tools, but they mostly employ prompt-based input and pretrained platforms like Stable Diffusion. One-fourth of the studies (No = 11, 26%) employ custom or off-the-shelf code, or macros. Without proper infrastructure, bespoke coding will remain the norm, forcing sophisticated AI workflows to heavily rely on machine learning experts.

Thus, in terms of Technical Skillset (TS), three-quarters of the studies (No = 31, 74%) demand highly specialized machine learning expertise; this would involve dataset curation, labeling, pre- and post-processing, GPU setup skills, and metadata management. Shifting Tier-0 studies to Tier-2 would need containerized components and plug-ins, robust user interfaces with schema-aware nodes for common tasks, and shared datasets that will allow architects to operate sophisticated methods with ordinary professional skills. Until then, most high-performing GenAI methods in architectural design will remain dependent on specialist expertise.

The column chart in Figure 3 presents the cumulative Tier scores across all five indicators for the 42 studies. Notably, Tier-2 scores—the level indicating the highest degree of practice readiness—are the lowest overall (total = 26). This underscores that current GenAI methods rarely achieve seamless integration into professional workflows. That is, only 12% of all indicator scores across the corpus reached Tier-2, highlighting the rarity of high-integration, practice-ready workflows. The stacked column chart in Figure 4 highlights that Tier-2 proportions are consistently low across all indicators, underscoring the limited practice-readiness of DL-based GenAI methods in architectural design workflows.

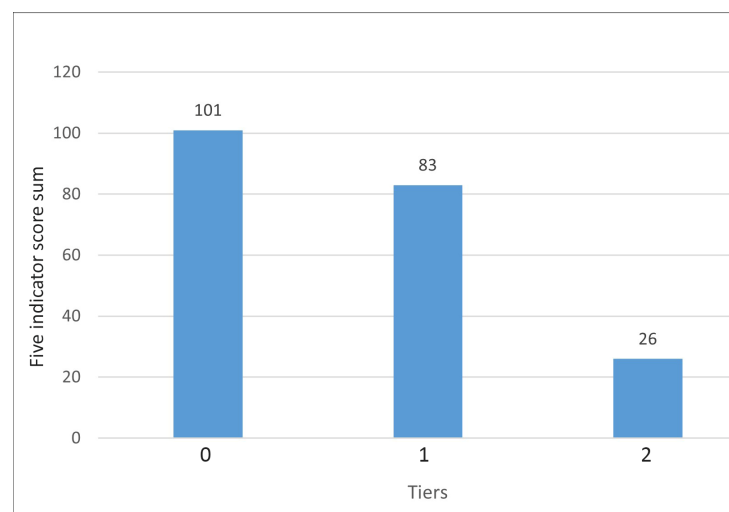


Figure 3. Cumulative tier scores across all five indicators for 42 studies. Tier-2, representing the highest level of workflow integration, accounts for only 26 of 210 possible scores ($\approx 12\%$), indicating that few GenAI workflows achieve seamless CAD/BIM integration and practice readiness.

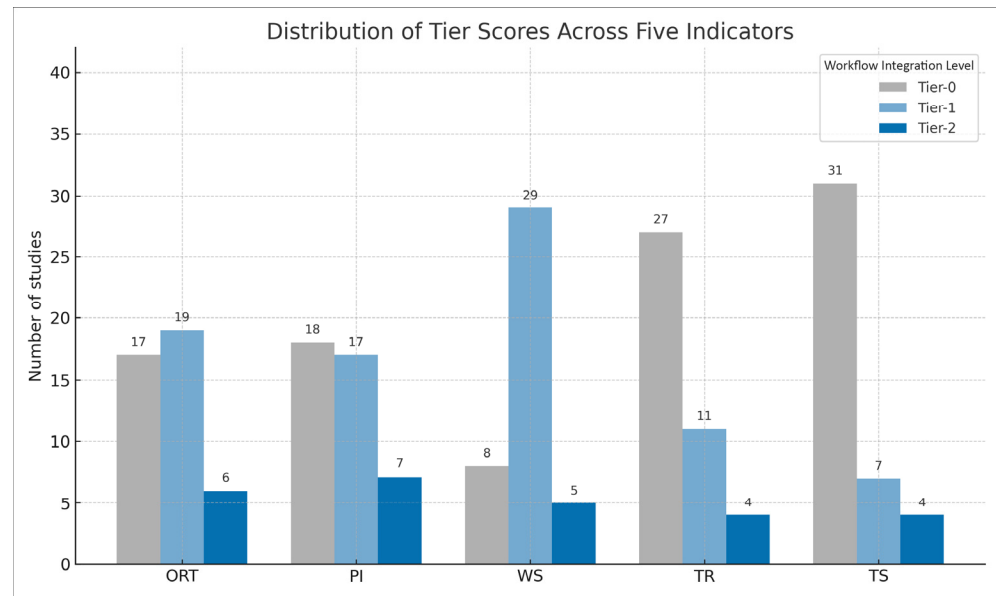


Figure 4. Proportional distribution of Tier scores (Tier-0 = low integration, Tier-1 = moderate integration, Tier-2 = high integration/practice-ready) across five indicators. The stacked format highlights that Tier-2 proportions are consistently low across all indicators, underscoring the limited practice-readiness of DL-based generative AI methods in architectural workflows.

Implications for Practice: From Academic Prototypes to Practice-Ready Tools

Most studies across the reviewed literature remain at the prototype stage and are hardly integrated with mainstream architectural workflows. Although they frequently demonstrate technical novelty, they demand skills and infrastructures that extend well beyond the competencies of most architectural practices. They usually generate raster images or non-native meshes, and rely heavily on bespoke scripting, machine learning expertise, and fragmented toolchains. This persistence of Tier-0 and Tier-1 studies in the corpus is due to several technical, professional, and systemic factors, most of which have already been highlighted in previous reviews on GenAI for architectural design.

- The high proportion of Tier-0 and Tier-1 studies is due to several technical reasons. As Abdelmoula et al. [24] showed, even promising raster-to-BIM pipelines require custom APIs and Python plug-ins, highlighting ongoing technical barriers. Further technical barriers include the demand for costly computational power able to store and process large, multi-layered, and semantically rich BIM models; the lack of standardization and interoperability of file formats; insufficient output control; stylistic bias; and so on. As already mentioned, Zhuang et al. [25] have noted that heterogeneous AEC data formats, insufficient data availability, and lack of controllability hinder workflow maturity for DL-based GenAI systems.
- In professional practice, architects tend to be cautious in adopting tools that are perceived as speculative, unreliable, or opaque. Concerns about authorship, intellectual property, and the interpretability of “black-box” models often reinforce resistance, particularly when outputs cannot be fully controlled or validated. As Li et al. [59] observed, architects remain hesitant to deploy AI because of algorithmic complexity and the need for specialized expertise. Also, practitioners may fear legal liability if GenAI that was trained on their BIM data produces unsafe or non-code compliant outputs. Indeed, Lystbæk [17] argued that many firms prefer experimenting with general-purpose AI tools (e.g., Midjourney, ChatGPT) rather than committing to domain-specific systems, partly due to concerns about authorship, reliability, and IP issues.

- Systemic factors include the lack of large, high-quality, and openly shared architectural datasets, which prevents the training of robust, domain-specific models comparable to those in computer vision or NLP. Confidentiality of design data, proprietary standards, and the absence of coordinated data-sharing frameworks exacerbate this gap. These factors help explain why, despite rapid academic interest and experimentation, most GenAI applications in architecture remain confined to early-stage ideation, rather than advancing toward CAD/BIM-ready, multi-stage workflow integration. Zhuang et al. [25] and Li et al. [59] pointed out that the absence of large, high-quality, openly shared architectural datasets is a major bottleneck for robust GenAI training and reproducibility. Some reviews have also mentioned confidentiality of design data and proprietary standards as obstacles to dataset sharing [16,17].

Only a minority of studies in the corpus were Tier-2 industry-ready cases. These encompass either direct CAD/BIM-native outputs or tightly integrated pipelines via plug-ins and APIs such as Grasshopper/Revit bridges. Tier-2 studies are significant because they reduce the translation gap between generative exploration and professional project delivery. For practice, this distinction highlights both the potential and the limitations of current GenAI research. Architects can readily use existing diffusion and GAN-based tools for early ideation and visual exploration, producing concept imagery, schematic massings, or exploratory floor plan layouts. Yet, meaningful integration into the full design workflow will require packaging deep learning modules into environments architects already use (Revit, ArchiCAD, Rhino/Grasshopper), and accessible user interfaces that minimize the need for custom coding and technical expertise. It will also require exported CAD/BIM models that preserve parametric editability and metadata, as well as tightly integrated API bridges and plug-ins, so that generative outputs flow directly into established workflows without manual conversion.

Recognizing the divide between academic experimentation and practice is essential for charting future development. The assessment indicators employed in this review and the tier score outcomes reveal the potential of experimental methods to be applied in professional practice. They suggest that academic researchers and practitioners should take further steps to make GenAI a dependable part of architectural design workflows. These steps would not only minimize adoption barriers but also enable GenAI to support a broader range of tasks across the design-to-delivery pipeline. For software developers, aligning GenAI outputs with CAD/BIM standards and embedding generative capabilities into existing platforms will be critical to ensure practical uptake in professional workflows. Many of the AI features currently embedded in mainstream design platforms provide only raster-based outputs or operate as assistive modules, rather than tools for generating CAD or BIM geometry from user prompts. For example, Graphisoft's AI Visualizer, which is integrated into ArchiCAD, can generate images from mass objects, sketch inputs, and text prompts. Moreover, ArchiCAD's forthcoming AI Assistant by Nemetschek will blend model queries, support documentation, and integrate with its AI visualizer. Also, such AI-augmented features are already embedded in Autodesk software (such as smart assistants with automation, suggestion, and predictive features). Although these AI modules do not provide generative components for fully supported and integrated CAD/BIM-native output, some indications of future scenarios are already in place; Graphisoft experiments with modules capable of generating parametric BIM objects from 2D drawings or catalog data, while Autodesk has proposed the "Neural CAD" concept, an experimental module for generating BIM or CAD elements from images, sketches, mass geometry objects, and text prompts. Although promising, these are in the proof-of-concept stage, not yet widely deployed as part of ArchiCAD's and Revit's core workspace.

5. Conclusions

This review evaluated 42 studies of DL-based GenAI methods for architectural design using a five-indicator framework—Output Representation Type (ORT), Pipeline Integration (PI), Workflow Standardization (WS), Tool Readiness (TR), and Technical Skillset (TS)—to assess how well these methods map onto architectural practice workflows. The corpus was assembled from Scopus and leading volumes, journals, and conferences proceedings (2015–2025) under clear eligibility criteria focused on form-generation.

Across studies, most GenAI outputs are not CAD/BIM-native representations: 40% of the outcomes are raster imagery (ORT Tier-0), 45% meshes/voxels/graphs, and only 15% are CAD/BIM-native geometry. Toolchain pipelines are typically fragmented: 43% of cases use more than four loosely coupled steps (PI Tier-0), 40% link two to three tools, and 17% operate in a single platform or embedded plug-in. Most studies map onto the schematic design phase (WS Tier-1 = 69%), with rare multi-stage, BIM-compatible pipelines (WS Tier-2 = 12%). The dominant pattern is heavy bespoke coding (TR Tier-0 = 65%) and specialist skill requirements (TS Tier-0 = 74%). The cumulative Tier-2 score across all five indicators for all 42 studies was 26, compared to a score of 101 for Tier-0. That is, only 12% of all five indicator scores across the corpus reached Tier-2, which suggests that practice readiness is low.

Overall, this review has shown that while most DL-based generative methods remain fragmented and limited to schematic ideation, only a small fraction (Tier-2 studies) achieve multi-stage CAD/BIM integration. Therefore, findings substantiate a significant chasm between ideation-oriented experimentation and mainstream CAD/BIM-based practice and delivery. The persistence of Tier-0 and Tier-1 studies in the corpus is firstly due to technical factors, such as fragmented toolchains, image-based output, demand for advanced engineering skills, computational cost, lack of controllability, heterogeneous file formats, or stylistic bias. It is also due to cultural and systemic factors: professional skepticism against the complex algorithmic nature of DL-based models, authorship and reliability issues, legal liability and confidentiality, and the absence of incentives for high-quality and openly shared domain-specific datasets.

Unlike prior reviews, which primarily describe the growth of GenAI applications in architecture, our study contributes a structured, five-indicator rubric that functions as a workflow maturity model. This model provides a replicable way to assess practice-readiness, highlighting where tools are stuck in Tier-0 experimental prototypes and where they approach Tier-2 integration into standard delivery workflows. By framing generative AI adoption through a maturity lens, our review offers a systematic benchmark for both researchers and practitioners to evaluate progress and identify gaps that must be closed for GenAI to move from experimentation to everyday practice. Limitations of this study include a focus on academic sources (likely undercounting proprietary industry deployments), selection of mostly early-stage form-exploration studies in GenAI, and a fast-moving domain of design research. Nonetheless, the five-indicator rubric offers a practical workflow integration maturity index for tracking progress over time and across domains.

Future Directions

As discussed, adoption and full integration of GenAI in architectural design depends to some extent on resolving technical barriers. Closing the gap between ideation-oriented experimentation and CAD/BIM-based delivery will require in-practice studies, studies that track GenAI downstream along the design development stages, and methods that minimize the demand for overly advanced engineering skills. This means shifting outputs from pixels to CAD/BIM geometry and metadata; compiling GenAI models into embedded modules, plug-ins, and API bridges (e.g., Rhino.Inside/Revit add-ins) that minimize hand-

offs; containerized components and schema-aware GUIs that can be managed by typical architectural skillsets; and shared datasets focused on practice limitations such as code compliance, environmental and structural metrics, etc. Two Tier-2 archetypes already visible—prompt-only ideation (accessible but raster-bound) and tightly embedded add-ins (practice-ready but engineering-intensive)—suggest a pragmatic development path.

Also, the evolution of GenAI in architecture will inevitably depend on overcoming major cultural and systemic barriers, professional skepticism, and ecosystem challenges. The ethical and socio-political concerns that these challenges raise cannot be overlooked. Issues of data ownership and intellectual property complicate both training and deployment, as most models rely on datasets whose provenance and licensing are often opaque. Bias and access to energy are also problems faced by the AI ecosystem. Bias embedded in training data risks reproducing narrow stylistic conventions or excluding culturally diverse design vocabularies, thereby limiting rather than expanding architectural imagination. Additionally, the environmental costs of training and running large-scale models—measured in energy consumption and carbon emissions—pose further challenges for a discipline increasingly concerned with sustainability. These issues, while not the primary focus of this review, are increasingly relevant for practitioners because they underscore the need for critical, responsible integration of GenAI into architectural workflows. Overcoming these problems would require shared data infrastructures with anonymization protocols; domain-specific CAD/BIM datasets built by industry consortia and academic bodies to overcome data scarcity; collaboration between firms, universities, and professionals to secure common licensing and metadata standards for these datasets; and governance frameworks to resolve authorship and liability issues. Professional bodies should set standards for ethical use, dataset sharing, or carbon accounting of AI tools in architecture.

Having said that, several possible scenarios ahead are plausible. While, as we showed, text-to-BIM or text-to-CAD GenAI is an active area of research [19,21,22,64], some preliminary off-the-shelf applications are already in use; platforms like BlenderGPT, Hypar, and Spline AI are currently capable of generating CAD drawings and 3D models from text prompts. So, one likely future scenario is the embedding of text-to-plan and text-to-façade GenAI modules directly within mainstream CAD/BIM platforms—such as Revit, ArchiCAD, or Rhino/Grasshopper—to allow designers to generate and iteratively refine parametric floorplans, façades, or massing studies. Thus, existing workflows will be able to offer rapid iteration without leaving the primary design environment, reduce data loss across hand-offs, and make generative exploration accessible to architects without coding expertise. This scenario could also include no-code GenAI pathways. In this case, Grasshopper or Dynamo plug-ins could act as GenAI bridges, enabling architects to experiment with generative models through visual scripting rather than bespoke programming. This would democratize experimentation and accelerate adoption in practice.

Another future scenario would require shared benchmark datasets that could allow for the standardized evaluation of design models, along with integrated AI modules to assist with performance analysis and regulation checks. A consortium model—linking universities, professional practices, and software vendors—could establish open repositories of BIM-native geometry, annotated floorplans, and performance-linked models. This would parallel the role of ImageNet in computer vision and provide the foundation for robust, generalizable GenAI models in architecture.

Furthermore, another scenario would involve mainstream CAD/BIM environments hosting “AI co-pilots” capable of generating regulation-aware, editable, and semantically rich models, enabling true end-to-end AI-augmented workflows. Finally, the evolution of ethical and governance frameworks will shape GenAI’s trajectory. Professional bodies and industry standards could establish guidelines on dataset provenance, energy use, and

intellectual property, ensuring that generative tools are aligned with architectural values of sustainability, inclusivity, and accountability.

Taken together, these scenarios suggest that the next stage of GenAI in architecture will not be driven by technical novelty alone, but by integration, collaboration, and governance—factors that can translate research advances into actionable tools for everyday design practice. By framing GenAI in this way we can envision concrete pathways from today's experimental prototypes to tomorrow's practice-ready design partners.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/architecture5040094/s1>, Table S1: ORT score and rationale for each study; Table S2: PI score and rationale for each study; Table S3: WS score and rationale for each study; Table S4: TR score and rationale for each study; Table S5: TS score and rationale for each study.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data produced by the initial paper search for this review will be made available by the author upon request.

Acknowledgments: During the preparation of this study, the author used ChatGPT-5 for the purposes of producing part of Figure 1 (Review method diagram). The author has reviewed and edited the output and takes full responsibility for the content of this publication.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GenAI	Generative Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
RL	Reinforcement Learning
ANN	Artificial Neural Network
DNN	Deep Neural Network
CNN	Convolutional Neural Network
GAN	Generative Adversarial Network
VAE	Variational Autoencoder
IWGAN	Improved Wasserstein GAN
cGAN	Conditional GAN
DDQN	Double Deep Q-Network
NLP	Natural Language Processing
BIM	Building Information Modeling
CAD	Computer Aided Design
UI	User Interface
GUI	Graphical UI
GPU	Graphics Processing Unit
API	Application Programming Interface
PPO	Proximal Policy Optimization
LoRA	Low-Rank Adaptation

References

- Alexander, C. *A Pattern Language: Towns, Buildings, Construction*; Oxford University Press: Oxford, UK, 1977.
- Stiny, G.; Gips, J. Shape Grammars and the Generative Specification of Painting and Sculpture. In *Information Processing 71*; NorthHolland: Amsterdam, The Netherlands, 1972; pp. 1460–1465.
- Gullichsen, E.; Chang, E. Generative Design in Architecture Using an Expert System. *Vis. Comput.* **1985**, *1*, 161–168. [\[CrossRef\]](#)
- Gero, J.S. Architectural Optimization-A Review. *Eng. Optim.* **1975**, *1*, 189–199. [\[CrossRef\]](#)
- Negroponte, N. *The Architecture Machine: Toward a More Human Environment*; MIT Press: Cambridge, MA, USA, 1972.
- Grobman, Y.J.; Yezioro, A.; Capeluto, I.G. Computer-Based Form Generation in Architectural Design—A Critical Review. *Int. J. Archit. Comput.* **2009**, *7*, 535–553. [\[CrossRef\]](#)
- Frazer, J. *An Evolutionary Architecture*; Architectural Association: London, UK, 1995.
- Caldas, L.G.; Norford, L.K. A Design Optimization Tool Based on a Genetic Algorithm. *Autom. Constr.* **2002**, *11*, 173–184. [\[CrossRef\]](#)
- Renner, G.; Ekárt, A. Genetic Algorithms in Computer Aided Design. *Comput.-Aided Des.* **2003**, *35*, 709–726. [\[CrossRef\]](#)
- Holland, B. Computational Organicism: Examining Evolutionary Design Strategies in Architecture. *Nexus Netw. J.* **2010**, *12*, 485–495. [\[CrossRef\]](#)
- Coates, P. *Programming Architecture*; Routledge: London, UK; New York, NY, USA, 2010.
- Herr, C.M.; Kvan, T. Adapting Cellular Automata to Support the Architectural Design Process. *Autom. Constr.* **2007**, *16*, 61–69. [\[CrossRef\]](#)
- Jacob, C.; Von Mammen, S. Swarm Grammars: Growing Dynamic Structures in 3D Agent Spaces. *Digit. Creat.* **2007**, *18*, 54–64. [\[CrossRef\]](#)
- Von Mammen, S.; Jacob, C. Swarm-Driven Idea Models—From Insect Nests to Modern Architecture. In *Eco-Architecture II*; WIT Press: Algarve, Portugal, 2008; Volume I, pp. 117–126. [\[CrossRef\]](#)
- Castro Pena, M.L.; Carballal, A.; Rodríguez-Fernández, N.; Santos, I.; Romero, J. Artificial Intelligence Applied to Conceptual Design. A Review of Its Use in Architecture. *Autom. Constr.* **2021**, *124*, 103550. [\[CrossRef\]](#)
- Vissers-Similon, E.; Dounas, T.; De Walsche, J. Classification of Artificial Intelligence Techniques for Early Architectural Design Stages. *Int. J. Archit. Comput.* **2024**, 14780771241260857. [\[CrossRef\]](#)
- Lystbæk, M.S. Machine Learning-Driven Processes in Architectural Building Design. *Autom. Constr.* **2025**, *178*, 106379. [\[CrossRef\]](#)
- Bölek, B.; Tatal, O.; Özbaşaran, H. A Systematic Review on Artificial Intelligence Applications in Architecture. *DRArch* **2023**, *4*, 91–104. [\[CrossRef\]](#)
- Newton, D. Generative Deep Learning in Architectural Design. *Technol. Archit. Des.* **2019**, *3*, 176–189. [\[CrossRef\]](#)
- Sebestyen, A.; Özdenizci, O.; Legenstein, R.; Hirschberg, U. Generating Conceptual Architectural 3D Geometries with Denoising Diffusion Models. In Proceedings of the 41st Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe 2023), Graz, Austria, 20–23 September 2023; Graz University of Technology Faculty of Architecture: Graz, Austria, 2023; Volume 2, pp. 451–460. [\[CrossRef\]](#)
- Ennemoser, B.; Mayrhofer-Hufnagl, I. Design across Multi-Scale Datasets by Developing a Novel Approach to 3DGANs. *Int. J. Archit. Comput.* **2023**, *21*, 358–373. [\[CrossRef\]](#)
- Pouliou, P.; Horvath, A.-S.; Palamas, G. Speculative Hybrids: Investigating the Generation of Conceptual Architectural Forms through the Use of 3D Generative Adversarial Networks. *Int. J. Archit. Comput.* **2023**, *21*, 315–336. [\[CrossRef\]](#)
- Mueller, L.-M.; Andriotis, C.; Turrin, M. Using Generative Adversarial Networks to Create 3D Building Geometries. In Proceedings of the 42nd Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe 2024), Nicosia, Cyprus, 11–13 September 2024; University of Cyprus Department of Architecture: Nicosia, Cyprus, 2024; pp. 479–488. [\[CrossRef\]](#)
- Abdelmoula, I.; Schulz, J.-U.; Da Silva Lopes Vieira, T. SketchPLAN Recognition and Vectorization of Floor Plan Sketches for Building Information Modelling Design Environment. In *Advancements in Architectural, Engineering, and Construction Research and Practice*; Olanrewaju, A., Bruno, S., Eds.; Advances in Science, Technology & Innovation; Springer Nature: Cham, Switzerland, 2024; pp. 63–79. [\[CrossRef\]](#)
- Zhuang, X.; Zhu, P.; Yang, A.; Caldas, L. Machine Learning for Generative Architectural Design: Advancements, Opportunities, and Challenges. *Autom. Constr.* **2025**, *174*, 106129. [\[CrossRef\]](#)
- Li, C.; Zhang, T.; Du, X.; Zhang, Y.; Xie, H. Generative AI Models for Different Steps in Architectural Design: A Literature Review. *Front. Archit. Res.* **2025**, *14*, 759–783. [\[CrossRef\]](#)
- Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ* **2021**, *372*. [\[CrossRef\]](#)
- As, I.; Pal, S.; Basu, P. Artificial Intelligence in Architecture: Generating Conceptual Design via Deep Learning. *Int. J. Archit. Comput.* **2018**, *16*, 306–327. [\[CrossRef\]](#)

29. Cai, C.; Li, B. Training Deep Convolution Network with Synthetic Data for Architectural Morphological Prototype Classification. *Front. Archit. Res.* **2021**, *10*, 304–316. [\[CrossRef\]](#)
30. Veloso, P.; Krishnamurti, R. An Academy of Spatial Agents—Generating Spatial Configurations with Deep Reinforcement Learning. In Proceedings of the 38th International Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe), Berlin, Germany, 16–17 September 2020; Volume 2, pp. 191–200. [\[CrossRef\]](#)
31. Huang, J.; Johanes, M.; Kim, F.C.; Doumptioti, C.; Holz, G.-C. On GANs, NLP and Architecture: Combining Human and Machine Intelligences for the Generation and Evaluation of Meaningful Designs. *Technol. Archit. Design* **2021**, *5*, 207–224. [\[CrossRef\]](#)
32. Zheng, H.; Yuan, P.F. A Generative Architectural and Urban Design Method through Artificial Neural Networks. *Build. Environ.* **2021**, *205*, 108178. [\[CrossRef\]](#)
33. Veloso, P.; Krishnamurti, R. Self-Learning Agents for Spatial Synthesis. In *Formal Methods in Architecture*; Eloy, S., Leite Viana, D., Morais, F., Vieira Vaz, J., Eds.; Advances in Science, Technology & Innovation; Springer International Publishing: Cham, Switzerland, 2021; pp. 265–276. [\[CrossRef\]](#)
34. Danchenko, E. The AI-Teration Method and the Role of AI in Architectural Design. In *Proceedings of the Future Technologies Conference (FTC) 2020, Volume 1*; Arai, K., Kapoor, S., Bhatia, R., Eds.; Advances in Intelligent Systems and Computing; Springer International Publishing: Cham, Switzerland, 2021; Volume 1288, pp. 525–538. [\[CrossRef\]](#)
35. Wang, D.; Snooks, R. Intuitive Behavior—The Operation of Reinforcement Learning in Generative Design Processes. In Proceedings of the 26th CAADRIA Conference, Hong Kong, China, 29 March–1 April 2021; Volume 1, pp. 101–110. [\[CrossRef\]](#)
36. Sun, C.; Zhou, Y.; Han, Y. Automatic Generation of Architecture Facade for Historical Urban Renovation Using Generative Adversarial Network. *Build. Environ.* **2022**, *212*, 108781. [\[CrossRef\]](#)
37. Eroğlu, R.; Gül, L.F. Architectural Form Explorations through Generative Adversarial Networks—Predicting the Potentials of StyleGAN. In Proceedings of the International Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe), Ghent, Belgium, 13–16 September 2022; KU Leuven Faculty of Architecture: Brussels, Belgium, 2022; Volume 2, pp. 575–582. [\[CrossRef\]](#)
38. Zhuang, X.; Ju, Y.; Yang, A.; Caldas, L. Synthesis and Generation for 3D Architecture Volume with Generative Modeling. *Int. J. Archit. Comput.* **2023**, *21*, 297–314. [\[CrossRef\]](#)
39. Veloso, P.; Krishnamurti, R. Spatial Synthesis for Architectural Design as an Interactive Simulation with Multiple Agents. *Autom. Constr.* **2023**, *154*, 104997. [\[CrossRef\]](#)
40. Paananen, V.; Oppenlaender, J.; Visuri, A. Using Text-to-Image Generation for Architectural Design Ideation. *Int. J. Archit. Comput.* **2024**, *22*, 458–474. [\[CrossRef\]](#)
41. Chen, J.; Wang, D.; Shao, Z.; Zhang, X.; Ruan, M.; Li, H.; Li, J. Using Artificial Intelligence to Generate Master-Quality Architectural Designs from Text Descriptions. *Buildings* **2023**, *13*, 2285. [\[CrossRef\]](#)
42. Li, Y.; Xu, W.; Liu, X. Research on Architectural Generation Design of Specific Architect’s Sketch Based on Image-To-Image Translation. In *Hybrid Intelligence*; Yuan, P.F., Chai, H., Yan, C., Li, K., Sun, T., Eds.; Computational Design and Robotic Fabrication; Springer Nature: Singapore, 2023; pp. 314–325. [\[CrossRef\]](#)
43. Çelik, T. The Role of Artificial Intelligence for The Architectural Plan Design: Automation in Decision-Making. In Proceedings of the 2023 8th International Conference on Machine Learning Technologies, Stockholm, Sweden, 10–12 March 2023; ACM: Stockholm Sweden, 2023; pp. 133–138. [\[CrossRef\]](#)
44. Horvath, A.-S.; Pouliou, P. AI for Conceptual Architecture: Reflections on Designing with Text-to-Text, Text-to-Image, and Image-to-Image Generators. *Front. Archit. Res.* **2024**, *13*, 593–612. [\[CrossRef\]](#)
45. Peng, Z.; Zhang, Y.; Lu, W.; Li, X. Data-Driven Generative Contextual Design Model for Building Morphology in Dense Metropolitan Areas. *Autom. Constr.* **2024**, *168*, 105820. [\[CrossRef\]](#)
46. Tono, A.; Huang, H.; Agrawal, A.; Fischer, M. Vitruvio: Conditional Variational Autoencoder to Generate Building Meshes via Single Perspective Sketches. *Autom. Constr.* **2024**, *166*, 105498. [\[CrossRef\]](#)
47. Wang, L.; Zhou, X.; Liu, J.; Cheng, G. Automated Layout Generation from Sites to Flats Using GAN and Transfer Learning. *Autom. Constr.* **2024**, *166*, 105668. [\[CrossRef\]](#)
48. Jo, H.; Lee, J.-K.; Lee, Y.-C.; Choo, S. Generative Artificial Intelligence and Building Design: Early Photorealistic Render Visualization of Façades Using Local Identity-Trained Models. *J. Comput. Des. Eng.* **2024**, *11*, 85–105. [\[CrossRef\]](#)
49. Lee, J.-K.; Yoo, Y.; Cha, S.H. Generative Early Architectural Visualizations: Incorporating Architect’s Style-Trained Models. *J. Comput. Des. Eng.* **2024**, *11*, 40–59. [\[CrossRef\]](#)
50. Shi, M.; Seo, J.; Cha, S.H.; Xiao, B.; Chi, H.-L. Generative AI-Powered Architectural Exterior Conceptual Design Based on the Design Intent. *J. Comput. Des. Eng.* **2024**, *11*, 125–142. [\[CrossRef\]](#)
51. Çelik, T. Generative Design Experiments with Artificial Intelligence: Reinterpretation of Shape Grammar. *OHI* **2024**, *49*, 822–842. [\[CrossRef\]](#)

52. Eisenstadt, V.; Langenhan, C.; Bielski, J.; Bergmann, R.; Althoff, K.-D. Autocompletion of Architectural Spatial Configurations Using Case-Based Reasoning, Graph Clustering, and Deep Learning. In *Case-Based Reasoning Research and Development*; Recio-Garcia, J.A., Orozco-del-Castillo, M.G., Bridge, D., Eds.; Lecture Notes in Computer Science; Springer Nature: Cham, Switzerland, 2024; Volume 14775, pp. 321–337. [\[CrossRef\]](#)
53. Tam, H.I.; Chen, Y.; Zheng, L.; Huang, L. Research on Machine Learning-Assisted Floor Plan Generation in Old-Style Residential Buildings: Taking Tong Lau in Macau as an Example. In Proceedings of the 3rd International Conference on Computer, Artificial Intelligence and Control Engineering, Xi'an, China, 26–28 January 2024; ACM: Xi'an, China, 2024; pp. 470–475. [\[CrossRef\]](#)
54. Zhang, F.; Sun, Z.; Chen, Q. Research on Interior Intelligent Design System Based on Image Generation Technology. *Procedia Comput. Sci.* **2024**, *243*, 690–699. [\[CrossRef\]](#)
55. Chen, L.; Zhang, Y.; Zheng, Y. A Performance-Based Generative Design Framework Based on a Design Grammar for High-Rise Office Towers during Early Design Stage. *Front. Archit. Res.* **2025**, *14*, 145–171. [\[CrossRef\]](#)
56. Zheng, H. A Diffusion-Based Machine Learning Method for 3D Architectural Form-Finding. *Front. Archit. Res.* **2025**. [\[CrossRef\]](#)
57. Zeng, P.; Gao, W.; Li, J.; Yin, J.; Chen, J.; Lu, S. Automated Residential Layout Generation and Editing Using Natural Language and Images. *Autom. Constr.* **2025**, *174*, 106133. [\[CrossRef\]](#)
58. Yang, F.; Qian, W. Generative Architectural Design from Textual Prompts: Enhancing High-Rise Building Concepts for Assisting Architects. *Appl. Sci.* **2025**, *15*, 3000. [\[CrossRef\]](#)
59. Li, Y.; Xu, W. A Deep Learning-Based Framework for Intelligent Modeling: From Architectural Sketch to 3D Model. *Front. Archit. Res.* **2025**, S2095263525000627. [\[CrossRef\]](#)
60. Kakooee, R.; Dillenburger, B. Enhancing Architectural Space Layout Design by Pretraining Deep Reinforcement Learning Agents. *J. Comput. Des. Eng.* **2024**, *12*, 149–166. [\[CrossRef\]](#)
61. Okonta, E.D.; Okeke, F.O.; Mgbemena, E.E.; Nnaemeka-Okeke, R.C.; Guo, S.; Awe, F.C.; Eke, C. An Intelligent Natural Language Processing (NLP) Workflow for Automated Smart Building Design. *Buildings* **2025**, *15*, 2413. [\[CrossRef\]](#)
62. Lee, E.J.; Park, S.J. A Structured Prompt Framework for AI-Generated Biophilic Architectural Spaces. *J. Build. Eng.* **2025**, *111*, 113326. [\[CrossRef\]](#)
63. Wang, Y.; Zhu, Y.; Wang, K.; Li, X. A Hybrid Deep Learning Approach to Investigating Architectural Morphology: A Workflow Combining Graph and Image Data to Classify High-Rise Residential Building Floorplans. *J. Build. Eng.* **2025**, *111*, 113255. [\[CrossRef\]](#)
64. Galanos, T.; Liapis, A.; Yannakakis, G.N. Architext: Language-Driven Generative Architecture Design. *arXiv* **2023**. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.