



LATENT ARCHITECTURE

A Journey through the control of Generative AI's Hallucinations

240842

Author: Ahmed Tareef Arafa
Supervisor: Prof. Simone Giostra
Co-Supervisor: Arch. Mariela Tsopanova

A.Y. 2025-2026

Politecnico di Milano
Master's Architecture Built Environment Interiors
School of Architecture Urban Planning Engineering

Acknowledgments

I wish to express my sincere gratitude and appreciation to my supervisor, Prof. Simone Giostra, whose constant guidance, unwavering patience, and interesting conversations shaped this thesis from its earliest uncertainties to its final form. His willingness to engage with such an experimental line of inquiry gave me the confidence to pursue it.

To my mother, Hanan, who matters most of all: your love, prayers, and belief in me carried this work as much as any idea within it. To my father, Tareef, the unsung hero, for his steady support and example, I'm truly thankful. To my sisters, Reem and Ghadeer, for their encouragement and warmth at every stage. To my cousin Alaa, and to all members of my extended family, thank you for being a constant source of grounding and joy. To my friend Loai, thank you for your companionship and for keeping me steady when the work felt heaviest.

I am also grateful to Politecnico di Milano for providing the academic environment, resources, and rigor that made this work possible.

Finally, I thank myself, for every unseen battle I fought, for always picking my self up after each fall, and for the persistence and quiet determination that brought this thesis to life.

This Thesis is dedicated to all of you.

Riassunto

Questa tesi indaga il ruolo dell'intelligenza artificiale generativa nella progettazione architettonica e si chiede cosa accada all'agency del progettista quando sistemi probabilistici entrano nel processo. Affronta tre difficoltà che ne limitano l'adozione: la perdita di controllo legata a modelli opachi a scatola chiusa; il divario tra le immagini raster prodotte da questi strumenti e la precisione vettoriale della documentazione architettonica; e l'assenza di un metodo strutturato per valutare un insieme di strumenti in rapido mutamento. A fronte di ciò, classifica gli strumenti generativi per funzione, sperimenta flussi di lavoro che bilanciano esplorazione creativa e controllo geometrico e traccia un percorso per il progetto assistito dall'IA.

L'approccio è il Research through Design e combina revisione della letteratura, analisi tecnica, sperimentazione e pratica progettuale. Gli strumenti generativi sono stati prima ordinati per modalità e interfaccia, poi valutati nelle fasi concettuale, schematica e di realizzazione. Test comparativi hanno valutato modelli di generazione e di editing rispetto a quattro criteri: usabilità, aderenza al prompt, coerenza e versatilità. I risultati sono confluiti in un quadro di valutazione, esaminato in un caso studio di una tipologia inconsueta, un centro islamico contemporaneo.

Lo studio rileva che l'IA generativa è più efficace come partner aumentativo che come progettista autonomo. I modelli attuali eccellono nell'esplorazione concettuale, nella visualizzazione e nell'iterazione rapida, ma restano inaffidabili per la documentazione ortografica e il lavoro tecnico di precisione. I flussi di lavoro più produttivi uniscono il giudizio umano a una generazione progressivamente vincolata, trasformando l'output stocastico in uno strumento controllato. Sull'autorialità la tesi non giunge a un verdetto definitivo, mostrando solo che è mantenuta o perduta secondo come l'interazione è strutturata. Il contributo duraturo è metodologico: un quadro trasferibile per introdurre l'IA generativa nella pratica alle condizioni stabilite dall'architetto, mantenendo in vista controllo e intento.

Abstract

This thesis investigates the role of generative artificial intelligence in architectural design and asks what becomes of the designer's agency as probabilistic systems are brought into the design process. It addresses three difficulties limiting their adoption: the loss of control associated with opaque, black-box models; the gap between the raster imagery these tools generate and the vector precision of architectural documentation; and the absence of a structured way to evaluate a tool set that changes rapidly. Against these, it categorizes generative tools by function, tests workflows that balance creative exploration with geometric control, and charts a route for AI-assisted design.

The approach is Research through Design, combining literature review, technical analysis, experimental testing, and design practice. Generative tools were first sorted by modality and interface, then assessed across the conceptual, schematic, and realization phases. Comparative testing measured image-generation and image-editing models against four criteria: usability, prompt adherence, consistency, and versatility. These findings were consolidated into an evaluation framework and examined through a case study of an unusual typology, a contemporary Islamic center.

The study finds generative AI most effective as an augmentative partner rather than an autonomous designer. Current models excel at concept exploration, visualization, and rapid iteration, yet remain unreliable for orthographic documentation and precise technical work. The most productive workflows pair human judgment with progressively constrained generation, turning stochastic output into a controlled instrument. On authorship the thesis reaches no final verdict, showing only that it is held or lost according to how the engagement is structured. Its enduring contribution is methodological: a transferable framework for bringing generative AI into practice on terms the architect sets, with control and intent kept in view.

Contents

Introduction	8	Applied Synthesis: EMBER	121
1.1. Technological Evolution	10	4.1. Project Brief	124
1.1.1. The Epistemic Shift	10	4.1.1. Context and Design Intent	125
1.1.2. Wild West	10	4.1.2. Program	125
1.2. Problem Statement	12	4.2. Site Analysis	126
1.2.1. Slot Machine Effect	12	4.2.1. Climate, Light, and Shelter	126
1.2.2. The Integration Gap	12	4.2.2. Architectural and Cultural Context	126
1.2.3. Validation Crisis	13	4.3. Concept	129
1.3. Research Objectives	15	4.3.1. Concept-Generation Skill	129
1.3.1. Categorization of AI Tools	15	4.3.2. Concept Catalogue	130
1.3.2. Hallucination to Constraint	15	4.3.3. Selected Concept	135
1.3.3. Proposing the Roadmap	15	4.3.4. Spatial Development	136
1.4. Scope & Limitations	15	4.4. Iterative Process	138
Literature Review & Methodology	17	4.4.1. Spatial Logic	139
2.1. Technological Context	18	4.4.2. Iterative Refinement	140
2.1.1. Foundational Concepts	18	4.4.3. Extending the View	154
2.1.2. Technical Hierarchy	18	4.4.4. Productive Hallucinations	182
2.1.3. Operational Parameters	20	4.4.5. The Projection Gap	183
2.1.4. Mechanics of Creation	20	Conclusion & Findings	186
2.2. Theoretical Framework	22	5.1. Practical Outcome	189
2.2.1. Augmentation v Automation	22	5.1.1. Diffusion Models in Practice	190
2.2.2. Concept of "Hallucination"	22	5.1.2. Interfaces in Practice	191
2.3. Methodology: Research through Design	24	5.1.3. Roadmap	192
Tools Landscape	26	5.2. Theoretical Realizations	196
3.1. Stage Application	30	5.2.1. The Hype Gap	197
3.1.1. Conceptualization	31	5.2.2. AI vs. Architect	197
3.1.2. Design & Iteration	50	5.2.3. Architect and Machine	197
3.1.3. Visualization	66	5.2.4. Shift in Thinking	197
3.2. Interface Spectrum	73	5.2.5. Design Beyond the Drawing	198
3.2.1. Black Box	74	5.3. Limitations	199
3.2.2. Integrated Plugins	76	5.3.1. Pace and Standardization	199
3.2.3. Local Standalone	78	5.3.2. Hardware	199
3.3. Experimental Testing	81	5.3.3. Training Data and Bias	199
3.3.1. Modal Categorization	83	5.3.4. Open or Closed Source	199
3.3.2. Image Generation	85	5.4. Conclusion	200
3.3.3. Editing Models	105		
3.3.4. Evaluation Matrix	116		

chapter one

Introduction



1.1.Technological Evolution

This thesis does not set out to compile an exhaustive history of architectural computing, nor to catalog every software release of the last decade. Some historical context is unavoidable, however. The shift that **"Neural Architecture"** represents becomes legible only against the technological lineage that precedes it: from the absolute precision of CAD, to the information density of BIM, and now to the probabilistic ambiguity of AI. That trajectory explains why the current wave of generative models marks such a radical departure from established professional norms.

1.1.1. The Epistemic Shift

The history of the architectural profession over the last half-century can be viewed as a continuous **"epistemic transition"** driven by the evolution of its primary tools. Architecture is a technology-intensive discipline. The medium of production shapes both how a building is represented and the cognitive process of design itself. This evolution falls into three phases.

The First Digital Turn: Explicit Representation (CAD)

For centuries, architectural education and practice were rooted in the **"architecture of the hand."** The introduction of CAD in the 1980s digitized the drafting board, yet it largely retained the board's underlying logic. Mario Carpo describes this early era pragmatically as **"Cheaper and Faster Drawing"** (Carpo, 2013). Empirical data supports this claim, with retrospective studies indicating that specialized CAD tools reduced drafting task execution time by approximately 50% compared to manual methods (MDPI, 2024). Early systems were skeuomorphic; they replicated the logic of physical drafting within a digital interface. The defining characteristic here was determinism. The architect explicitly drew every line and defined every coordinate. Adoption, even so, was a **"slow burn."** As late as 2008, the ratio of CAD to BIM usage remained at 3:1 even in technologically advanced regions (Kiviniemi et al., 2008).

The Second Digital Turn: Associative Modeling (Parametricism & BIM)

The second phase shifted the architect's role from **"drawing lines"** to **"defining relationships."** Patrik Schumacher defined this era through **"Parametricism,"** where form is driven by variables and algorithms (Schumacher, 2016). Parallel to this was the adoption of Building Information Modeling (BIM), where the building became a database of objects enriched with non-geometric attributes. Unlike the slow adoption of CAD, BIM saw rapid acceleration driven by mandates. In the UK, for instance, adoption among practices surged from 13% in 2011 to 62% in 2017 (NBS, 2017). Here, the architect became a system designer. That complexity also created a resource gap: small practices were found to be roughly 65% less likely to adopt BIM than large firms (Vidalakis et al., 2019). Despite the complexity, the process remained logic-driven. The architect remained the sole author of the system's rules.

The Third Digital Turn: Probabilistic Generation (Artificial Intelligence)

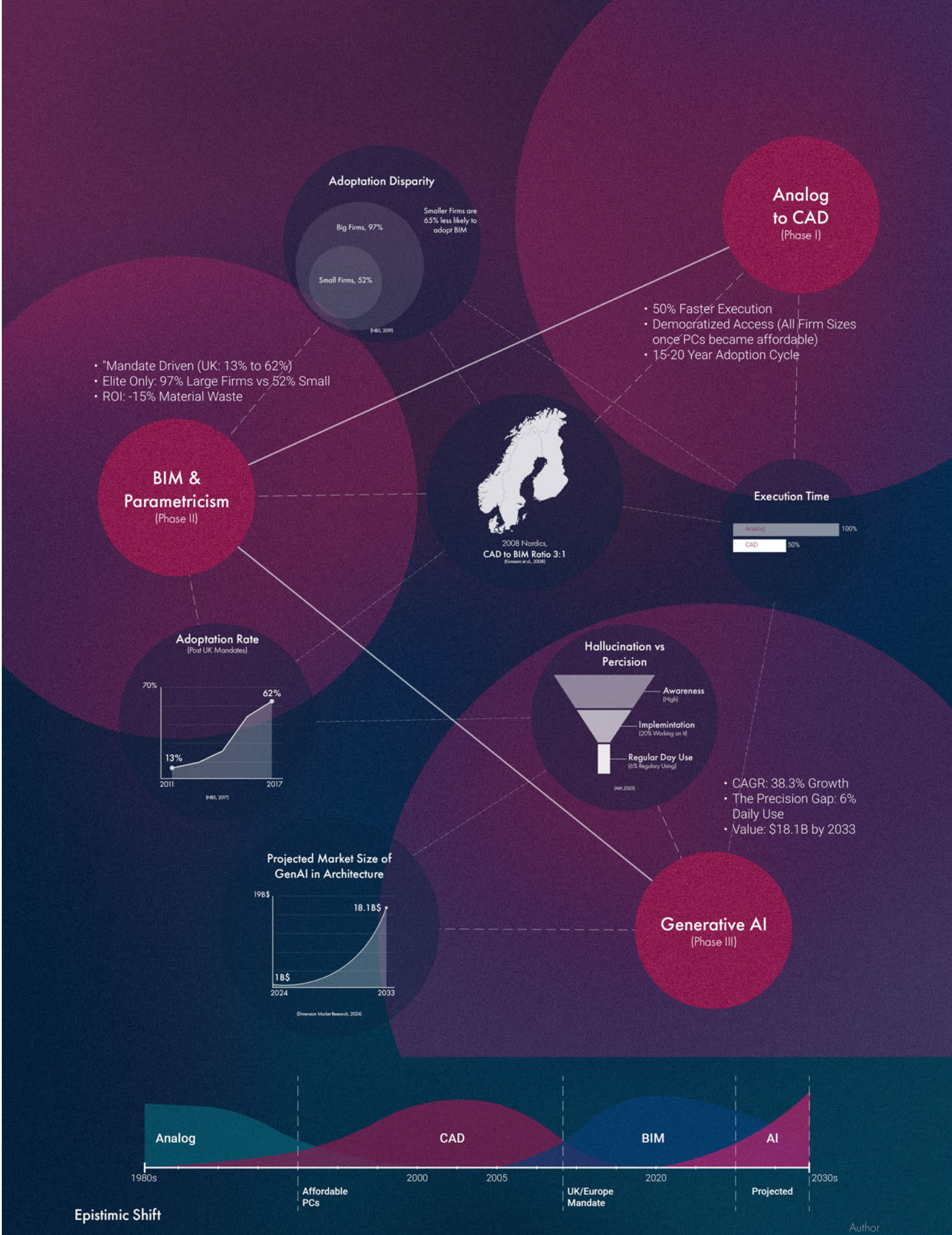
The discipline is now entering a transition toward **"Probabilistic Generation."** Unlike the deterministic logic of CAD and Parametricism (where the architect explicitly defines the rules), generative AI operates on probabilistic statistics. The economic magnitude of this shift is projected to be seismic, with the global market for Generative AI in architecture expected to grow from \$1.0 billion in 2024 to \$18.1 billion by 2033 (Dimension Market Research, 2024). Tools like diffusion models do not **"know"** geometry. They predict pixel distributions based on vast datasets of existing imagery. This represents a fundamental ontological rupture. The architect no longer models the form directly; instead, the architect prompts the system to **"hallucinate"** a solution. This shift poses the central challenge addressed by this thesis: how to impose necessary architectural precision (determinism) onto a system that is inherently stochastic (probabilistic).

1.1.2. Wild West

The contemporary field of architectural software is undergoing what Stanislas Chaillou describes as a **"Cambrian Explosion"** of generative tools (Chaillou, 2019). Following the public release of DALL-E 2, Midjourney, and Stable Diffusion in 2022, the barrier to entry for high-fidelity visualization collapsed. This led to an unregulated proliferation of software platforms.

Unlike the transition to BIM, which was stabilized by ISO standards and open protocols like IFC (Industry Foundation Classes), the current wave of AI adoption is characterized by a distinct lack of standardization. It is, effectively, a **"Wild West."** Recent industry data shows a wide gap between hype and reality: while 20% of firms report actively working on implementation, only 6% of the profession report **"regularly using"** AI in their daily workflows (AIA, 2025).

This environment is fragmented between proprietary **"Black Box"** models and open-source modular systems. Commercial platforms like Midjourney operate as closed systems where the training data and algorithms are opaque to the user. This limits the architect's ability to understand or control the bias within the generated results. Conversely, local ecosystems like Stable Diffusion offer granular control but require technical literacy akin to software engineering. This dichotomy creates a **"workflow fracture."** Architects are forced to rely on a disjointed stack of unverified tools that do not communicate with standard CAD/BIM environments, resulting in a loss of data integrity as the project moves from concept to development.



1.2. Problem Statement

Despite the exponential adoption of generative AI tools within the architectural discourse, their application in professional practice remains largely superficial. While the capacity of these models to generate high-fidelity imagery is undisputed, their utility as rigorous design instruments is compromised by fundamental methodological conflicts. The current state of "AI in Architecture" is defined by friction. This thesis identifies three barriers that currently keep generative AI from becoming a necessity rather than a novelty in the design studio.

1.2.1. Slot Machine Effect

The current engagement with "Black Box" AI tools is defined by a lack of direct agency, a phenomenon often described as the "Slot Machine" effect. In this workflow, the architect inputs a text prompt and "pulls the lever," initiating a stochastic diffusion process. As Neil Leach argues in *Architecture in the Age of Artificial Intelligence*, this fundamentally alters the design ontology, shifting the architect from a "creator" to a "curator" (Leach, 2022).

The core problem is the decoupling of input intent from output resolution. In a traditional parametric definition (like Grasshopper), a change in parameter X results in a predictable change in geometry Y. In contrast, generative AI models operate within a "Latent Space" of statistical correlations. A prompt for a "modern library" triggers a probabilistic synthesis of what the model's training data defines as "modern" and "library," often overriding specific user constraints. Matias del Campo notes that while this process can produce "defamiliarized" aesthetics (del Campo, 2022), it strips the architect

of precise control. The design process is reduced to a game of chance, where the architect hopes for a "lucky seed" rather than executing a deliberate strategy.

1.2.2. The Integration Gap

The fundamental barrier to the adoption of AI in the technical design studio is the ontological schism between the probabilistic nature of generative models and the deterministic requirements of architectural documentation. Architectural production relies on precision: specific dimensions, defined material properties, and spatial coordinates (Vectors). In contrast, current text-to-image (TTI) models operate exclusively in the domain of the raster image (Pixels). Mario Carpo describes this friction as the conflict between "visual fidelity" and "geometric validity" (Carpo, 2023).

This gap manifests primarily as Dimensional Amnesia. Diffusion models generate pixels based on statistical likelihood, not geometric rules. The AI does not "understand" scale or topology. It creates a simulation of a building where a door might be visually convincing but proportionally 4 meters tall relative to the adjacent steps. This "Scale Blindness" renders the output resistant to translation into BIM environments. The integration gap is widened further by Iterative Volatility. Attempting to modify a single architectural element often triggers a "butterfly effect," altering the global composition because the model generates a new image from scratch rather than modifying the existing object. This lack of "disentanglement" breaks the standard feedback loop of the design studio.



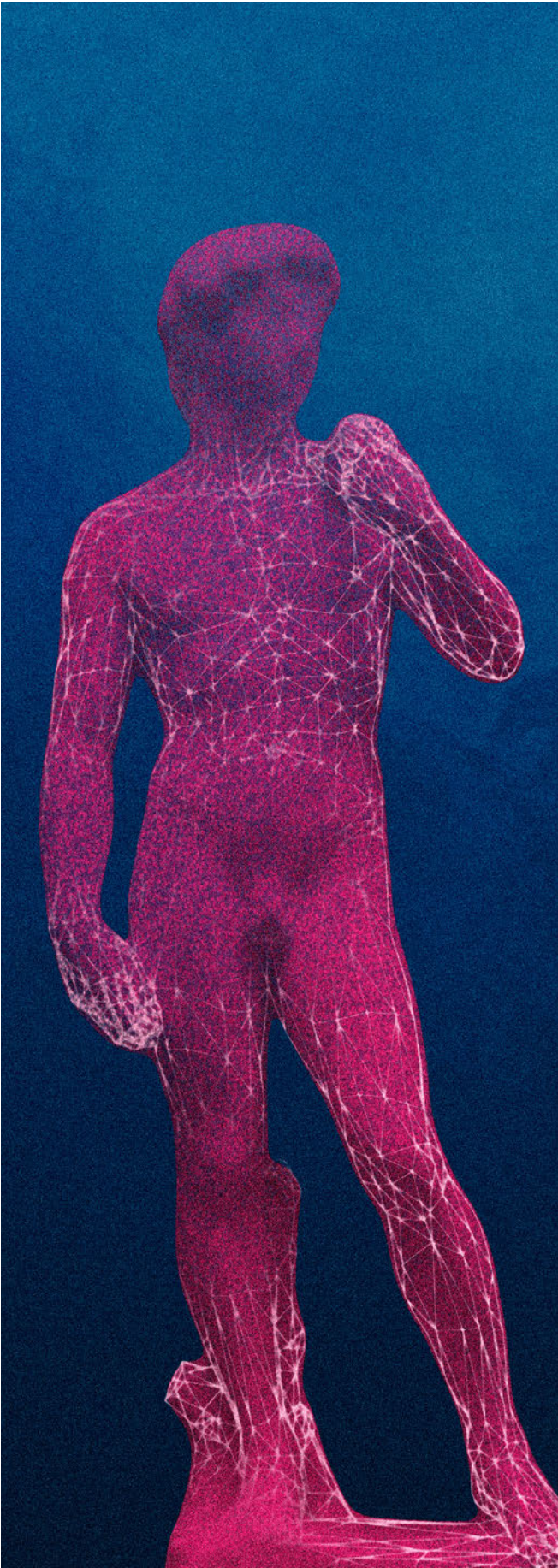
[Slot Machine]
Author, Flux

1.2.3. Validation Crisis

The architectural software market is presently experiencing what industry analysts classify as an "AI Bubble," paralleling the "Peak of Inflated Expectations" defined in the Gartner Hype Cycle (Gartner, 2023). This phase is characterized by an explosion of startups and platforms vying for market share, often prioritizing speed to market over functional depth. For the practicing architect, this oversaturation creates a condition of "analysis paralysis," where the sheer volume of available tools (each promising to revolutionize the design workflow) makes the selection of a viable technology stack nearly impossible.

This crisis is compounded by the prevalence of "Vaporware" in the AI sector: tools that are marketed with curated, cherry-picked demonstrations that far exceed their reliability in actual production environments. Many emerging platforms act merely as "wrappers" (User Interface skins) around standard open-source models like Stable Diffusion, yet they are marketed as proprietary "Architectural Intelligence." This marketing obfuscates the true capabilities of the software, leading firms to invest in tools that lack the reliability required for professional practice. As Phil Bernstein notes in his analysis of architectural technology adoption, the profession is historically slow to validate new tools empirically, often relying on marketing narratives rather than rigorous stress-testing (Bernstein, 2018).

This proliferation also produces Subscription Fatigue and workflow fragmentation. Instead of a unified environment like the BIM ecosystem, architects maintain disparate subscriptions for ideation (e.g., Midjourney), rendering (e.g., Veras), and upscaling (e.g., Magnific), with no interoperability between them. This "Tool-Siloing" blocks the free flow of data, turning the design process into a disjointed series of file exports and manual overrides. The research problem, therefore, is one of validation: which tools possess the "Production-Readiness" to survive the inevitable "Trough of Disillusionment" that follows the hype, rather than merely which tools exist.



[Pseudo]
Author, Nano Banana Pro



[Chaos into Order]
Author, Midjourney

1.3. Research Objectives

In response to the methodological fissures identified in the Problem Statement, this thesis establishes a set of targeted research objectives designed to operationalize generative AI within the professional design studio. Moving beyond the passive consumption of "black box" outputs, these objectives seek to construct a methodological framework that restores architectural agency.

1.3.1. Categorization of AI Tools

The primary objective is to establish a rigorous taxonomy for generative AI tools, moving beyond commercial branding to a classification based on functional application. Currently, architects face a "mismatch of agency," where they attempt to use divergent tools for convergent tasks, leading to inefficiency.

To resolve this, the thesis maps the AI tool field onto the RIBA Plan of Work stages, using the "Double Diamond" design process as a framework. This distinguishes between "Divergent Thinking" (ideation/expansion) and "Convergent Thinking" (refinement/execution). The research seeks to categorize tools as either "Semantic Explorers": tools like Midjourney that excel in RIBA Stage 2 by generating broad conceptual iterations; or "Geometric Rationalizers": approaches such as reference-image conditioning, Veras, or ControlNet that support RIBA Stage 3 by converging raster data onto vector constraints. By evaluating tools according to their plasticity (their ability to adhere to rigid inputs), this objective gives architects a strategic method for selecting the right tool for the specific maturity of the design.

1.3.2. Hallucination to Constraint

A critical objective is to demonstrate technical methods that manage the transition from "unconstrained hallucination" to "constrained synthesis." This research acknowledges that the "Black Box" stochastic nature of tools like Midjourney is highly beneficial during the early concept phases, where ambiguity feeds creativity. As the design matures, however, that lack of control becomes a liability.

To bridge this gap, this research aims to demonstrate workflows for "Geometric Locking": constraining the generative process with the architect's own geometry, the massing and dimensioned model views, so that the AI respects the design's form while manipulating surface attributes. This locking can be pursued by more than one route, principally reference-image conditioning, in which the model is supplied with views of the architect's own massing, or dedicated adapters such as ControlNet; the case study tests the former, since the strongest current models are closed and do not expose adapter-based control. Fidelity, this thesis argues, is not only visual; it also includes logical coherence. The investigation therefore validates the use of Large Language Models (LLMs) for "Logic Synthesis," such as structuring design narratives or generating Python code for parametric environments (e.g., Grasshopper). The goal is to prove that AI acts as a flexible co-pilot: offering "loose" agency for brainstorming and "tight" agency for definition.

1.3.3. Proposing the Roadmap

The final objective is to synthesize these findings into a structured "Roadmap": a pedagogical and professional guide for integrating AI into the design studio. Currently, AI adoption is often ad-hoc and disjointed, lacking a formal protocol for deployment. This results in workflows where AI is used as a post-rationalization device rather than a driver of design inquiry.

To counter this, the thesis proposes a structured roadmap for integrating AI into the design studio. This roadmap defines specific entry and exit points for AI intervention, establishing a protocol for "Human-in-the-Loop" (HITL) workflows. It categorizes interaction modes into "Prompting" (ideation), "Conditioning" (definition), and "Curating" (resolution). By mapping these modes against the requirements of each project phase, the roadmap aims to transform the "Black Box" of AI into a "Glass Box," where the logic of generation is transparent, reproducible, and subservient to the architect's vision.

1.4. Scope & Limitations

The scope of this research is explicitly bounded by the creative and iterative phases of the architectural design process, specifically corresponding to RIBA Stages 2 (Concept Design) and 3 (Spatial Coordination). This timeframe encapsulates the transition from abstract ideation to resolved massing and final visual communication. By focusing on these early to mid-stages, the research investigates the "frontend" of architectural production, where the generative ambiguity of AI tools acts as a productive virtue for exploring form, style, and atmosphere.

The research therefore excludes the later documentation and construction phases (RIBA Stages 4 and 5), which encompass Technical Design, Construction Documentation, BIM Data Integration, and Mechanical, Electrical, and Plumbing (MEP) systems. While the application of machine learning to these technical fields is a valid area of inquiry, it falls outside the boundaries of this specific investigation into the creative design studio.

This limitation follows from the "tricky" and stochastic nature of current generative models, which struggle with the deterministic precision required for fabrication and engineering. Bridging the gap between probabilistic AI and technical documentation is theoretically feasible, but it carries a magnitude of complexity that warrants a dedicated study of its own. A full investigation into AI-driven construction documentation would require a separate thesis, ideally conducted by researchers with specialized experience in computational engineering and systems integration. To keep the work focused, this study stays with the design and visualization workflows

chapter **two**

Literature Review & Methodology

2.1. Technological Context

Using artificial intelligence effectively in the design studio depends on first demystifying the vocabulary that surrounds it. The discourse is often clouded by marketing hyperbole and anthropomorphic metaphors that obscure what the software can actually do. For the architect, AI is best understood as a complex system of interacting components rather than a single, monolithic entity. This section breaks down these components to establish a clear operational framework for the research, moving from broad definitions to specific control mechanisms

2.1.1. Foundational Concepts

In the rigorous definition provided by Russell and Norvig, an Agent is any system that perceives its environment and takes actions to achieve a goal. In the context of generative design, the core of this agent is the Model. Think of the Model as the "brain" or the "engine" which is the complex mathematical file (often gigabytes in size) that has been trained on billions of images or texts. It contains the learned patterns (such as what a "brick" looks like). A common confusion arises between the product and the engine. For example, ChatGPT is merely a User Interface (UI), a chat window that allows humans to interact with an underlying model (e.g., GPT-4). Similarly, when an architect generates an image using Gemini, they are using a web interface to send instructions to a specific diffusion model (e.g. Nano Banana) running on a remote server. This distinction matters because different interfaces (e.g. ComfyUI vs. Gemini) can run similar models yet offer very different levels of control. The Prompt is the input stimulus. It is the bridge between human intent and machine execution. In text-to-image systems, the prompt is not just a description; it is a set of coordinates. It steers the model through its internal "latent space" toward a specific visual outcome. Mastering the prompt is equivalent to learning how to frame a view in photography; the subject exists, but the user must define how it is captured.

Two pervasive myths often hinder the architectural adoption of these tools. The Collage Myth is a widespread belief that Generative AI acts like a search engine, "copy-pasting" fragments of existing images (e.g., a window from Zaha Hadid and a door from Le Corbusier) to create a collage. This is incorrect. The model does not store images; it stores patterns and relationships (weights). When it generates an image, it synthesizes new pixels from scratch based on those learned probabilities, much like an artist drawing from memory rather than cutting up photographs. The second is the Physics Fallacy. Architects often assume that because an image looks photo-realistic, the building portrayed is structurally sound. Current diffusion models, however, prioritize Aesthetics (light, texture, composition) over Physics (gravity, load paths, material strength). The AI does not "know" that a concrete slab needs a column; it only knows that concrete slabs usually appear above columns in its training data. This leads to "hallucinations" which are architectural errors that look convincing but are structurally impossible.

2.1.2. Technical Hierarchy

Historically, the discipline has evolved from early knowledge-based systems that relied on hard-coded logical rules defined by humans to modern data-driven architectures where the system extracts its own rules and patterns from vast datasets. Within this evolution, artificial intelligence systems are categorized based on their capability and scope relative to human intelligence. The vast majority of existing architectural tools fall under Artificial Narrow Intelligence (ANI), also known as Weak AI, which is designed to perform a specific task such as image recognition or language translation with superhuman proficiency but lacks consciousness or versatility. This contrasts with the theoretical concept of Artificial General Intelligence (AGI), or Strong AI, which would possess the ability to understand, learn, and apply knowledge across a

wide variety of tasks much like a human being. The trajectory of the field points toward a hypothetical future state known as Artificial Super Intelligence (ASI), where machine cognition would surpass human intellect across all fields, including creativity and social skills.

The technological context of this thesis is best pictured as a nested hierarchy, in which Artificial Intelligence is the broad umbrella discipline encompassing all intelligent agents. Within AI resides Machine Learning (ML), a subset focused on algorithms that learn from data. Inside Machine Learning is Deep Learning (DL), which uses multi-layered neural networks. Finally, Generative AI exists as a specialized application within Deep Learning, focused on the synthesis of new data rather than the analysis of existing data.

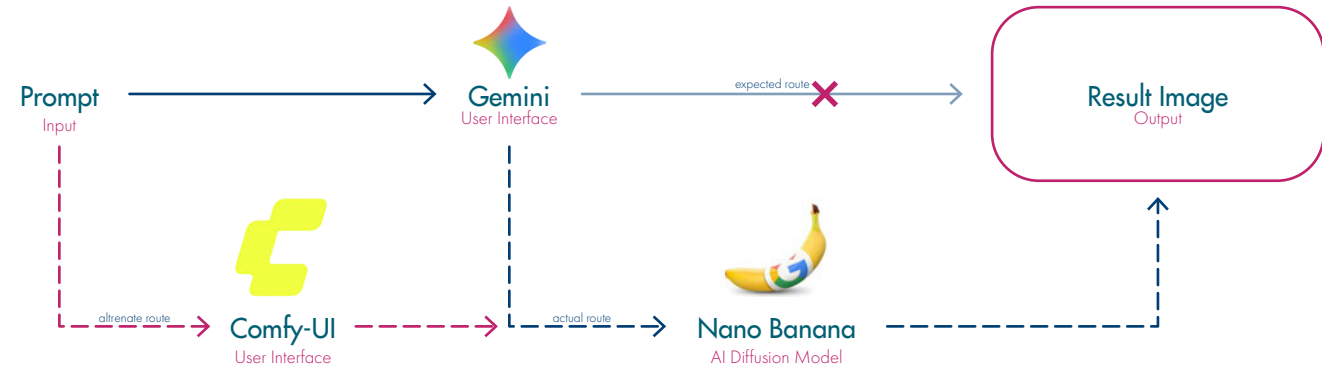
At the heart of this hierarchy lies Machine Learning, a field that shifts the programming paradigm from explicit instruction to inductive inference. Arthur Samuel, a pioneer in the field, defined it as the field of study that gives computers the ability to learn without being explicitly programmed (Géron, 2022). A more formal and engineering-oriented definition was provided by Tom Mitchell, who stated that a computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E (Mitchell, 1997). This process effectively automates automation by allowing the algorithm to analyze data, such as thousands of floor plans, to discern the rules of architecture itself rather than having a programmer define specific dimensions for a wall.

Deep Learning represents a sophisticated evolution of machine learning inspired by the neural structures of the biological brain. Traditional machine learning often struggles with raw, unstructured data like images because it relies on humans to manually define relevant features. Deep Learning solves

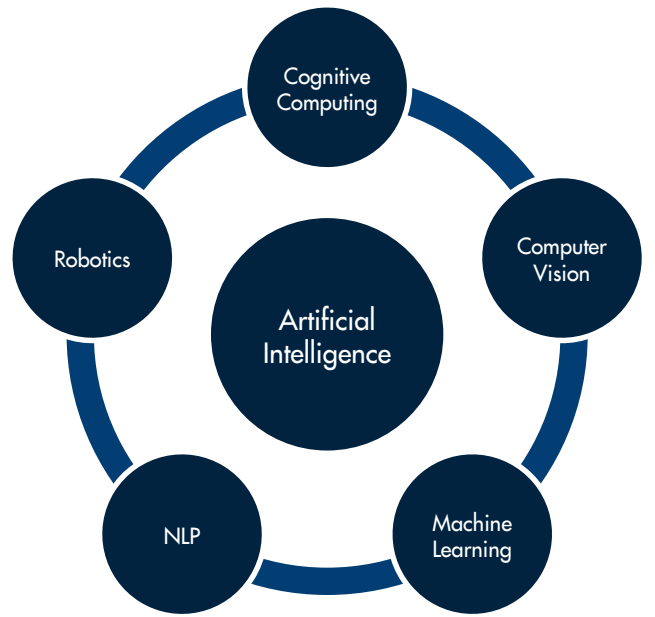
this through representation learning, allowing the computer to build a hierarchy of concepts. As explained by Goodfellow, Bengio, and Courville (2016), a deep neural network learns to represent the world as a nested hierarchy of concepts, with each concept defined through its relation to simpler ones. In analyzing an image, the initial layers might detect simple edges, subsequent layers combine these edges to identify shapes, and deeper layers assemble these shapes into complex objects like windows or columns.

Generative AI marks the transition from discriminative models that classify data to generative models that synthesize it. While discriminative AI draws boundaries to distinguish between a cat and a dog, Generative AI learns the underlying distribution of the data to create entirely new samples that plausibly belong to the original set. This field is further subdivided based on input and output modalities, including Text-to-Text models for language generation, Text-to-Image models for visualization, Text-to-Audio for sound synthesis, and Text-to-Video for motion. In the architectural context, Text-to-3D and Image-to-3D models are emerging as critical tools for spatial generation.

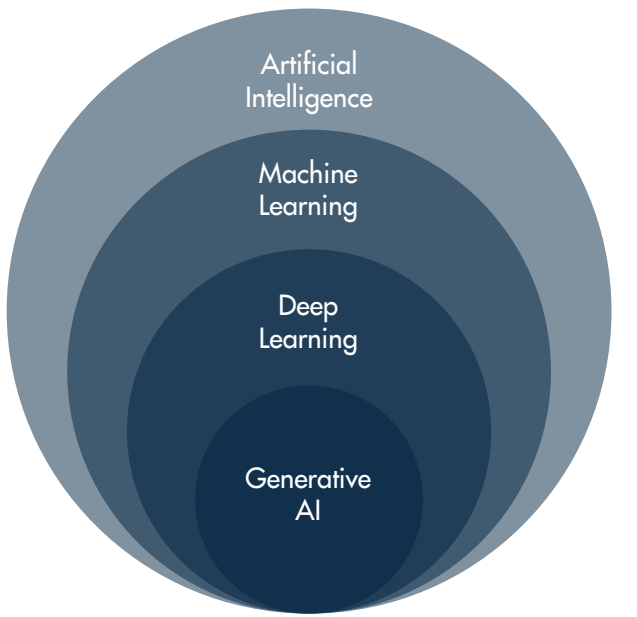
The engine driving these capabilities is the Artificial Neural Network (ANN), a computational model mimicking the biological neuron. An ANN consists of an Input Layer that receives raw data, multiple Hidden Layers that process this data through mathematical transformations, and an Output Layer that delivers the final prediction or generation. The connections between these neurons are governed by parameters known as weights, which determine the strength and influence of the signal passing between them. Training a neural network involves a process of forward propagation where data moves through the network to produce a result, followed by backpropagation where the error is calculated and the weights are adjusted to minimize that error over time.



[Generation Process]
Author



[AI Subsections]
Author



[Types of AI]
Author

[Neural Networks]
Author

2.1.3. Operational Parameters

The more advanced operation of generative AI (e.g. ComfyUI) calls for a defined technical vocabulary. These terms represent the "levers" available to the architect for controlling the stochastic output of the model.

Weights and Parameters: In the context of a neural network, Parameters refer to the total number of internal variables (primarily weights and biases) that the model learns during training. A model with more parameters generally possesses a higher capacity to learn complex patterns (Kaplan et al., 2020). Weights are the specific learnable values that determine the strength of the connection between neurons across layers. During the training of a diffusion model, these weights are iteratively adjusted to minimize the error in noise prediction (Goodfellow et al., 2016).

Tokens: Neural networks cannot process raw text. Tokens are the fundamental units of linguistic data (i.e. words, sub-words, or characters) that are converted into numerical vectors (embeddings) for the model to process. In architectures like CLIP, the text prompt is tokenized and mapped into a high-dimensional space where semantic relationships are preserved mathematically (Vaswani et al., 2017).

Seed: Generative AI is inherently stochastic, meaning it relies on randomness. The Seed is the initial integer used to initialize the pseudo-random number generator that creates the starting noise field. Fixing the seed allows for reproducibility; if the same prompt and settings are used with the same seed, the model will generate the exact same image, allowing for controlled iteration (Dürr et al., 2020).

Denoising Strength: This parameter is specific to Image-to-Image (Img2Img) workflows. It controls the magnitude of noise added to the input image before the reverse diffusion process begins. A low denoising strength (e.g., 0.3) retains the original geometry and structure, resulting in minor stylistic changes. A high denoising strength (e.g., 0.8) allows the model to deviate widely from the input, hallucinating new

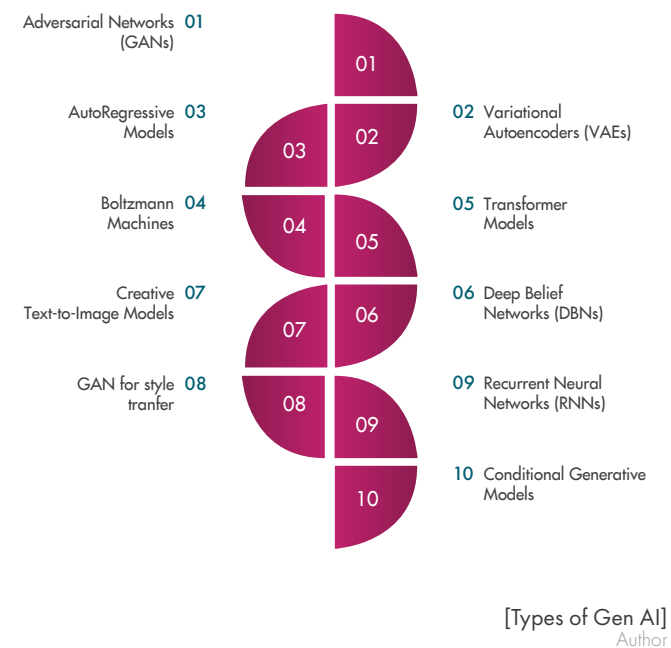
forms and compositions (Rombach et al., 2022).

CFG Scale (Classifier-Free Guidance): The CFG Scale controls the "obedience" of the model to the text prompt. It balances the generation between the unconditional distribution (what the model "wants" to generate based on its training) and the conditional distribution (i.e. what the user asked for). A higher CFG scale forces the image to align more strictly with the prompt but can lead to visual artifacts or "frying," while a lower scale produces more creative but less relevant results (Ho & Salimans, 2022).

2.1.4. Mechanics of Creation

The specific architecture powering the current revolution in architectural visualization is the Denoising Diffusion Probabilistic Model (DDPM), a generative framework inspired by non-equilibrium thermodynamics. Unlike Generative Adversarial Networks (GANs), which attempt to produce an image in a single conceptual leap (Goodfellow et al., 2014), diffusion models operate through a methodical, iterative process of refinement. This process splits into two phases: forward diffusion and reverse diffusion. During training (forward diffusion), the system systematically destroys data by adding Gaussian noise to an image over a series of steps until the visual information is completely obliterated into pure, random static. The model essentially learns the statistical gradient of this destruction. The generation phase (reverse diffusion) exploits this learned gradient to reverse the arrow of time. Starting with a canvas of pure random noise, the AI iteratively subtracts the noise to recover a coherent image, effectively "sculpting" structure out of chaos (Ho et al., 2020).

To make this computationally feasible on standard consumer hardware, modern architectures like Stable Diffusion do not operate directly on the high-dimensional data of raw pixels. Instead, they use Latent Diffusion Models (LDMs). This

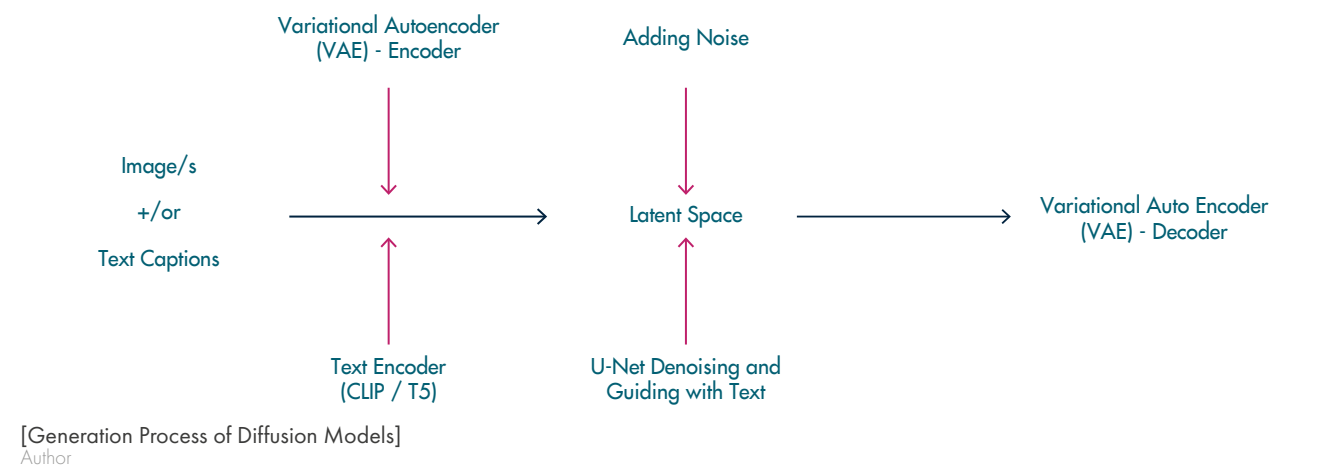
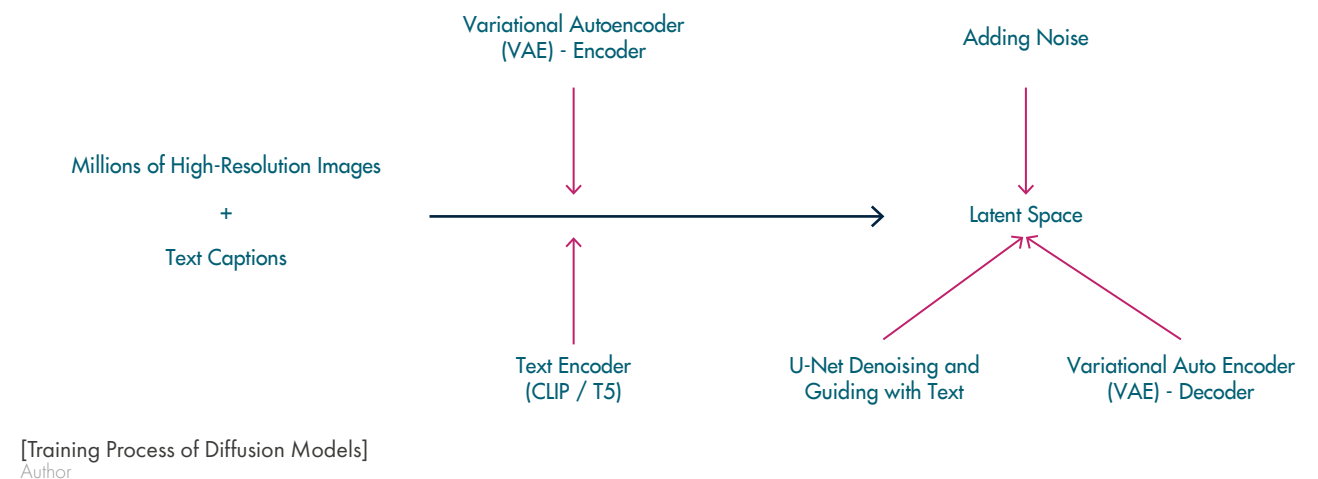


approach involves a Variational Autoencoder (VAE), a neural component that compresses the visual data into a lower-dimensional representation known as "Latent Space" (Kingma & Welling, 2013). This space retains the perceptual essence of the image while discarding redundant information, allowing the heavy mathematical lifting to occur efficiently. Once the diffusion process is complete within this compressed space, the VAE acts as a decoder to translate the latent mathematics back into a high-resolution pixel image (Rombach et al., 2022). This separation of "perceptual compression" (VAE) from "semantic generation" (Diffusion) is what allows architects to generate 4K renderings on a local GPU.

The engine driving the actual denoising process is typically a U-Net neural network. The U-Net is tasked with a specific prediction problem: given a noisy image and a specific time step, it must predict exactly how much noise exists so it can be subtracted. It accomplishes this through a series of down-sampling layers that capture the broad context of the image (composition and geometry) and up-sampling layers that resolve the fine details (texture and edges). Crucially, this neural network does not work in isolation. It is connected to the user's intent through "Cross-Attention" layers. These layers act as bridges where the visual processing intersects with linguistic data, allowing the prompt to steer the noise prediction (Ronneberger et al., 2015).

This linguistic steering is managed by CLIP (Contrastive Language-Image Pre-training), a separate neural network developed by OpenAI. CLIP acts as a semantic translator that converts the architect's text prompt into numerical vectors (embeddings) that the U-Net can understand. It ensures that when an architect types "concrete facade," the U-Net steers the denoising process toward the mathematical cluster representing concrete textures and planar geometry within the latent space. Without CLIP, the diffusion model would generate coherent images, but they would be random and unrelated to the user's request (Radford et al., 2021).

Finally, the speed and quality of this iterative process are governed by the Scheduler (or Sampler). If the U-Net is the engine and CLIP is the steering wheel, the Scheduler is the transmission. It determines how big of a step to take during the denoising process. Classical schedulers might require 1000 steps to produce a clean image, while modern solvers like Euler Ancestral or DPM++ can achieve the same result in as few as 20 steps by making smarter approximations of the noise gradient. This component matters for the architect, since different schedulers impart subtle stylistic differences to the final render, ranging from softer, painterly outputs to sharper, more photorealistic definitions (Karras et al., 2022).



2.2.Theoretical Framework

The integration of Artificial Intelligence into architecture is not merely a technical upgrade; it challenges the ontological status of the designer. To navigate this shift, this thesis grounds its methodology in two primary theoretical constructs: the distinction between Augmentation and Automation, and the reframing of algorithmic error as Productive Hallucination.

2.2.1. Augmentation v Automation

The discourse surrounding AI in labor is often dominated by the fear of Automation: the replacement of human agency with machine efficiency. The theoretical lineage of architectural computing offers an alternative trajectory, Augmentation. This concept traces back to J.C.R. Licklider's 1960 paper, *Man-Computer Symbiosis*, where he argued for a system that thinks with the human rather than for the human. Licklider envisioned a partnership where the machine handles "routinizable, clerical operations" while the human provides "insights and decisions" (Licklider, 1960).

In the architectural context, this framework was expanded by Nicholas Negroponte in *The Architecture Machine* (1970). Negroponte argued against the computer as a mere drafting tool (a "slave"), proposing instead a "design partner" capable of dialogue. He famously stated, "The dialogue would be so intimate that the two were phenomenologically one" (Negroponte, 1970). This thesis adopts this cybernetic stance. It positions generative AI as an Intelligence Amplifier (IA) rather than an "Automated Architect" that generates finished buildings from a single prompt.

In this framework, the AI acts as a high-speed "stochastic lateral thinker." It provides the Breadth (generating hundreds of variations, styles, and typologies in seconds) while the architect provides the Depth (curating, refining, and applying structural logic). This aligns with Randy Deutsch's definition of the "Superuser," where the architect's value shifts from manual production to the strategic orchestration of digital processes (Deutsch, 2019). Success in the workflow is therefore measured by how far the architect's design intent is expanded beyond cognitive bias, not by how much human labor is removed.

2.2.2. Concept of "Hallucination"

In computer science, a "hallucination" is a failure state: an instance where a model generates false or nonsensical information. Within the theoretical framework of this thesis, however, hallucination is reframed as a generative virtue. This draws upon the concept of "Defamiliarization" (or *Ostranenie*), originally coined by Viktor Shklovsky (Shklovsky, 1917), which argues that the function of art is to make objects "unfamiliar" to force the viewer to see them with fresh perception.

Matias del Campo applies this to AI in *Neural Architecture*, arguing that because diffusion models operate on statistical correlations rather than tectonic rules, they inherently produce "alien" geometries that bypass human convention. An AI might merge a column into a window, or dissolve a floor slab into a fluid canopy. While structurally incorrect, these "glitches" often contain novel aesthetic qualities that a human architect, trained in the rigid logic of gravity and construction, would never intuitively draw (del Campo, 2022).

This thesis posits that the "Integration Gap" identified in the Problem Statement can be bridged by treating these hallucinations as "Provocations" rather than "Solutions." In the early design phases (RIBA Stage 2), the AI's lack of physics allows it to explore a "latent space" of pure form, unencumbered by the immediate constraints of constructability. The method proposed in Chapter 3, therefore, aims to tame hallucination rather than eliminate it: constraining the model with the architect's own geometry to preserve the "strange" aesthetic qualities of its output while enforcing the geometric discipline the project needs to exist in reality.



2.3. Methodology: Research through Design

To investigate the operational role of Artificial Intelligence in architecture, this thesis adopts a Research through Design (RtD) methodology. Originally articulated by Christopher Frayling in his 1993 paper *Research in Art and Design*, this approach distinguishes itself from research into design (historical study) or research for design (technical data gathering). Instead, RtD posits that the act of designing itself is a method of knowledge production (Frayling, 1993). In this context, the design artifacts (the renders, the node graphs, and the workflows) are not merely illustrations of the research. They are the research results.

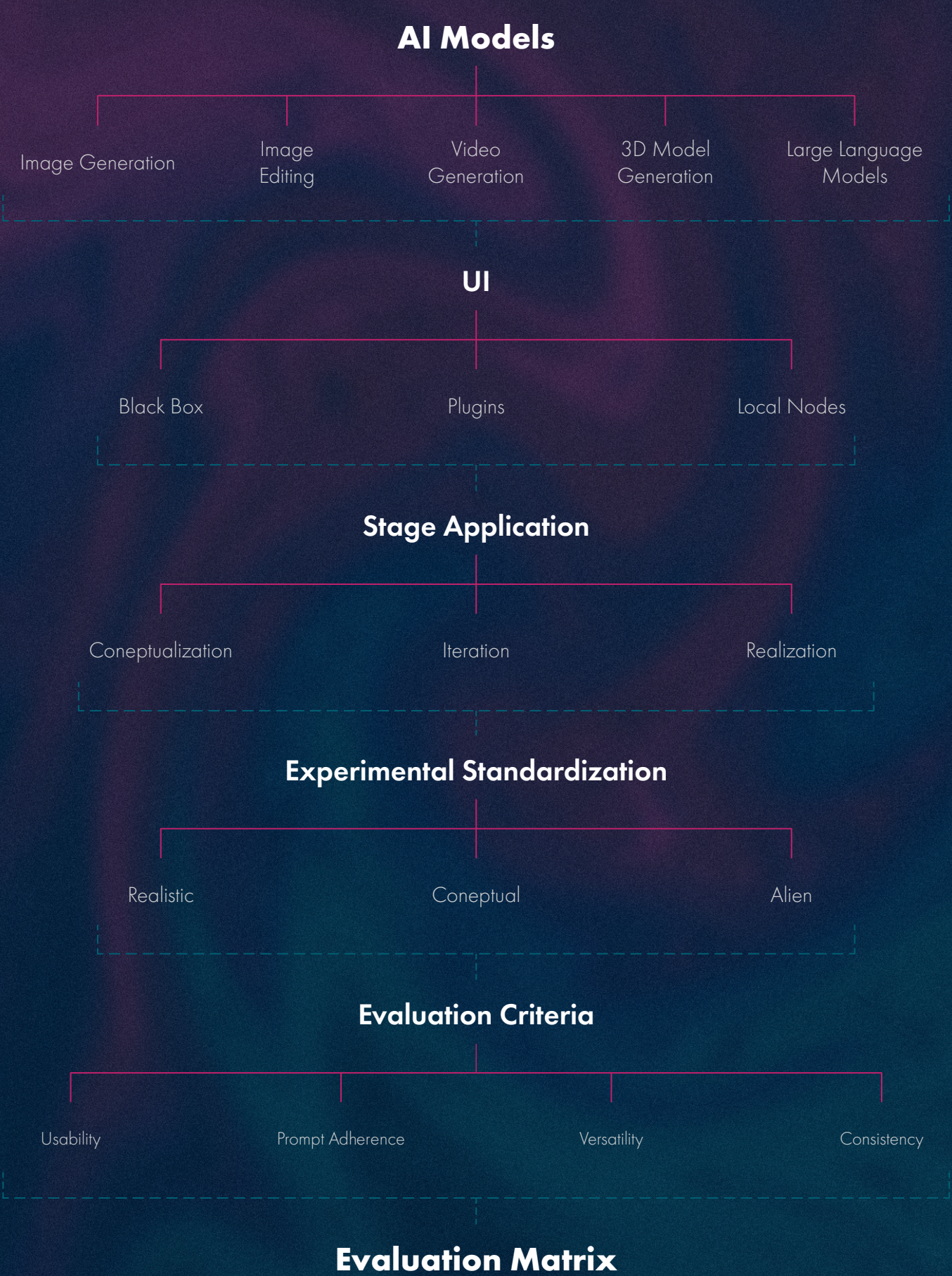
This methodology is particularly suited to interaction design and emerging technologies where standard protocols do not yet exist. As Zimmerman, Forlizzi, and Evenson argue, researchers using RtD "make the right thing" to transform the current state of the world into a preferred state (Zimmerman et al., 2007). By actively engaging with the software, failing, modifying parameters, and testing again, this thesis generates knowledge about the limitations and affordances of generative AI that could not be discovered through literature review alone. The "artifact" produced by this thesis is therefore the Workflow Framework itself, not only the final building design presented in Chapter 4.

The research design begins with a mapping of the AI field in Chapter 3. This initial phase categorizes available models based on their primary output modalities, specifically distinguishing between image generation, image editing, video synthesis, 3D model generation, and text-based logic. This scope intentionally excludes modalities such as audio generation or speech recognition to maintain relevance to the visual nature of architectural design. To further refine this taxonomy, the tools are subsequently classified by their User Interface architectures into Black Box systems, Integrated Plugins, and Local Nodes. This distinction allows for a critical analysis of the accessibility and control offered by each platform type, weighing the immediate ease of use against the depth of customization.

Following this categorization, the research investigates specific applications across the three primary design stages, the conceptual, schematic, and realization phases. At each stage, distinct use cases are identified to demonstrate how different tools interface with the evolving maturity of a design project. To enable a standardized comparison across these diverse tools, the experimental output is simplified into three architectural styles. These are defined as Realistic for standard visualization, Conceptual for abstract representation, and Alien for defamiliarized aesthetics. This stylistic triangulation ensures that the performance of a model is tested against varying degrees of abstraction and precision.

The outputs from these experiments are then assessed against four criteria. Usability checks the output for artifacts and defects, Prompt Adherence measures the model's obedience to the given instructions, Consistency gauges its stability across repeated iterations, and Versatility records its ability to handle diverse tasks. These assessments are synthesized into a structured Evaluation Matrix that highlights which models excel in specific design contexts.

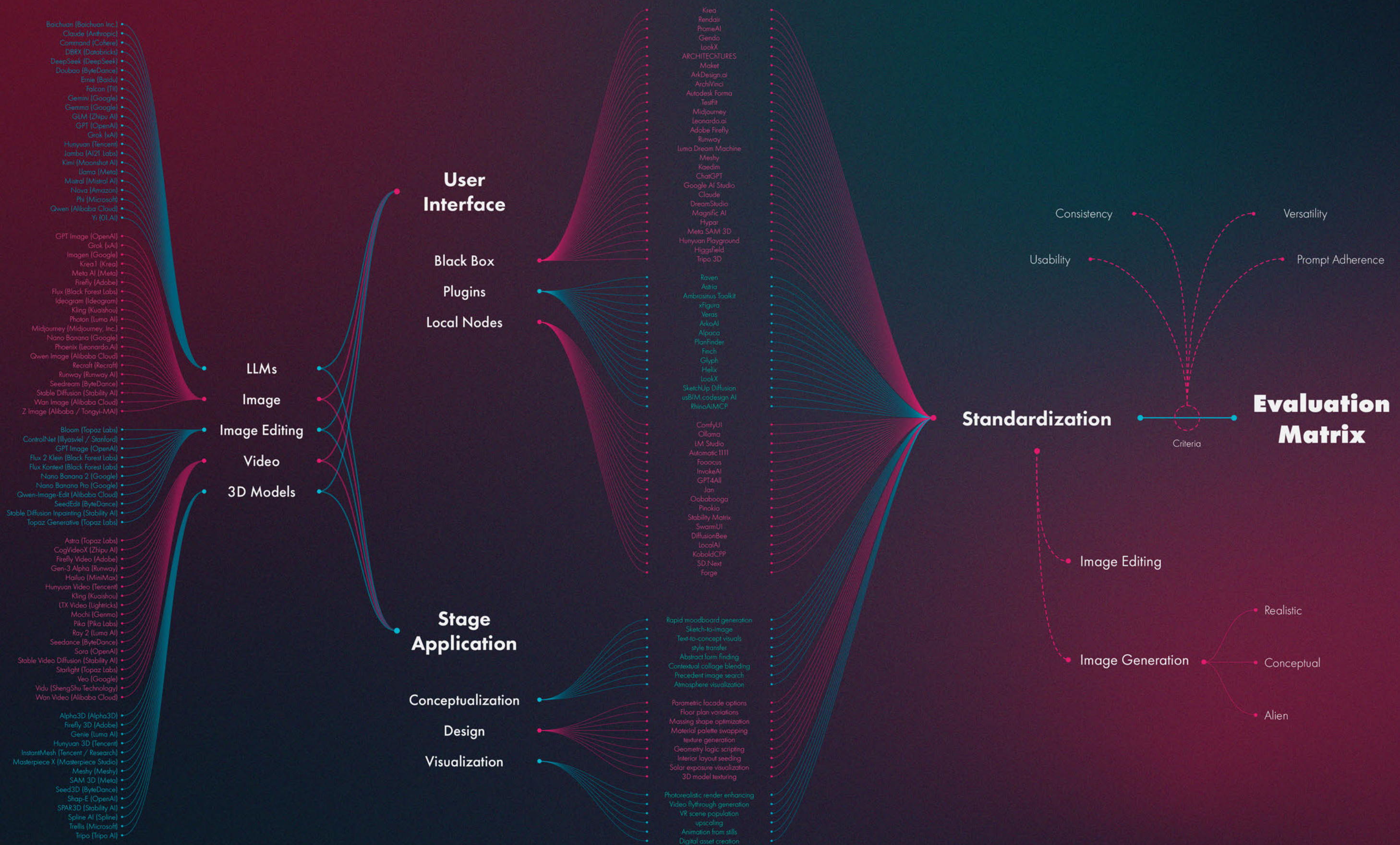
Building upon the comparative testing, Chapter 3 concludes with evaluation matrices that establish which models perform best within each task. Chapter 4 then integrates the prior chapters into a full case study. This architectural project serves as a demonstration of the proposed framework, applying the roadmap from the initial concept sketch through to the final presentation render, and showing how the architect's own geometry is used to constrain the generative output.



chapter **three**

Tools Landscape





3.1. Stage Application

This section investigates specific use cases for AI tools across three primary stages of the design process. The applications detailed here are only those explicitly tested within the scope of this research, and they are by no means an exhaustive list, since the potential uses of these tools are vast. The grouping of applications into phases is itself a subjective choice. A tool used for ideation could equally serve as a mechanism for final realization, depending on the user's intent. The three phases are themselves a speculative framework, proposed to help structure the integration of these emerging technologies into the architect's daily workflow.

The proposed framework delineates the workflow into three sequential but overlapping territories. The first phase, the conceptual phase, addresses the initial genesis of the project where the primary goal is the rapid expansion of the solution space through divergent thinking and narrative exploration. Following this, the schematic phase represents the core of the architectural struggle, focusing on the translation of abstract concepts into defined geometries through a cyclical process of testing, refining, and rejecting options. Finally, the realization phase concerns the production of polished visual artifacts intended for communication, shifting the focus from internal discovery to external persuasion through high-fidelity rendering and contextualization.

3.1.1. Conceptualization

Concept Generation

Large Language Models (e.g. ChatGPT) were utilized to generate preliminary architectural concepts and narratives. These textual concepts were subsequently fed into image generation models to visualize the spatial and atmospheric qualities. This workflow proves highly useful in the early stages of a project for rapidly expanding the initial solution space before any formal drawing begins.



Hanging School Clipped to a Cliff Face.

Author, ChatGPT, Flux

A school suspended on a sheer mountain cliff face, modular classrooms hanging from steel anchors drilled into rock, exposed walkways and bridges connecting volumes, dramatic vertical drop beneath, clouds and mist moving through the scene, raw structural expression, students as small figures emphasizing scale, high contrast lighting, tense and gravity aware composition



Data Center Monastery in a Remote Forest.

Author, ChatGPT, Midjourney

A remote forest monastery reimagined as a data center, massive stone walls, cloistered courtyards, long silent corridors, server halls replacing chapels, subtle glow of technology within austere spaces, misty woodland surroundings, restrained color palette, calm symmetrical composition, atmosphere of ritual and quiet power.



Tidal Archive Library on a Mangrove Coast

Author, ChatGPT, Krea 1

Architectural visualization of a public library built within a coastal mangrove landscape, elevated structural cores above shallow tidal water while floating reading platforms drift between them, water partially flooding the lower levels, reflections of timber and concrete structures on calm water, dense mangrove roots integrated into foundations, humid atmosphere, soft diffused daylight, emphasis on coexistence between architecture and tidal cycles, cinematic wide angle composition.



Refugee Housing System Embedded in a Border Wall.

Author, ChatGPT, Krea 1

A massive border wall transformed into an inhabitable structure, thickened wall containing housing units, clinics, schools, and markets carved into its depth, long linear perspective, repetitive openings with signs of daily life, harsh landscape on both sides, architecture expressing tension between control and habitation, documentary style realism.



Floating Prison in International Waters.

Author, ChatGPT, Krea 1

A stark floating prison platform in open international waters, rigid grid of identical cell blocks, repetitive modular architecture, no land visible on the horizon, cold industrial materials, overcast sky, controlled circulation paths, sense of isolation and political ambiguity, aerial oblique view emphasizing detachment and scale.



Vertical Cemetery in a Dense Megacity.

Author, ChatGPT, Krea1

A dense urban high rise cemetery designed as a vertical tower, stacked burial chambers and memorial niches visible behind perforated concrete and stone facades, slow spiral circulation ramp wrapping the interior void, solemn lighting with shafts of light cutting through darkness, no greenery, heavy materiality, surrounding city skyline pressing closely around the building, moody and introspective architectural rendering.



Mountain Observatory That Slowly Rotates Over Decades.

Author, ChatGPT, Flux

A mountain observatory anchored to a heavy stone base with a rotating upper structure, circular forms aligned with the sky, snow and rock surrounding the building, clear night atmosphere with visible stars, long exposure feel, architecture responding to celestial movement, minimal human presence, timeless and monumental composition.



Underground Market Beneath a Historic City Square.

Author, ChatGPT, NanoBanana Pro

Cutaway architectural perspective of a historic city square revealing a dense underground market below, layered spaces carved beneath stone paving, light shafts piercing from above, crowded stalls and circulation paths organized organically, contrast between calm classical surface above and chaotic vibrant interior below, warm artificial lighting underground, strong sectional clarity.



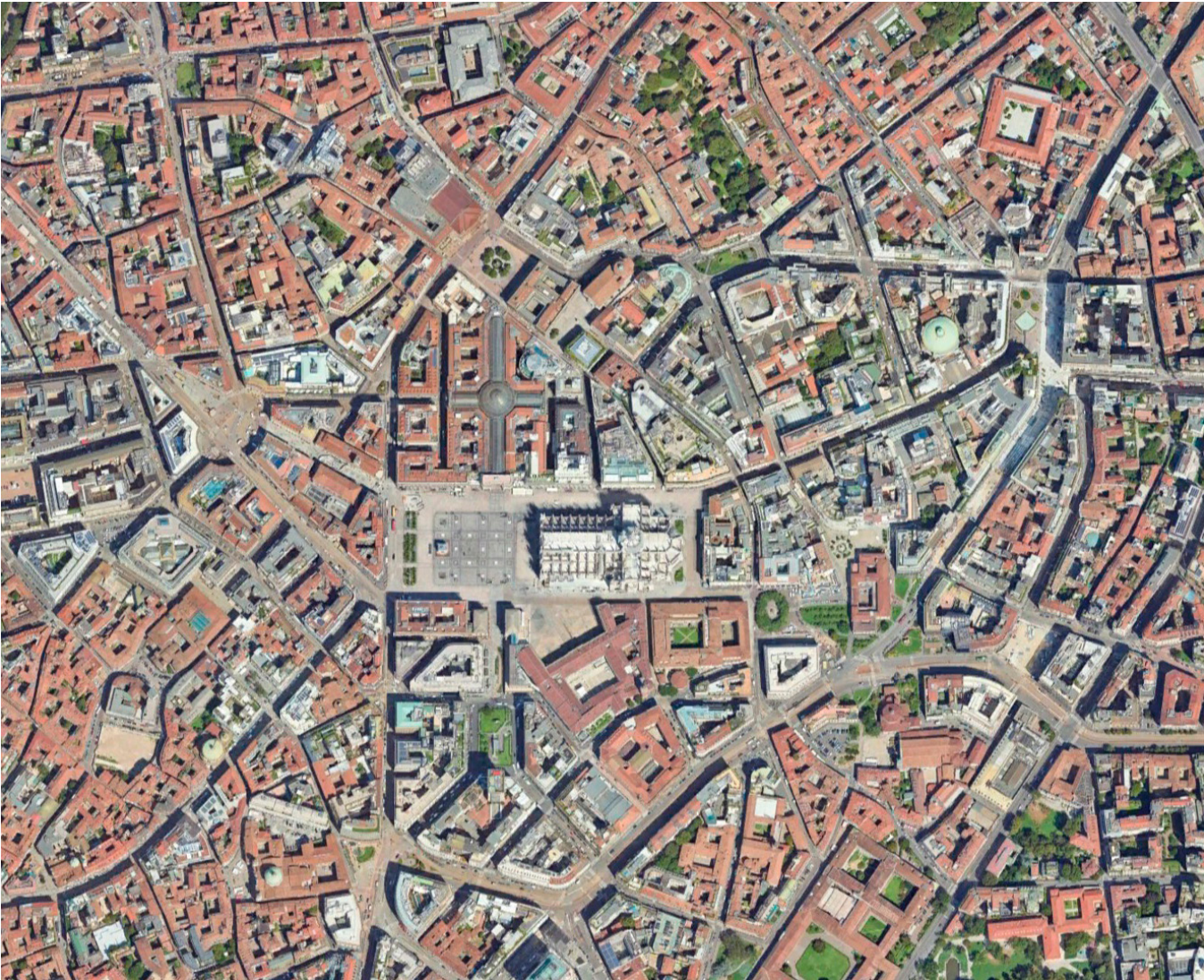
Desert Research Station That Buries Itself Over Time.

Author, ChatGPT, NanoBanana Pro

A desert research facility partially swallowed by sand dunes, exposed upper structures eroding and collapsing into buried subterranean volumes, thick walls and angular forms resisting heat, wind carved surfaces, minimal openings, harsh sunlight and deep shadows, vast empty desert context, atmosphere of isolation and entropy, architectural photography style realism.

Map Generation

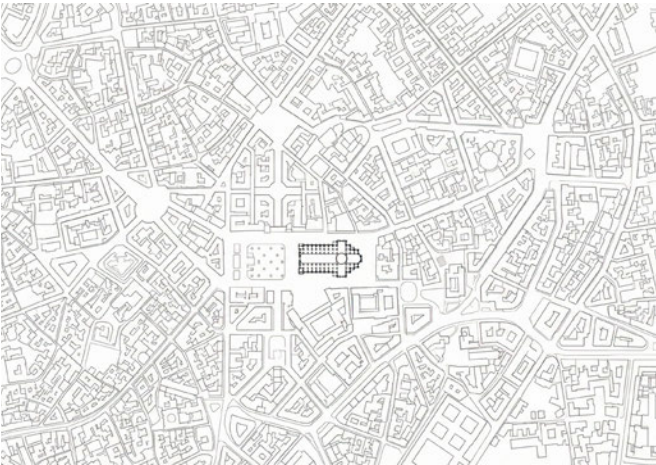
Image generation models were employed to transform standard Google Earth satellite imagery into usable, stylized line drawings. This allows for the rapid generation of site maps and context plans in various graphical styles without the need for tedious manual tracing.



Google Earth



Convert the attached satellite imagery to a realistic perspective top-view drone image.



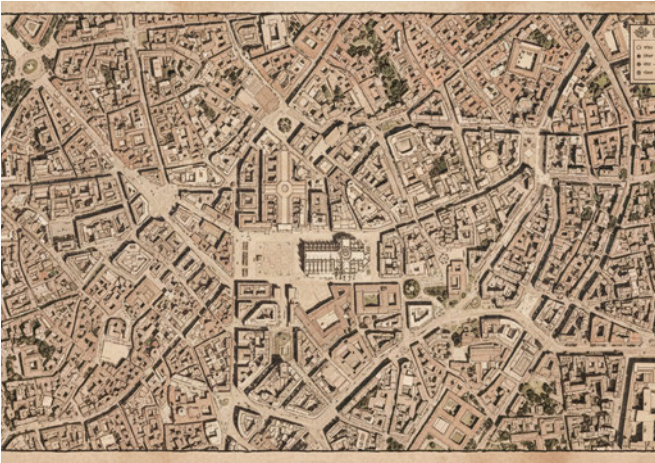
Convert the attached satellite imagery to a CAD Line Drawing.



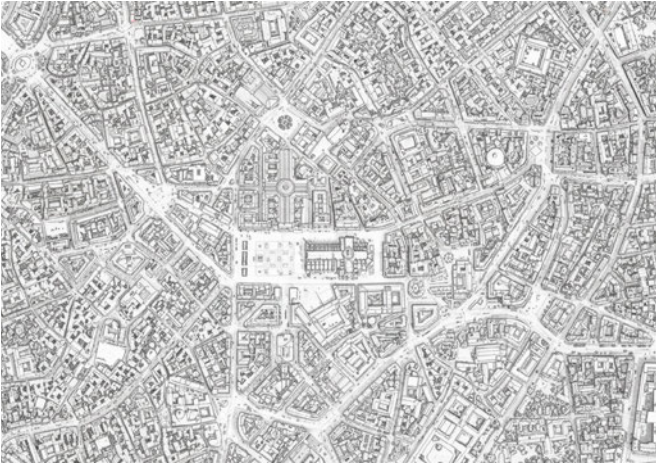
Convert the attached satellite imagery to a flat colored top-view orthographic image.



Convert the attached satellite imagery to a Solid & Void Map.



Convert the attached satellite imagery to an old map drawing.

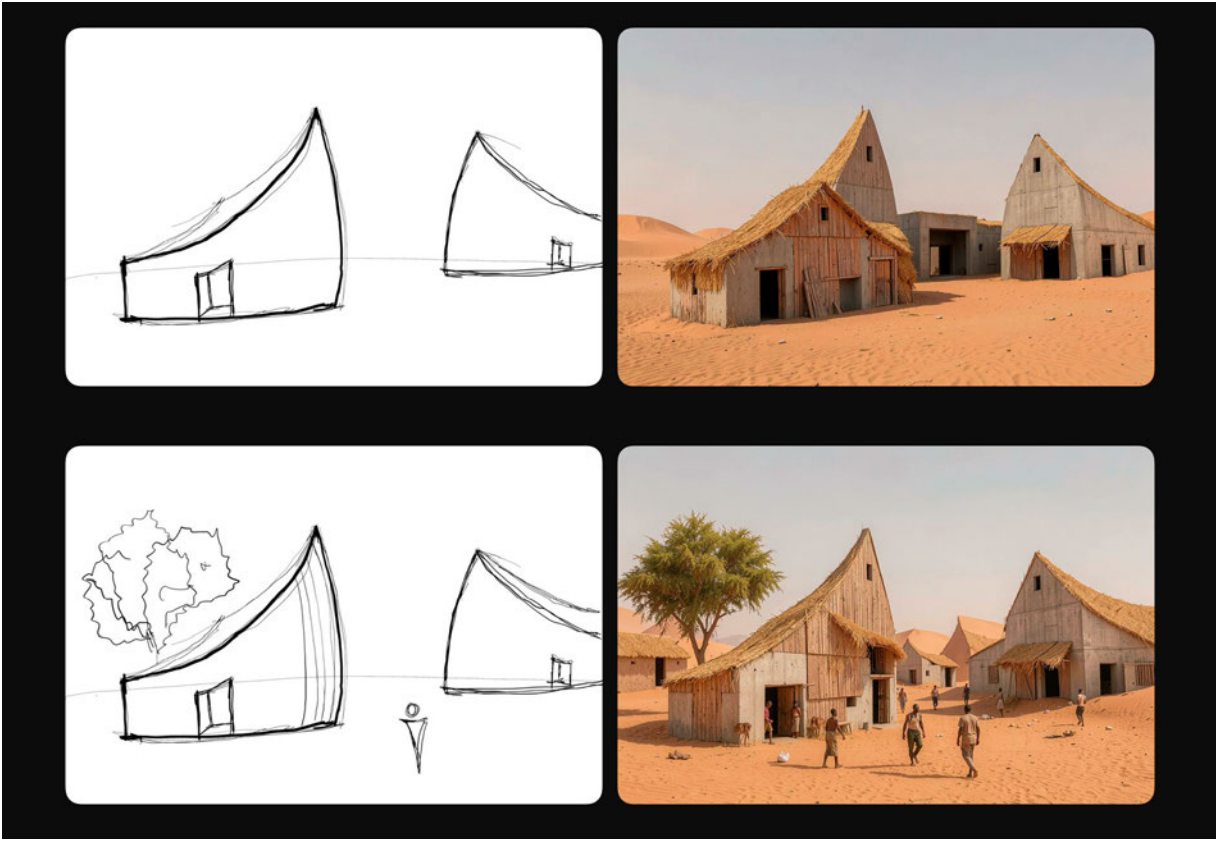
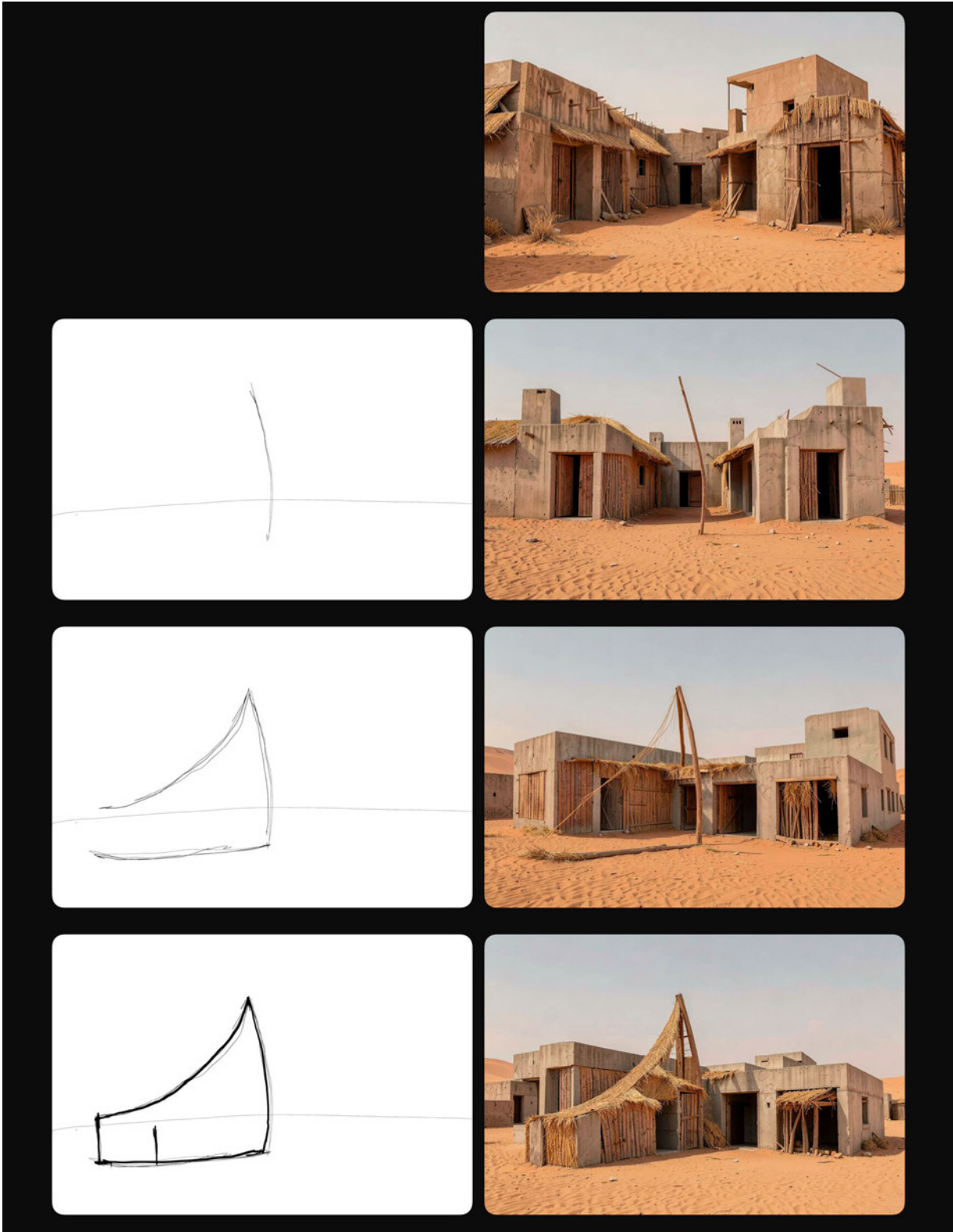


Convert the attached satellite imagery to a CAD Line Drawing.

Author, NanoBanana Pro

Realtime Sketch Realization

Utilizing realtime image generation tools (such as Krea), rough conceptual sketches were instantly translated into realized visualizations. As the user draws or modifies strokes, the AI continuously updates a high-fidelity render, offering maximum flexibility and immediate visual feedback during the form-finding process.



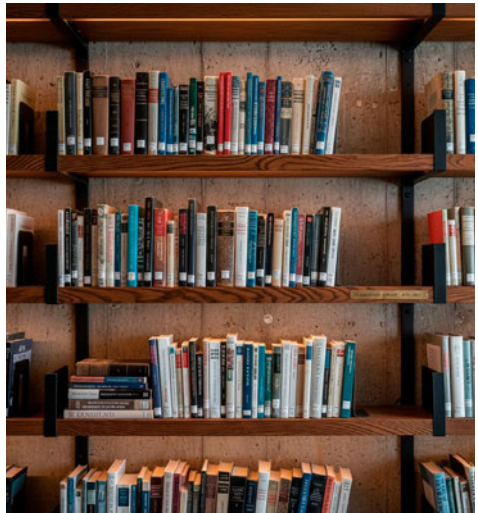
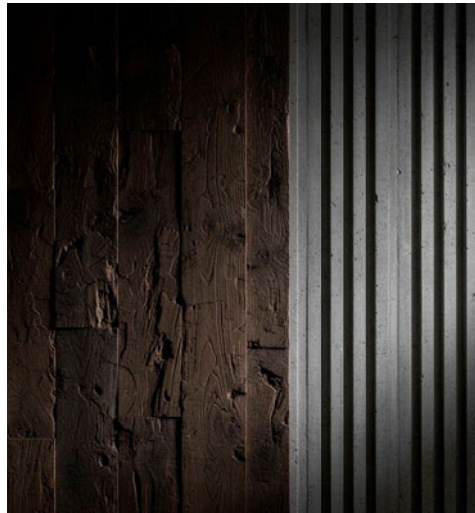
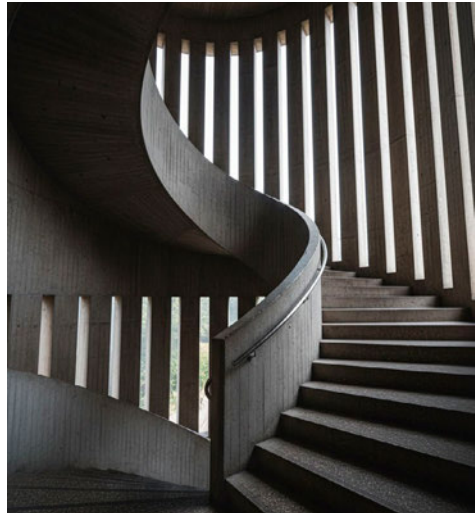
Author, Krea



Author, Krea

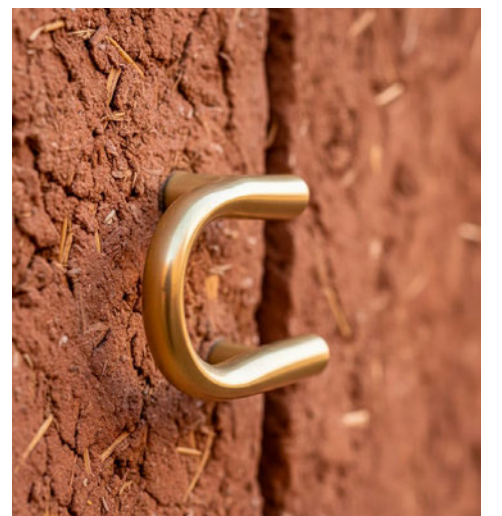
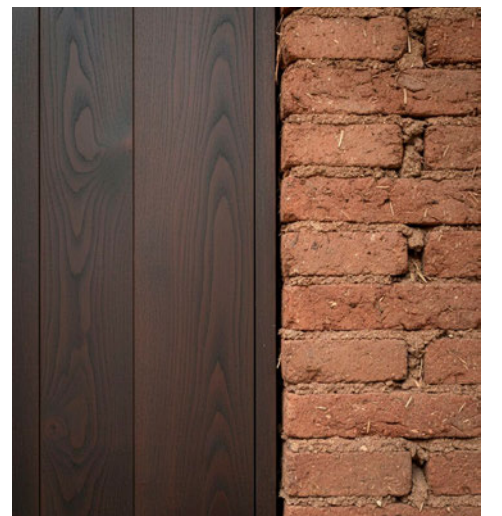
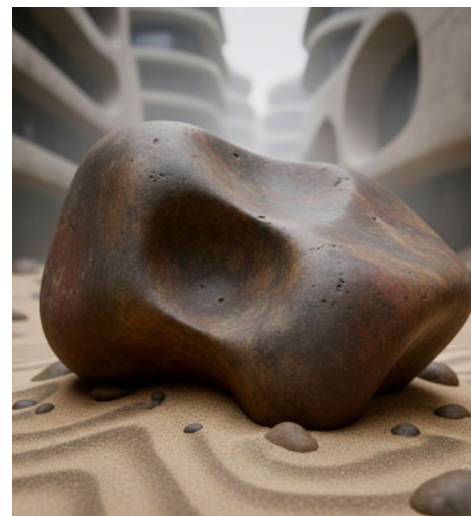
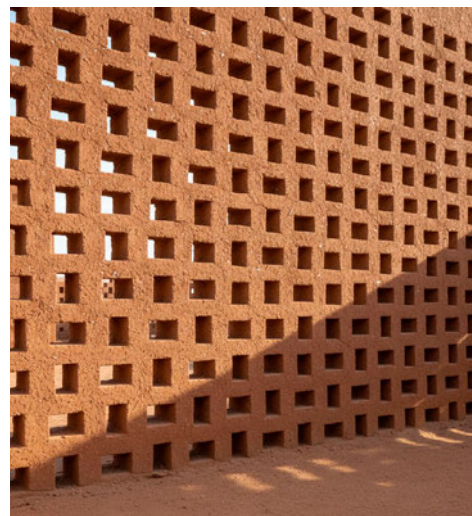
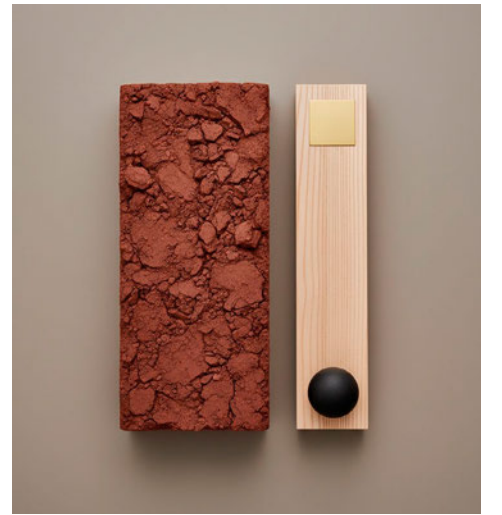
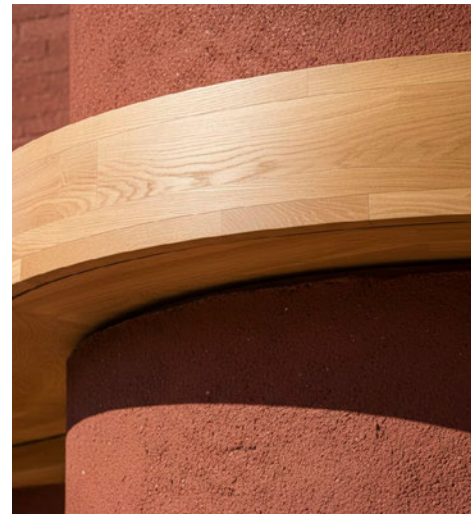
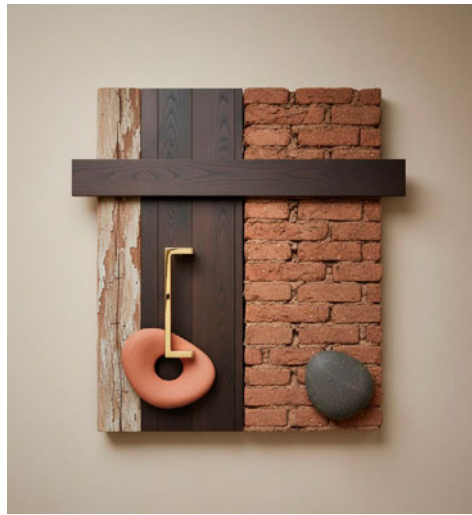
Rapid Moodboarding

Generative layout tools (such as Google Mix Board) were used to synthesize comprehensive architectural mood boards. With a single text prompt, the system can output a cohesive collage detailing the building massing, interior atmospheres, and specific material textures, establishing a project's visual identity almost instantly.



Brutalist Library
Author, Google's MixBoard, NanoBanana Pro

Create a highly detailed mood board of the architectural elements of a Library. The Building has a Brutalist style to it, set in a snowy mountain top. The building has a huge cantilever structure on the mountain's side. The mood board should cover all details from door knobs to the entire mass of the building.



Vernacular Museum
 Author, Google's MixBoard, NanoBanana Pro

Generate a mood board of a modern museum that is influenced by organic curve-linear geometry lines. The museum will be built with vernacular materials like red mudbricks and wood. Give inspiration to all levels of detail, from door knobs to the entire building mass.

Sketch Realization

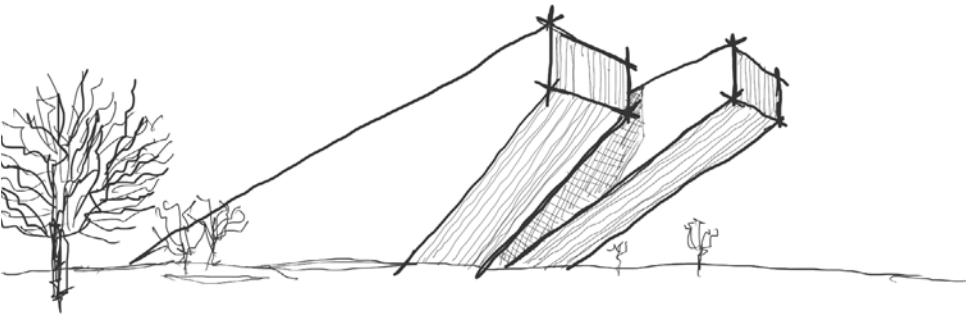


Image editing models were applied to transform a single conceptual sketch into a wide array of atmospheric variations. By simply adjusting prompts, the sketch evolved from a hyper-realistic visualization into speculative environments (such as a dystopian post-apocalyptic world), allowing the designer to test the emotional resonance of a form under vastly different conditions.

Author



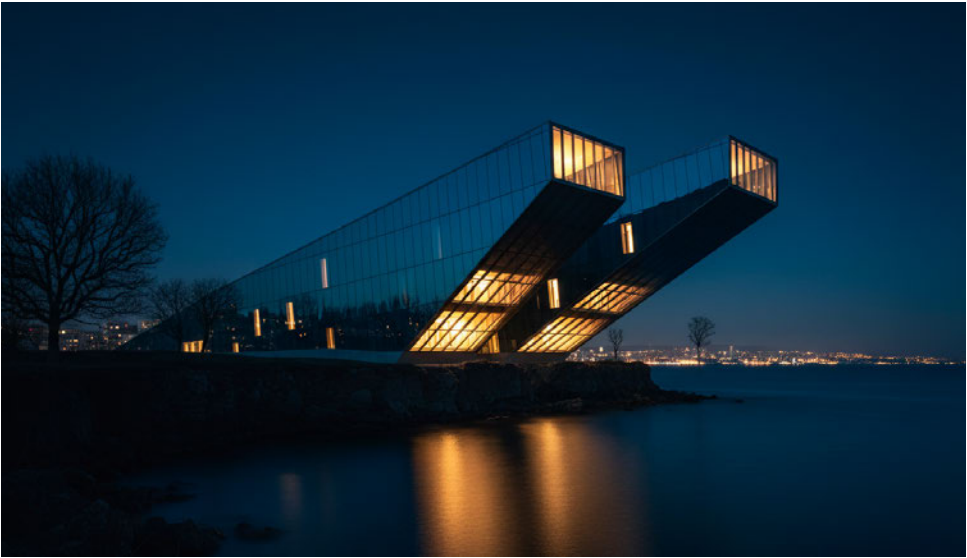
Turn the attached sketch into a photorealistic image of a monumental brutalist building made of two massive concrete slabs leaning sharply to one side, rising straight from flat snow-covered ground. The surfaces are raw board-formed concrete with visible plank texture and faint corner bracing where the slabs meet. Set in a cold open winter landscape under a flat overcast sky, with bare leafless trees clustered in the foreground on the left. Keep the exact same eye-level frontal viewpoint, wide horizontal framing, and low horizon line as the sketch. Muted grey palette, soft diffuse daylight, architectural photography.

Author, NanoBanana Pro



Turn the attached sketch into a photorealistic image of a serene minimalist building formed by two smooth white monolithic volumes leaning at a sharp angle, growing out of flat mossy ground. The walls are clean matte white plaster with no visible windows, crisp edges catching soft light. The building sits in a quiet pine and birch forest, with thin bare trees in the foreground on the left and dappled daylight filtering through. Maintain the identical eye-level frontal camera position, wide composition, and flat horizon of the sketch. Calm, bright, and contemplative atmosphere.

Author, NanoBanana Pro



Turn the attached sketch into a photorealistic image of a postmodern building made of two leaning glass curtain-wall volumes tilted to one side, their reflective blue facades glowing warmly from interior light. The structure stands on a dark coastal edge overlooking a calm sea at night, with the lit windows and distant city glow reflecting across the water. Bare silhouetted trees sit in the foreground on the left against the dark sky. Preserve the same eye-level frontal viewpoint, wide horizontal framing, and low horizon as the sketch. Cinematic, moody, deep blues with golden window light.

Author, NanoBanana



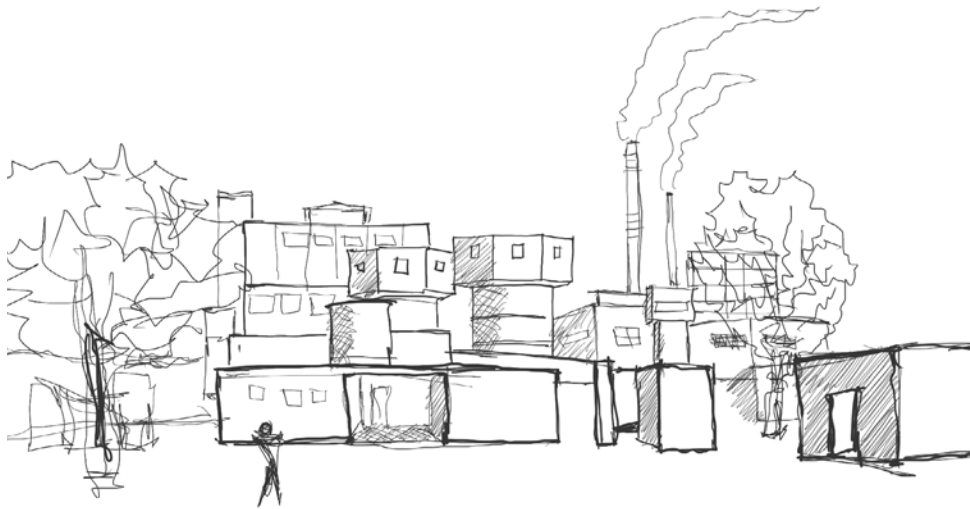
Turn the attached sketch into a photorealistic image of a bold building built from two enormous leaning Corten steel slabs canted to one side, rising from cracked dry desert ground. The panels are rich rusted orange-brown with streaked weathering and faint welded seams at the corners. Warm low dusk sunlight rakes across the surfaces, casting long shadows, with sparse dry shrubs and a few bare branches in the foreground on the left. Keep the exact eye-level frontal camera, wide framing, and flat horizon line of the sketch. Dramatic, warm, golden-hour mood.

Author, NanoBanana



Turn the attached sketch into a photorealistic image of a quiet building made of two leaning monolithic volumes clad in vertical charred black timber slats, tilted sharply to one side and rooted in flat grassy ground. Thin soft fog wraps the base of the structure in a muted Nordic landscape, with slender bare birch trees in the foreground on the left. Cool grey morning light, gentle mist, subdued earthy tones. Hold the same eye-level frontal viewpoint, wide horizontal composition, and low horizon as the sketch throughout. Atmospheric and tranquil.

Author, NanoBanana



Author



Turn the attached sketch into a richly painterly illustration in the style of the animated series Arcane, a sprawling industrial undercity of salvaged concrete and riveted metal, masses stacked and terracing upward in irregular cubic layers with small windows glowing warm amber from within. A tall soot-stained chimney rises right of center, leaking smoke into a hazy teal sky lit by a hidden source. Painterly hand-brushed textures on every surface, dramatic chiaroscuro lighting, a moody palette of deep teal shadows against glowing copper and orange highlights, drifting smog and atmospheric haze. Overgrown trees push in from both edges, a lone figure stands small in the foreground for scale, and low blocks close the front. Stylized 2D-3D hybrid look, oil-paint brushstrokes, cinematic and grungy, dieselpunk mood.

Author, Flux 2 [Pro]



Turn the attached sketch into a cold, cinematic image of brutalist dystopian social housing, heavy grey concrete blocks stacked in repetitive cubic tiers with tiny uniform windows, stained by damp and time. A tall factory chimney looms right of center, belching dark smoke into an overcast leaden sky, with industrial structures crowding the back. Bare sickly trees frame both sides, and a single isolated figure stands in the empty foreground. Bleak, oppressive, muted blue-grey palette, slightly grainy film look. Preserve the same eye-level frontal camera, wide composition, and flat horizon of the sketch.

Author, GPT Image 2



Turn the attached sketch into a bright, optimistic vision of a utopian terraced city, clean stacked white and pastel volumes cascading upward with greenery spilling from every level and small glowing windows. The tall vertical element right of center becomes a slender solar tower releasing soft white vapor into a clear blue sky. Lush flourishing trees frame both edges, and a small contented figure stands in a sunny open plaza in the foreground. Warm sunlight, soft shadows, hopeful and serene, gentle painterly realism. Hold the identical eye-level frontal viewpoint, wide horizontal framing, and low horizon as the sketch.

Author, Flux 2 [Pro]



Turn the attached sketch into a photorealistic image of a working industrial complex, blocky concrete and corrugated-metal buildings stacked and clustered across uneven ground with small grimy windows. A tall smokestack rises right of center, twin plumes of smoke drifting across a pale hazy sky, with pipes and structures filling the background. Dense trees frame both sides, and a lone worker stands small in the foreground for scale. Overcast daylight, realistic textures, muted industrial tones, architectural photography. Keep the exact same eye-level frontal viewpoint, wide framing, and low horizon line of the sketch.

Author, NanoBanana



Turn the attached sketch into a photorealistic image of a dense Brazilian favela climbing a hillside, small self-built homes of exposed red brick, raw concrete, and patched plaster stacked and terracing tightly upward in irregular cubic layers, with small mismatched windows, corrugated metal roofs, tangled electrical wires, and laundry strung between walls. The tall vertical element right of center reads as a slim utility tower or water tank against a warm hazy sky. Lush tropical trees crowd in from both edges, and a lone resident stands small in the foreground for scale, with low houses closing the front. Sun-bleached pastel paint, weathered surfaces, layers of life and improvisation, soft warm afternoon light. Keep the exact same eye-level frontal viewpoint, wide horizontal framing, and low horizon as the sketch.

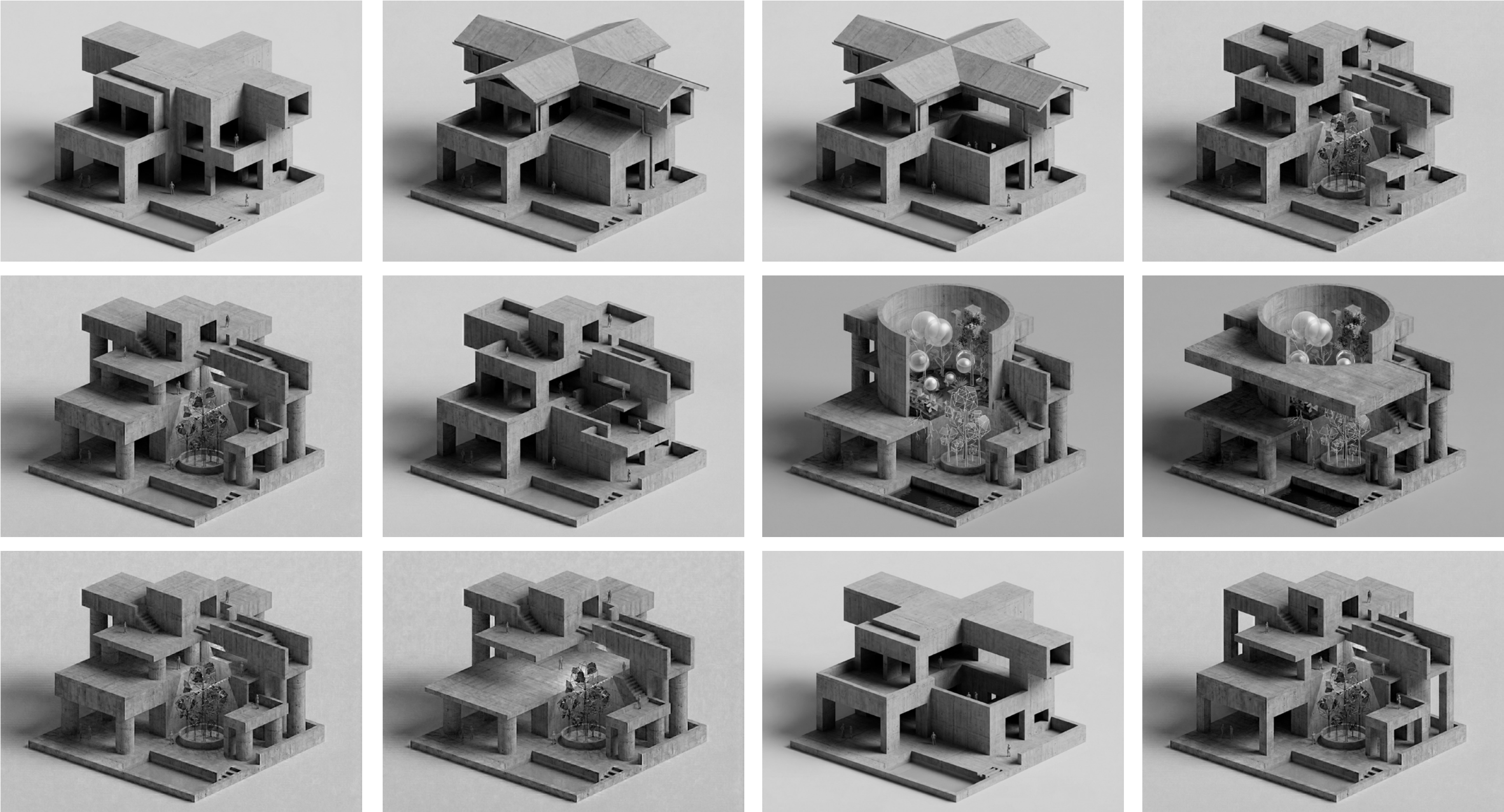
Author, Flux 2 [Pro]

Massing Iteration

Image generation models were used to process raw photographs of basic physical massing models. By applying simple text prompts, the underlying geometry of the physical model was manipulated and refined digitally, allowing the designer to test numerous volumetric variations iteratively to reach the final massing strategy.

Perform a massing study. Your goal is to explore volumetric variations. Ignore all fenestration, materials, and fine details. Focus solely on the overall form, silhouette, and relationship of solid volumes. Using the general scale and footprint of the provided image as a starting constraint, manipulate the volume by applying additive principles. I want you to extrude the mass to create a structure that feels branched and creates moments of overhangs and permeable spaces. Render the final output as an isometric conceptual model study. The entire structure must be visualized as monolithic, rough grey concrete with a brutalist texture, showing subtle imperfections and aggregate. It should be set against a seamless, infinite light grey studio background with soft, diffused lighting that casts gentle, grounding shadows. Include a couple of tiny monochromatic scale figures next to the mass and inside the mass as well to emphasize its monumental scale. The base volume is a villa type.

Author, NanoBanana Pro



3.1.2. Design & Iteration

Material Iteration

Through the use of image editing models, existing rendered images were modified using specific text prompts to swap out surface materials. This allows the designer to evaluate different material palettes (such as changing a concrete wall to timber) dynamically, without enduring the time-consuming process of re-rendering the entire 3D scene.



Author



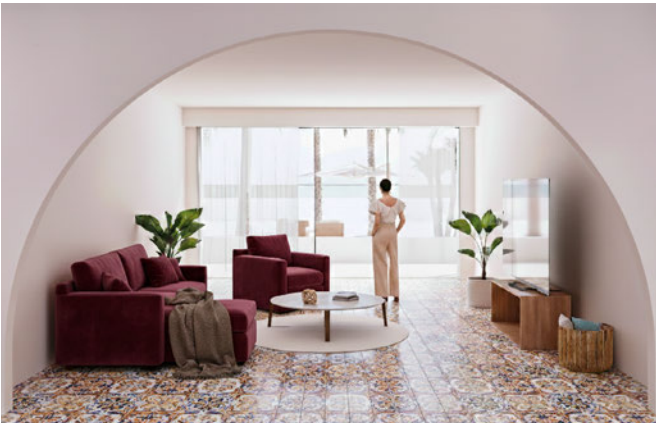
Change the flooring material to a navy blue carpet.



Change the floor materials to Italian-patterned ceramic tiles.



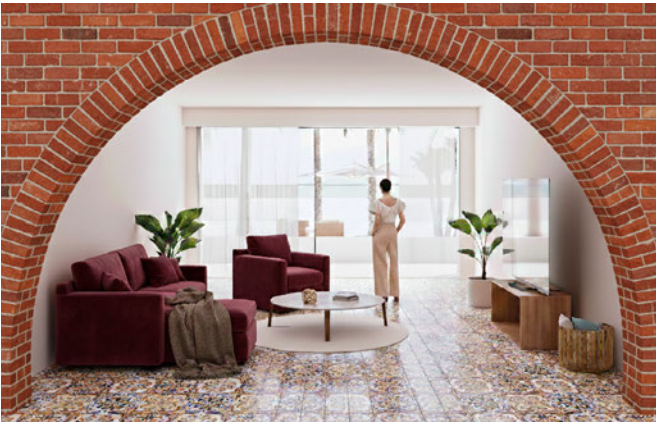
Change the sofa material to a flower pattern texture.



Change the sofa and chair materials to wine velvet material.



Change the foreground Arch material to irregularly shaped lime fieldstones.



Change the foreground Arch material to red bricks.

Author, NanoBanana Pro

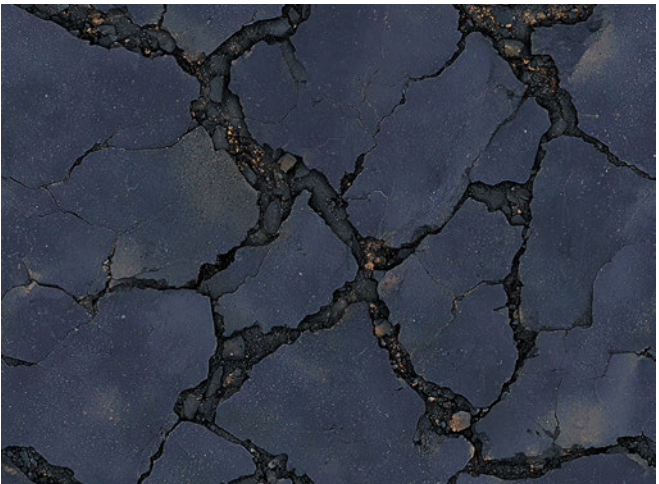
Material Generation from Text

Text-to-image models were prompted to create custom, seamless texture maps for materials that are not commonly found in standard architectural material libraries. This grants the designer infinite control over the bespoke texturing and detailing of their project.

Author, Midjourney



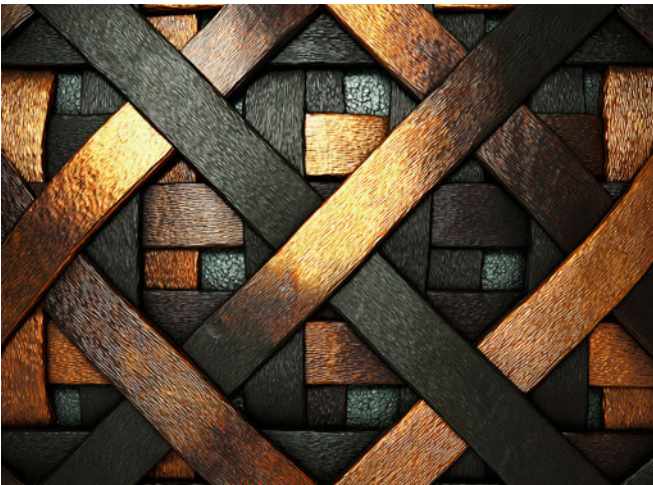
Seamless texture of bio-grown coral limestone with irregular cellular pores, soft erosion patterns, organic density shifts, natural randomness, tileable material map, neutral lighting removed



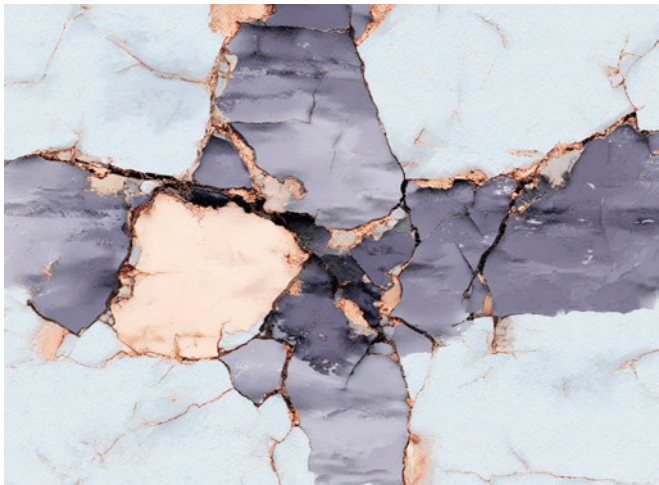
Seamless texture of compressed volcanic ash fused with translucent obsidian veins, micro-cracks and subtle granular variation, physically plausible surface depth, evenly distributed detail, tileable, no lighting, no shadows



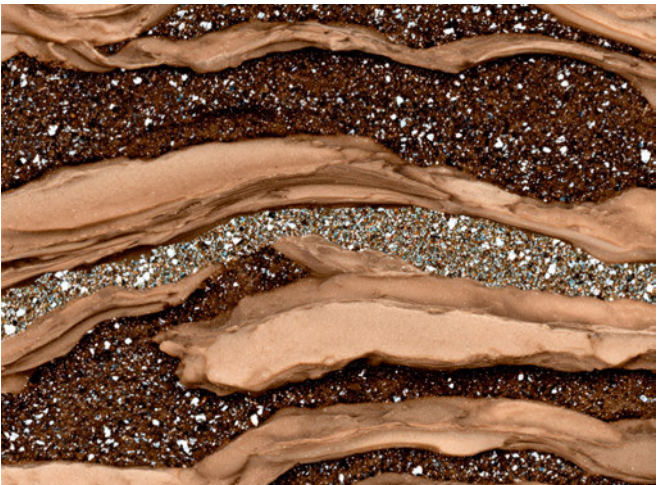
Seamless texture of fungal mycelium grown into dense sheets, branching vein networks, soft organic transitions, porous surface logic, seamless, no lighting artifacts



Seamless texture of woven carbon fiber interlaced with oxidized copper threads, subtle tarnish gradients, high-frequency weave detail, perfectly tileable, flat material scan style



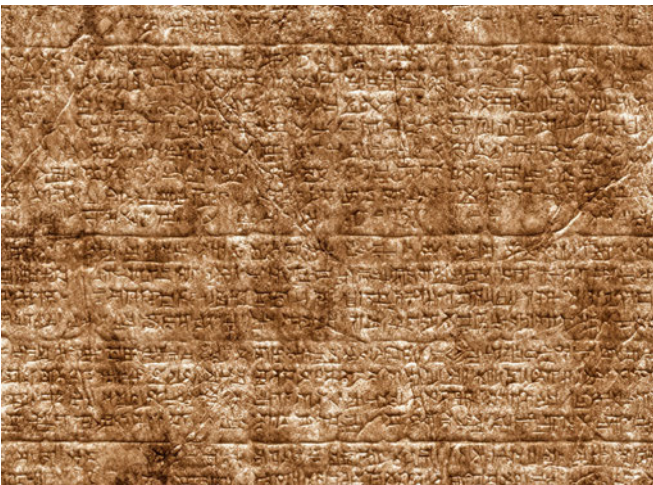
Seamless texture of cracked salt flats infused with metallic mineral deposits, crystalline edges, matte base with sporadic reflective inclusions, tileable surface, no directional light



Seamless texture of eroded sandstone mixed with glass dust, soft grain contrast with sharp reflective specks, wind-carved micro patterns, tileable texture map



Seamless texture of hand-cast resin with trapped air pockets and flow marks, semi-transparent appearance translated into surface detail only, seamless, neutral capture



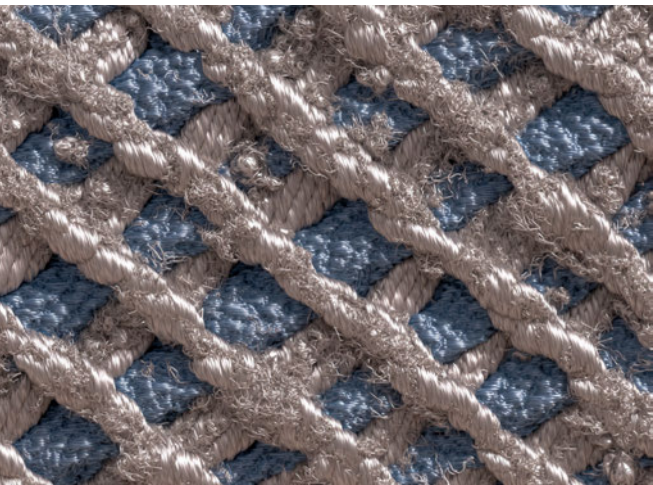
Seamless texture of ancient clay tablets ground into a continuous surface, faint cuneiform-like impressions, worn edges, dusty micro-noise, seamless and repeatable



Seamless texture of liquid metal frozen mid-flow, smooth tension lines and pooled ripples, alien but physically coherent, seamless tiling, uniform surface response



Seamless texture of digital noise translated into physical relief, layered pixel-like ridges with smooth erosion, abstract yet material-believable, perfectly tileable



Seamless texture of woven basalt fibers, rough mineral threads with slight irregular tension, matte volcanic finish, uniform pattern continuity, tileable material

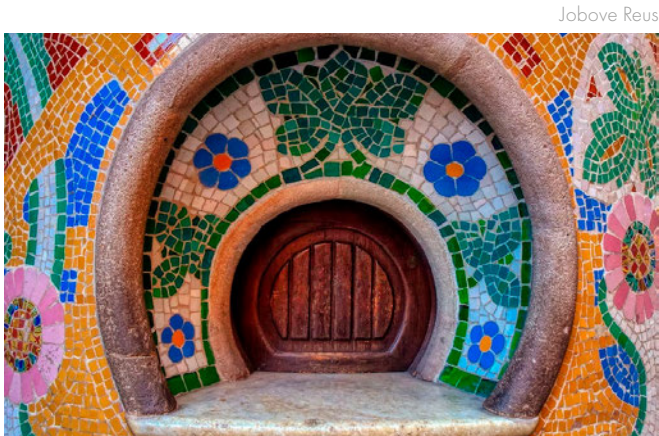
Material Generation from Reference

A seamless texture map was generated by uploading a reference photograph (e.g. unique stone pattern) to an image generation model. This custom material was subsequently applied to a separate architectural visualization using an image editing model, successfully bridging the gap between physical inspiration and digital application.



Author, NanoBanana

Apply the attached mosaic material to the central cylindrical mass of the building in the attached image.



Jobove Reus

Create a seamless tileable texture map of the mosaic pattern in the attached image.



Author



Author, NanoBanana

Create a seamless tileable texture map of the mosaic pattern in the attached image.



Author, NanoBanana

Apply the attached mosaic material to the central cylindrical mass of the building in the attached image.



atkinandthyme.co.uk



valenciatheaterseating.com



Author, NanoBanana

Mood Variation

A single baseline image was uploaded to an image editing model to create several environmental and atmospheric variations. By adjusting the prompt, the same architecture was visualized at different times of day or within entirely new contextual backgrounds (e.g. moving an urban building into a desert landscape) to test its visual resilience.

Author, Flux 2 Klein



Make the weather conditions in the spring at noon, remove all fog, and turn off artificial lights.

Make the weather conditions snowy in the winter at dawn.

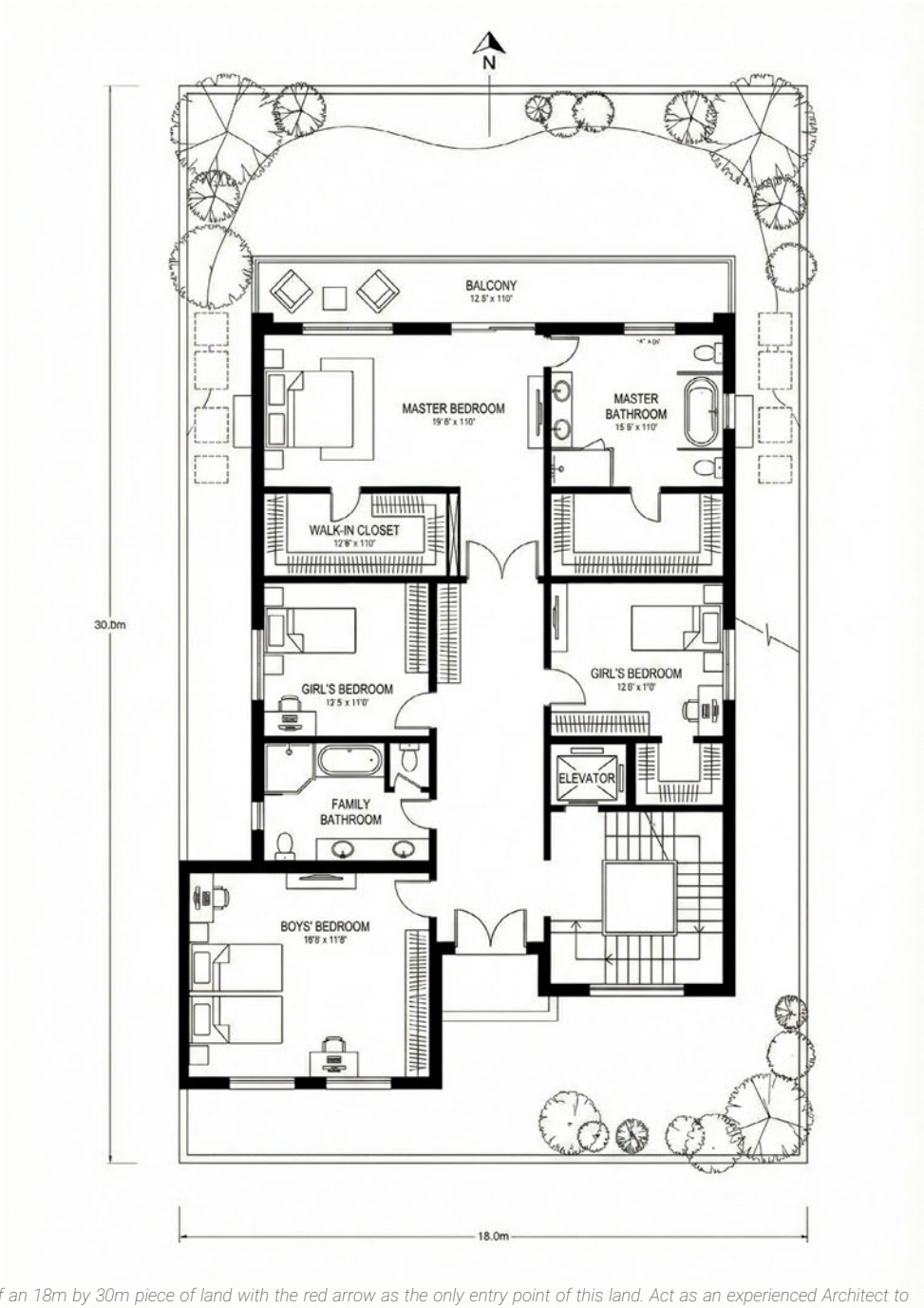
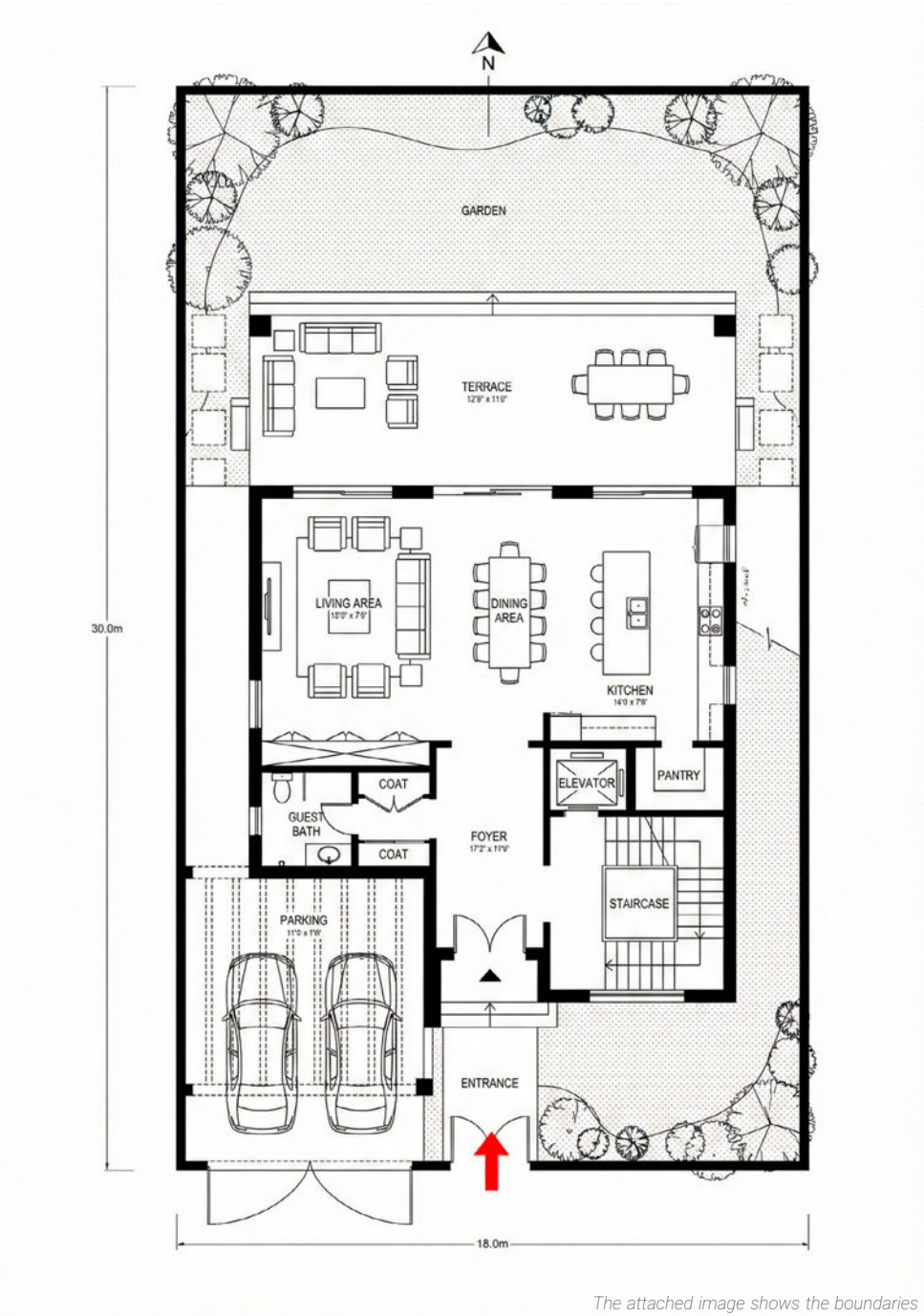
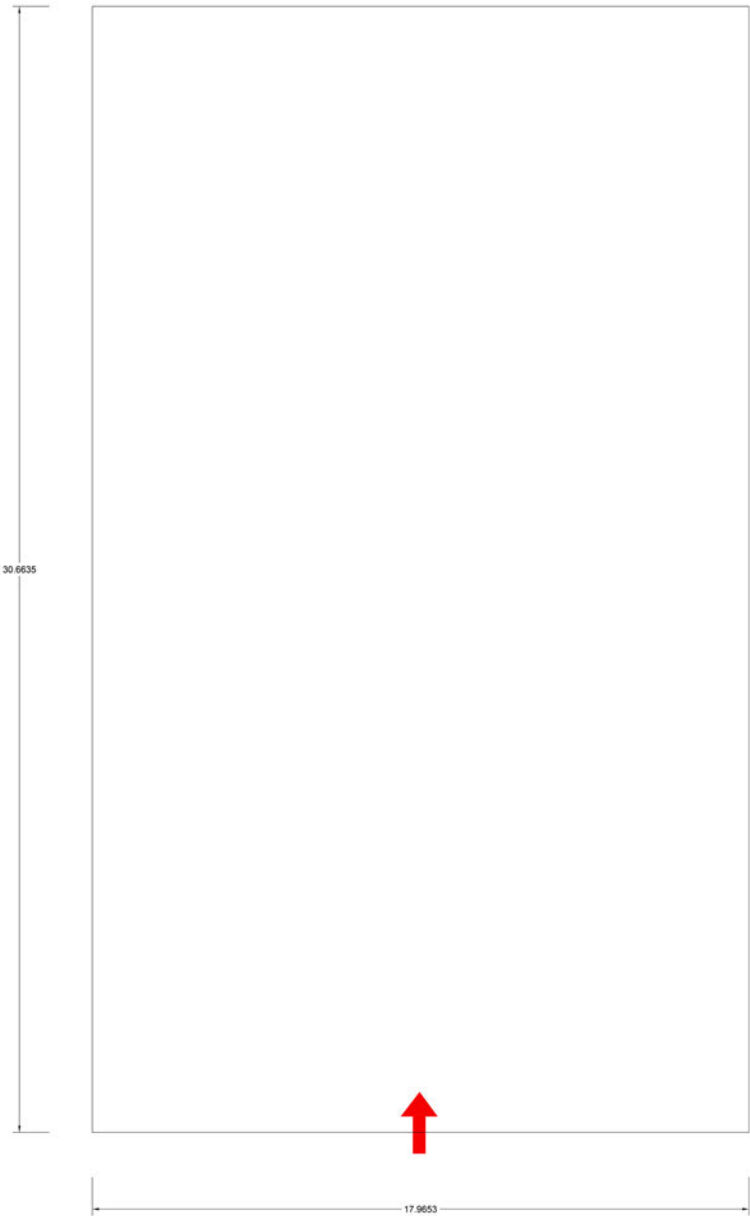
Reinterpret this scene as a futuristic dystopian world.

Rinterpret this scene as an Oasis with structures built with white limestone in a desert environment with minimalist Islamic patterns with accent blue patterned openings. Replace trees with palm trees.

Plan Generation

An empty land plot with specific boundary dimensions was provided as an input to an image generation model, which successfully generated an entire floor plan layout in seconds. While the raw output is not completely accurate or strictly usable for construction documents, it serves as a highly effective visualization tool for brainstorming spatial arrangements and programmatic zoning.

Author, NanoBanana Pro



The attached image shows the boundaries of an 18m by 30m piece of land with the red arrow as the only entry point of this land. Act as an experienced Architect to create an architectural line drawing of the ground floor plan of a house that needs to be built on this land. Some information you might need is: 1. The spaces required in the ground floor are a Foyer, guest bathroom, a big open kitchen with a pantry, a living area, dining area, a U-shaped staircase to go to the first floor and the basement, An Elevator next to the staircase, A big terrace that overlooks a big garden, parking spaces for two cars next to the entrance. 2. The north direction is upwards of the layout. 3. The house location is in Cairo, Egypt (so you can consider the openings, light, and weather conditions). 4. you can only build on 40% of the land Area. 5. The entry point is at the south, and their are neighboring houses on the east and west. The final output should be a furnished monochrome floor plan of the ground floor of the house in the site.

Based on the plan you provided. Create the first floor plan. The first floor contains a master bedroom with a balcony and two bedrooms for children (one girl and two boys), and a big bathroom.

Realtime Rapid Iteration

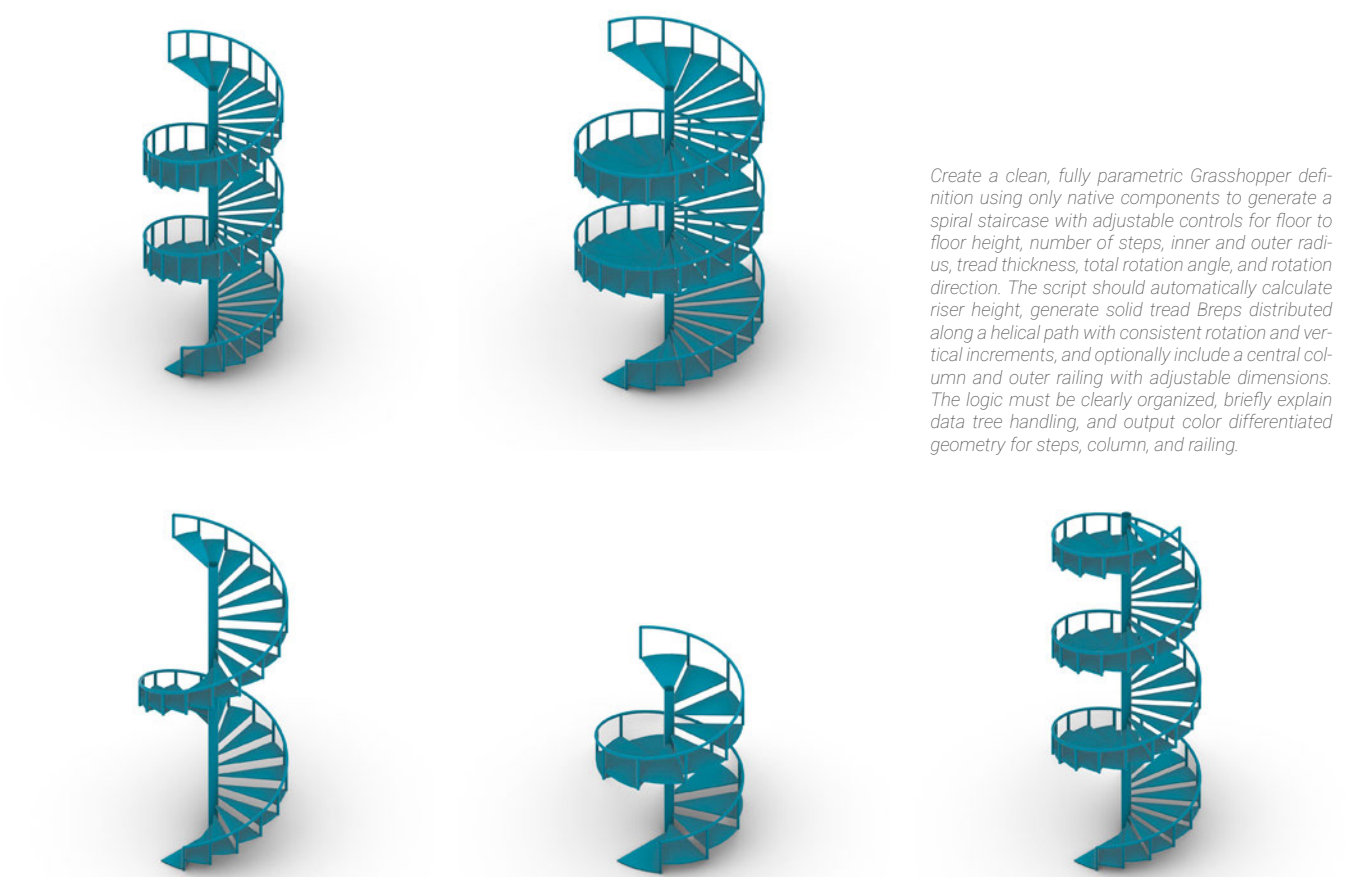
Integrating realtime image generation (e.g. Krea) with a parametric Grasshopper model allowed for instant, high-fidelity visualization of algorithmic changes. As the parameters of the 3D model were adjusted, the AI generated realized renders in realtime, effectively closing the feedback loop between abstract parametric logic and tangible visual atmosphere.

Author, Grasshopper, Krea Realtime

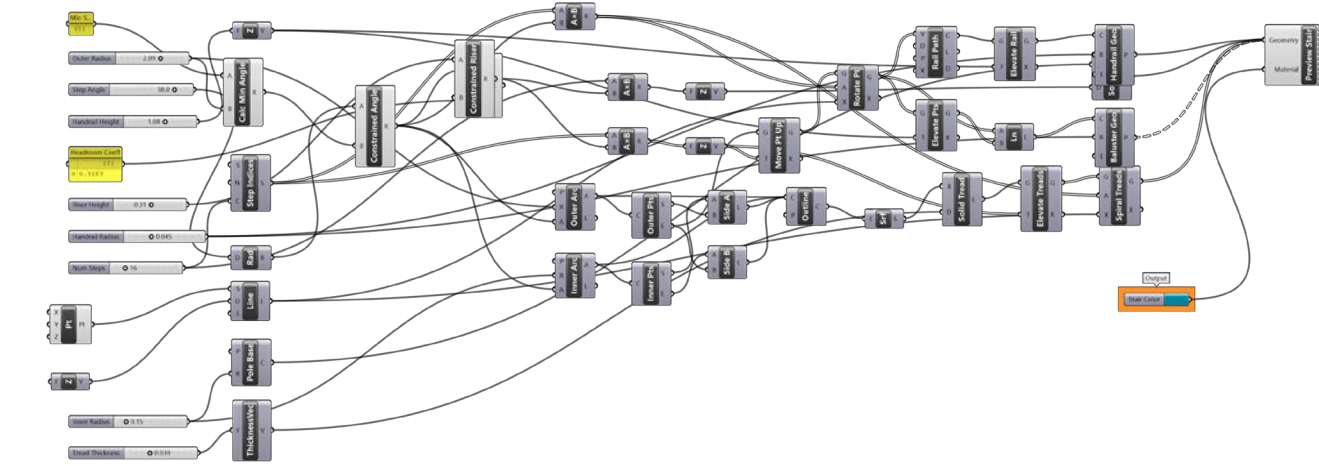


LLM to Grasshopper Script

Using integrated scripting plugins (e.g. Raven), a fully parametric script for a complex staircase was generated in Grasshopper using only a natural language prompt, completely bypassing the manual placement of components. This workflow was also successfully employed to mimic and reconstruct a parametric facade system based purely on a provided reference image.



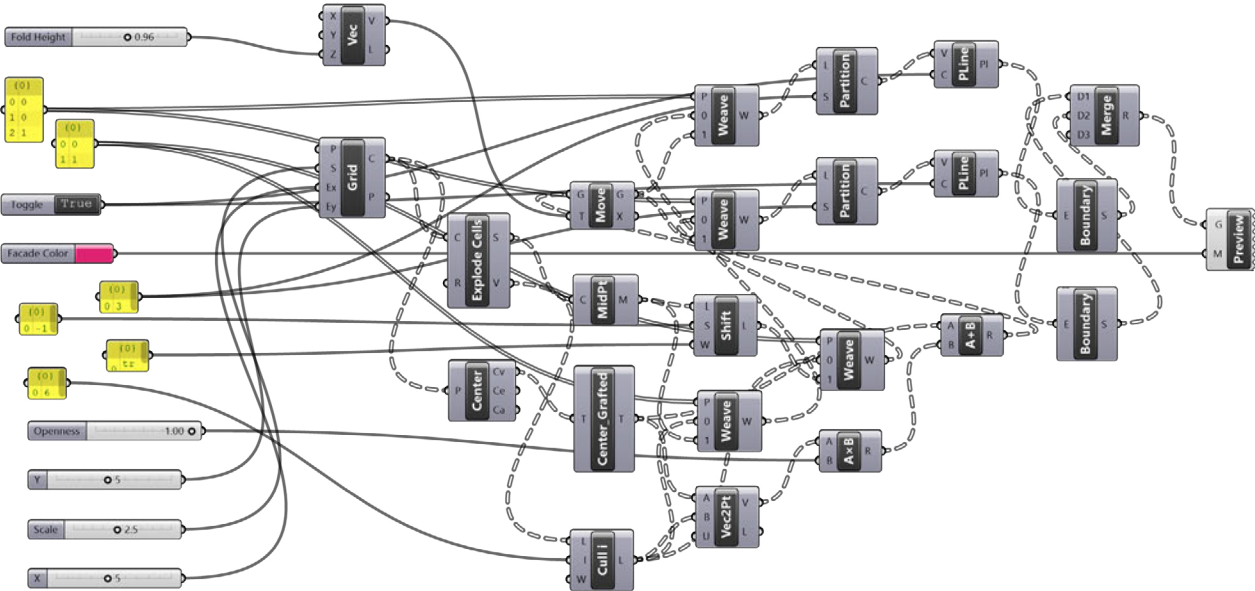
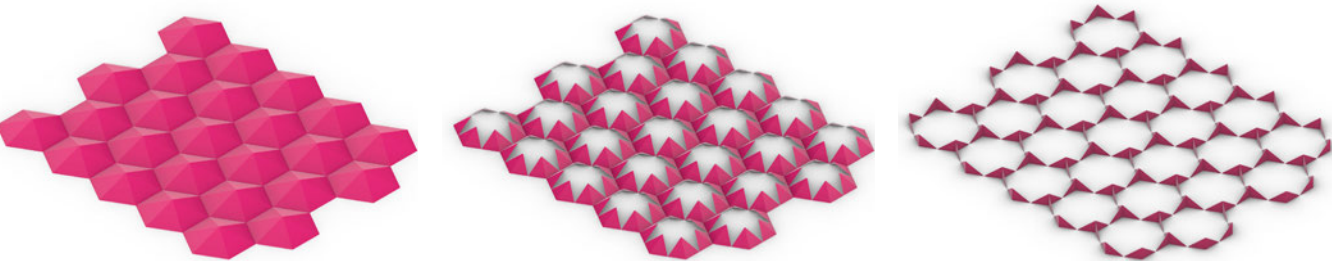
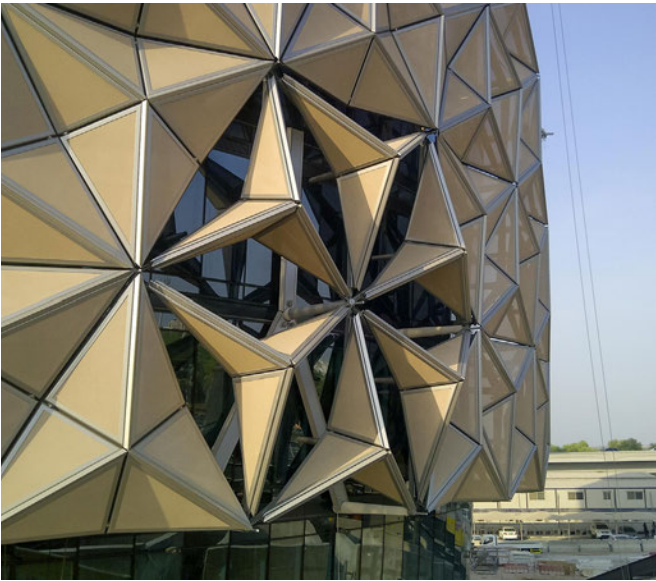
Create a clean, fully parametric Grasshopper definition using only native components to generate a spiral staircase with adjustable controls for floor to floor height, number of steps, inner and outer radius, tread thickness, total rotation angle, and rotation direction. The script should automatically calculate riser height, generate solid tread Breps distributed along a helical path with consistent rotation and vertical increments, and optionally include a central column and outer railing with adjustable dimensions. The logic must be clearly organized, briefly explain data tree handling, and output color differentiated geometry for steps, column, and railing.



Author, Grasshopper, Raven

AEDAS

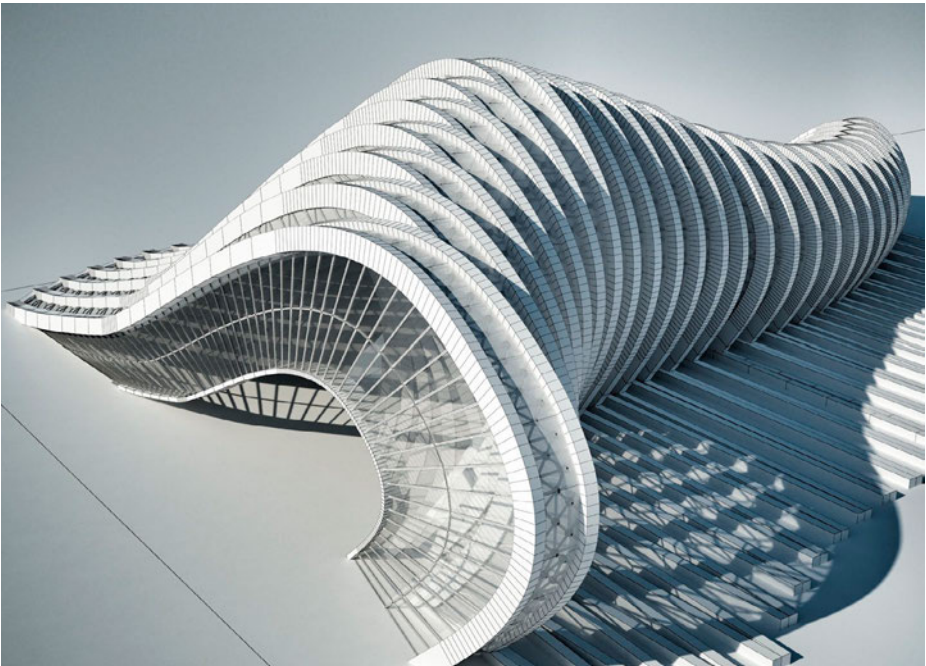
Create a high detail parametric architectural façade inspired by a kinetic mashrabiya system composed of repeating triangular modules arranged in a hexagonal grid. Each module should consist of three dimensional folded triangular panels framed in aluminum, forming a star like geometric pattern that protrudes outward from a glass curtain wall. The panels should appear operable, with some units partially open to reveal the glazing and structural system behind. The façade must have depth, shadow variation, and precise metal joints at each node connection. Use a warm beige fabric or composite infill material inside slim metallic frames, contrasted against reflective dark glass behind. Emphasize repetition, modular logic, and geometric precision. Lighting should be strong daylight from a side angle to enhance shadows and highlight the three dimensional folding. The final image should feel contemporary, engineered, and climatically responsive, with a clean urban background and realistic material detailing.



Author, Grasshopper, Raven

LLM to Grasshopper Python Script

A Large Language Model was explicitly employed to write custom Python code intended to mimic a specific geometric reference image. This generated code was then executed directly within the Grasshopper environment. While this process required a few trial-and-error iterations to debug, it highlights the immense potential for AI to bridge the gap between visual intent and complex algorithmic scripting.



Miroslav Naskov

```
Grasshopper Python Script to generate a parametric structure.
Notes:
1. Replaced 'GetPerpendicularFrame' with 'PerpendicularFrame' (the correct API method).
2. Includes previous integer and type hint fixes.
...
import Rhino.Geometry as rg
import Rhino
import System
import math

# --- Helper function to fix 'Null' errors ---
def ensure_curve(obj):
    if isinstance(obj, rg.Curve):
        return obj.DuplicateCurve()
    elif isinstance(obj, System.Collections.Generic.IEnumerable):
        obj_obj = Rhino.RhinoDoc.ActiveDoc.Objects.Find(obj)
        if obj_obj and isinstance(obj_obj.Geometry, rg.Curve):
            return obj_obj.Geometry.Duplicate()
    return None

# --- Main Script ---
# Initialize output lists
Ribs_Outter = []
Ribs_Inner = []
Glass_Panels = []
Skin_Srf = None

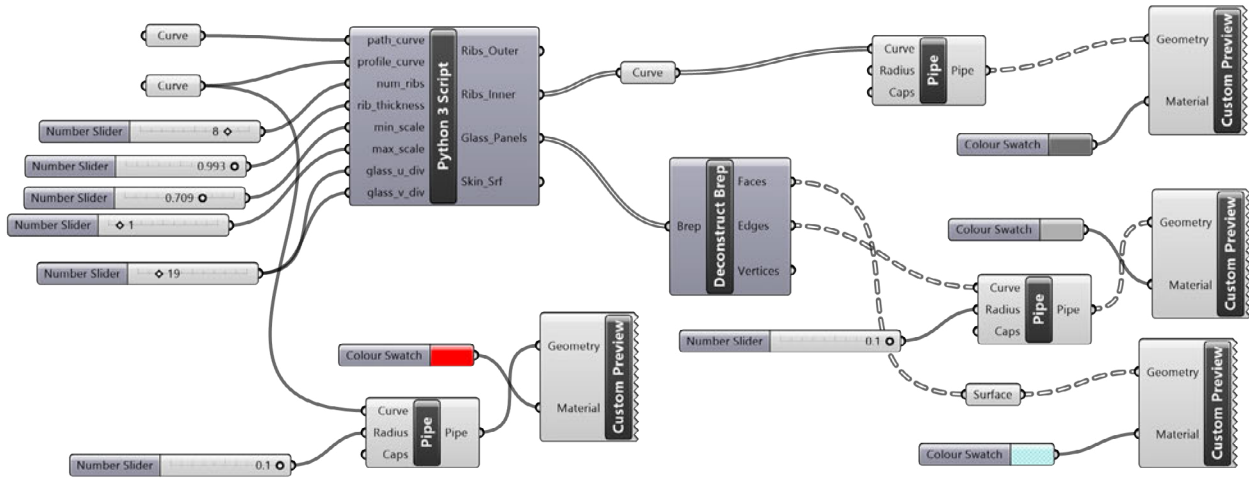
# 1. Safe Input Retrieval
safe_path = ensure_curve(path_curve)
safe_profile = ensure_curve(profile_curve)

# 2. Force Inputs to Integers
n_ribs = int(num_ribs)
w_div = int(glass_w_div)
v_div = int(glass_v_div)

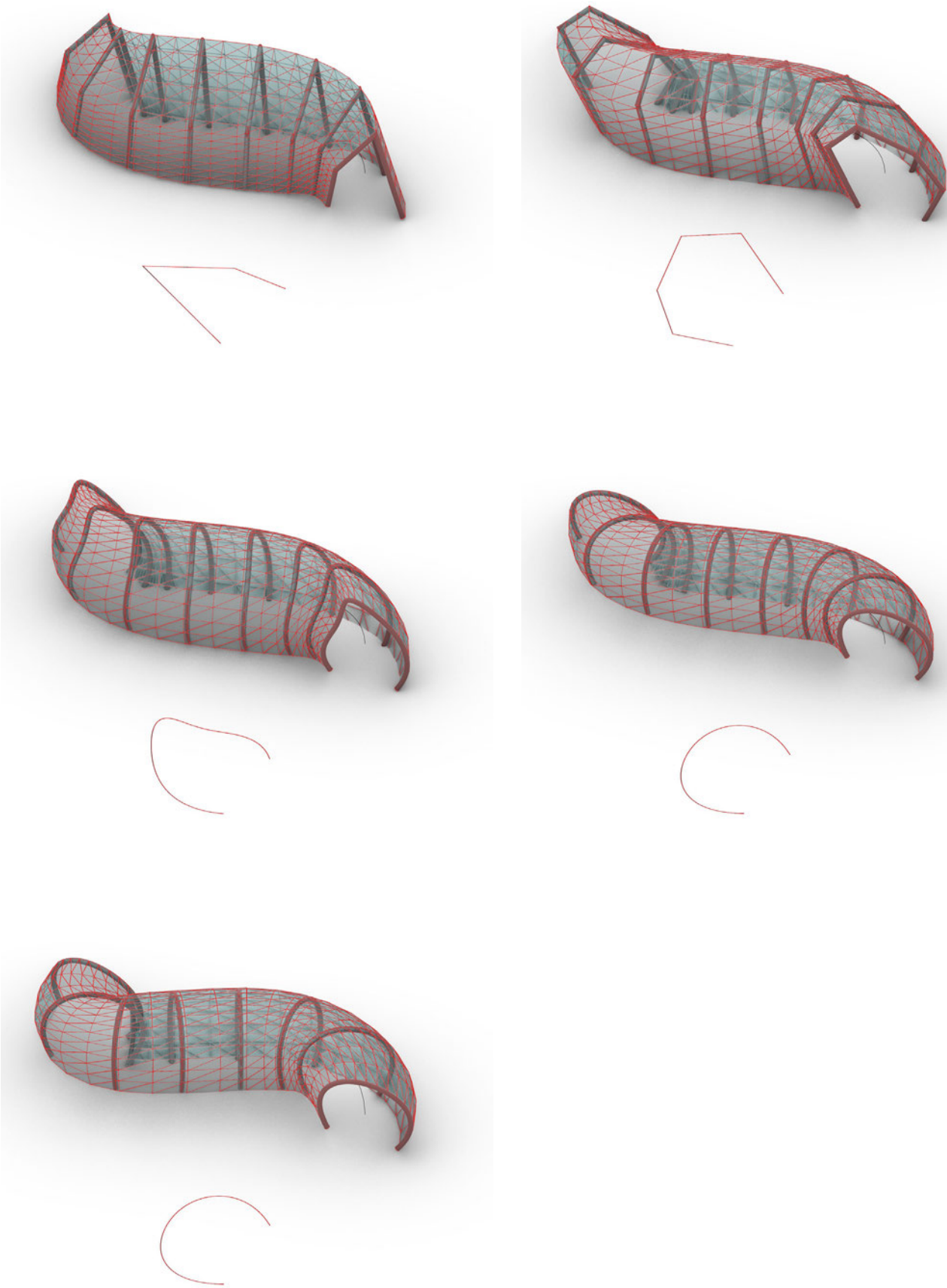
# 3. Validation & Exception
if safe_path and safe_profile and n_ribs > 0:
    # Prepare Geometry
    safe_path.Domain = rg.Interval(0, 1)
    safe_path.Domain = rg.Interval(0, 1)
    # Center the profile curve on the World XYZ origin
    Move = safe_profile.Translate(rg.Vector3d.Zero - Move.Center)
    safe_profile.Translate(rg.Vector3d.Zero - Move.Center)
    # Generate Ribs
    for i in range(1, n_ribs + 1):
        t = i / (n_ribs + 1.0)
        # Fix 12 HERE: Use PerpendicularFrame
        success, frame = safe_path.PerpendicularFrame(t)
        if success:
            # Calculate (dynamic) scale factor
            sine_val = math.sin(t * math.pi)
            current_scale = min_scale + sine_val * (max_scale - min_scale)
            # Create Transformations
            xform_rot = rg.Transform.PlaneToWorld(rg.Plane.WorldXY, frame)
            xform_scale = rg.Transform.Scale(frame.Origin, current_scale)
            xform_final = xform_rot * xform_scale * xform_rot
```

Author, Gemini

Create a Python code that I can use in Grasshopper to generate a structure similar to the attached image. Keep in mind that the only geometrical inputs that I'll provide are the curve making the path and the curve that constructs the profile of the mass (which the script should scale and deform to create variation)



Author, Grasshopper



Author, Grasshopper

3.1.3. Visualization

Built Environment to Isometric

Using a standard Google Street View image of an existing building, an image generation model was used to extract and generate a clean, isolated isometric projection of that specific structure. Additionally, an isometric urban tile was created from a single aerial view image (sourced from Google Earth's scanned buildings), demonstrating the model's ability to abstract and formalize raw contextual data for diagrams.



Google Maps

Using the attached building image as the only source, reconstruct the building as a clean orthographic isometric. Place the isometric model in a minimal white studio environment with no horizon line or visible walls, only a soft continuous backdrop. The building should read as a physical architectural scale model rather than a real building. Use matte, slightly desaturated materials, subtle edge chamfers, and simplified details typical of a 1:100–1:200 study model. Apply a single dominant directional light from one side to create crisp, legible shadows, supported by very soft ambient fill. Shadows should fall cleanly onto the white base plane, reinforcing volume and depth without drama. No background objects, no textures, no realism cues, no reflections, no people unless reduced to abstract scale figures. The result should feel like a photographed architectural critique studio model used for architectural critique.



Author, NanoBanana Pro



GoogleEarth

Using the attached Google Earth aerial image as the only reference, extract a clean urban tile and reconstruct it as a strict orthographic isometric axonometric. Remove all photographic perspective, lens distortion, atmospheric haze, and satellite artifacts. Convert the city block into a precise isometric tile with consistent scale, perfectly vertical buildings, and flattened terrain. Place the isometric tile in a minimal white studio environment with a seamless backdrop and a clean base plane. The city should read as a physical urban scale model rather than a real city. Simplify geometry and details to massing-level accuracy, muted colors, matte surfaces, and crisp edges. Apply a single directional studio light to create clear, readable shadows on the base plane, supported by soft ambient fill. No sky, no context beyond the tile boundary, no labels, no roads extending beyond the cut, no realism effects, no depth of field. The final result should resemble a photographed architectural city model used for urban analysis or exhibition.



Author, NanoBanana Pro, Magnific Percise

Image to 3D Model

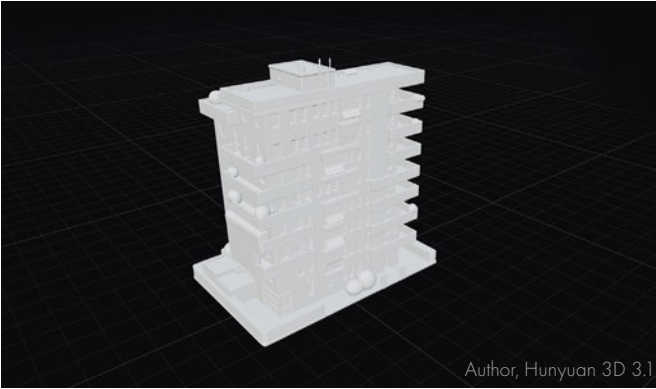
A single isometric image of a building was processed through a 3D generation AI model to create a high-fidelity 3D mesh. The resulting model was automatically separated into logical components, seamlessly imported into Rhino, and cleaned in a matter of minutes, showcasing a rapid pipeline from 2D raster imagery to workable 3D vector geometry.



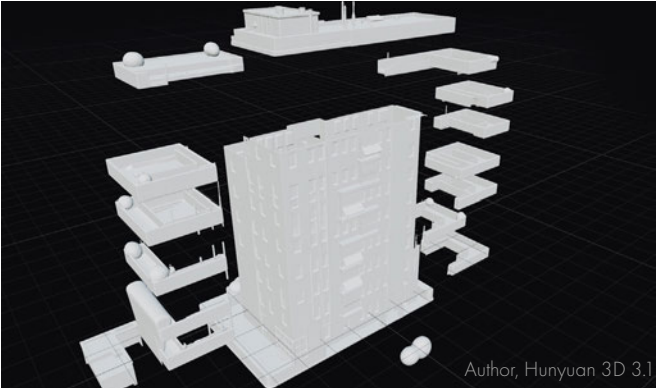
Author, NanoBanana Pro



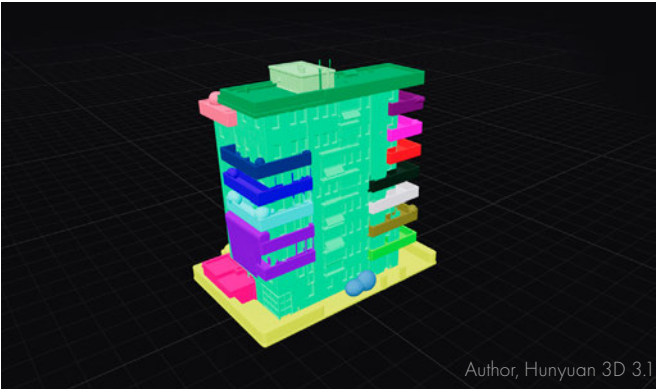
Author, Hunyuan 3D 3.1



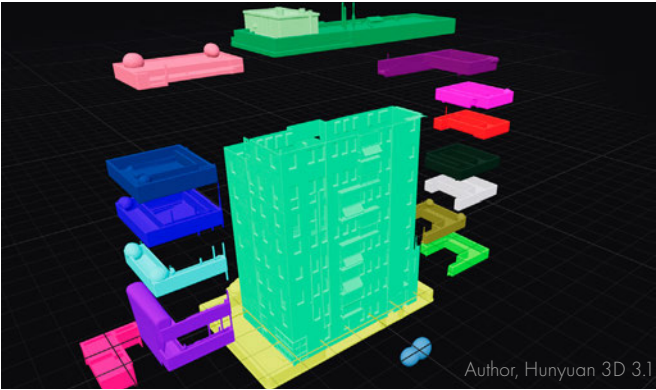
Author, Hunyuan 3D 3.1



Author, Hunyuan 3D 3.1



Author, Hunyuan 3D 3.1



Author, Hunyuan 3D 3.1



Author, Rhino



Author, Rhino

Image to CAD

A single perspective image of a building was fed into a specialized model to generate a flattened, orthographic line drawing. This successfully translates complex visual data into standard architectural elevation formats suitable for further drafting or presentation.



Convert the attached Isometric image of the building to a front elevation orthographic line drawing.



Author, NanoBanana

HDRI Generation

Using a simple descriptive text prompt, a custom High Dynamic Range Image (HDRI) was generated. This allows the architect to create highly specific, 360-degree environmental lighting setups that can be natively used to illuminate scenes in any standard 3D rendering software.



create an HDRI map of a modern city so I can use it in 3Ds Max to be visible from the window of my scene. The POV should be from an altitude of 10 meters. The lighting conditions should be around 10 am during Autumn.

Author, NanoBanana Pro

Rendered Images Enhancement

Renders were processed through an image editing model equipped with an upscaler. This significantly enhanced all micro-details of the image, refining material textures, sky fidelity, lighting accuracy, and overall realism, effectively functioning as an intelligent and automated post-production tool.



Author



Author, Krea



Create a photo-realistic HDRI map of a modern city so I can use it in 3Ds Max to be visible from the window of my scene. The POV should be from an altitude of 6 meters. The lighting conditions should be around 1 pm. The HDRI map should resemble a 360-degree image. Don't add too many skyscrapers. The sky should be partly cloudy.

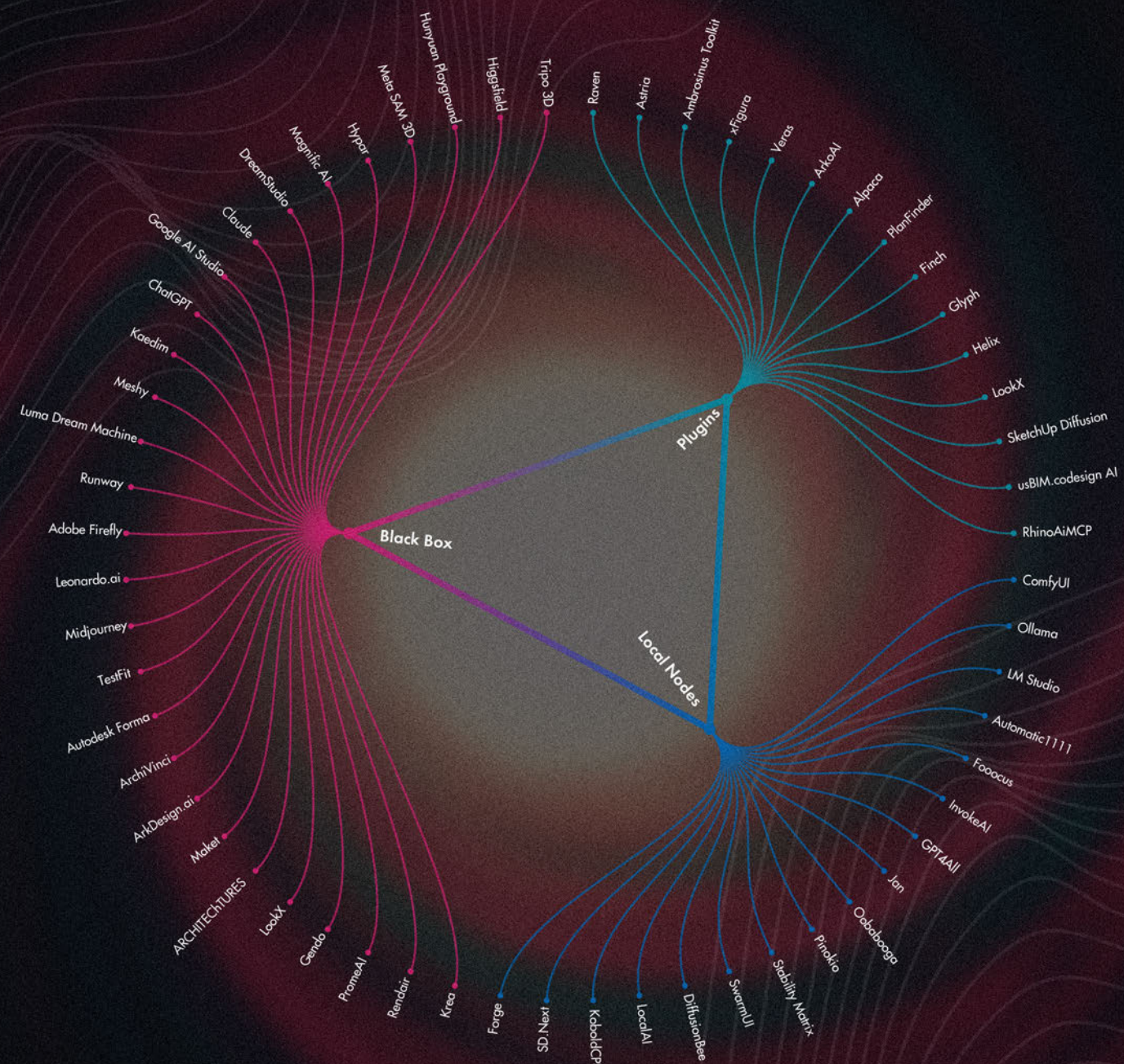
Author, NanoBanana Pro



Author



Author, Krea



3.2. Interface Spectrum

To evaluate the practical usability of generative models, this section categorizes the software field by User Interface (UI) architecture. The UI is the bridge between human intent and machine execution, and it determines whether the architect is a passive prompter or an active director. A model's underlying power is frequently secondary to this interaction framework; a highly capable model restrained by a rigid, text-only interface often provides less utility than an older model integrated with a flexible, geometry-aware system.

UI categorization establishes the technical barrier to entry and exposes the primary trade-off between rapid aesthetic generation and granular geometric control. Across this spectrum, the analysis divides the ecosystem into three categories: Black Box / Commercial Cloud platforms that prioritize high accessibility and rapid visualization but lack structural control, Integrated Plugins that directly embed generative capabilities within familiar BIM/CAD environments, and Local Nodes that offer deep customization and spatial accuracy but demand steep learning curves and heavy hardware investments.

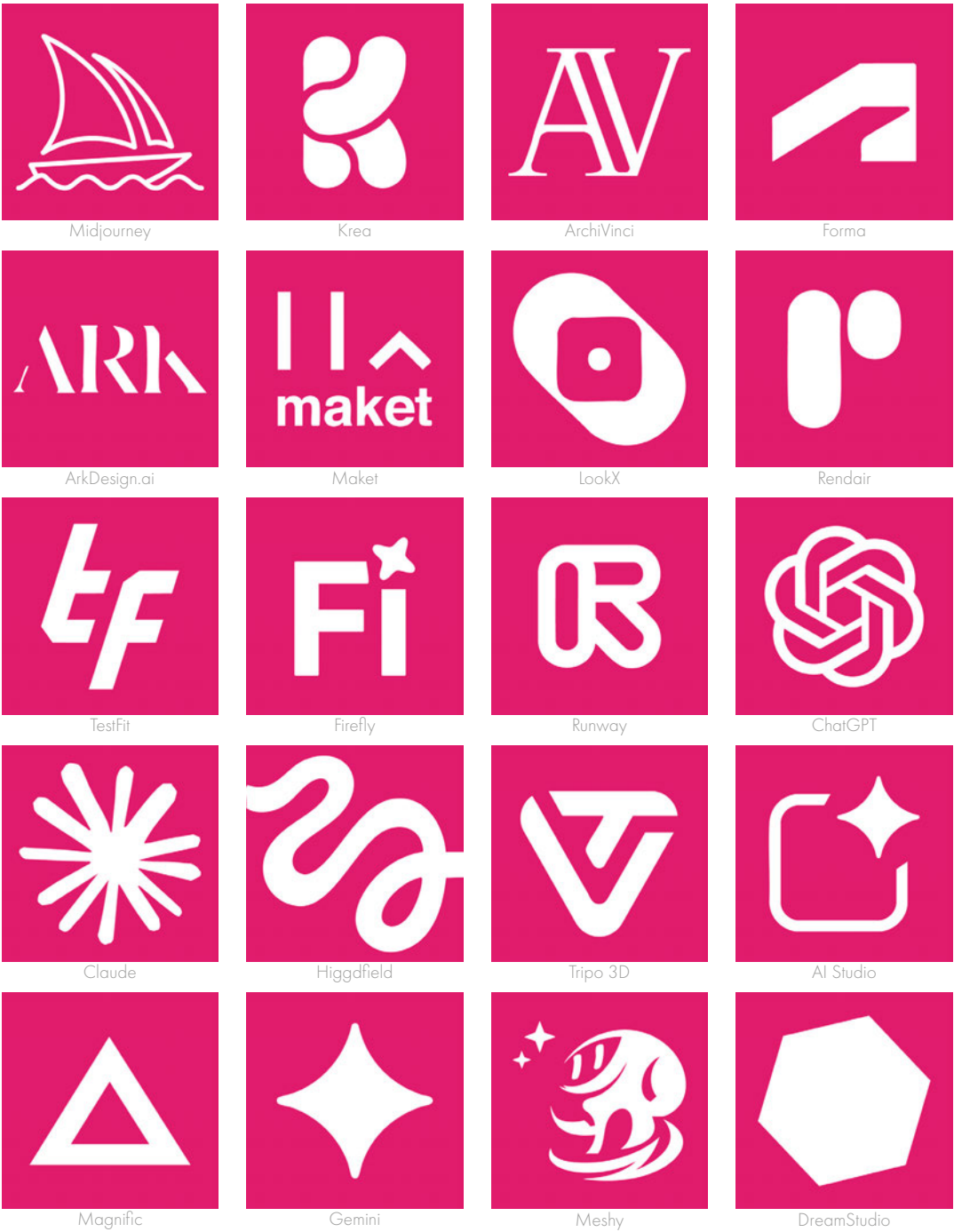
3.2.1. Black Box

The Black Box category refers to proprietary, cloud-based artificial intelligence platforms designed primarily for mass commercial appeal. In these systems, the underlying architecture, the training data, and the algorithmic processing are completely obscured from the user. The architect interfaces with the tool entirely through a web browser or a simple chat application, inputting natural language prompts and receiving a finished output without ever witnessing the internal mechanisms of generation. These platforms are deliberately streamlined to remove technical friction, effectively hiding the complex mathematics of diffusion models behind highly polished, accessible consumer interfaces.

The primary advantage of Black Box systems is their unmatched accessibility and exceptionally low barrier to entry. Because all the heavy computational lifting is performed on remote server farms, the architect requires absolutely no specialized hardware or high-end graphics processing units. These platforms are also heavily fine-tuned by their developers to produce polished, photorealistic results out of the box. That makes them well suited to rapid conceptualization, letting designers generate high-fidelity mood boards and speculative atmospheres in seconds without any background in computer science or programming.

This accessibility, however, sacrifices agency and geometric control. The user is relegated to the role of a passive consumer pulling a digital slot machine lever, relying heavily on stochastic chance rather than deliberate architectural design. Practitioners have zero control over the backend processes, making it nearly impossible to enforce specific spatial dimensions, tectonic rules, or precise material constraints. These platforms also operate on recurring paid subscription models and raise real data-privacy concerns, since confidential project concepts must be uploaded to third-party servers to be processed.

Examples of Black Box platforms span every generative modality. For image generation, Midjourney and DALL-E 3 represent the industry standard for commercial visualization. For video synthesis, cloud platforms like Runway Gen-2 and OpenAI's Sora generate temporal flythroughs from text. For text and logic, ChatGPT and Anthropic's Claude dominate as highly accessible large language models. For early-stage 3D model generation, cloud services such as Luma AI provide rapid mesh generation from simple text prompts entirely within a web browser.



3.2.2. Integrated Plugins

Integrated Plugins represent a bridging technology that embeds generative artificial intelligence directly into the architect's existing digital environment. Rather than forcing the user to navigate to a separate website, these tools exist as extensions within standard host applications such as Revit, Rhino, or SketchUp. By operating within the native modeling workspace, these plugins function as context-aware assistants. They can read the active viewport, interpret the camera angle, and analyze the basic geometry of the 3D model, using this extracted spatial data as the foundational baseline for the AI generation process.

The main benefit of integrated plugins is workflow continuity. Architects are spared the disruptive process of constantly exporting files, uploading them to external sites, generating results, and importing them back into their drawing space. Because the AI is directly linked to the Building Information Model or the 3D canvas, the generated visualizations are inherently tethered to the correct scale, perspective, and massing of the actual project. This accelerates the design-development phase, letting practitioners test material palettes, facade treatments, and contextual lighting while actively modeling.

Despite these workflow enhancements, plugins carry notable disadvantages related to dependency and interface restriction. Users are bound both by the limitations of the AI model and by the specific, often simplified, graphical interface provided by the plugin developer. If the AI hallucinates a structural error, the architect usually lacks the granular, backend tools to correct it manually. These tools also frequently rely on cloud compute credits, creating a meter-running cost structure. They also suffer from versioning friction, as updates to the host software can easily break the plugin's functionality until the developer releases a patch.

Prominent examples of integrated plugins include Veras, which attaches to Revit and Rhino to provide instant, stylized renderings based on the active 3D view. Similarly, tools like ArkoAI and LookX offer context-aware architectural visualization directly inside standard CAD software. Beyond image generation, plugins like Raven for Grasshopper use large language models to generate parametric scripting nodes from text prompts, automating complex geometric coding. For image editing, Adobe Photoshop's Generative Fill works as an integrated plugin, letting architects inpaint entourage or modify contextual elements directly within their standard post-production software.



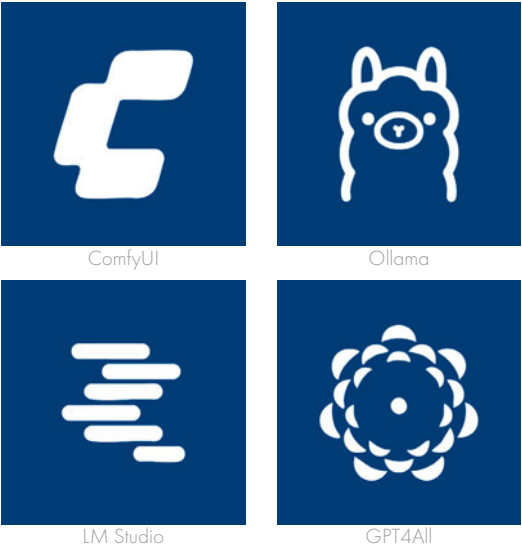
3.2.3. Local Standalone

The Local category encompasses open-source, standalone AI environments that run entirely on the user's personal hardware. Rather than relying on simplified text boxes, these systems frequently use complex, node-based interfaces that expose the entire generative pipeline to the user. In this setup, the architect manually connects data streams, linking text encoders, image samplers, noise schedulers, and specific model weights together on a visual canvas. This approach effectively unboxes the AI, transforming the black box into a transparent, mechanical switchboard where every variable of the diffusion process can be scrutinized and manipulated.

The defining advantage of local node-based systems is granular control. Architects evolve from passive prompters to active directors, using custom adapters like ControlNet (Zhang & Agrawala, 2023) to lock down strict geometric boundaries, force exact perspective lines, and dictate precise material applications. Because the software runs locally, it guarantees complete data privacy for sensitive projects and removes ongoing subscription costs or cloud computing fees. The open-source nature of these platforms also gives users a large, community-driven library of specialized models, supporting deep customization and the ability to train the AI on proprietary firm data.

Conversely, the barrier to entry for local systems is very high, which makes them intimidating for the average practitioner. Operating these platforms requires a steep learning curve closely resembling software engineering, demanding familiarity with Python environments, dependency management, and neural network architectures. The hardware constraints are equally severe, requiring expensive, high-end graphics processing units to perform the complex mathematical calculations locally. Maintaining these open-source ecosystems is also a constant burden, since the rapid pace of community development frequently leads to software conflicts, broken nodes, and unstable update cycles.

The defining example of this category for architectural image generation and editing is ComfyUI, specifically when used to operate underlying models like Stable Diffusion through advanced node-based interfaces. For local text generation, architects can run open-source large language models such as Llama 3 using standalone interfaces like LM Studio, ensuring data privacy while generating project narratives. For offline 3D model synthesis, local implementations of tools like TripoSR allow for the generation of meshes without cloud dependency. Local video generation is handled through specialized nodes like AnimateDiff, which integrate directly into the ComfyUI ecosystem to animate architectural concepts entirely on local hardware.

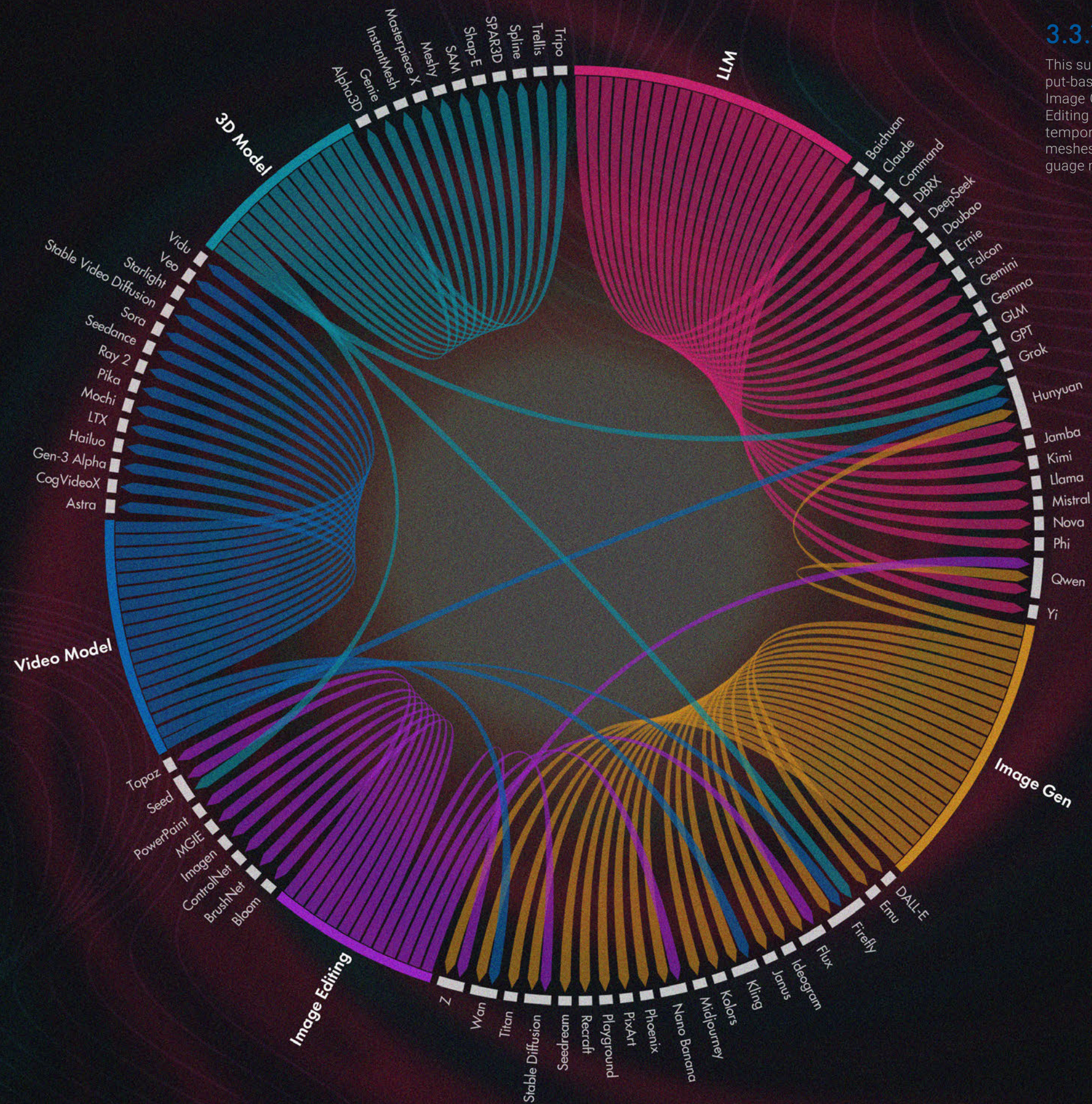


[Slot Machine]
Author, Krea 1

Create a vertical composition of a minimalistic, non-realistic scene viewed from above at a 45 degree angle. The camera is positioned high in a bird's eye perspective but tilted at 45 degrees, looking down from the direction opposite the light source. A solitary human figure sits on a simple chair in front of a plain desk placed at the center of a massive, empty, dark room. On top of the desk sits a small slot machine, and the human is captured in the act of pulling its lever. From the slot machine, an explosive burst of tiny paper fragments erupts outward, each small piece carrying abstract, random miniature images and symbols. The papers scatter dynamically through the air and across the floor, filling the vertical frame with motion. A single dramatic spotlight cuts through the darkness at a 45 degree angle, aimed at the figure from the opposite side of the camera, creating strong directional shadows stretching across the floor. The room remains almost entirely black beyond the cone of light. Use clean geometric forms, a simplified human silhouette, a limited color palette, high contrast between light and shadow, flat shading or subtle gradients, no textures, no realism, emphasizing negative space, graphic clarity, and bold composition.

3.3.Experimental Testing

To assess the practical utility of these tools, a structured testing protocol is applied throughout this section. The process begins by systematically laying out all relevant AI models and organizing them into distinct functional categories based on their output modalities. Following this categorization, the research establishes the core rating criteria and, for the generation models, defines the three stylistic outputs used for testing: Realistic, Conceptual, and Alien. Each model is run through the relevant prompts and its outputs rated against the criteria, with the generation models assessed within each of the three stylistic outputs and the editing models assessed across a set of representative editing scenarios. These ratings are consolidated into two evaluation matrices, one for the generation models and one for the editing models. As a methodological disclaimer, the rating of these outputs is a structured but ultimately subjective process that relies on the author's architectural judgment. Not all criteria apply to every type of output, just as not all stylistic outputs apply to every category of model tested.



3.3.1. Modal Categorization

This subsection categorizes relevant AI models into five output-based groups, excluding non-visual modalities like audio. Image Generation synthesizes new 2D images, while Image Editing manipulates existing ones. Video Generation creates temporal sequences. 3D Generation produces volumetric meshes. Lastly, Text and Logic Generation employs large language models to output text and code.

3.3.2. Image Generation

To assess the image generation models, a set of fixed architectural prompts is run through several of them, with each prompt prepared at three levels of parameter complexity so that performance can be observed as the demands on the model rise. Testing every model on realism alone would be unfair, since some are built for photorealistic output and others for looser, more exploratory work. The process therefore branches into three stylistic categories, each carried through the same three levels of complexity: Realistic (photorealistic visualization of a standard building typology), Conceptual (artistic, sketch-like, or abstract representation of a form, and Alien (defamiliarized structure that pushes beyond conventional logic).

Image Generation

Realistic - 8 Parameters

Photorealistic daytime wide shot of a modern museum: a single-story rectangular building with a glass curtain wall and exposed concrete frame, set on a clean paved plaza with one bench and two visitors walking by; clear blue sky, soft natural sunlight from the left, crisp shadows, neutral color palette, high resolution, realistic material finishes (reflective glass, rough cast concrete), camera at eye level, 35mm equivalent, slight wide-angle, minimal post-processing.



Firefly 4 Ultra



Firefly 5 Preview



Flux 1.1 Ultra



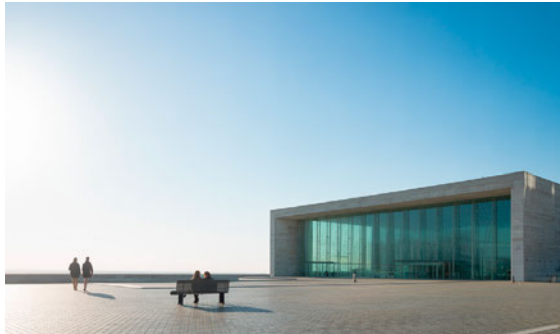
Flux 2 Max



GPT Image 1.5



Grok



Ideogram V3



Imagen 4



Kling 3.0 Omni



Krea 1



Meta AI



Midjourney



Nano Banana 2



Nano Banana Pro



Phoenix 1.0



Photon 1



Qwen Image 2.5



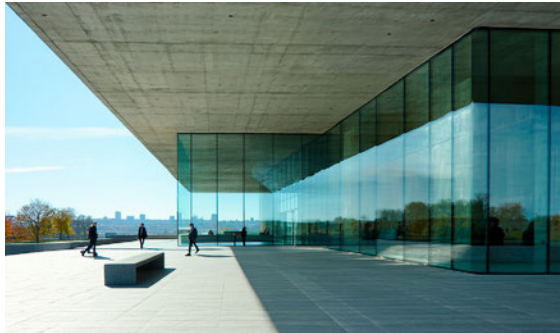
Recraft V4



Runway



Seedream 4.5



Stable Diffusion 3.5



Stable Image Ultra



Wan 2.5



Z-Image Turbo

Image Generation

Realistic - 16 Parameters

Photorealistic street-level view of a contemporary residential tower with staggered terraces: 12-stories, alternating concrete slabs and floor-to-ceiling glazing, planted green terraces on every third floor, corten-steel balcony screens, warm golden-hour lighting from the west casting long soft shadows, bustling urban sidewalk with cyclists and three groups of pedestrians, wet pavement reflections from a recent rain, visible building textures (concrete pores, rust patina on steel, subtle glazing reflections), depth of field with foreground foliage slightly soft, camera 24–35mm, f/5.6, subtle cinematic color grade, ultra-detailed realism.



Firefly 4 Ultra



Firefly 5 Preview



Flux 1.1 Ultra



Flux 2 Max



GPT Image 1.5



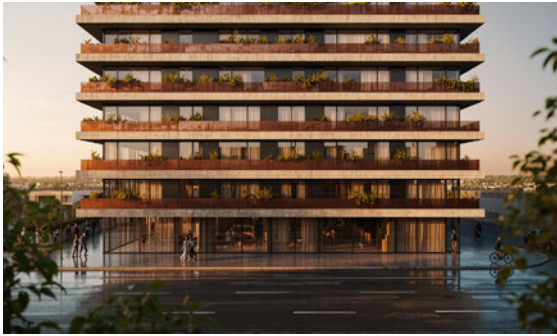
Grok



Ideogram V3



Imagen 4



Kling 3.0 Omni



Krea 1



Meta AI



Midjourney



Nano Banana 2



Nano Banana Pro



Phoenix 1.0



Photon 1



Qwen Image 2.5



Recraft V4



Runway



Seedream 4.5



Stable Diffusion 3.5



Stable Image Ultra



Wan 2.5

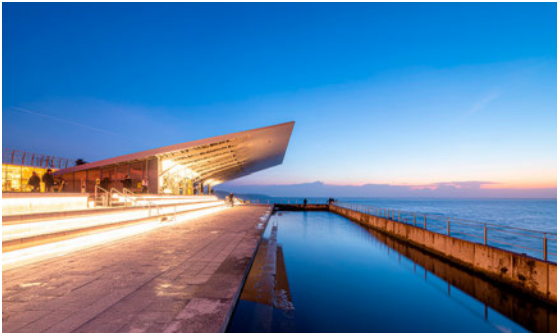


Z-Image Turbo

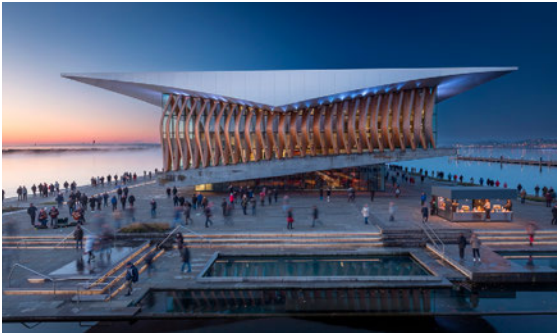
Image Generation

Realistic - 28 Parameters

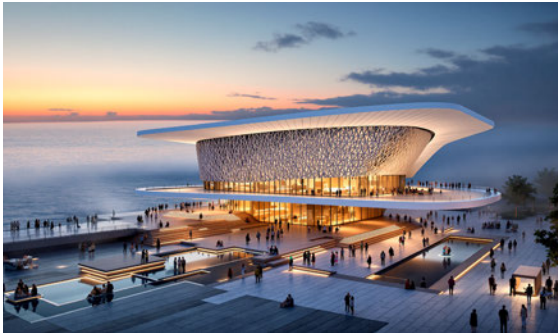
Generate a photorealistic image of a contemporary waterfront cultural centre with a cantilevered main volume over a public plaza and a sweeping white composite roof extending 6 meters beyond the façade. A parametric timber lattice wraps the auditorium with floor-to-ceiling glazing behind it. Stepped terraces face the water with linear LED bench lighting, stainless steel handrails, and visible timber joints. Two shallow reflecting pools mirror the sky and building. The stone plaza holds at least 40 distinct people, with street musicians on the left and a small food stall on the right. Subtle weathering at the base and minor concrete imperfections are visible. Light mist rises above the water under a dusk sky fading orange to deep blue. Warm interior light contrasts cool lattice uplighting. Motion blur on moving people, crisp reflections on wet pavement. Aerial 45-degree view, tilt-shift, 24mm lens, f/8, HDR.



Firefly 4 Ultra



Firefly 5 Preview



Flux 1.1 Ultra



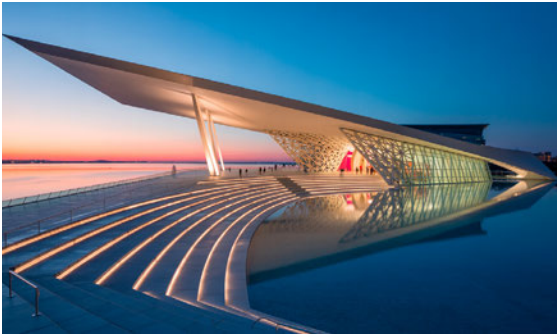
Flux 2 Max



GPT Image 1.5



Grok



Ideogram V3



Imagen 4



Kling 3.0 Omni



Krea 1



Meta AI



Midjourney



Nano Banana 2



Nano Banana Pro



Phoenix 1.0



Photon 1



Qwen Image 2.512



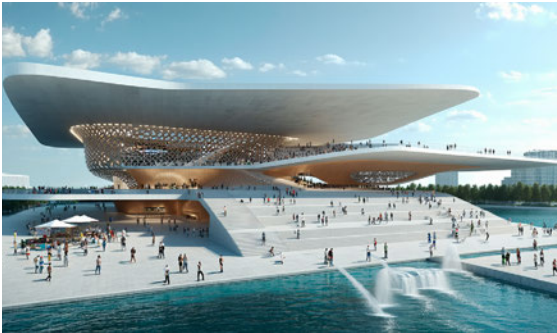
Recraft V4



Runway



Seedream 4.5



Stable Diffusion 3.5



Stable Image Ultra



Wan 2.5



Z-Image Turbo

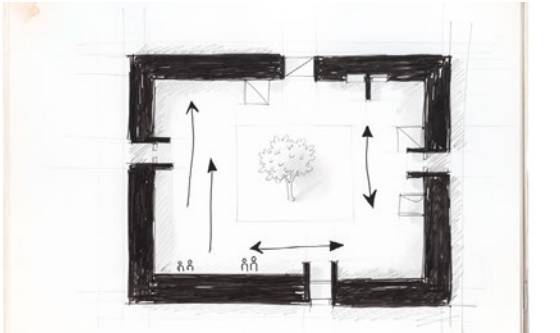
Image Generation

Coceptual - 14 Parameters

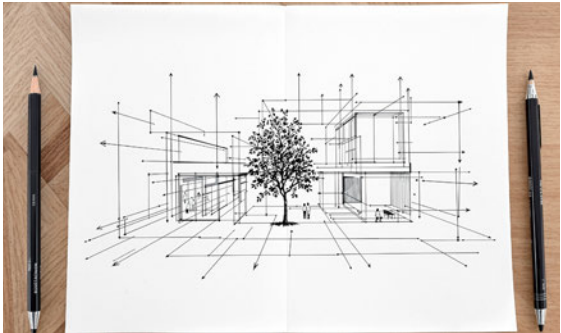
Generate a conceptual architectural sketch of a small courtyard house organized around a central void. The drawing is hand-sketched in black ink on white paper with loose linework and visible construction lines. Walls are shown as thick poche. A single tree occupies the courtyard. Arrows indicate natural ventilation across the plan. Minimal human silhouettes suggest scale. No shading except light hatching in shadows. Composition centered on the page with wide margins. The image should feel like an early design notebook study.



Firefly 4 Ultra



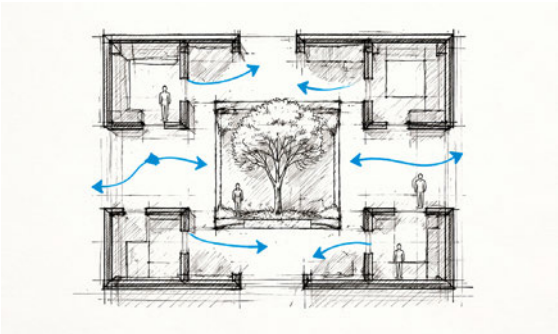
Firefly 5 Preview



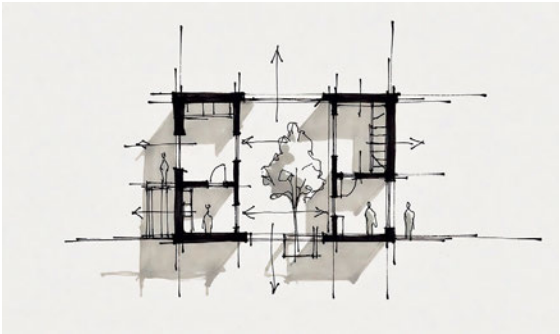
Flux 1.1 Ultra



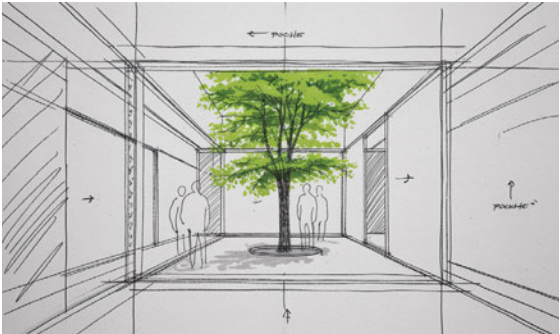
Flux 2 Max



GPT Image 1.5



Grok



Ideogram V3

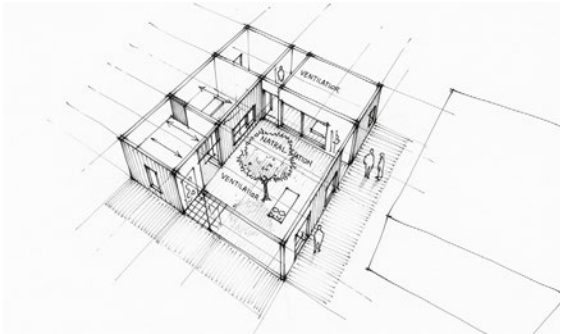


Imagen 4



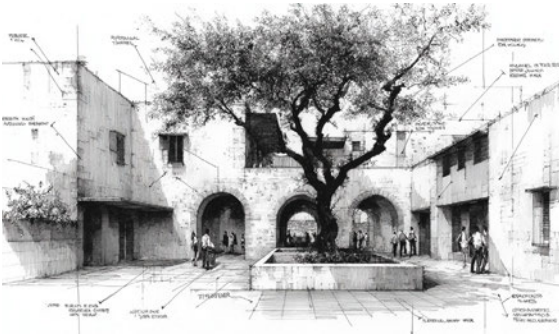
Kling 3.0 Omni



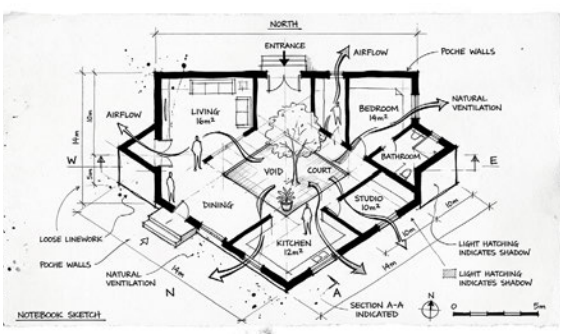
Krea 1



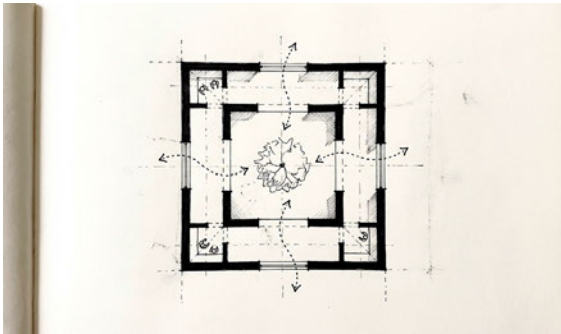
Meta AI



Midjourney



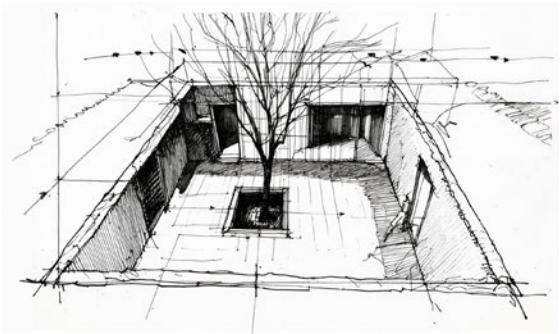
Nano Banana 2



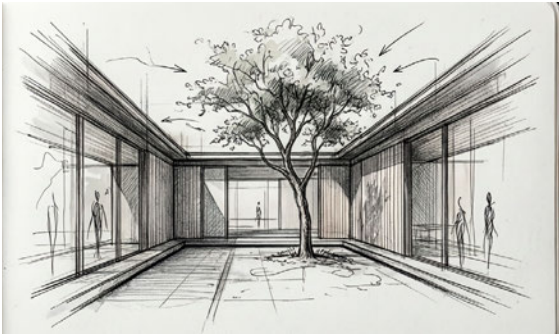
Nano Banana Pro



Phoenix 1.0



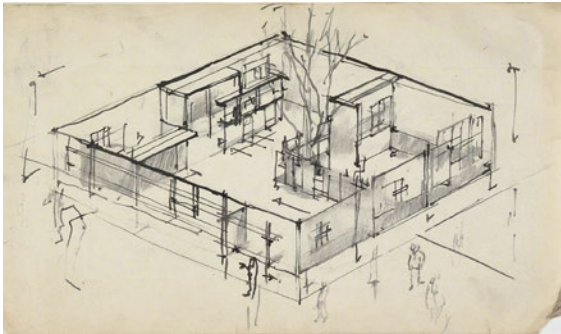
Photon 1



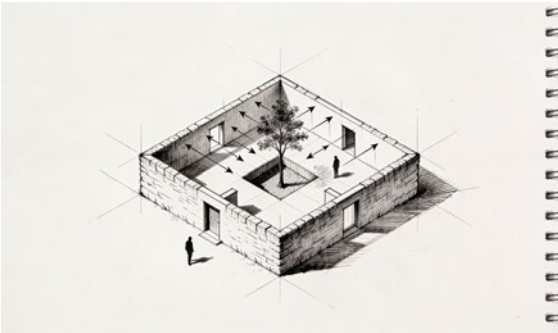
Qwen Image 2512



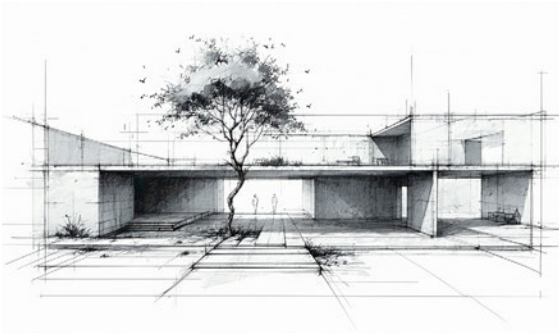
Recraft V4



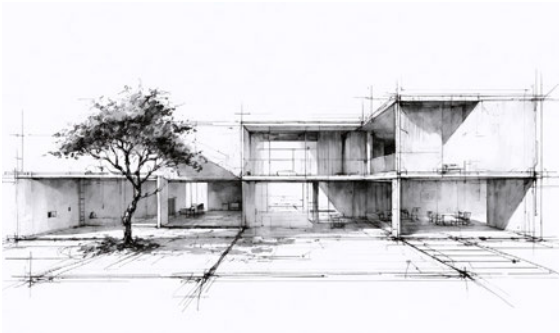
Runway



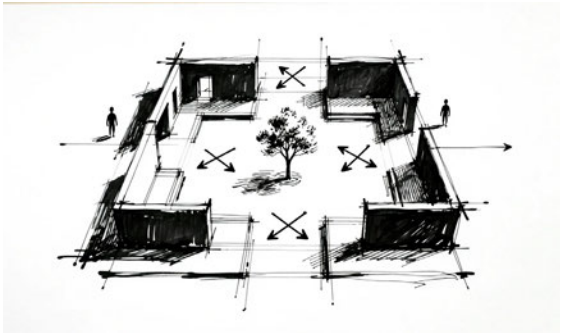
Seedream 4.5



Stable Diffusion 3.5



Stable Image Ultra



Wan 2.5

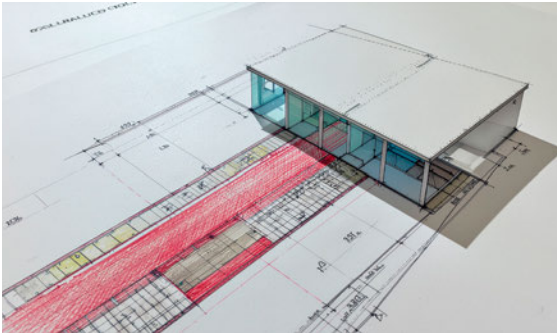


Z-Image Turbo

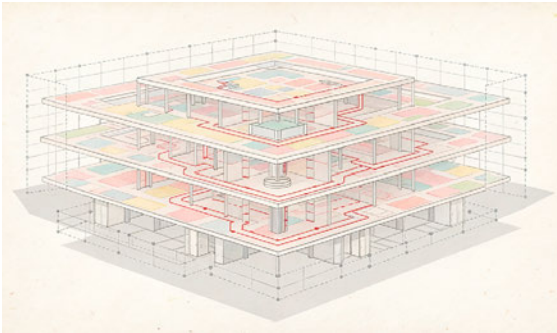
Image Generation

Coeptual - 15 Parameters

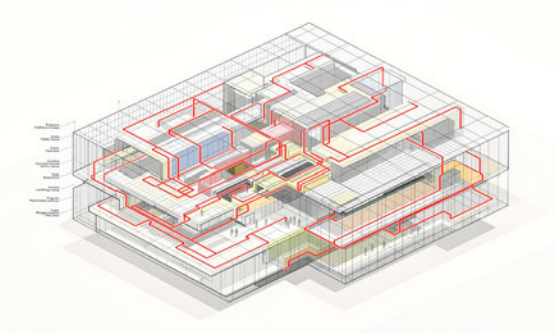
Generate a conceptual axonometric drawing of a cultural pavilion exploring layered public circulation. Use exploded axonometric projection with separated floor plates. Circulation paths highlighted in red linework. Program zones rendered in muted pastel blocks. Structural grid shown as thin grey lines. Transparent façade indicated with dashed outlines. Annotations and diagram labels in small uppercase typography. Background is off-white paper texture. Include shadow cast beneath the exploded layers for depth. Emphasize clarity over realism. Graphic, precise, analytical representation.



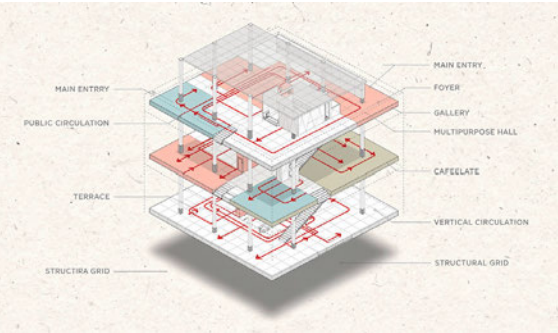
Firefly 4 Ultra



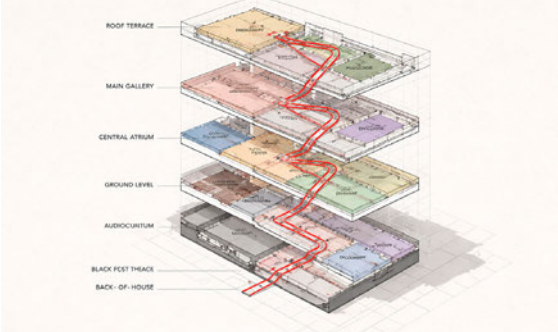
Firefly 5 Preview



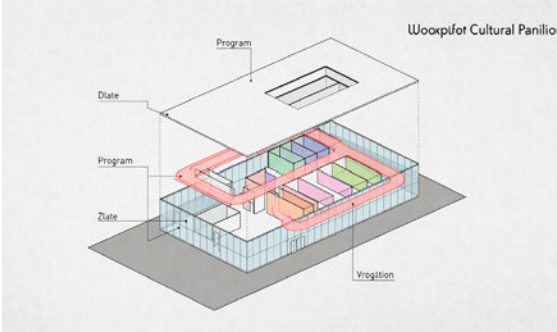
Flux 1.1 Ultra



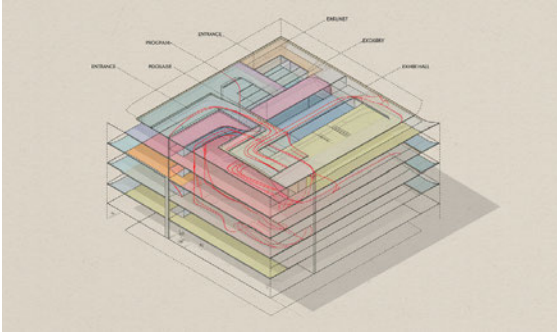
Flux 2 Max



GPT Image 1.5



Grok



Ideogram V3

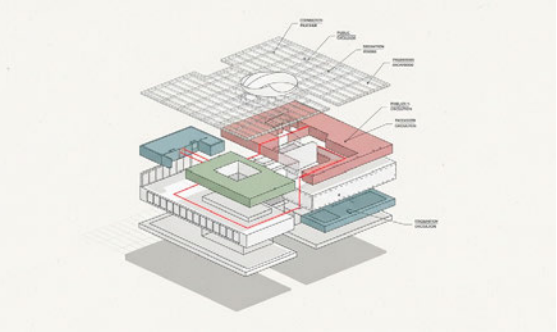
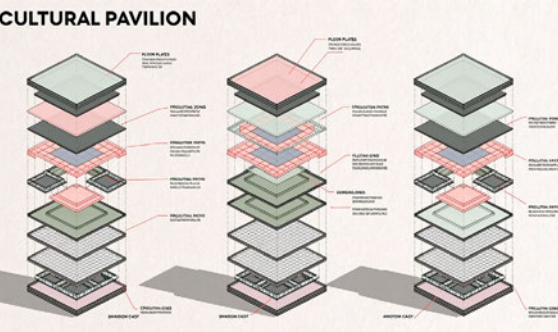
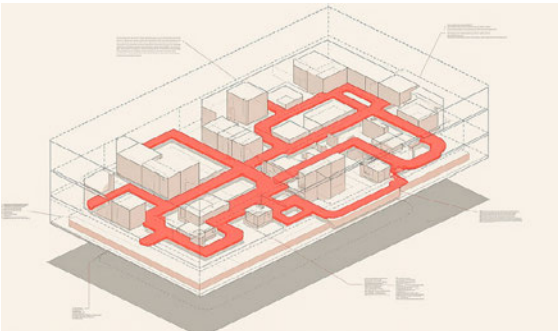


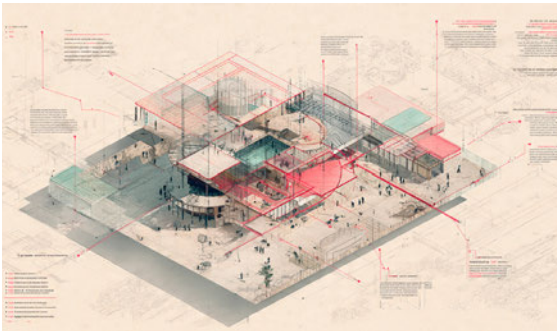
Imagen 4



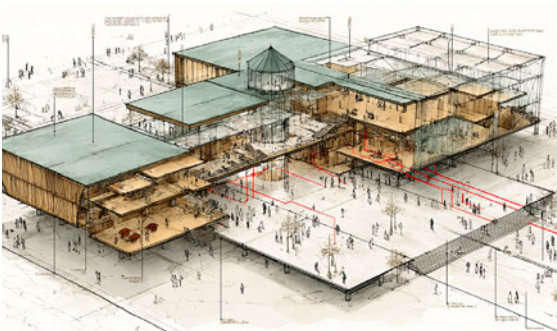
Kling 3.0 Omni



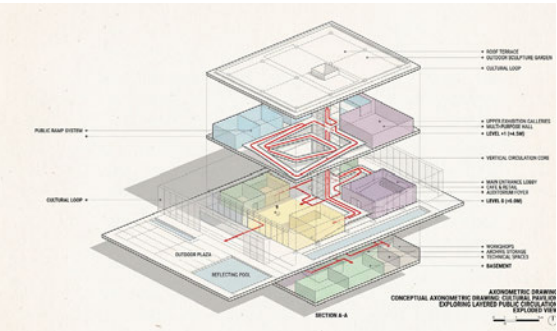
Krea I



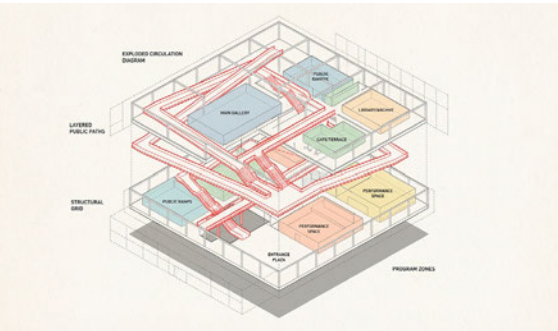
Meta I



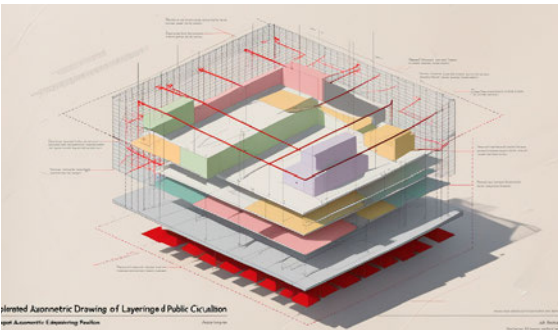
Midjourney



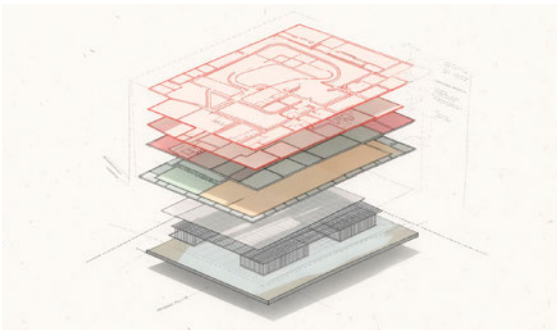
Nano Banana 2



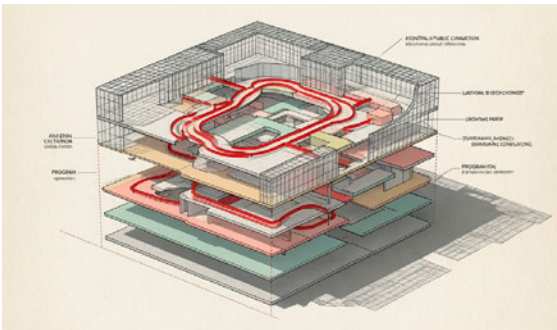
Nano Banana Pro



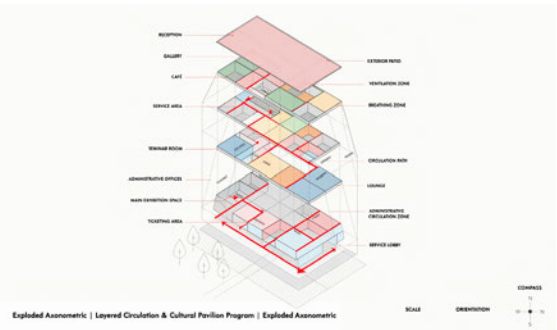
Phoenix 1.0



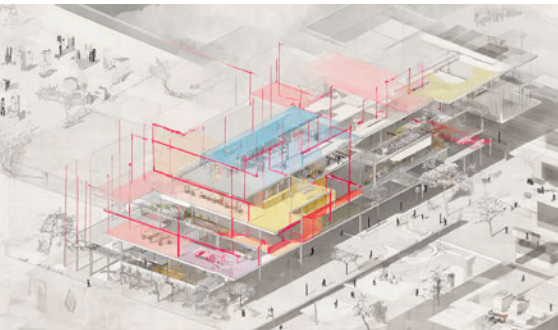
Photon 1



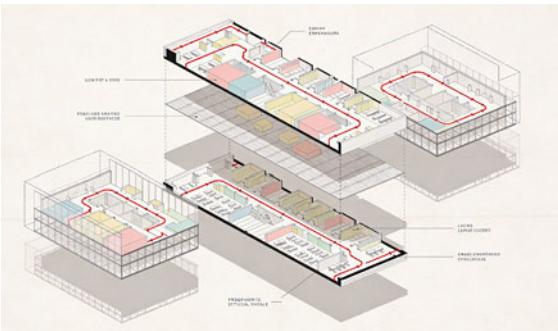
Qwen Image 2512



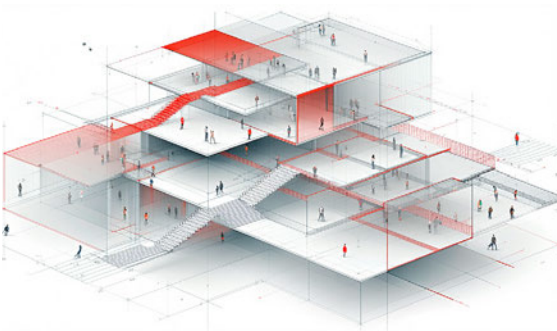
Recraft V4



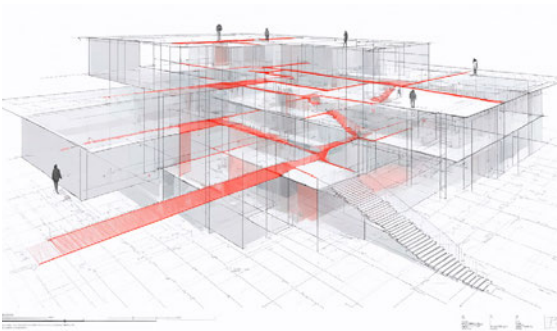
Runway



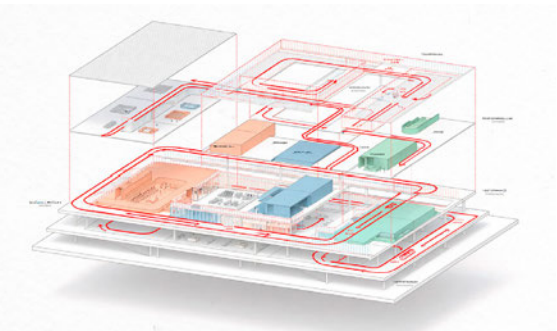
Seedream 4.5



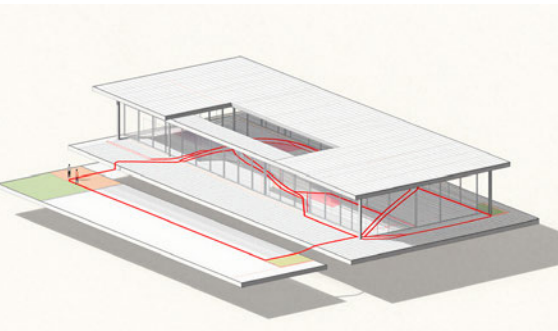
Stable Diffusion 3.5



Stable Image Ultra



Wan 2.5

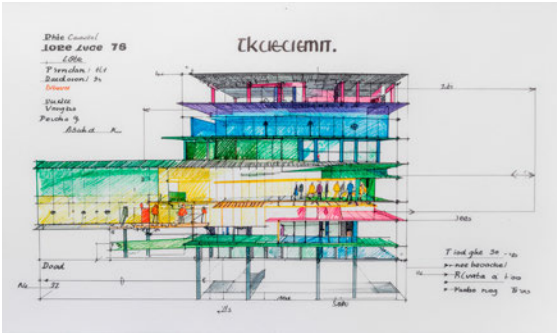


Z-Image Turbo

Image Generation

Coeptual - 23 Parameters

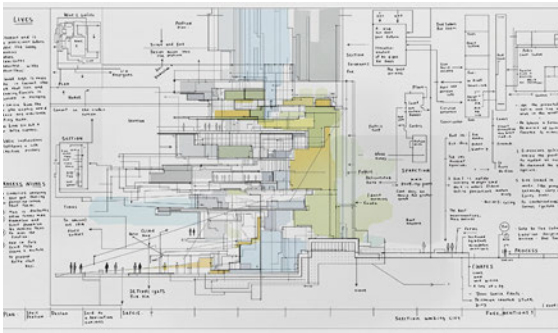
Generate a conceptual architectural design board combining sectional perspective, plan fragments, and process diagrams for a speculative vertical city prototype. Use mixed media style: graphite linework, soft watercolor washes, and digital overlays. Include layered tracing-paper effect with partially visible iterations beneath. Section cut reveals stacked public platforms and voids. Circulation mapped with gradient arrows. Environmental forces (wind, sun, density) diagrammed around the form. Scaled human figures vary intentionally to test spatial ambiguity. Include handwritten notes, diagram keys, and timeline sketches along the margins. Composition asymmetrical but structured on a clear grid. Emphasize process, experimentation, and unresolved design thinking.



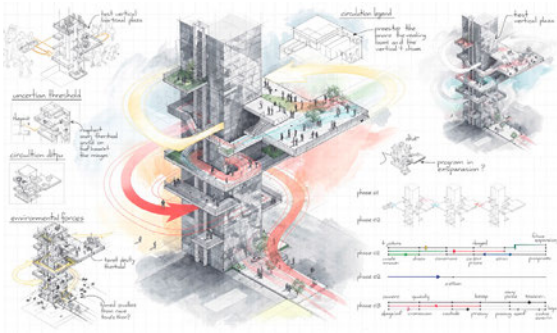
Firefly 4 Ultra



Firefly 5 Preview



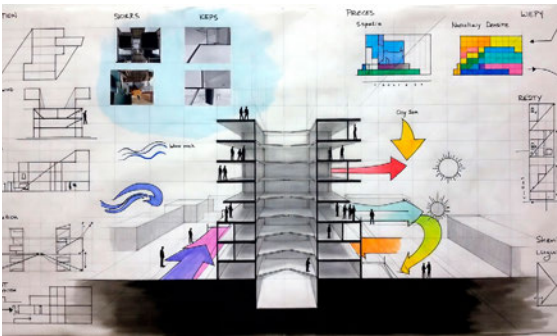
Flux 1.1 Ultra



Flux 2 Max



GPT Image 1.5



Grok



Ideogram V3

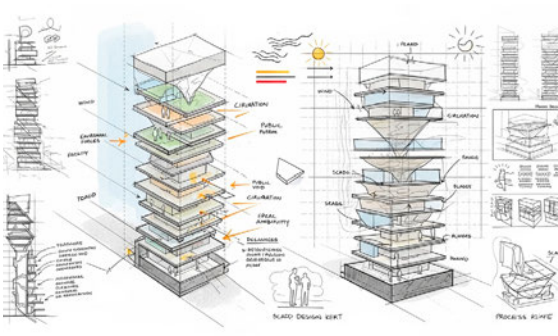
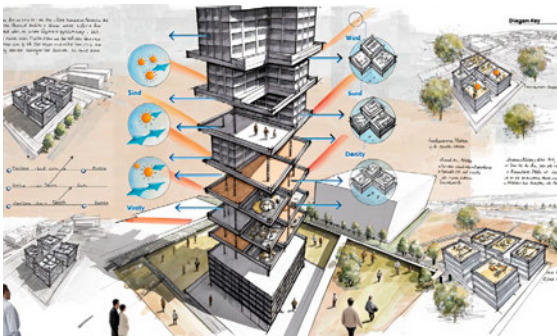
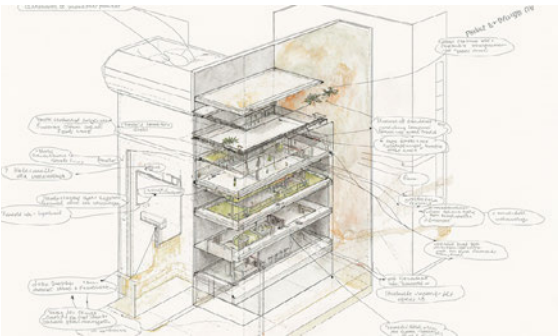


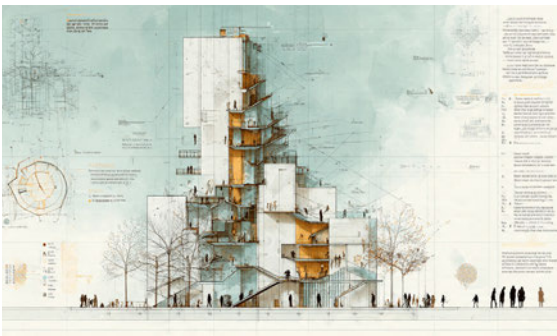
Imagen 4



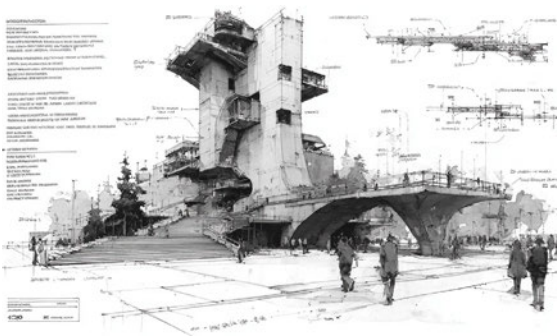
Kling 3.0 Omni



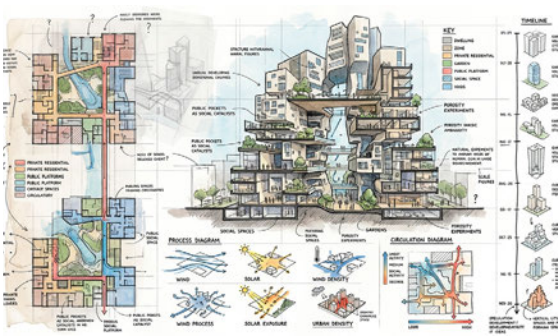
Krea 1



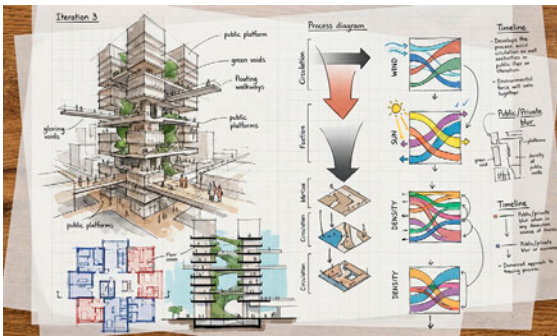
Meta AI



Midjourney



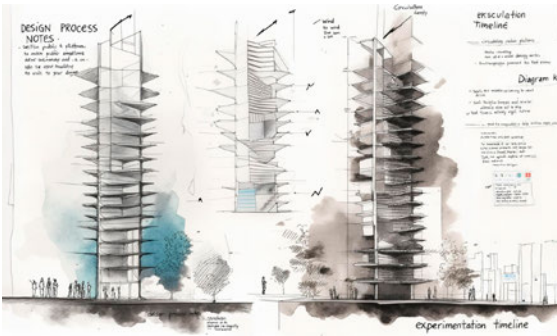
Nano Banana 2



Nano Banana Pro



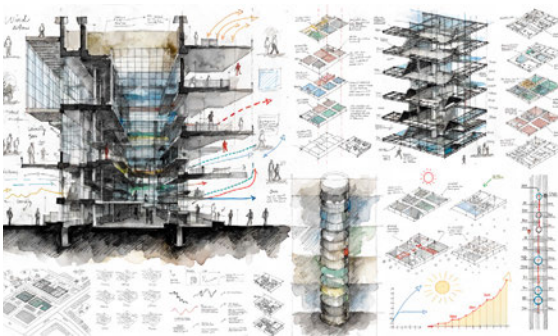
Phoenix 1.0



Photon 1



Qwen Image 2.5



Recraft V4



Runway



Seedream 4.5



Stable Diffusion 3.5



Stable Image Ultra



Wan 2.5



Z-Image Turbo

Image Generation

Alien - 13 Parameters

Generate a conceptual architectural image of a floating cubic pavilion suspended above a flat desert plane. The cube is hollow at its center, pierced by a vertical void of light. Its surface is matte white with no visible joints. A single narrow stair connects ground to the cube. No surrounding buildings, only horizon and sky. Harsh overhead sun creates sharp shadows. Composition is centered and symmetrical. Atmosphere is silent, minimal, and abstract.



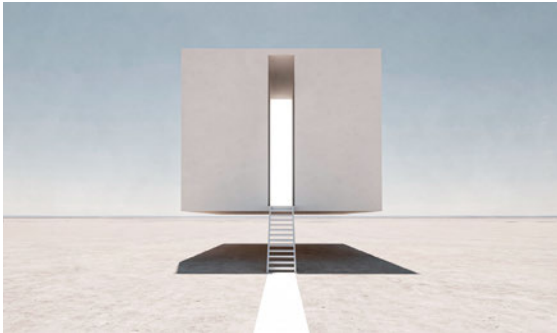
Firefly 4 Ultra



Firefly 5 Preview



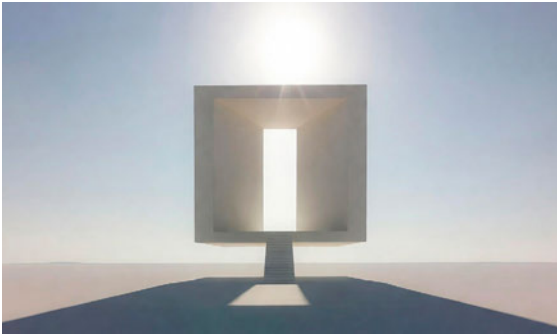
Flux 1.1 Ultra



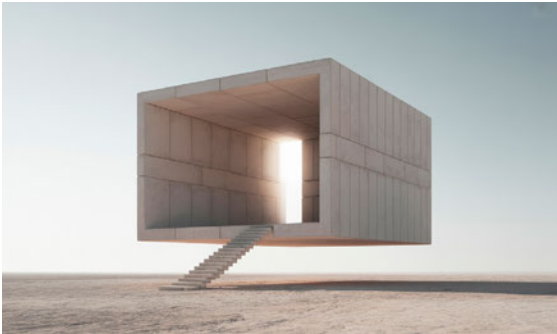
Flux 2 Max



GPT Image 1.5



Grok



Ideogram V3



Imagen 4



Kling 3.0 Omni



Krea 1



Meta AI



Midjourney



Nano Banana 2



Nano Banana Pro



Phoenix 1.0



Photon 1



Qwen Image 2.5 12



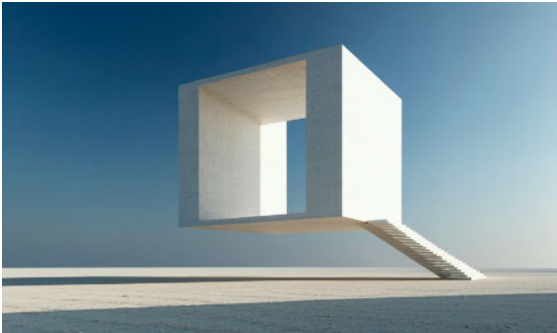
Recraft V4



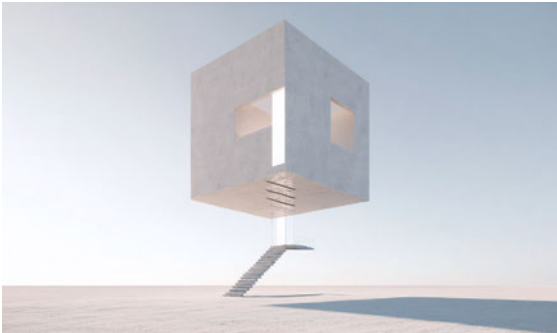
Runway



Seedream 4.5



Stable Diffusion 3.5



Stable Image Ultra



Wan 2.5



Z-Image Turbo

Image Generation

Alien - 23 Parameters

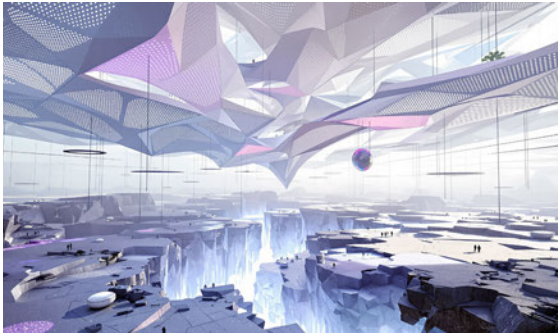
Generate a conceptual architectural megastructure hovering over a fractured landscape, composed of interlocking parametric membranes stretched between skeletal frames. Volumes fold, twist, and dissolve into perforated surfaces. Circulation is implied through suspended platforms and hanging capsules. The ground below is split into geometric fissures emitting diffuse light. Structural cables extend beyond the frame, disappearing into void. Interior and exterior conditions blur. Scale is ambiguous. Atmospheric particles drift through the space. Lighting is multi-directional and surreal, casting overlapping colored shadows. Perspective is aerial oblique with exaggerated depth. Composition is asymmetrical yet balanced. Focus on abstraction, diagrammatic clarity, and spatial complexity rather than realism.



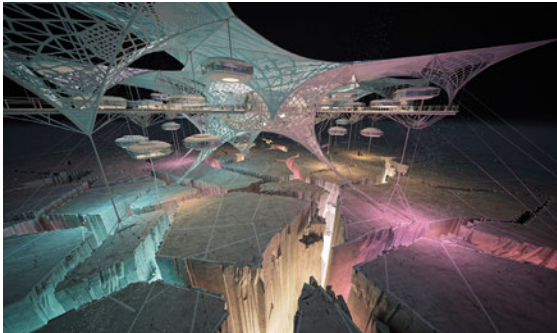
Firefly 4 Ultra



Firefly 5 Preview



Flux 1.1 Ultra



Flux 2 Max



GPT Image 1.5



Grok



Ideogram V3

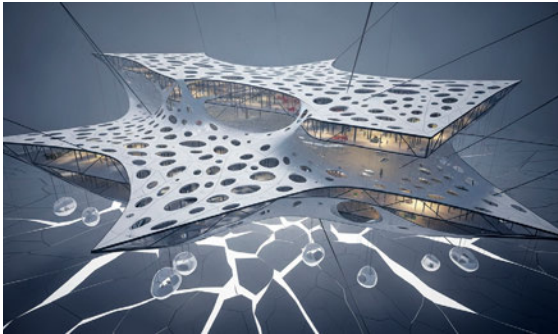
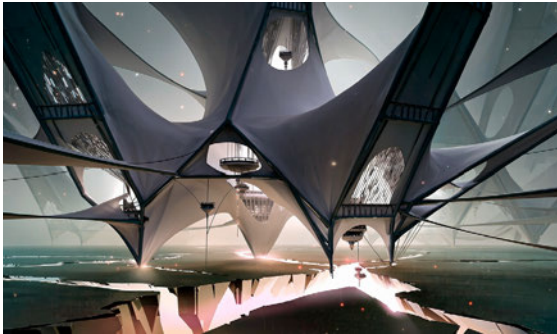


Imagen 4



Kling 3.0 Omni



Krea 1



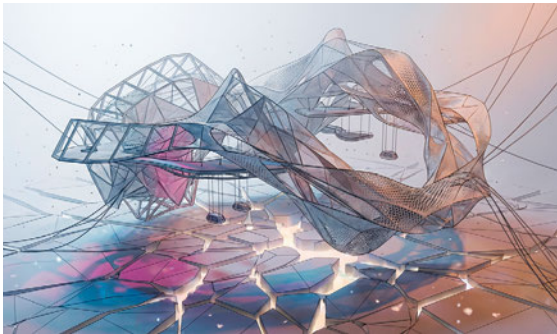
Meta AI



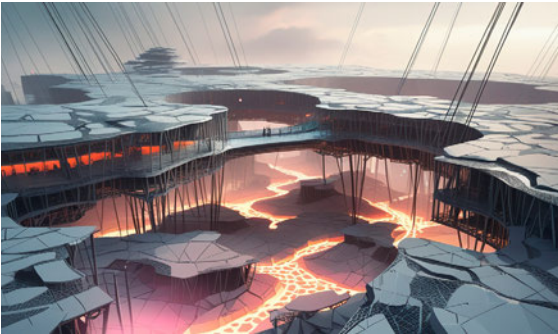
Midjourney



Nano Banana 2



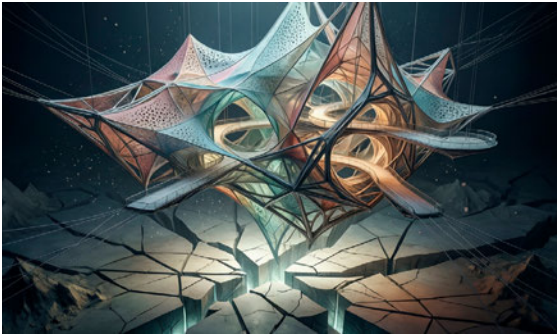
Nano Banana Pro



Phoenix 1.0



Photon 1



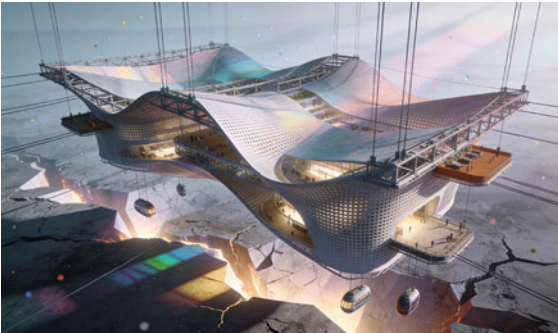
Qwen Image 2512



Recraft V4



Runway



Seedream 4.5



Stable Diffusion 3.5



Stable Image Ultra



Wan 2.5



Z-Image Turbo

3.3.3. Editing Models

The image editing models are assessed through the same protocol, with two differences that follow from the nature of their task. Because they manipulate an existing image rather than synthesize one from a prompt, neither the graded complexity nor the stylistic branching used for the generation models applies. In their place, the models are tested across a range of representative editing scenarios, such as changing the weather in a scene or turning a screenshot from 3D modeling software into a finished render, and are judged on how faithfully and convincingly they carry each operation out.

Image Editing

Material - 12 Parameters

Change the building complex facade from light stone to red adobe while preserving the original geometry, massing, bridges, perforated screens, and window proportions. Replace all exterior surfaces with textured red adobe featuring subtle tonal variation, clay imperfections, soft rounded edges, and slight corner erosion for authenticity. Reinterpret the v perforated panels as carved adobe mashrabiya-style screens while maintaining their depth and pattern complexity. Keep the warm interior lighting, stepped landscape, water canal, vegetation, and overall composition unchanged. Ensure realistic matte material behavior with light dust and soft shadows appropriate for a hot arid climate.



Author



FIBO



Flux.1 Kontext Dev



Flux.2 Klein



GPT Image 1.5



Grok Edit



Nano Banana 2



Nano Banana Pro



Qwen Edit 2511



Seedream 5.0 Lite



Wan 2.5

Image Editing

Time & Weather - 12 Parameters

Transform the image into a calm winter night scene while keeping the exact same composition, camera angle, perspective, architecture, framing, and all existing people in their precise positions and poses without removing or adding anything. Change the overcast daylight to nighttime with a dark sky softly illuminated by ambient city glow. Add warm yellow light from shop interiors and some residential windows, creating a cozy contrast with cool blue exterior tones. The light should spill naturally onto the wet cobblestone street with realistic reflections. Introduce gentle snowfall, not a blizzard, with visible snowflakes under the lighting and a thin natural layer of snow on cobblestones, rooftops, ledges, and bicycles while keeping textures visible. Keep the people unchanged, with only subtle snow dusting if appropriate. Maintain high photographic realism and cinematic atmosphere.



Author



FIBO



Flux.1 Kontext Dev



Flux.2 Klein



GPT Image 1.5



Grok Edit



Nano Banana 2



Nano Banana Pro



Qwen Edit 2511



Seedream 5.0 Lite



Wan 2.5

Image Editing

Staging - 28 Parameters

Stage this empty interior in a Japandi loft style while preserving the exact room geometry, window and door positions, ceiling height, and camera angle. Add warm light oak plank flooring. Paint the walls a soft warm beige with one muted clay accent wall at the far end. Install sheer linen curtains on all windows. Add a low modular linen sofa centered in the room with a textured wool area rug beneath it. Place a solid wood coffee table and a slim black metal side table. Add a built-in oak media unit along the back wall. Include a minimalist dining table with four wishbone chairs near the windows. Install sculptural pendant lighting above the dining area and a paper floor lamp near the sofa. Add recessed dimmable ceiling lights with warm temperature. Introduce indoor plants in ceramic pots. Layer natural textures such as rattan, linen, and wool. Add subtle wall art with abstract earth tones. Ensure balanced composition, realistic scale, soft daylight mixed with warm artificial lighting, and high photorealistic material detail.



Author



FIBO



Flux.1 Kontext Dev



Flux.2 Klein



GPT Image 1.5



Grok Edit



Nano Banana 2



Nano Banana Pro



Qwen Edit 2511



Seedream 5.0 Lite



Wan 2.5

Image Editing

Render - 27 Parameters

Transform this architectural screenshot into a hyper realistic photograph while strictly preserving the exact building geometry, proportions, façade pattern, structural elements, landscaping layout, street width, and the original camera angle, height, focal length, and perspective. Maintain the same composition and framing. Set the scene at late afternoon around 5:30 PM with soft golden hour sunlight casting long natural shadows and warm highlights. Use clear, slightly hazy weather with realistic atmospheric depth. Replace all flat materials with photorealistic ones: textured concrete, brushed metal panels, reflective low-e glass with subtle interior reflections, realistic stone paving, asphalt road with surface imperfections, and detailed water in the fountains. Convert all people into real human figures with natural skin tones, clothing textures, and accurate scale. Turn vegetation into detailed trees with natural leaf variation and light interaction. Add physically accurate lighting, global illumination, soft reflections, subtle lens characteristics, high dynamic range, and true-to-life color grading for a cinematic architectural photograph.



Author



FIBO



Flux.1 Kontext Dev



Flux.2 Klein



GPT Image 1.5



Grok Edit



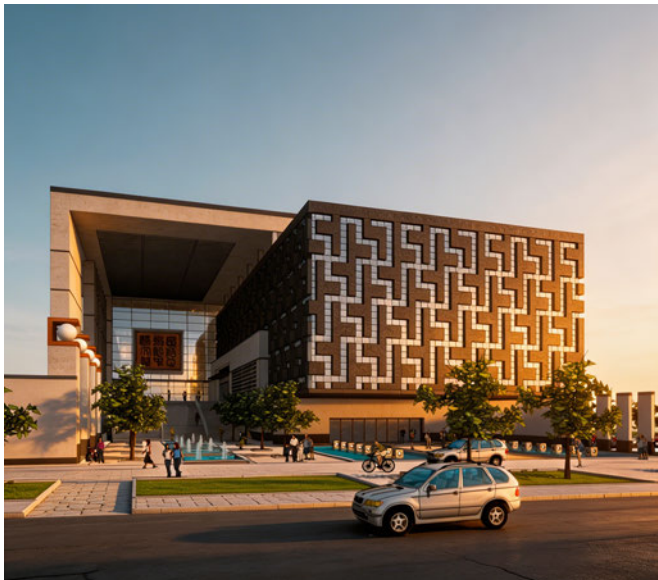
Nano Banana 2



Nano Banana Pro



Qwen Edit 2511



Seedream 5.0 Lite



Wan 2.5

Image Editing

Placement - 13 Parameters

Using the empty urban land plot image as the base and the modern concrete mid-rise building image as the reference, seamlessly insert the building into the fenced lot, removing the fence, debris, and vegetation. Align it flush with the adjacent brick façades, maintaining the existing street wall and sidewalk edge, and scale it accurately to match neighboring height and width proportions. Match the original camera perspective, eye level, focal length, and vanishing points. Adapt lighting and shadows to the warm late-afternoon conditions, add realistic reflections of surrounding buildings in the glazing, and integrate the transparent ground floor naturally with the sidewalk. Apply subtle urban weathering to the concrete and blend color grading, contrast, and depth of field to create a fully photorealistic, seamless architectural composite with no distortions or perspective mismatch.



Author



Author



FIBO



Flux.1 Kontext Dev



Flux.2 Klein



GPT Image 1.5



Nano Banana 2



Nano Banana Pro



Gwen Edit 2511



Seedream 5.0 Lite



Wan 2.5

3.3.4.Evaluation Matrix

The isolated tests of the preceding two subsections are synthesized here into evaluation matrices that show which model is best suited to which task. The picture is consistent across both generation and editing: a small group of models performs well across every scenario, a broad middle is competent but uneven, and a short tail performs poorly.

For generation, baseline quality is now common, and the models no longer separate on whether they can produce a good image. They separate on whether they stay good as the task shifts from realistic to conceptual to alien. The two Nano Banana models, the Pro and the second version, lead the field by holding their quality across all three categories. A group of strong all-rounders sits just below them, among them GPT Image 1.5, Imagen 4, Seedream 4.5, Midjourney, Recraft V4, and Z-Image Turbo. A broad and competent middle follows, and a softer tail of Firefly 5 Preview, Phoenix 1.0, Photon 1, Runway, and Stable Image Ultra trails at the bottom. What distinguishes the leaders is not peak quality on any single image but the absence of degradation across the range.

The editing matrix takes the same shape across a narrower field. Nano Banana Pro and GPT Image 1.5 occupy the top, near-complete across every scenario, with Nano Banana 2 a short step behind. Seedream 5.0 Lite and Wan 2.5 form a solid second tier. FIBO, Flux 2 Klein, Grok Edit, and Qwen Edit 2511 are middling specialists, each strong in some scenarios and weak in others. Flux 1 Kontext Dev is the clear floor, performing poorly almost everywhere. General-purpose editing quality, in other words, has consolidated among a handful of models at the top, while most of the field remains uneven and one model is genuinely weak.

Read together, the two matrices point to a single conclusion. The same model family, Nano Banana, tops both tasks, which indicates that what now separates the best tools from the merely competent is robustness across scenarios rather than peak quality on any one of them.

AI Models/Criteria	Material		Time & Weather		Staging		Render		Placement		All Image Editing			
	Usability	Prompt Adherence	Usability	Prompt Adherence	Usability	Prompt Adherence	Usability	Prompt Adherence	Usability	Prompt Adherence	Usability	Prompt Adherence	Consistency	Versatility
FIBO	●	◐	●	●	◐	○	○	○	○	○	◐	◐	◐	◐
Flux 1 Kontext Dev	○	○	○	○	◐	◐	◐	◐	○	◐	○	○	○	○
Flux 2 Klein	○	◐	●	◐	●	●	◐	○	◐	◐	◐	◐	◐	◐
GPT Image 1.5	●	●	●	●	●	●	●	◐	●	◐	●	●	●	●
Grok Edit	●	●	◐	○	◐	◐	●	●	○	○	◐	◐	◐	◐
Nano Banana 2	●	●	◐	◐	●	●	●	●	●	●	●	●	●	●
Nano Banana Pro	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Qwen Edit 2511	◐	○	○	◐	◐	◐	◐	●	●	●	◐	◐	◐	◐
Seedream 5.0 Lite	●	◐	●	●	◐	◐	◐	○	●	●	●	◐	●	●
Wan 2.5	●	◐	◐	○	●	◐	◐	◐	●	●	●	◐	●	●

- Poor
- ◐ Passable
- Good

AI Models \ Criteria	Realistic Image Generation									Conceptual Image Generation									Alien Image Generation												All Image Generation			
	IG_R_01		IG_R_02		IG_R_03		Average			IG_C_01		IG_C_02		IG_C_03		Average			IG_A_01		IG_A_02		IG_A_03		Average									
	Usability	Prompt Adherence	Usability	Prompt Adherence	Usability	Prompt Adherence	Usability	Prompt Adherence	Consistency	Usability	Prompt Adherence	Usability	Prompt Adherence	Usability	Prompt Adherence	Usability	Prompt Adherence	Consistency	Usability	Prompt Adherence	Usability	Prompt Adherence	Usability	Prompt Adherence	Usability	Prompt Adherence	Consistency	Usability	Prompt Adherence	Consistency	Versatility			
Firefly 4 Ultra	🟢	🟡	🟢	🟡	🟡	🟡	🟢	🟡	🟡	🟢	🟡	🟡	🟡	🟡	🟢	🟡	🟡	🟡	🟡	🟡	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟡	🟡	🟡			
Firefly 5 Preview	🟡	🟢	🟡	🟡	🟡	🟢	🟡	🟢	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡			
Flux 1.1 Ultra	🟡	🟡	🟢	🟢	🟢	🟡	🟢	🟡	🟢	🟡	🟡	🟢	🟡	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟢	🟡	🟡	🟡	🟢	🟡	🟡	🟡	🟡	🟡	🟡			
Flux 2 Max	🟢	🟢	🟢	🟡	🟡	🟡	🟢	🟡	🟢	🟡	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟡	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟢	🟡			
GPT Image 1.5	🟢	🟢	🟢	🟢	🟢	🟡	🟢	🟢	🟢	🟡	🟡	🟡	🟡	🟢	🟢	🟡	🟡	🟡	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢			
Grok	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟢	🟡	🟡	🟡	🟢	🟢	🟡	🟡	🟢	🟡	🟢	🟡	🟢	🟡	🟡	🟡	🟡			
Ideogram V3	🟡	🟢	🟡	🟡	🟢	🟡	🟡	🟡	🟡	🟡	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟢	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡			
Imagen 4	🟢	🟡	🟢	🟡	🟡	🟢	🟢	🟡	🟢	🟢	🟢	🟢	🟢	🟡	🟡	🟢	🟢	🟢	🟢	🟡	🟢	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟢	🟡	🟢			
Kling 3.0 Omni	🟢	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟢	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟢	🟡	🟢	🟢	🟡	🟡	🟢	🟡	🟢	🟢	🟡	🟡	🟡			
Krea1	🟢	🟡	🟢	🟡	🟢	🟡	🟢	🟡	🟢	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢			
Meta AI	🟡	🟡	🟢	🟢	🟡	🟡	🟡	🟡	🟡	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟡	🟡	🟢	🟡	🟢	🟡	🟢	🟡	🟢	🟢	🟢	🟢	🟢			
Midjourney	🟡	🟡	🟢	🟢	🟢	🟡	🟢	🟡	🟢	🟢	🟢	🟢	🟡	🟢	🟢	🟢	🟢	🟢	🟢	🟡	🟢	🟢	🟢	🟡	🟢	🟡	🟢	🟢	🟢	🟢	🟢			
Nano Banana 2	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟡	🟡	🟢	🟢	🟢	🟢	🟡	🟢	🟢	🟢	🟡	🟢	🟡	🟢	🟢	🟢	🟡	🟢	🟢	🟢	🟢	🟢			
Nano Banana Pro	🟢	🟢	🟢	🟡	🟢	🟢	🟢	🟢	🟢	🟡	🟡	🟢	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡			
Phoenix 1.0	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟢	🟡	🟡	🟡	🟡	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡			
Photon 1	🟡	🟢	🟡	🟢	🟡	🟡	🟡	🟢	🟢	🟢	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡			
Qwen Image 2512	🟢	🟢	🟡	🟡	🟡	🟢	🟡	🟢	🟢	🟡	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟢	🟢	🟡	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢			
Recraft V4	🟢	🟡	🟢	🟢	🟢	🟡	🟢	🟡	🟢	🟡	🟡	🟢	🟢	🟢	🟢	🟡	🟡	🟡	🟢	🟡	🟢	🟢	🟡	🟡	🟢	🟡	🟢	🟢	🟢	🟢	🟢			
Runway	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟢	🟡	🟢	🟡	🟡	🟡	🟢	🟡	🟢	🟡	🟡	🟢	🟡	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡			
Seedream 4.5	🟢	🟢	🟡	🟡	🟡	🟢	🟡	🟢	🟢	🟡	🟡	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟢	🟡	🟡	🟡	🟡	🟢	🟡	🟢	🟢	🟡	🟡	🟢	🟡			
Stable Diffusion 3.5	🟢	🟢	🟡	🟢	🟡	🟡	🟡	🟡	🟡	🟢	🟡	🟢	🟡	🟢	🟢	🟢	🟡	🟢	🟢	🟡	🟢	🟢	🟡	🟢	🟡	🟢	🟡	🟢	🟢	🟢	🟢			
Stable Image Ultra	🟡	🟡	🟢	🟡	🟢	🟡	🟢	🟡	🟢	🟢	🟡	🟢	🟡	🟢	🟢	🟢	🟡	🟢	🟢	🟡	🟢	🟡	🟢	🟡	🟢	🟡	🟢	🟢	🟢	🟢	🟢			
Wan 2.5	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟢	🟢	🟢	🟢	🟡	🟡	🟡	🟡	🟡	🟡	🟡	🟢	🟢	🟢	🟡	🟢	🟢	🟢	🟢	🟢			
Z-Image Turbo	🟢	🟡	🟡	🟡	🟢	🟢	🟢	🟡	🟢	🟡	🟡	🟡	🟡	🟢	🟢	🟡	🟡	🟡	🟢	🟢	🟢	🟢	🟢	🟡	🟢	🟡	🟢	🟢	🟢	🟢	🟢			

○ Poor
◐ Passable
● Good

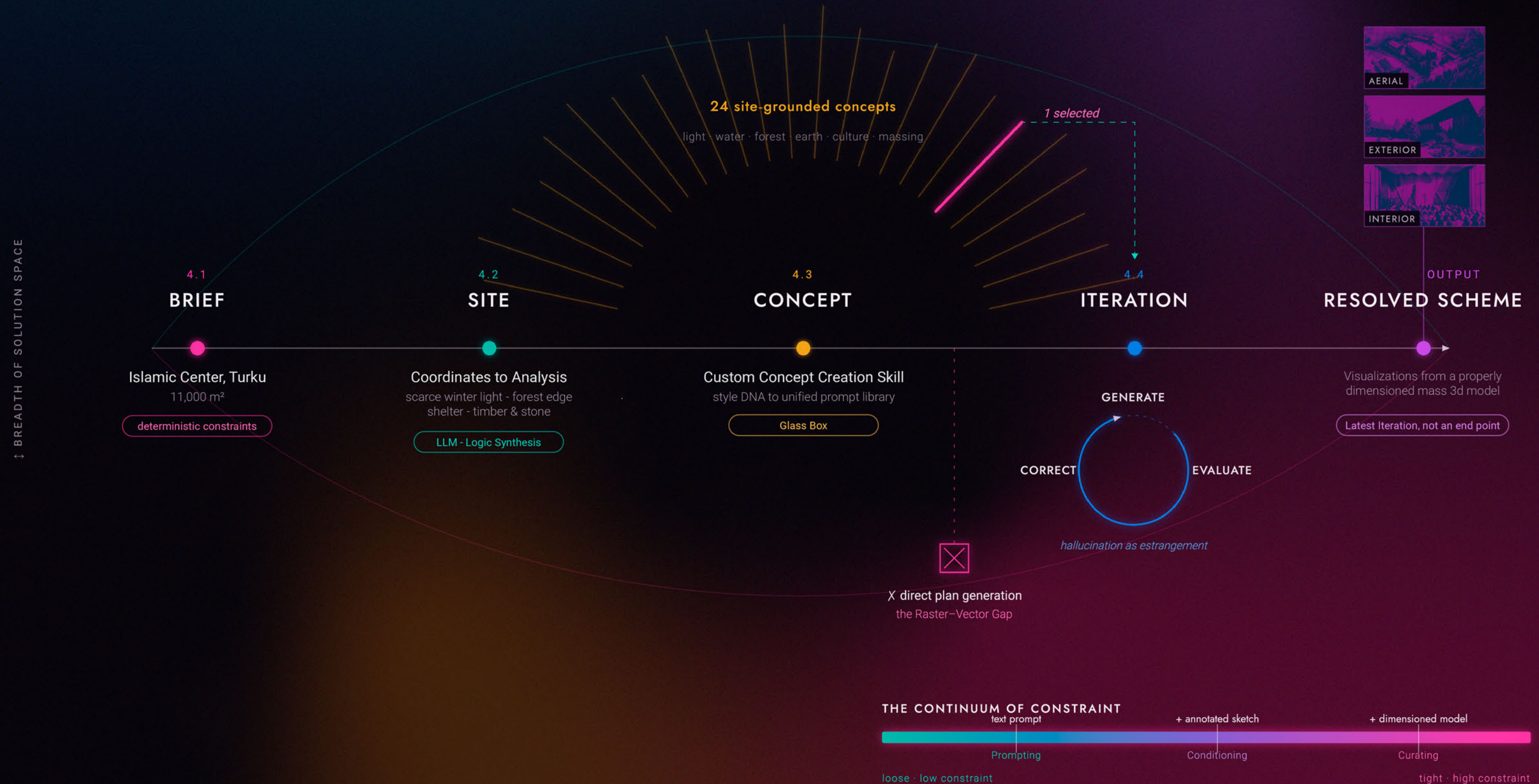
IV

Applied Synthesis: EMBER

EMBER

AI-augmented design process

A single variable governs the workflow: the quantity & percision of information given to the models to range from *Loose* to *Locked* geometry.



4.1. Project Brief

EMBER is a self-directed design exercise: the design of a new Islamic Center on a site in Turku, Finland, conceived as the proving ground for the framework built in the preceding chapters. The brief centers on a mosque supported by educational, cultural, social, and commercial facilities, a spiritual and communal hub for the city's Muslim population that stays open to the wider public. Its governing ambition is a contemporary Islamic architecture native to the Nordic setting rather than transplanted from elsewhere, a building that supports cultural dialogue and a sense of belonging within the Finnish urban fabric. This mix of programmatic breadth, symbolic weight, and contextual specificity was chosen deliberately. It demands both the divergent, atmosphere-driven exploration that generative tools accelerate and the geometric discipline the controlled workflow of Section 4.4 enforces.

The typology reinforced the choice. An Islamic center is an uncommon building type, thinly represented in the datasets these models are trained on, and a contemporary Nordic mosque is rarer still. The scheme therefore pushes the tools past the familiar typologies they reproduce fluently, into the margins of their training distribution. That same scarcity of data makes the brief a useful probe for bias. Where a model has seen few examples of a type, it falls back on the dominant stereotypes within that thin sample; a mosque for a Scandinavian setting is well placed to reveal whether the tools default to generic imagery or answer the specific context. The sections that follow set out the constraints adopted for the study: context, program, and design parameters.

4.1.1. Context and Design Intent

The building is conceived as more than a place of worship: a multifunctional institution that brings religious, educational, cultural, and social activity under one roof and encourages exchange among diverse communities.

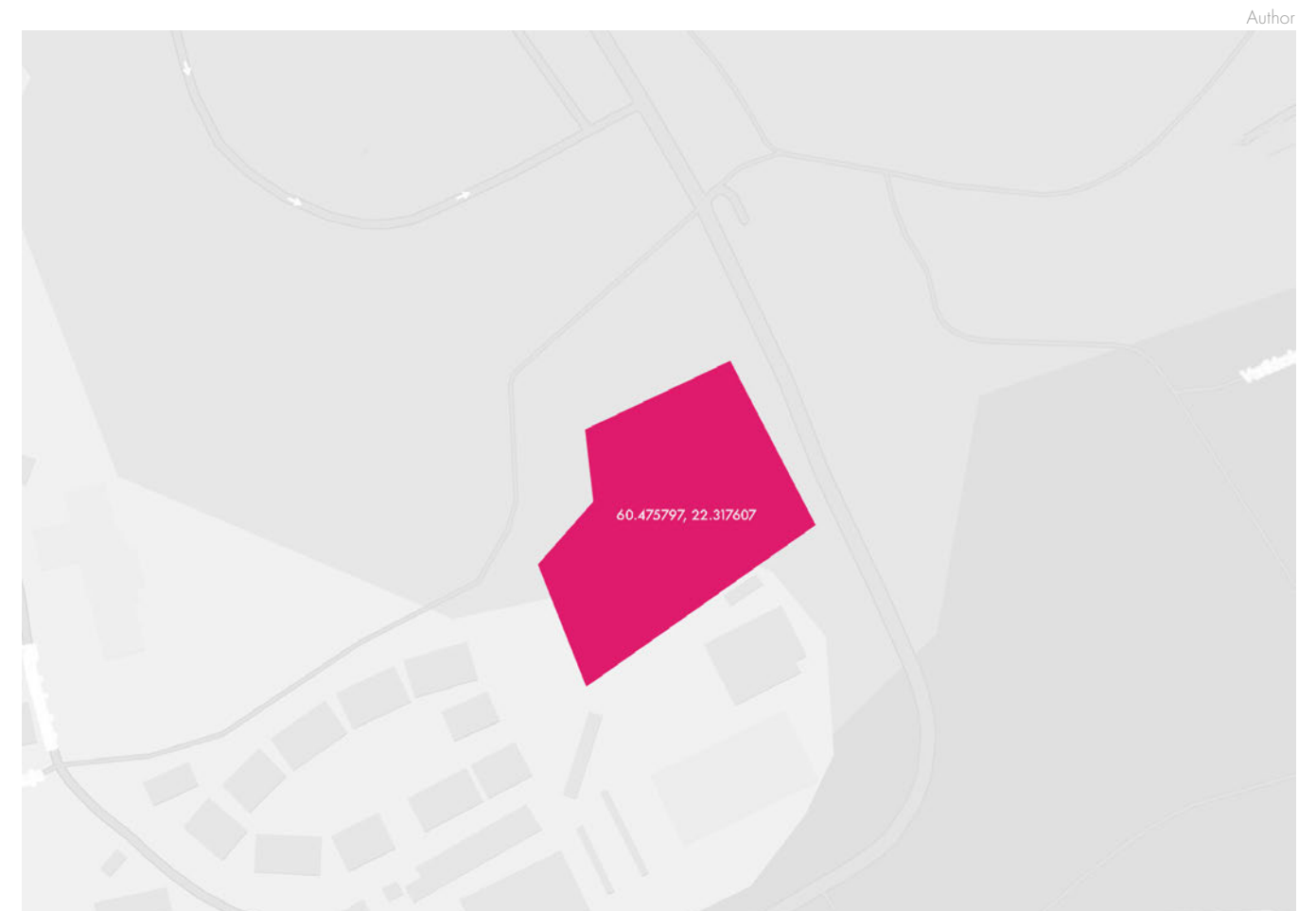
Several intentions structure the design problem. The Center should carry the spiritual and cultural values of Islam in a contemporary architectural language fitted to the Scandinavian setting, without pastiche or direct historical quotation. It should read as a distinctive civic landmark that enriches Turku's urban fabric and signals openness, inclusivity, and harmony. Its design and its eventual operation should prioritize accessibility, sustainability, and energy efficiency. It should also stay cost-effective to build and to run over the long term, and compliant with local planning and building regulations. Rather than a checklist, these intentions form the value system against which every later design and generative decision is measured.

4.1.2. Program

The program is broad and spread over two floors. It centers on the mosque: a men's prayer hall for some 800 to 1,000 worshippers, a women's prayer hall for about 300, separate ablution and sanitary facilities with their own entrances for men and women, and flexible multipurpose spaces for gatherings, lectures, and social events. The rest takes in classrooms for about twenty students each, administrative offices, a library

and reading area, a children's playroom, a kitchen and dining hall seating about 120, commercial rental units whose rents help cover the institution's operating costs, a kindergarten, a youth center, parking with accessible bays and drop-off zones, landscaped gardens and shaded pedestrian paths, and a playground and gathering spaces for children and families.

Beyond the schedule of spaces, a set of qualitative parameters shapes what counts as an acceptable solution. The design must carry spiritual symbolism, environmental responsibility, and architectural innovation together, and read as forward-looking rather than nostalgic. The brief also asks for careful handling of daylight and ventilation, in the service of comfort and energy use. Sustainability is pursued strategically, through renewable energy, recycled or local materials, and water-saving systems. Accessibility and safety are non-negotiable, with universal-design principles, emergency egress, and elevators built in from the start. The building must also sit well with its surroundings, respecting the existing urban fabric, topography, and cultural context, and comply with Turku's municipal regulations. Together these parameters set the boundary conditions the generative workflow has to operate within: the deterministic constraints that the probabilistic tools of Chapter 3 must finally be made to respect.



4.2. Site Analysis

Before the site could inform the design, it had to be understood, and the way that understanding was assembled is itself part of the workflow this thesis examines. The entire site analysis below was compiled with a large language model, prompted with little more than the site's coordinates. The conventional route would take hours or days, drawing on scattered, unconnected sources: meteorological records for climate, solar tables for sun angles, geological surveys for ground conditions, the building code and municipal plan for the regulatory envelope, and separate literature for regional precedent. The model gathered and cross-referenced these strands into a single coherent, architecturally filtered briefing in minutes, leaving the designer to verify, interpret, and select. Not all of it proved equally decisive. The account below foregrounds the conditions that went on to shape the design, and that triage is itself part of the architect's work, a clear case of the augmentation this thesis argues for rather than automation. It is the "Logic Synthesis" use of large language models from Section 1.3.2 applied to site research.

The site lies in the coastal lowlands of south-western Finland near Turku, at roughly 60.5° N: a flat, open meadow of dry grass, bounded by dense boreal forest and edged by a quiet rural road. Its defining spatial frame is the treeline. That forest edge proved the single most consequential condition on the site, and the analysis returns to it below.

4.2.1. Climate, Light, and Shelter

The climate is humid continental (Köppen Dfb), Nordic but moderated by the Baltic into one of Finland's milder regions. Its defining feature is the long, cold, dark winter, with heavy snowfall and continuous cover from roughly late December to late March. Two consequences carried straight into the design. The first is snow: the snow load, reinforced by the deep Finnish tradition of the pitched roof, calls for steep roof planes that shed accumulation rather than hold it. The second, and the most design-critical condition on the site, is light.

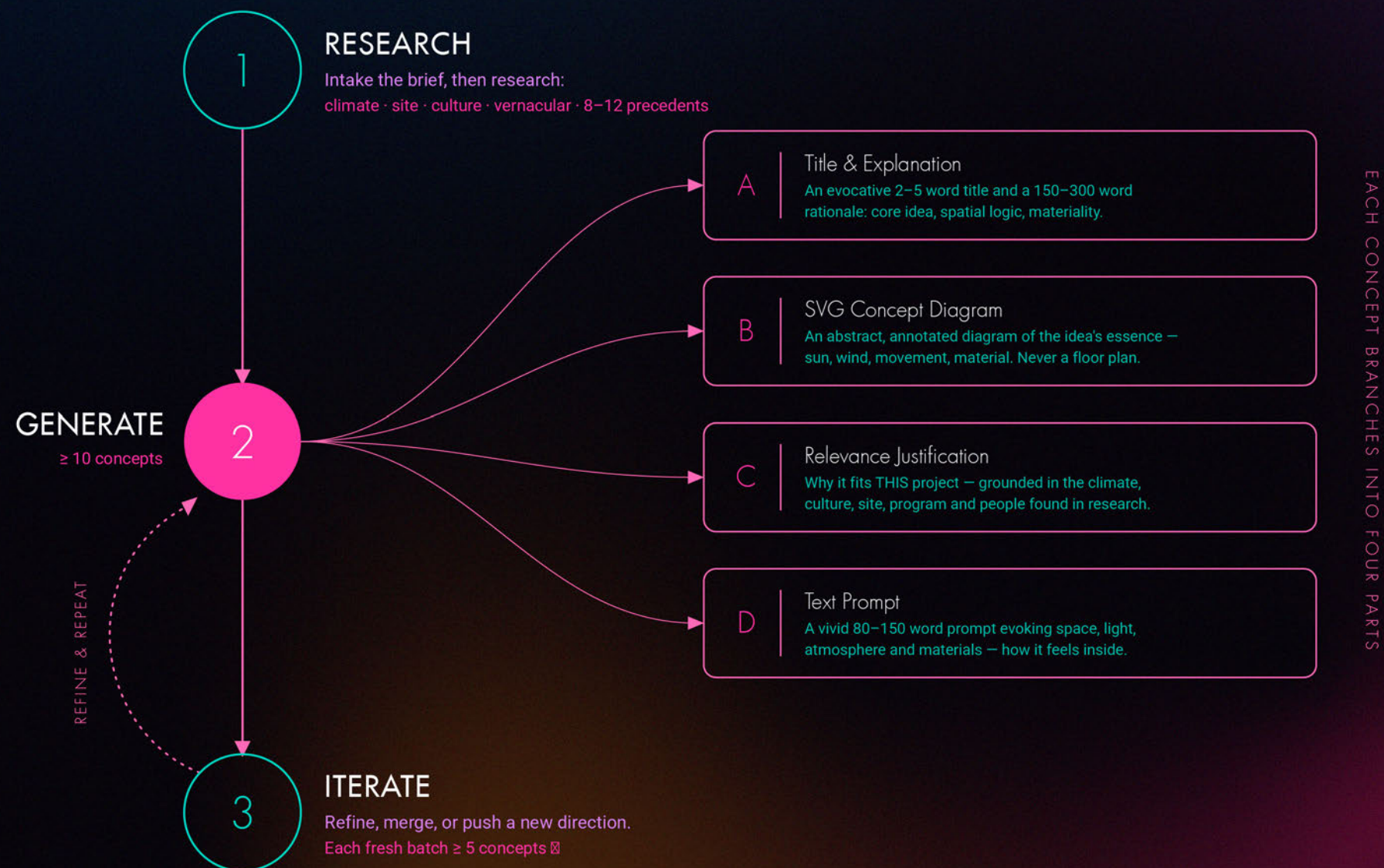
At this latitude the daylight swing is extreme, from nearly nineteen hours at the summer solstice to under six at midwinter, when the sun barely clears the horizon and the treeline throws long shadows across the clearing. Winter daylight is scarce, low, and precious. The opportunity it offers is qualitative rather than quantitative: not maximizing a resource but admitting and choreographing a limited one. For a building of spiritual character, that scarce light can become a carrier of meaning.

The open meadow gives little shelter, which makes the surrounding forest the main modifier of the site's microclimate. Rather than clear it, the analysis pointed toward keeping the treeline as natural shelter, settling the building within it, and organizing the program around protected courtyards that turn the exposed clearing into usable outdoor space.

4.2.2. Architectural and Cultural Context

The site also sits within one of the world's most distinctive building cultures, and the intent set out in Section 4.1, a contemporary Islamic architecture native to the Nordic setting, depends on engaging that tradition. Finnish architecture is shaped by a hard climate, abundant forest, and a deep regard for nature and craft. Its line runs from vernacular log construction and the medieval stone churches of the south-west, through the granite-and-timber National Romanticism of Saarinen, to the daylight-attuned modernism of Alvar Aalto and on to a contemporary practice that leads the world in engineered mass timber. Its recurring qualities (siting among trees, the careful handling of scarce daylight, tactile natural materials, restrained forms) gave EMBER a direct vocabulary, as did its palette of pine and spruce, increasingly as cross-laminated and laminated veneer timber, alongside stone and patinating metal roofing. In this culture, generous daylight, durable natural materials, and low operational energy are baseline expectations, not aspirations. For a flat, forest-framed clearing in southern Finland, the tradition points to a grounded, timber-and-stone architecture that addresses its setting directly: the contextual logic that, more than any other, shaped the design that follows.





4.3. Concept

With the site understood as a set of climatic, topographic, and cultural constraints, the concept phase turns to the divergent half of the process: expanding the solution space before any geometry is fixed. Instead of the ad-hoc, single-prompt ideation criticized in Chapter 1, this phase ran on a purpose-built instrument that makes concept generation systematic and repeatable. What follows describes that instrument, presents the concepts it produced, identifies the direction taken forward, and sets out the prompt system used to visualize the chosen scheme space by space.

4.3.1. Concept-Generation Skill

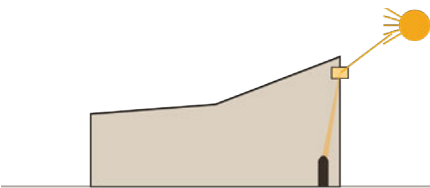
The first step was to build a custom Claude skill, the architecture-concept-creator, a reusable protocol for producing complex architectural concepts on demand. It is an instance of the "Logic Synthesis" use of large language models from Section 1.3.2 and of the augmentation paradigm of Section 2.2.1: it widens the range of ideation rather than replacing the designer's judgment, and keeps the logic of generation transparent and reproducible.

The skill encodes three things. First, it embeds a working definition of a strong architectural concept: a buildable design logic that answers to place, people, climate, culture, or program, not a style, a shape, or a floor-plan gimmick. It also holds the principle that the best concept is "clever but simple" yet carries real depth, and it forbids shape-driven or cliché proposals. Second, it specifies a structured intake of project information (location, site, climate, sun and wind, culture, typology, program, users, vision, constraints, references) and an instruction to research and fill any gaps, so each concept rests on real context rather than generic association. Third, it fixes the output: every concept arrives as a titled idea with an explanation, an abstract diagram, a relevance justification tied to the gathered data, and an image-generation prompt, produced in batches large enough to force genuine variety across contextual, tectonic, phenomenological, cultural, and other types. In operation, the skill turns the "slot machine" of unstructured prompting into a controlled, auditable procedure: the same shift from Black Box to Glass Box the thesis argues for in image generation, applied here to ideas.

4.3.2. Concept Catalogue

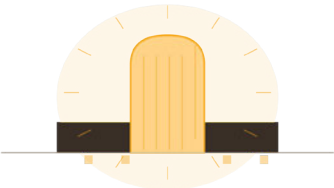
Applied to EMBER, the skill produced a large field of directions across several sessions, later consolidated to twenty-four distinct concepts by merging close relatives. Each is a design logic rooted in the conditions established in Section 4.2 (the low winter sun, the freezing and thawing year, the boreal forest edge, the bedrock and its thermal mass, and the Finnish building culture a contemporary Islamic architecture must enter) rather than a stylistic motif.

The Qibla Light: a horizontal slot in the qibla wall calibrated so the low winter-solstice sun strikes the mihrab as a slow-travelling band of gold, so that light, not ornament, becomes the mihrab.



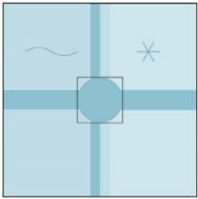
Author,
Claude Design,
Nano Banana 2

The Winter Lantern: a translucent prayer-hall shell that glows from within across the dark fields like a paper lantern in the snow, projecting girih light-patterns outward.



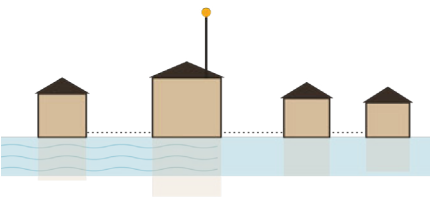
Author,
Claude Design,
Midjourney

The Freeze-Thaw Garden: a Persian chahar bagh re-tuned to a climate where water is liquid half the year and ice the other, its rills freezing into a walkable sheet and its snow reservoirs refilling them in spring.



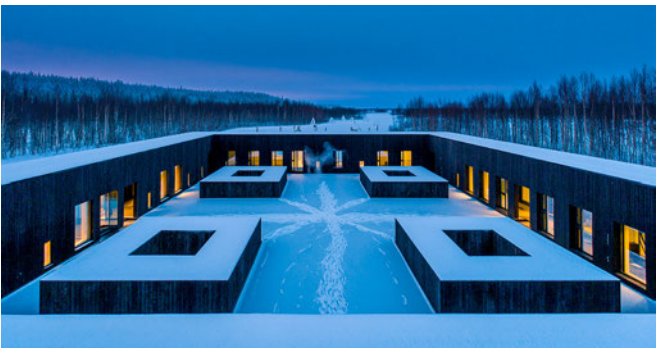
Author,
Claude Design,
Midjourney

The Archipelago: discrete pavilions set as islands in a shallow reflecting pool that freezes into a walkable winter plaza, modeled on the Turku archipelago.



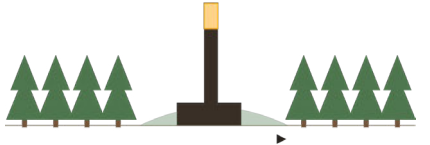
Author,
Claude Design,
Midjourney

Taken together, the catalogue shows the breadth the skill was built for: one consistent procedure yielding two dozen distinct, site-grounded design logics across phenomenological, tectonic, cultural, programmatic, and compositional lines. This is the divergent expansion the framework assigns to this phase.

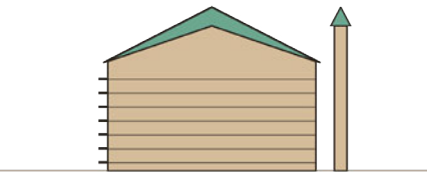


The Boreal Clearing: the building as a void carved into a forest the project itself plants, so that over decades the forest, not the building, becomes the landmark.

Author,
Claude Design,
Nano Banana 2



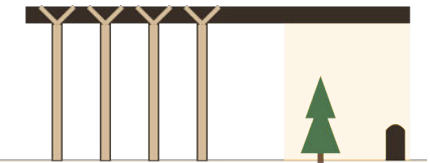
Author,
Claude Design,
Nano Banana 2



The Continuing Loghouse: the contemporary chapter of Finland's Tatar wooden-mosque tradition, built at civic scale in cross-laminated timber with exposed end-grain joinery.



Author,
Claude Design,
Midjourney

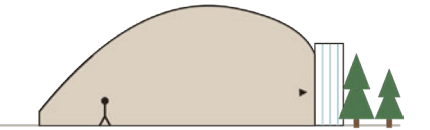


The Hypostyle Grove: the column-forests of Córdoba and Kairouan rendered as a glulam grove that dissolves toward the qibla into an open clearing.



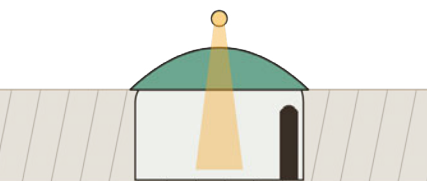
The Forest Mihrab: a glulam shell whose fully glazed qibla wall faces the densest forest, so worshippers face living trees rather than an ornamented wall.

Author,
Claude Design,
Midjourney

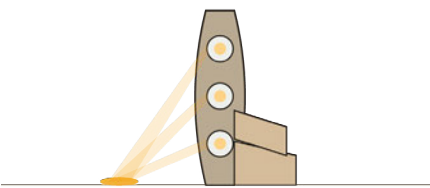


The Rock Mosque (Kalliomoskeija): a prayer hall half-excavated into the glacial bedrock, in dialogue with Tempeliaukio and the Cave of Hira, its earth-sheltering cutting heating demand by 40–60%.

Author,
Claude Design,
Midjourney



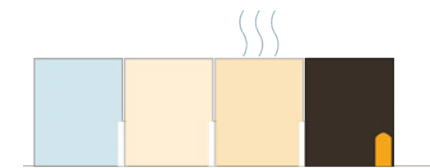
The Breathing Wall: a single thick, undulating rammed-earth or hempcrete wall as a passive climate machine, its circular oculi bridging Islamic jali screens and the Finnish reikäleipä.



Author,
Claude Design,
Nano Banana 2



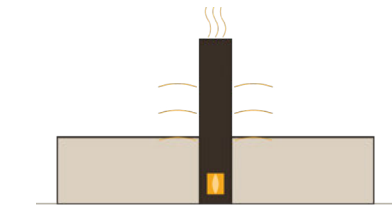
The Sauna Mosque (Löyly Masjid): the ablution sequence reimagined as a cool-to-warm progression of sauna-like timber chambers that doubles as a heat-saving air-lock.



Author,
Claude Design,
Nano Banana 2



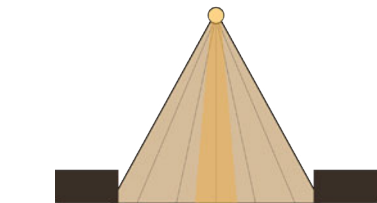
The Hearth at the Heart: a great Finnish-soapstone hearth at the literal center of the complex, storing summer heat for winter release and firing at Eid and iftar.



Author,
Claude Design,
Midjourney



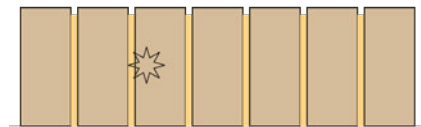
The Kota and the Qubba: the Sámi kota with its smoke-hole and the Islamic dome with its oculus treated as a single archetype in a conical glulam structure with a glazed apex.



Author,
Claude Design,
Midjourney



The Illuminated Manuscript: the building as a manuscript of seven CLT "pages" separated by glass slits, revealing its interior page by page on approach.

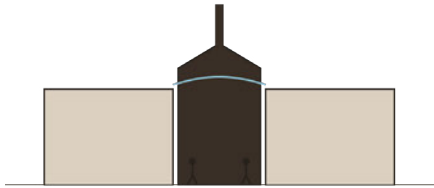


Author,
Claude Design,
Midjourney



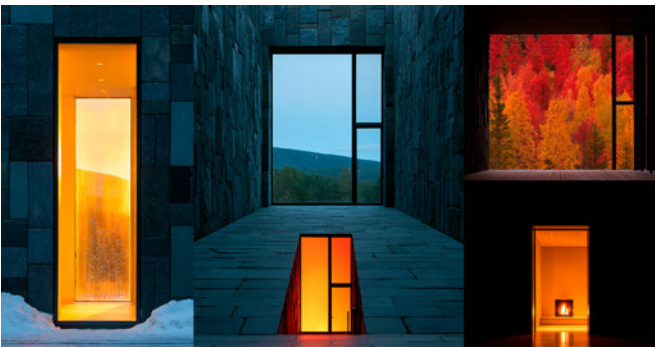
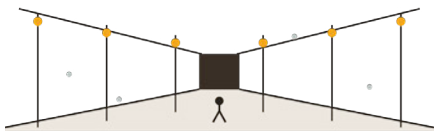
The Civic Threshold: a public, weatherproofed Finnish "street" running through the complex, allowing daily public use without entering the mosque proper.

Author,
Claude Design,
Nano Banana 2



The Long Approach: an 80–120 m covered, slowly warming walkway that turns the arduous Nordic winter arrival into a meditative crossing into prayer.

Author,
Claude Design,
Nano Banana 2



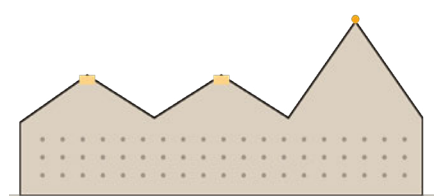
The Mosque of Many Seasons: a system of rooms each tuned to a different climatic condition, with worship migrating between them across the Finnish year.

Author,
Claude Design,
Midjourney



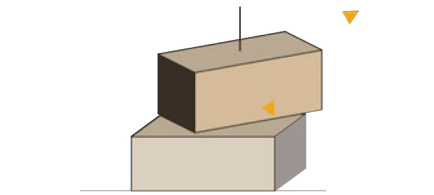
The Folded Veil: one continuous folded timber surface drawn from both mashrabiya and origami, sheathed in a patinating bronze veil that ages toward Turku Cathedral's green roofs.

Author,
Claude Design,
Midjourney

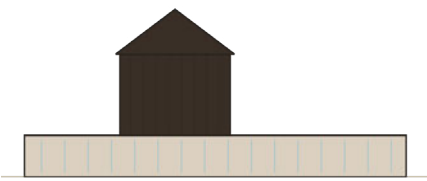


Cardinal Rotation: a street-aligned stone plinth supporting an upper volume rotated about 135° to face Mecca, so the building visibly turns toward something.

Author,
Claude Design,
Midjourney



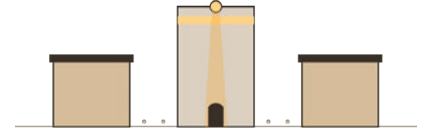
The Anchor and the Plinth: a near-cubic charred-timber prayer hall set as figure on a long, low stone-and-glass plinth holding the civic program.



Author,
Claude Design,
Nano Banana 2



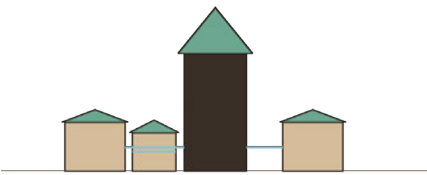
The Cloister: a granite prayer-hall cube ringed by a continuous low timber wrapper around an enclosed, roofed Nordic cloister courtyard.



Author,
Claude Design,
Nano Banana 2



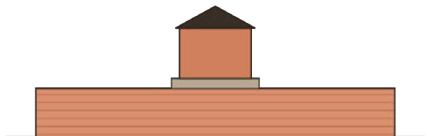
The Mother and Her Children: a tall dark-timber prayer hall with three low pavilions clustered around it like a Finnish farmstead.



Author,
Claude Design,
Midjourney



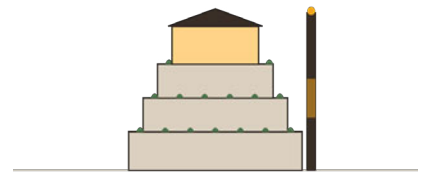
The Shepherd: a compact brick prayer hall emerging from the center of a long single-storey wing, dominating by position rather than size, its brick conversing with Turku Cathedral.



Author,
Claude Design,
Midjourney



The Stacked Garden Minaret: a vertical scheme with terraced green roofs spiraling upward and the prayer hall lifted into the tree canopy.



Author,
Claude Design,
Midjourney



4.3.3. Selected Concept

An architectural project is far more complicated than any single concept, and EMBER was no exception: its design was driven by several ideas rather than one, each carrying a different weight. The twenty-four concepts of the catalogue are best understood in this light, less as a menu from which one option is chosen than as a tailor-made mood board, specific to this brief and this site, against which ideas could be tested and combined. The direction that emerged is a cluster of a few mutually reinforcing concepts, dominant in some cases and faint in others, that together set the character of the scheme. The design itself is resolved in Section 4.4; what follows traces which concepts drove it, and how strongly.

Five concepts shaped the project most directly. The Continuing Loghouse is its substrate: charred shou-sugi-ban timber and cross-laminated and glulam construction at civic scale, exposed timber throughout, placing the building squarely in the Finnish Tatar wooden-mosque lineage. Every exterior and the prayer-hall interior rest on it. The Mother and Her Children supplies the governing massing logic: one dominant, tall, dark-timber prayer hall and its tower preside over several lower subordinate pavilions and wings gathered around it like a farmstead. This reads as a primary mosque mass with complementary subordinate masses, and follows the brief's own language almost word for word. The Folded Veil does much of the visible work: the faceted standing-seam roofs fold and pitch, and the prayer-hall ceiling resolves into a single continuous folded timber surface that reads directly out of the meeting of origami and mashrabiya, among the most photogenic moves in the scheme. The Winter Lantern survives in transposed form. Rather than the prayer hall becoming a translucent, glowing shell, it is the minaret that glows amber through a girih-perforated screen across the dark setting, while the hall itself stays dark timber and glazing; the lantern idea is realized, relocated to the campanile. The Boreal Clearing, finally, governs the building's relation to its surroundings: the scheme sits consistently as a clearing carved into boreal pine and birch, and whether or not the forest is project-planted, the figure-against-forest relationship is unmistakable.

A second tier of concepts is present as undertone rather than driver. The Anchor and the Plinth registers in the way the figural prayer hall sits above lower civic wings that act as a base, though it reads as a pitched figure rather than a pure cube. The Cloister can be felt where a U-shaped wing wraps an enclosed courtyard, here left open rather than roofed. Cardinal Rotation appears in the main hall's visible rotation against the lower volumes and the site lines, consistent with turning toward the qibla. The Sauna Mosque lingers in the ablution hall, whose dark timber, top light, and steam-like atmosphere carry a genuinely sauna-like progression. The Archipelago and the Freeze-Thaw Garden are atmospheric at most: reflecting pools and discrete pavilions are present, but the pavilions are not islands in the water, and no clear chahar-bagh rill structure survives. The Kota and the Qubba is a loose gesture, glimpsed in the converging faceted ceiling and the open-apex tower. The Civic Threshold appears in the commercial spine and public plaza, which carry the semi-public street idea.

Several concepts were set aside, and one was actively contradicted. The Forest Mihrab is ruled out by the resolved interior: worshippers face a solid carved marble mihrab, not a glazed wall onto living trees. Absent altogether are the Rock Mosque (the building sits above grade rather than excavated), the Breathing Wall (no rammed earth or hempcrete), the Hearth at the Heart, the Hypostyle Grove (a folded shell rather than a column forest), the Shepherd (timber rather than brick, and a tall rather than a low hall), the Stacked Garden Minaret, the Long Approach, the Illuminated Manuscript, the Mosque of Many Seasons, and the Qibla Light (no calibrated horizontal solar slot is evident, and the mihrab is lit ambiently rather than by a travelling band of gold).

Stated in a single line, the project's conceptual DNA is a Continuing Loghouse organized as a Mother and Her Children, wrapped in a Folded Veil, with the minaret as a Winter Lantern, set in a Boreal Clearing. The five reinforce one another rather than fight, which is what makes the cluster legible: a coherent direction assembled from the mood board the catalogue provided.

4.3.4. Spatial Development

With a general direction set, the work shifted from generating ideas to visualizing them, and from a few images to a coherent identity across the whole program. A structured prompt library was built to cover every space, organized as an interactive reference by program category (exterior, prayer halls, ablution, classrooms and library, dining and kitchen, kindergarten and play, youth center, commercial units, and outdoor and gardens), with several framed shots per space (e.g. aerial, eye-level, detail, and night views) and alternative prompt variants for different angles of the same room.

Turku Islamic Center

MOOD BOARD PROMPT LIBRARY — GOOGLE IMAGEN / WHISK

STYLE DNA — INCLUDE IN EVERY PROMPT FOR VISUAL CONSISTENCY

Contemporary Nordic Islamic architecture, dark charred timber (shou sugi ban) cladding, perforated COR-TEN steel panels with geometric Islamic lattice patterns glowing warm amber from within, Finnish pine forest backdrop, crystalline Nordic diffuse light, granite and stone paving, minimalist Scandinavian spatial logic fused with Islamic spatial tradition, architectural photography, Hufton+Crow editorial quality

Copy DNA

Exterior

Prayer Halls

Ablution

Classrooms & Library

Dining & Kitchen

Kindergarten & Play

Youth Center

Hero Aerial — Golden Hour

AERIAL / WIDE

V1V2

Aerial drone view of contemporary Islamic center complex, Turku Finland, dark charred timber volumes of varied heights, perforated minaret tower with geometric light slots, Finnish pine forest surrounding the site, stone-paved courtyard with native birch trees and autumn foliage, parking area with structured timber pergola planting, warm golden hour light raking across dark facades, warm amber glow from interior through perforated panels, high-altitude architectural drone photography, cinematic, 8K. Contemporary Nordic Islamic architecture, dark charred timber (shou sugi ban) cladding, perforated COR-TEN steel panels with geometric Islamic lattice patterns glowing warm amber from within, Finnish pine forest backdrop, crystalline Nordic diffuse light, granite and stone paving, minimalist Scandinavian spatial logic fused with Islamic spatial tradition, architectural photography, Hufton+Crow editorial quality

Copy prompt

Main Entrance Facade

GROUND LEVEL / EYE LEVEL

V1V2

Street-level view of contemporary mosque entrance facade, dark charred timber and perforated geometric metal panels, large floor-to-ceiling glazed ground floor lobby, geometric metal canopy casting intricate lattice shadow patterns on granite paving, Finnish birch trees at left periphery, late afternoon sun, Nordic blue sky with thin clouds, people approaching entrance from left, architectural photography. Contemporary Nordic Islamic architecture, dark charred timber (shou sugi ban) cladding, perforated COR-TEN steel panels with geometric Islamic lattice patterns glowing warm amber from within, Finnish pine forest backdrop, crystalline Nordic diffuse light, granite and stone paving, minimalist Scandinavian spatial logic fused with Islamic spatial tradition, architectural photography, Hufton+Crow editorial quality

Copy prompt

Author, Claude Opus 4.8

136

LATENT ARCHITECTURE

APPLIED SYNTHESIS: EMBER

137

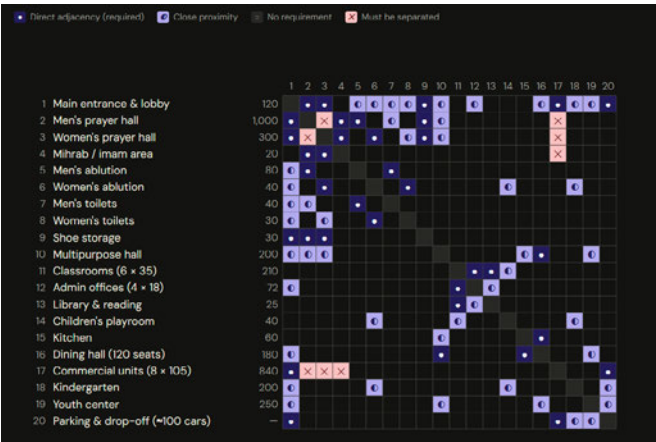
4.4. Iterative Process

The concept phase produced twenty-four design directions; schematic design reduced them to a single resolved proposal, with a coherent plan, massing, and atmosphere. This proved the most demanding phase of the project. A diffusion model generates images that are visually convincing but not dimensionally reliable, whereas a plan must hold accurate measurements and remain constructible, and the gap between the two is not readily closed. The design was therefore developed iteratively rather than linearly: each version was generated, assessed, and revised before being regenerated, through many successive cycles. This procedure is consistent with the Research through Design methodology set out in Chapter 2, in which knowledge is held to reside in the act of iteration rather than in any individual output. The final visualizations are accordingly understood as the most recent iteration within that cycle, not as an endpoint standing apart from it.

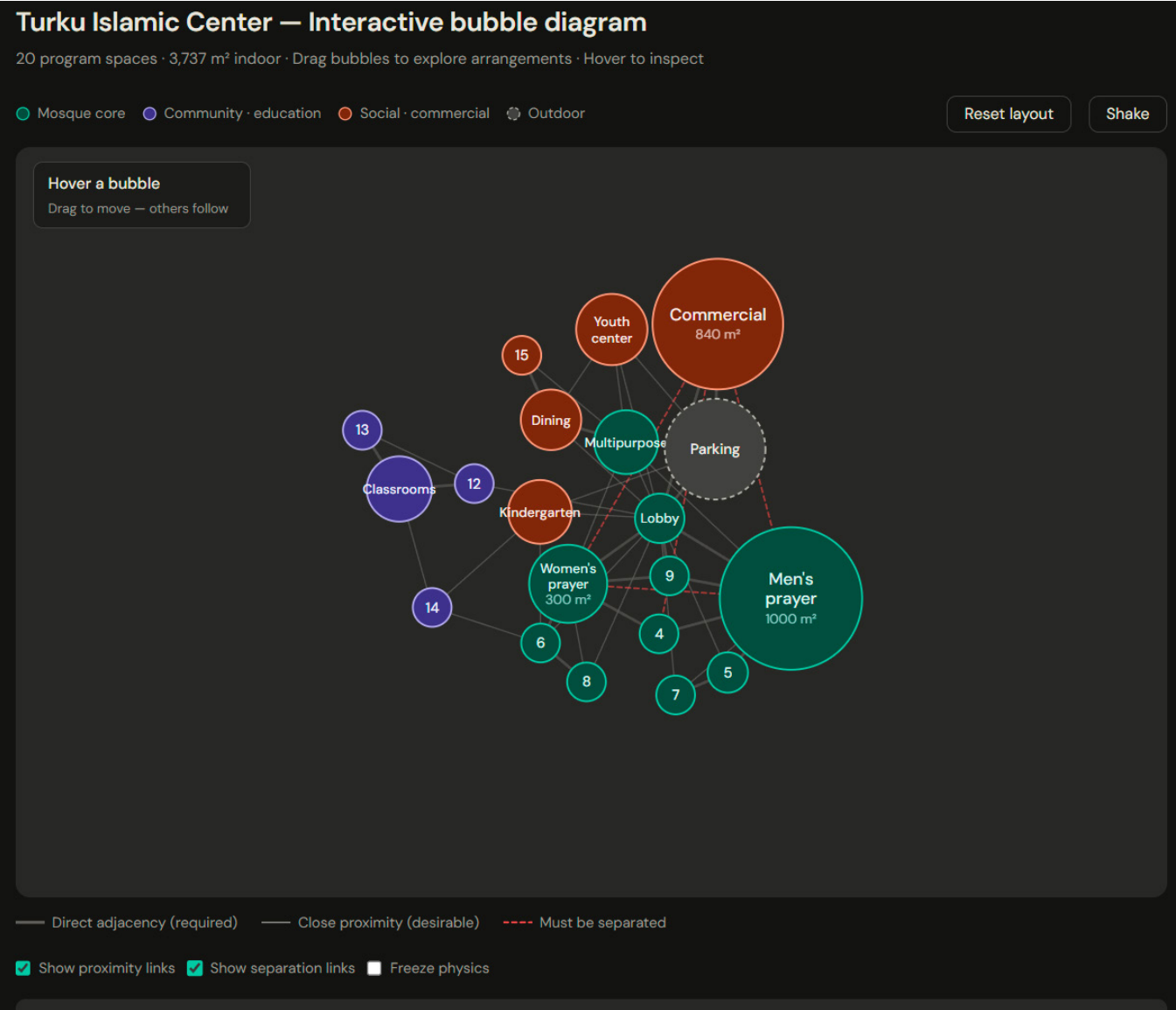
Across these cycles, a single variable proved decisive: the quantity and precision of the information supplied to the model. This may be represented as a continuum of constraint marked by three reference points. The first is a text prompt alone. The second is a text prompt accompanied by an annotated sketch. The third is a text prompt accompanied by a screenshot of a dimensionally accurate massing model. At each point the input may be specified with greater or lesser precision. These three positions correspond to the interaction modes defined in Section 1.3.3 (Prompting, Conditioning, and Curating), and the subsections that follow address them in turn. This reference-image conditioning is one of the two routes to geometric locking set out in Section 1.3.2; the other, adapter-based control such as ControlNet, was set aside because the models that performed best in Chapter 3 are closed and do not expose it, whereas reference conditioning works with open and closed models alike.

4.4.1. Spatial Logic

The process began before any image was generated, with the organization of the program itself. A large language model was used to write the code for two linked instruments: a relationship matrix and an interactive bubble diagram. The bubble diagram inventoried every space in the brief and estimated provisional areas for those the brief left unquantified, while the relationship matrix encoded which spaces should be adjacent and which kept apart, according to both functional and ritual requirements. Combined into a force-directed diagram, each space then attracted or repelled the others in proportion to its prescribed proximity, so that a coherent spatial structure emerged from the adjacency rules rather than from manual composition. This is the application of large language models to "Logic Synthesis" anticipated in Section 1.3.2: here the model authors the relational logic underlying the plan rather than generating visual atmosphere.



Author, Claude Opus 4.8



Author, Claude Opus 4.8

4.4.2. Iterative Refinement

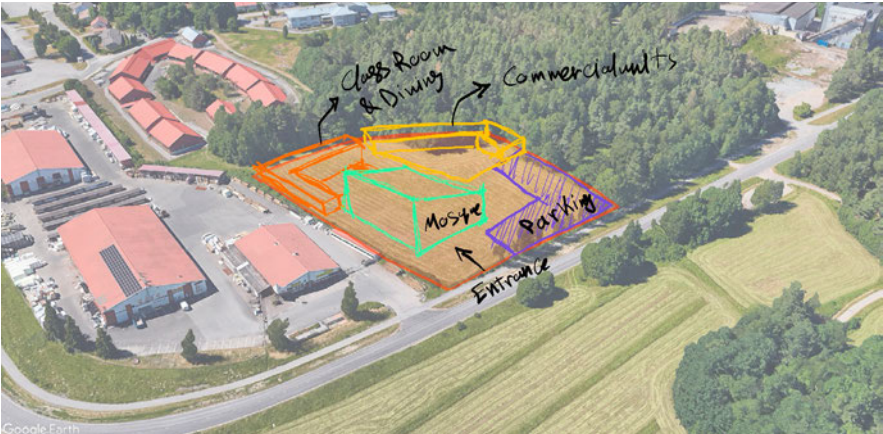
By this point, the general architectural direction had been set in the concept phase, and the approximate position of every space had been fixed by the spatial logic, yet no concrete floor plan existed. The work therefore proceeded by supplying the model with progressively more precise information as the vision became clearer, in the three stages set out below. This division into three is not fixed; the number of stages could be greater or smaller, depending on a range of variables, the most decisive of which is the point at which a satisfactory result is reached.

The principle that governs the stage is straightforward. As the input becomes more precise, the output grows steadily more faithful to the design and steadily less surprising. This is the image-level expression of the denoising behavior described in Section 2.1.3 and of the distinction between loose and tight agency in Section 1.3.2: low constraint purchases breadth at the cost of control, while high constraint purchases fidelity at the cost of serendipity. The constraint is best understood as a dial the designer turns deliberately. Looseness was in part a consequence of an unclear vision, but it was also, at times, a deliberate choice: information was withheld either to avoid burdening the model with detail irrelevant to the question at hand, or to preserve the creative latitude from which the early, productive deviations had come.

At every step of this stage, many images were generated and then filtered down to the most usable, and that selection was carried forward to the next step; the figures that follow reproduce both the range of iterations and the image chosen at each stage.

Stage I

The first stage used the loosest usable input. Rough masses were sketched directly onto an aerial screenshot of the site taken from satellite imagery, labeled by function, and paired with one of the concept images as an architectural style reference. Because the sketch was so approximate, the generated aerials varied widely. This breadth was useful rather than a defect: among the variants were unplanned features, strictly hallucinations, that proved attractive enough to be carried into the final design.



Input 1: Sketch
Author, Google Earth



Input 2: Style Reference
Author, Midjourney

Generate a photorealistic aerial render of a contemporary Islamic Center complex in Turku, Finland, using the attached sketch as a precise guide for building placement, massing, and zoning. The viewpoint should match the sketch exactly — a southwest aerial perspective looking northeast, at roughly a 45-degree angle. Site context: The site is surrounded by industrial/warehouse buildings to the west (with red/brown roofs), dense coniferous and birch forest to the north and east, a two-lane road running along the southeast edge, and open green fields to the south. Replicate the real surroundings visible in the sketch exactly as they appear. Architectural style and materiality: All buildings should be clad in dark-stained Finnish timber with vertical battens/slats, matching the reference image provided — a deep charcoal-brown wood cladding with narrow vertical reveals that create a rich textural rhythm. The wood should appear natural and warm despite its dark tone. All roofs should be slanted — the mosque and educational wing should feature opposing pitched roofs that slope inward, creating dramatic angular silhouettes similar to the reference image. The commercial units can have a simpler single-pitch shed roof sloping toward the rear (north). Building masses (follow sketch placement precisely): Mosque (green outline in sketch): The central and most prominent volume, roughly two stories in height with a tall, steeply pitched roof. Positioned in the middle-left of the site. Large glazed openings at ground level on the entrance-facing (south) side, with warm interior light faintly visible. The roof ridge line should run roughly east-west. Minaret/Tower: A tall, slender, freestanding tower positioned at the main entrance, adjacent to the mosque's south facade. It should be the tallest element on site — roughly 25–30 meters — clad in the same dark vertical timber slats with narrow slit openings that allow light to filter through, exactly as in the reference image. It should be square or slightly rectangular in plan and taper subtly toward the top. Educational wing (orange outline in sketch): A two-story L-shaped or rectangular volume to the west and northwest of the mosque, housing classrooms, kindergarten, dining hall, and offices. It should have opposing slanted roofs echoing the mosque's language but at a slightly lower height. Punctuated by regularly spaced rectangular windows with warm-toned interior glow. Commercial units (yellow outline in sketch): A linear block to the northeast of the mosque, with a public-facing plaza/courtyard between it and the mosque. Single or 1.5 story height with a simple mono-pitch roof. Ground-level shopfronts with generous glazing facing the plaza. Parking (purple hatched area in sketch): An open surface parking area to the southeast of the buildings, between the complex and the road. Approximately 100 cars, arranged in orderly rows with a few trees interspersed. A clear vehicular drop-off zone near the entrance. Entrance: Located on the south side of the complex, facing the road, with the minaret marking it as a vertical landmark. A landscaped pedestrian approach leads from the parking area to the entrance plaza. Landscaping: Birch trees and native Nordic vegetation scattered throughout the site, especially between the buildings and along pedestrian paths. A landscaped courtyard/plaza between the mosque and commercial block. A small playground visible near the educational wing. Pedestrian pathways connecting all areas with low ground-cover planting. Atmosphere and lighting: Bright, sunny Nordic summer day with long, soft shadows. Clear blue sky with a few high, thin clouds. Lush green grass on the surrounding fields. Trees fully leafed out. The dark timber facades should contrast beautifully against the green landscape and bright sky. The overall feeling should be calm, welcoming, and contemporary — a building that belongs to its Finnish landscape while carrying clear Islamic architectural identity through the minaret and spatial organization. Quality: Ultra-high resolution, photorealistic architectural visualization quality, comparable to a professional drone photograph. Sharp details on the timber cladding, accurate shadows, realistic vegetation, and natural color grading.

Input 3: Text Prompt
Author, Claude

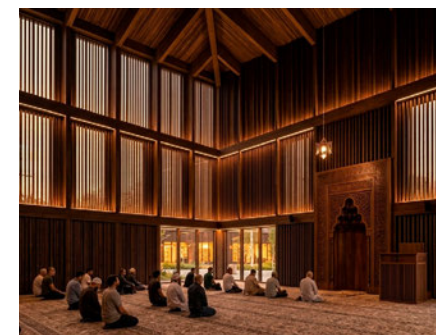
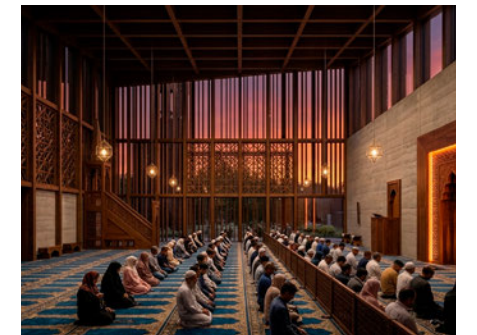


Generated Iterations
Author, Nano Banana, GPT Image 1.5

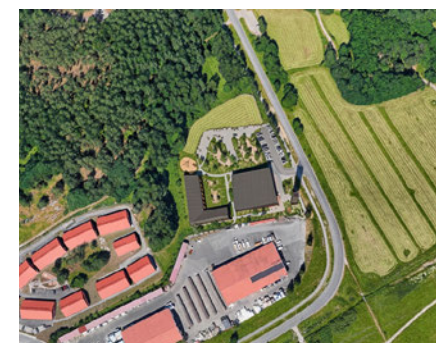
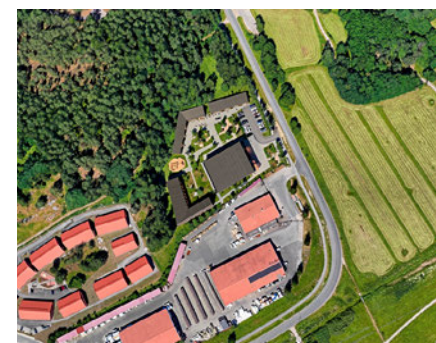


Final Result
Author, Nano Banana Pro

At this stage a further attempt was also made, anticipating the method later adopted in Section 4.4.3: one of the resulting aerials was taken as a base image and returned to the model to generate additional views of individual spaces, and to produce a top-view layout of the project. Both attempts were partially successful, but the results lacked consistency from one view to the next and showed little geometric locking, because the loose base image carried too little stable structure to hold the design steady across generations.



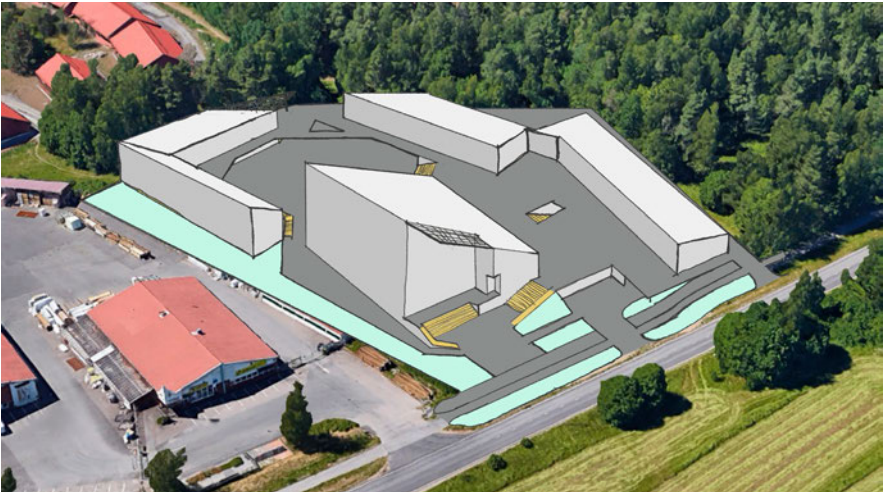
Spaces Iterations
Author, Nano Banana Pro



Layout Iterations
Author, Nano Banana pro, Flux 2 Max

Stage II

As the intended scheme became clearer, the input was tightened. A rough master plan was drawn to fix the location of each space, and a far more detailed massing sketch was laid over the same aerial screenshot. Generated against this clearer sketch and the same style reference, the aerials were markedly more consistent with the intended design, precisely because the input now carried more information. The iteration process also expanded at this stage: a custom ComfyUI workflow was built to run each generation through six different diffusion models simultaneously, producing a wider pool of candidates to filter at each step.



Input 1: Sketch
Author, Google Earth



Input 2: Style Reference
Author, Midjourney

Input 3: Text Prompt
Author, Claude

High-oblique aerial drone photograph of a contemporary Nordic Islamic center on an irregular pentagonal plot at the forested northeastern edge of Turku, Finland, clear summer late afternoon, low raking northern sun. READING THE REFERENCES: The attached plan and the 3D massing sketch are the SAME building from different views — match them to each other. On the plan, the RED edge marks the STREET side: that is the public road frontage. Orient the view so the camera looks across this street frontage toward the building and the forest behind it. The main mosque mass sits toward the street/red edge, the courtyard and the angled classroom/dining wings sit deeper into the plot away from the street. MASSING (follow the reference geometry exactly): a dominant central freestanding mosque volume — a tall single-storey prayer hall with a steep folded dark roof, a clerestory ridge, and a slender vertical minaret light-tower at its street-facing entrance. The tower face inset with a perforated COR-TEN girih panel glowing warm amber. This main mass faces the red/street edge. Framing it are lower two-storey subordinate bars: a long repetitive bar of eight commercial rental units along the rear edge; an angled, kinking bar of classrooms, dining hall and offices following the western boundary; a lower classroom wing stepping along the southwest. The volumes wrap a central landscaped courtyard (sahn) laid out as a quartered geometric chahar-bagh garden with stone paths, planting beds, native birches, heather and a small water element. PARKING (critical — do NOT add a surface car park): parking is fully UNDERGROUND. Show only a single vehicle ramp descending into underground parking on the right-hand (street) side of the plot, near the road. The only at-grade parking is a small handful of accessible / disabled spaces and a brief drop-off directly in front of the main entrance forecourt. Everywhere else at ground level is paving, garden, or forest — no rows of parked cars, no open asphalt lot. MATERIAL & STYLE: all volumes clad in dark charred timber (shou sugi ban / yakisugi) vertical boards; angular folded mono-pitch dark standing-seam roofs; floor-to-ceiling glazed ground floors glowing warm amber from within; perforated COR-TEN steel girih panels on the minaret and key facades casting arabesque light; granite and stone paving; native Finnish landscaping of birch, pine, ferns and heather. SITE CONTEXT (true to location): the public road runs along the red/street edge; light-industrial warehouses with red gable roofs sit to the west; dense boreal pine and birch forest wraps the north and east; open hay-stripped agricultural fields lie across the road. ATMOSPHERE: crystalline Nordic light, deep green forest, pale blue sky with thin high cloud, crisp long shadows, a few people on the plaza for scale. Photorealistic architectural drone photography, HufTon+Crow editorial quality, ultra-detailed, sharp, 8K.



Unusable Iterations
Author, Nano Banana, Nano Banana Pro, Nano Banana 2, GPT Image 2, Qwen Edit 2511, Flux 2 Max



Interesting Hallucinations

Author, Nano Banana, Nano Banana Pro, Nano Banana 2, GPT Image 2, Qwen Edit 2511, Flux 2 Max



Usable Iterations

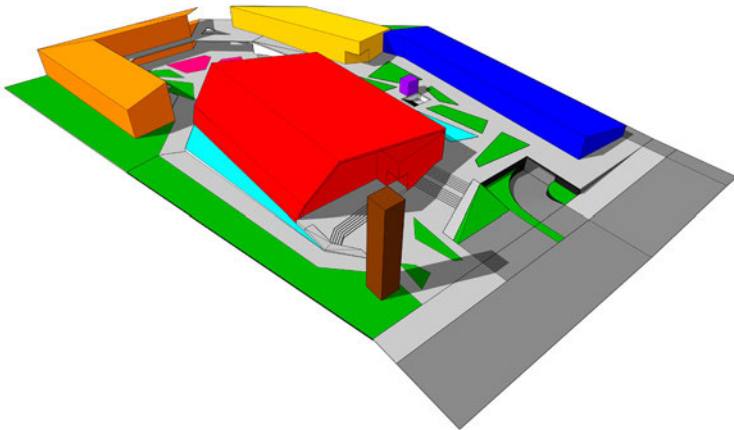
Author, Nano Banana, Nano Banana Pro, Nano Banana 2, GPT Image 2, Qwen Edit 2511, Flux 2 Max



Most Usable Iteration
Author, Nano Banana Pro

Stage III

At the final and most constrained stage, an actual floor plan was drafted in AutoCAD and a basic colored massing model built from it in SketchUp. A screenshot taken from roughly the same vantage as the original satellite view was then supplied to the model in place of the sketch, again with the style reference. The result was usable and approximately 80 percent faithful to the drawn design. A round of regional edits, enhancement, and upscaling, carried out with further AI tools, brought the final aerial to roughly 99 percent agreement with the floor plan and the intended scheme.



Input 1: Massing Model Screenshot
Author



Input 2: Site Aerial View
Author, Google Earth



Input 3: Style Reference
Author, Midjourney

Photorealistic aerial drone render of a contemporary Nordic Islamic center, generated by applying real materials and a real site to the provided 3D massing model. IMAGE 1 (color-coded 3D massing model) defines the EXACT geometry, volumes, positions, heights and camera angle. Reproduce this massing and viewpoint faithfully. The colors are a LABELING LEGEND ONLY — they are NOT real materials. Translate each color as follows: - RED volume = the main mosque prayer hall (tallest mass, steep folded roof). - BROWN tall slab = the minaret light-tower. Dark charred timber with a vertical perforated geometric girih panel glowing warm amber. - BLUE long volume = single-storey commercial rental units. Dark charred timber with large glazed shopfronts glowing warm. - YELLOW volume = single-storey restaurant. Dark charred timber with glazed walls, warm interior light. - ORANGE L-shaped volume = two-storey classroom wing. Dark charred timber, regular windows, warm interior glow. - GREEN areas = landscaped grass with native Finnish birch and pine trees, ferns and heather. - VIOLET area in the landscape = a children's playground with natural timber play equipment on a soft safety surface. - AQUA shapes = water fountains / reflecting pools. - PURPLE small cube = a glazed outdoor elevator. - LIGHT GREY = stone and granite plaza paving. - DARK GREY = asphalt road and ramp. - BLACK = walls / retaining edges in dark material. Do not leave any surface red, blue, yellow, orange, violet, purple or aqua-as-paint — every colored mass becomes dark timber architecture, landscape, or warm glazing. The mosque prayer hall remains the tallest and most dominant volume. IMAGE 2 (Google Earth site photo) is the REAL CONTEXT. Place the complex onto the plot outlined in red/orange and keep the surroundings exactly: red-roofed industrial warehouses and lumber yard to the left, the two-lane road along the front, dense green pine-and-birch forest wrapping behind and right, open hay fields across the road. Match this photo's bright summer daylight, sun direction and green season. IMAGE 3 (reference render) is the MATERIAL AND DETAIL reference ONLY — copy its architectural language (charred timber boards, folded dark roofs, perforated girih tower with amber glow, warm glazed ground floors, birch trees) but DO NOT copy its winter, snow, or grey overcast sky. PARKING: underground only — one ramp (the dark grey ramp in the model) descends to underground parking near the road; only a few accessible spaces and a drop-off at the entrance. No surface parking lot. WEATHER & MOOD: clear bright Finnish summer afternoon, blue sky, lush green forest, warm low sunlight, crisp shadows, a few people on the plaza for scale. Ultra-detailed, sharp, photorealistic architectural drone photography, 8K.

Input 4: Text Prompt
Author, Claude



Unusable Iterations
Author, Nano Banana, Nano Banana Pro, Nano Banana 2, GPT Image 2, Qwen Edit 2511, Flux 2 Max



Usable Iterations
Author, Nano Banana Pro, GPT Image 2



Most usable Iteration
Author, Nano Banana Pro



Region Edit (trees & Vegetation)
Author, Nano Banana Pro



Style Transfer
Author, Gendo AI



Style Transfer
Author, Gendo AI



Region Edit (curtain wall)
Author, Nano Banana Pro



Style Transfer
Author, Gendo AI



Style Transfer
Author, Gendo AI



Style Transfer
Author, Gendo AI

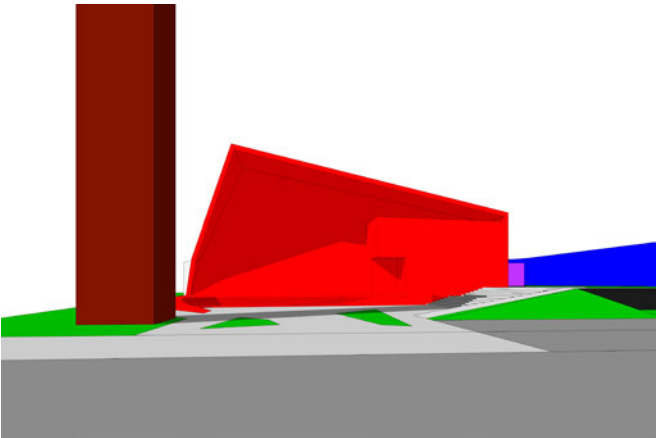


Final Result (upscale & Enhance)
Author: Topaz Generative, Magnific

4.4.3. Extending the View

With an accurate aerial view resolved, attention turned to the other views needed to communicate the design, and the reference strategy changed accordingly. In place of the concept image used earlier, the finished aerial now served as the principal reference, supplied together with further screenshots taken from other angles of the SketchUp massing model. Conditioned in this way, the model returned exterior perspectives and interior atmospheres consistent with the resolved scheme, since each new image was generated from the project's own geometry and its own established appearance. [Figures: exterior and interior views.] Every result still required refinement: a loop of regional editing, enhancement, and upscaling brought each view to a presentable standard. As before, each finished image was returned to the model alongside the earlier images and the relevant screenshot, so that the growing set reinforced the coherence of those still to come.

Image 01



Input 1: Massing Model
Author



Input 2: Architecture Reference
Author

Generate a hyperrealistic, eye-level architectural photograph of a contemporary Nordic Islamic center in Turku, Finland, by applying real materials and lighting to the provided color-coded 3D massing model. GEOMETRY AND CAMERA LOCK: The color-coded 3D massing screenshot defines the EXACT geometry, volumes, proportions, positions, silhouettes and camera angle. Reproduce this massing and this eye-level pedestrian viewpoint faithfully — do not move, rotate, add, remove or resize any volume, and do not change the camera angle, lens or composition under any circumstances. The colors in that screenshot are a LABELING LEGEND ONLY; they are NOT real materials. Translate each color into real architecture exactly where it sits in the frame — BROWN tall vertical slab (left foreground, the tallest element, ~28 m) = the minaret light-tower. Dark charred timber (shou sugi ban) with a full-height perforated COR-TEN steel girth lattice panel glowing warm amber from within. — RED angular mass (center, the primary and most dominant building, ~14 m at its peak, with a steep folded roofline dropping toward the front-left) = the main mosque prayer hall. Dark charred timber cladding, dark folded standing-seam metal roof, tall floor-to-ceiling warm-amber glazing, and a low entrance volume at its base. — PURPLE small cube (right of the mosque, ~4 m) = a glazed outdoor elevator. Slender steel-framed full-height glass, warm interior glow. — BLUE long low volume (right side, single-storey, ~5–6 m, sitting on a slightly raised granite terrace reached by the broad stone steps visible in front of it) = the single-storey commercial rental units. Dark charred timber piers with floor-to-ceiling glazed shopfronts glowing warm amber and individual geometric metal entry canopies. — GREEN wedges on the ground = landscaped planting beds and grass with native Finnish birch, pine, ferns and heather. — LIGHT GREY surfaces = granite and stone plaza paving, the stepped grey element on the right = granite steps rising from the plaza up to the commercial terrace. Render every colored mass as dark timber architecture, landscape or warm glazing — leave no surface red, brown, blue, purple or green as flat paint. MASSING HIERARCHY: the mosque is the primary datum and the minaret is the tallest vertical accent; every other volume is clearly subordinate and lower. Keep the broad open granite plaza in the foreground exactly as framed. SAME-BUILDING MATERIAL REFERENCE: the photorealistic aerial render provided is NOT a generic style image — it is a finished aerial render of THIS EXACT SAME building and site, seen from above. Treat it as the source of truth for the real appearance of every element: the dark charred timber cladding and its tone, the dark folded standing-seam roofs, the perforated COR-TEN minaret screen and its amber glow, the granite paving, the warm glazing, and the surrounding boreal pine and birch forest. The minaret, mosque, commercial units and landscape you render at eye level must read as the very same structures shown in that aerial — identical materials, finishes and colors — simply viewed from ground level instead of from the air. Use the aerial ONLY for material and finish truth, NOT for camera angle, weather or composition. ATMOSPHERE (independent of the reference): clear, bright Finnish summer late afternoon, warm low golden sunlight raking across the dark facades, crisp shadows on the granite, blue sky with thin high cloud, deep green pine forest behind the rooflines, a few people on the plaza for scale. Do not carry over fog, mist, snow, winter or overcast conditions from any reference image. Do not add a dome, classical arches, white marble, golden ornament, surface car parking, extra towers, or any building beyond the plot. Keep the geometry and the eye-level camera identical to the massing screenshot. Ultra-detailed, sharp, photorealistic architectural photography, Hufon+Crow editorial quality, 8K.

Input 3: Text Prompt
Author, Claude

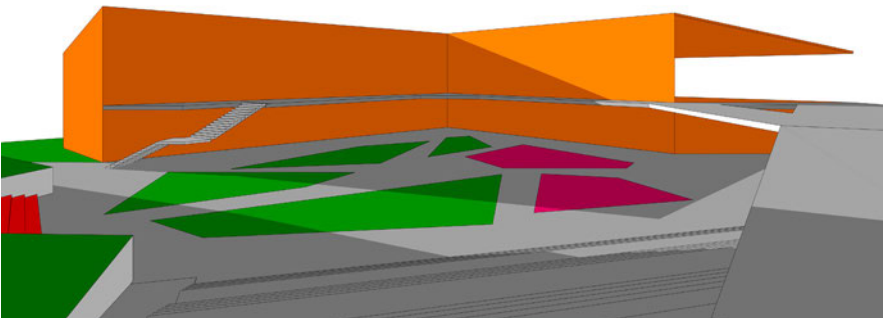


Output: Iterations
Author, Nano Banana 2, GPT Image 2, Gendo AI



Final Result
Author: Gendo AI

Image 02



Input 1: Massing Model
Author



Input 2: Architecture Reference 1
Author



Input 3: Architecture Reference 2
Author

Convert the supplied color-coded 3D massing screenshot into a photorealistic architectural rendering. Hold the exact composition, camera angle, volume positions, and proportions of the color-coded screenshot — do not move, rotate, re-frame, or re-interpret any geometry. The viewpoint is a low, slightly elevated eye-level looking across a sunken landscaped courtyard toward a long building behind it; keep this viewpoint locked under all circumstances. Image roles: (1) the color-coded 3D screenshot is the GEOMETRY AND COMPOSITION SOURCE — every mass, plane, level change, and the camera are defined entirely by it; (2) the warm-lit entrance photograph with the tall perforated glowing tower and dark timber gabled volume is a MATERIAL AND ATMOSPHERE REFERENCE ONLY — copy its charred dark timber cladding, warm amber interior glow, glazing detailing, granite paving, Finnish pine/birch planting and golden-hour light from it, but take none of its geometry, and do not import its tower, gable, or canopy into this frame; (3) the misty aerial photograph of the dark timber complex is also a MATERIAL AND ATMOSPHERE REFERENCE ONLY — match its dark shou sugi ban timber tone, dark folded standing-seam metal roofs and Nordic light quality, but take none of its geometry. Translate the color zones as follows. The orange / dark-orange mass is a single building; a two-storey classroom and library block, approximately 8–9 m high, clad in dark charred timber (shou sugi ban / yakisugi) vertical boarding with a dark folded standing-seam metal roof, and a highly glazed ground and upper floor — generous floor-to-ceiling windows and glass classroom doors set into the dark timber, warm amber light glowing from the interiors. This block is the dominant mass in this particular view and reads as the backdrop datum behind the courtyard. The grey planes are hardscape: granite / pale concrete paving for the courtyard floor, the descending stairs at the bottom-foreground that step down into the sunken courtyard, and — on the far right — a gently rising accessible ramp that climbs to a raised platform at the building's upper/first-floor level; render slim dark metal handrails along both the ramp and the full edge of the raised platform. The green patches are landscaping: mown grass beds, low native shrubs, and several mature Finnish birch and pine trees with full summer foliage. The magenta / violet zones are a children's playground surfaced in impact-absorbing pale terracotta-red rubber flooring, with low timber play elements. The small red folded forms at lower-left are low courtyard seating clusters in charred timber and granite, roughly bench height. Atmosphere: soft golden-hour Finnish summer light, warm low raking sun, clear crystalline Nordic sky, long gentle shadows across the granite paving, warm amber glow spilling from the glazed classroom interiors — matching the warm evening mood of the reference renders. Render in high-end architectural photography quality, Hufton+Crow editorial realism, sharp focus, natural materials, tactile texture. Do not add a minaret, dome, tower, prayer hall, or any mosque element to this frame. Do not add snow, winter conditions, fog interior to the courtyard, water reflecting pools, cars, or text/signage. Do not change the camera angle, the number of floors on the building, or the position of the stairs, ramp, playground, or planting beds. Keep the sunken-courtyard relationship and the level change between the foreground stairs and the raised right-hand platform clearly legible.

Input 4: Text Prompt
Author, Claude

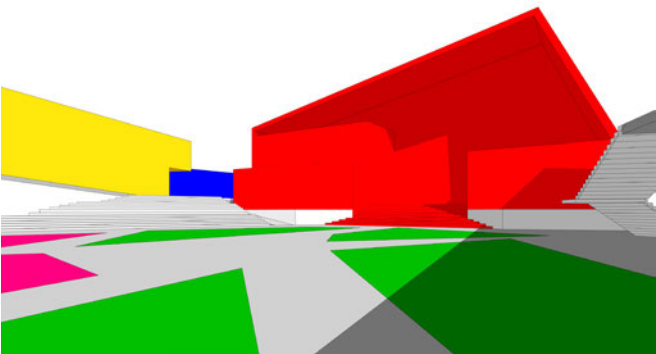


Output: Iterations
Author, Nano Banana 2, GPT Image 2, Gendo AI



Final Result
 Author: GPT-Image 2, Gensai AI, Magnific

Image 04



Input 1: Massing Model
Author



Input 2: Architecture Reference 1
Author



Input 3: Architecture Reference 2
Author



Input 4: Architecture Reference 3
Author

Convert the supplied color-coded 3D massing screenshot into a photorealistic architectural rendering. Hold the exact composition, camera angle, volume positions, proportions, and perspective of the color-coded screenshot — do not move, rotate, re-frame, or re-interpret any geometry. The viewpoint is a low eye-level positioned down inside a sunken courtyard, looking back and slightly up toward a cluster of buildings sitting on the raised level around the courtyard, with a large mass looming on the right side of the frame; keep this viewpoint locked under all circumstances. Image roles: (1) the color-coded 3D screenshot is the GEOMETRY AND COMPOSITION SOURCE — every mass, plane, roofline, level change, stair, and the camera are defined entirely by it; (2) the entrance photograph showing an angular folded dark-timber roof over a full-height glazed curtain wall with a faintly visible warm interior and people inside is a MATERIAL AND ATMOSPHERE REFERENCE ONLY — copy from it the charred dark timber cladding, the folded/angular roof treatment, the glazed curtain-wall detailing, the warm amber interior glow seen faintly through glass, and the golden-hour light, but take none of its geometry and do not import its tower into this frame; (3) the courtyard photograph with descending granite stairs, landscaped beds, a rubber-floored playground and a long dark-timber building behind is a MATERIAL AND ATMOSPHERE REFERENCE ONLY — match its granite/concrete paving, charred timber, planting, handrails, and warm evening light, but take none of its geometry; (4) the misty aerial photograph of the dark-timber complex is also a MATERIAL AND ATMOSPHERE REFERENCE ONLY — match its dark shou sugi ban timber tone and Nordic light quality, but take none of its geometry. Translate the color zones as follows. The RED mass is the MOSQUE and is the dominant, tallest primary volume in this view: a double-height prayer hall with a dramatic angular, cantilevering folded roof, roughly 10–12 m at its high point, clad in dark charred timber (shou sugi ban / yakisugi). From this angle its courtyard-facing facade reads as mostly solid, opaque charred timber, EXCEPT for one full-height floor-to-ceiling glazed curtain-wall slot, through which the prayer hall interior is faintly visible — neat rows of worshippers praying, all oriented consistently toward the qibla wall, with a soft warm interior glow. A flight of broad entrance steps in the center-foreground rises from the courtyard floor up to the mosque. The YELLOW mass is the RESTAURANT: a single-storey volume, approximately 5 m high, clad in dark charred timber with a full-height glazed curtain-wall facade set into the timber, warm interior light behind the glass. The BLUE mass is the COMMERCIAL STORES: also single-storey, approximately 5 m high, dark charred timber with a glazed curtain-wall shopfront facade, sitting between the restaurant and the mosque. The mosque must clearly remain the tallest mass; the restaurant and commercial volumes are subordinate and low. The grey planes are hardscape: granite and pale concrete paving for the sunken courtyard floor, and the broad ascending stairs on the left and right that climb out of the sunken courtyard up to the raised building platform level — render slim dark metal handrails along these stairs. The green patches are landscaping: mown grass beds, low native shrubs, and several mature Finnish birch and pine trees in full summer foliage. The magenta / pink patches at the lower-left foreground are a children's playground surfaced in impact-absorbing pale terracotta-red rubber flooring, with low timber play elements. Atmosphere: soft golden-hour Finnish summer light, warm low raking sun, clear crystalline Nordic sky, long gentle shadows across the granite paving, warm amber glow spilling from the glazed restaurant, commercial, and mosque interiors — matching the warm evening mood of the reference renders. High-end architectural photography quality, Hufton+Crow editorial realism, sharp focus, natural materials, tactile texture. Do not add a minaret, dome, or any traditional Islamic element to this frame. Do not make the restaurant or commercial volumes more than one storey. Do not glaze the rest of the mosque facade — keep it opaque charred timber apart from the single curtain-wall slot. Do not add snow, winter conditions, water reflecting pools, cars, or text/signage. Do not change the camera angle, the perspective, the number of floors on any mass, or the position of the stairs, playground, or planting. Keep the sunken-courtyard relationship and the level change between the foreground courtyard floor and the raised buildings clearly legible.

Input 5: Text Prompt
Author, Claude

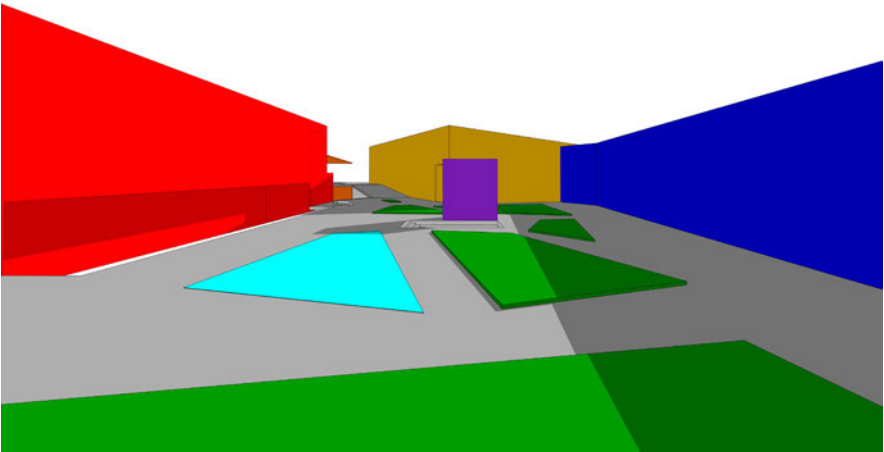


Output: Iterations
Author, Nano Banana 2, GPT Image 2, Gendo AI



Final Result
 Author: GPT-Image 2, Gendo AI, Magnific

Image 05



Input 1: Massing Model
Author

Input 2: Architecture Reference 1
Author



Input 3: Architecture Reference 2
Author



Input 4: Architecture Reference 3
Author

Input 5: Architecture Reference 4
Author

Convert the supplied color-coded 3D massing screenshot into a hyper-realistic architectural rendering. Hold the exact composition, camera angle, volume positions, proportions, and perspective of the color-coded screenshot — do not move, rotate, re-frame, or re-interpret any geometry. The viewpoint is a low eye-level standing in an open plaza, looking down its length between a tall building filling the left edge of the frame and a long building filling the right edge, with smaller volumes closing the space in the middle distance; keep this viewpoint locked under all circumstances. Image roles: (1) the color-coded 3D screenshot is the GEOMETRY AND COMPOSITION SOURCE — every mass, plane, roofline, level change, paving edge, and the camera are defined entirely by it; (2) the entrance photograph showing an angular folded dark-timber roof over a full-height glazed curtain wall with a faintly visible warm interior and people inside is a MATERIAL AND ATMOSPHERE REFERENCE ONLY — copy from it the charred dark timber cladding, the glazed curtain-wall detailing, the warm amber interior glow seen faintly through glass, and the evening light, but take none of its geometry and do not import its tower into this frame; (3) the courtyard/plaza photograph with granite paving, landscaped beds, low timber benches, mature trees and a rubber-floored playground is a MATERIAL AND ATMOSPHERE REFERENCE ONLY — match its granite/concrete paving, charred timber, planting, and warm light, but take none of its geometry; (4) the misty aerial photograph of the dark-timber complex is also a MATERIAL AND ATMOSPHERE REFERENCE ONLY — match its dark shou sugi ban timber tone and Nordic light quality, but take none of its geometry. Translate the color zones as follows. The RED mass filling the left side of the frame is the MOSQUE and is the dominant, tallest primary volume in this view, clad in dark charred timber (shou sugi ban / yakisugi). From this angle its plaza-facing facade is mostly solid charred timber but carries TWO long horizontal glazed curtain-wall bands: one at the upper/first-floor level running horizontally along the facade, and a second at the ground floor set into a RECESSED entrance zone — this entrance is both recessed back into the timber mass AND sits on a slightly lower level than the plaza floor, reached by a few steps down, with warm amber interior light glowing out from behind the glazing. The BLUE mass filling the right side of the frame is the COMMERCIAL STORES: a single-storey volume, approximately 5 m high, clad in dark charred timber with a continuous glazed curtain-wall facade set into the timber, warm interior light behind the glass, and absolutely no signage, logos, lettering, or text anywhere on it. The YELLOW / tan mass closing the middle distance is the RESTAURANT: also single-storey, approximately 5 m high, dark charred timber with a full-height glazed curtain-wall facade, warm interior glow. The mosque must clearly remain the tallest mass; the restaurant and commercial volumes are subordinate and low. The grey planes are hardscape: granite and pale concrete plaza paving, with the slightly sunken paved level at the recessed mosque entrance and the small steps down to it clearly legible. The AQUA / cyan zone in the plaza floor is a still, shallow reflecting fountain / water feature set flush into the paving, its surface mirroring the sky and surrounding timber. The green patches are landscaping: low planted beds of mown grass and native shrubs, plus a few mature Finnish birch and pine trees in full summer foliage. Render slim dark metal handrails along any level change or steps where one is needed. Atmosphere: soft golden-hour Finnish summer light, warm low raking sun, clear crystalline Nordic sky, long gentle shadows across the granite paving, warm amber glow spilling from the glazed mosque, restaurant, and commercial interiors, faint warm reflections on the fountain water — matching the warm evening mood of the reference renders. Hyper-realistic high-end architectural photography quality. Hutton+Crow editorial realism. Sharp focus, natural materials, tactile texture. Do not add a minaret, dome, or any traditional Islamic element to this frame. Do not make the restaurant or commercial volumes more than one storey. Do not glaze the rest of the mosque facade beyond the two stated curtain-wall bands — keep the remainder opaque charred timber. Do not add any signs, text, logos, or lettering on the commercial stores or anywhere else. Do not add snow, winter conditions, cars, or paving over the fountain. Do not change the camera angle, the perspective, the number of floors on any mass, or the position of the fountain, planting, steps, or recessed entrance. Keep the level change between the plaza floor and the lower recessed mosque entrance clearly legible.

Input 6: Text Prompt
Author, Claude



Output: Iterations
Author, Nano Banana 2, GPT Image 2, Gendo AI

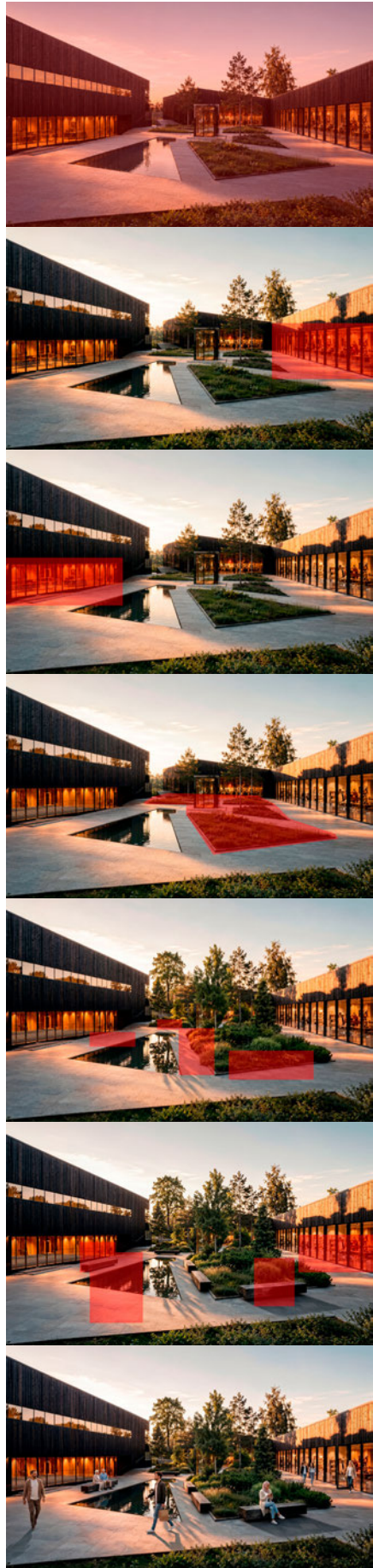
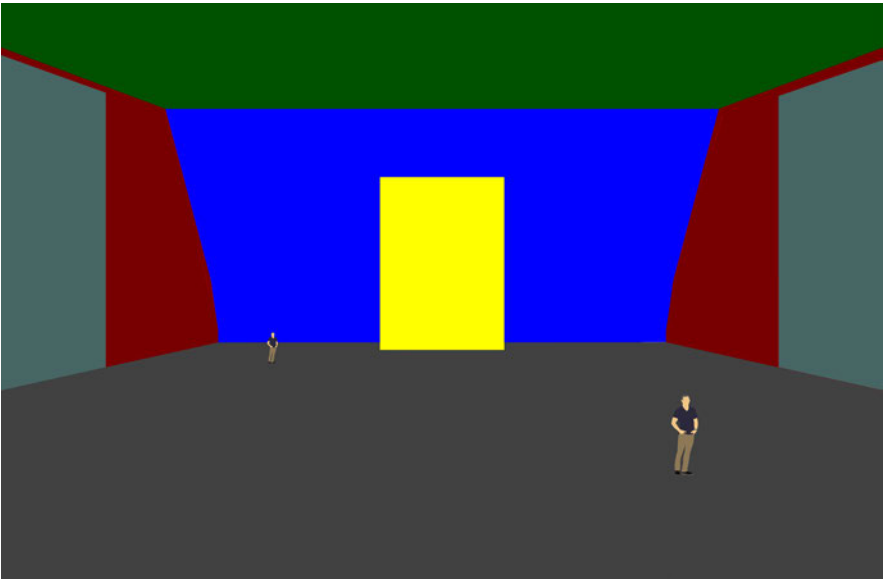


Image 06



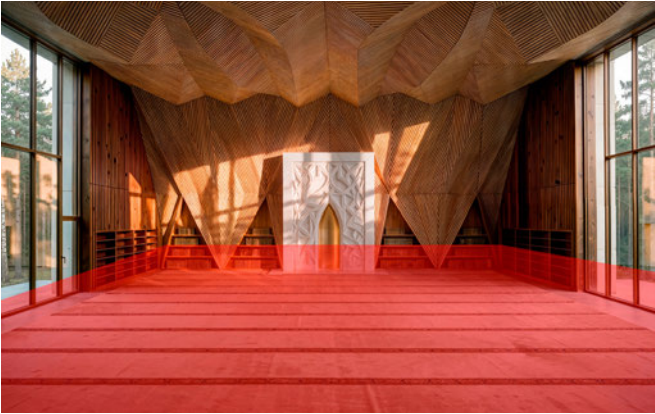
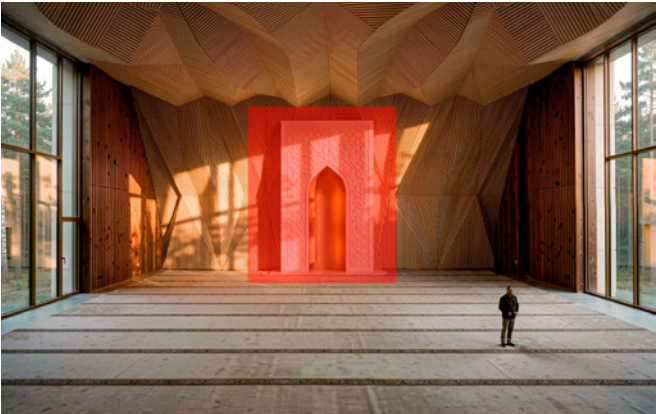
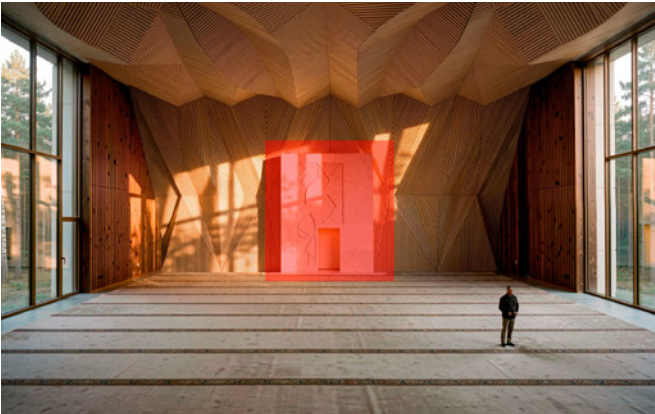
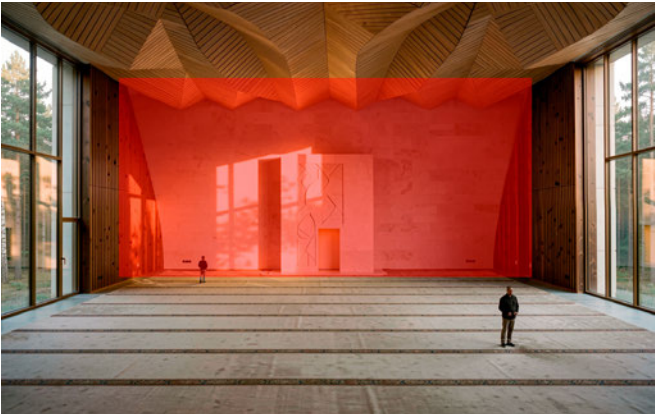
Input 1: Massing Model
Author

GEOMETRY + CAMERA LOCK You are given one image: a color-coded 3D interior screenshot of a mosque prayer hall, modelled in SketchUp and shown in a symmetric, near-frontal one-point perspective looking straight toward the qibla wall. Treat this image only as a fixed geometry, layout and camera reference — it is a flat-shaded massing diagram, not a finished render, and carries no material, lighting or texture information of its own. Reconstruct the exact same room as a hyper-realistic architectural photograph. Do not change the camera angle, position, height, focal length, framing or composition under any circumstances; keep the identical vanishing point, identical symmetry, the same side-wall splay angles, the same proportions, and the same placement of every element. The output must read as a photograph taken from precisely the camera that produced this screenshot. SCALE A large single-volume prayer hall, 15 m clear floor-to-ceiling height and 35 m wide. The two small standing human figures in the screenshot are scale references — render them as two adult people standing quietly on the carpet, correctly sized against the 15 m height, to convey the monumental scale. Do not add crowds, furniture or objects beyond these two figures. MATERIAL MAPPING (replace each color zone with the material below, keeping its exact shape and position) - RED zones (the two side walls splaying back toward the viewer): brownish-black charred timber (shou sugi ban / yakisugi) cladding in fine vertical boards, deep matte blackened grain with subtle warm undertones. - COOL-GREY zones set within / flanking the red side walls: full-height floor-to-ceiling glazed curtain walls with slim dark metal framing, soft Nordic daylight pouring in and a boreal Finnish pine-and-birch forest gently visible outside. - BLUE zone (the rear qibla wall facing the camera): a broad wall clad in pale, warm light limestone in fine ashlar courses, smooth and softly lit. - YELLOW zone (centered on the qibla wall): the mihrab — a single monumental, full-height niche of the same pale light limestone, towering and monolithic, its great height dominating the qibla wall, restrained abstract geometric relief, washed by a soft directional shaft of warm late-afternoon daylight that deepens its shadows. - GREEN zone (the ceiling): a sculptural deformed / faceted acoustic timber ceiling — folded, angular pale-timber slatted panels with shifting planes, warm in tone, soft light catching the facets. - GREY zone (the floor): fully laid with prayer carpet in a muted, low-contrast geometric pattern (deep blue with muted sage and cream), running in straight prayer rows toward the qibla. ATMOSPHERE Contemporary Nordic Islamic architecture, Scandinavian minimalism fused with Islamic spatial tradition; calm, serene, monumental, spiritual. Clear Finnish late-afternoon daylight entering through the flanking curtain walls, raking warm light across the pale stone qibla wall and the charred-timber sides, gentle natura shadows. Architectural photography, Hufton+Crow editorial quality, ultra-sharp, high dynamic range, 8K. EXCLUSIONS Keep the room as one clean uninterrupted volume: do not add a dome, columns, chandeliers or hanging fixtures, calligraphy, signage or text, additional windows or skylights, decorative furniture, plants or clutter. Do not introduce traditional mosque ornament or color beyond the materials specified. Do not alter the geometry, the wall angles or the camera.

Input 2: Text Prompt
Author, Claude



Output: Iterations
Author, Nano Banana 2, GPT Image 2, Gendo AI



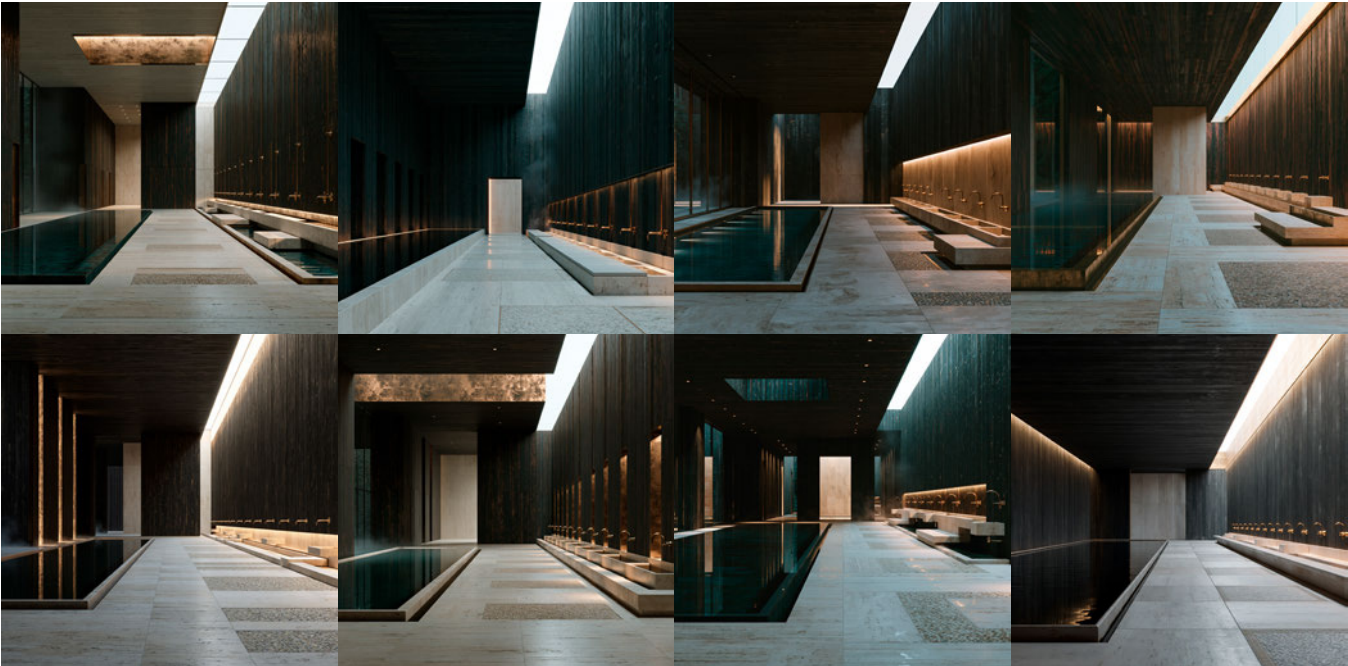
Output: Region Edit
Author, GPT Image 2, Gendo AI



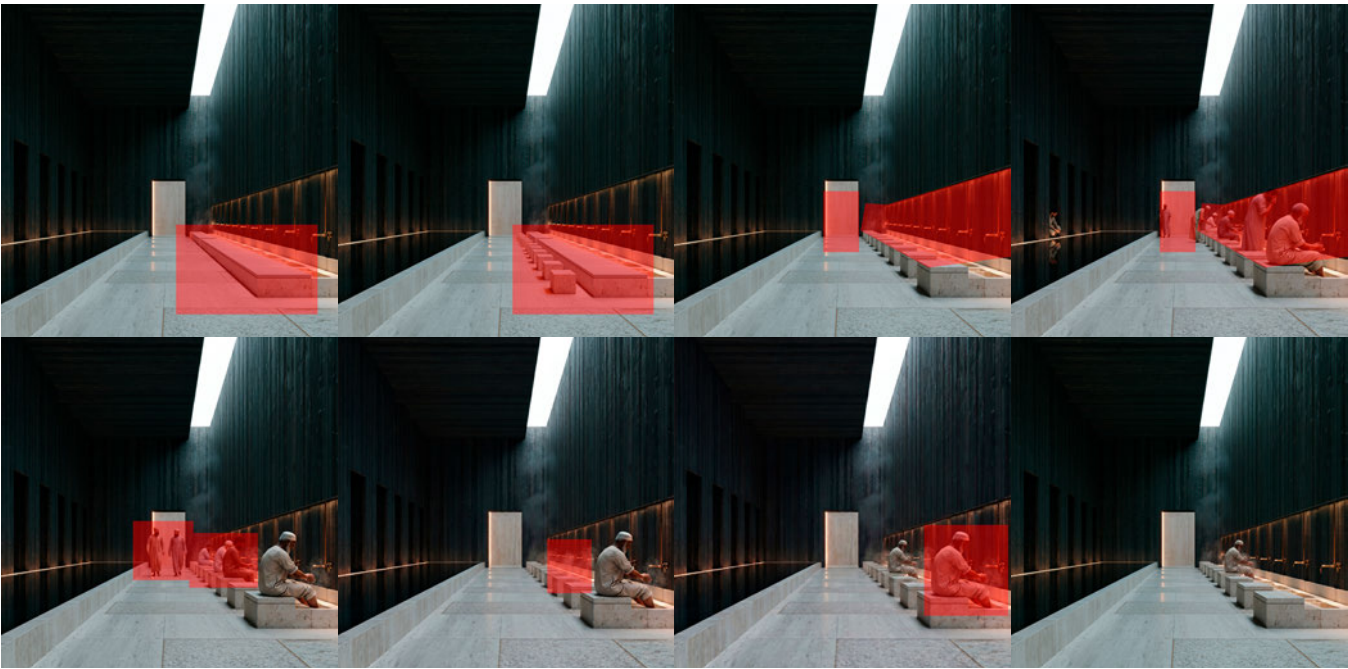
Image 07

Input 1: Text Prompt
Author, Claude

Hyper realistic architectural interior, 8K, Hufton+Crow editorial quality, ultra sharp, high dynamic range. Contemporary Nordic Islamic ablution (wudu) hall. Long rectangular room, 5m ceiling, one point perspective toward solid far end wall, side walls receding symmetrically. LAYOUT: RIGHT WALL: Full length continuous wudu washing wall, slim brushed brass wall mounted faucets, pale stone basins and channel, low integrated seating bench/platform for foot washing. SKYLIGHT: Long continuous linear skylight above right washing zone, soft daylight falls precisely onto faucets, basins, and bench. FAR WALL: Solid, full height charred timber or pale stone, no openings. LEFT SIDE: Full length rectilinear reflecting fountain, rim 30cm above floor, perfectly still water mirroring space and skylight. MATERIALS: Yakisugi vertical boards on walls and ceiling soffit, aged brushed brass on faucets, basins, channel edges; large pale travertine/limestone floor slabs with subtle inset pebble mosaic panels. ATMOSPHERE: Dimly, intimately lit - skylight dominant, brass glows warm, charred timber in deep shadow. Meditative, sacred, tranquil, soft steamy air. Boreal forest faintly visible only where glazing naturally occurs, dark and subordinate. EXCLUSIONS: No people, mirrors, signage, bright artificial lighting, colourful tiles, plants, towels, clutter, dome, columns, ornament, or bubbling fountain.



Output: Iterations
Author, Midjourney



Output: Region Edit
Author, Gendo AI



Final Result



Image 08

Input 1: Text Prompt
Author, Claude

ROLE OF THE ATTACHED IMAGE: The attached image is a 2D top-down architectural floor plan (orthographic line drawing) of the library inside a contemporary Nordic Islamic center. Use it ONLY as a spatial-layout and proportion reference — to read the wedge-shaped room footprint, the line of the slanted glazed wall, the bookshelf wall, the information desk, the reading-table arrangement, and the adjacent outdoor terrace beyond the glass. Do NOT reproduce the line-drawing or plan style; produce a photorealistic interior architectural photograph in perspective. SCENE: Eye-level interior point-of-view photograph standing among the reading tables of a contemporary Nordic Islamic community-center library, looking straight ahead toward a tall, slanted (raked, leaning) floor-to-ceiling curtain wall that dominates the far end of the room. The library is a single-storey volume, approximately 4–5 m clear ceiling height. One-point perspective, camera held at standing eye level (~1.6 m), lens roughly 24–28mm, the diagonal of the slanted glazing reading clearly as the focal feature. LEFT WALL: full-height built-in bookshelves in pale Finnish birch, filled with books, integrated warm strip lighting along the shelves, set against a backdrop of dark charred shou-sugi-ban timber. RIGHT WALL: a low information / reference desk in birch and charred timber with a single staff position, clean and uncluttered, no visible signage or logos. FOREGROUND: library reading seating — a long communal reading table in solid birch with slim chairs, plus a cluster of soft lounge armchairs and a low side table on a muted geometric-patterned wool rug; one or two quiet readers for scale, unobtrusive. THE SLANTED CURTAIN WALL: a leaning floor-to-ceiling glass wall with fine dark metal framing and warm amber-toned glazing, through it, an outdoor terrace with café-style outdoor seating (tables and chairs) sheltered under fabric and timber shade canopies / parasols, and a boreal pine-and-birch forest edge beyond. Late-afternoon golden-hour Finnish summer light rakes through the glazing, casting long warm reflections across a pale timber or honed granite floor and warming the timber and birch surfaces. Calm, scholarly, contemplative atmosphere. MATERIALS & ATMOSPHERE: dark charred timber (shou sugi ban), pale Finnish birch joinery, warm amber glazing, honed granite or warm timber flooring, an acoustic timber ceiling with a subtle faceted geometry, optional single perforated COR-TEN girih lattice screen detail as an accent, glowing warm from behind. Crisp clear Finnish summer day, golden / late-afternoon light, Hufton+Crow editorial architectural photography quality, natural color, high dynamic range. CONSTRAINTS: Keep bookshelves on the LEFT and the information desk on the RIGHT — do not swap them. Do not add a dome, arches, chandeliers, or traditional mosque ornament. Do not add snow, winter weather, mist, or overcast sky — keep it clear summer golden hour. Do not add signage, brand logos, or text. Do not add a second floor, mezzanine, or floating volumes. Do not add stairs. Keep the slanted curtain wall leaning/raked, not vertical. Do not change the camera angle, framing, or composition under any circumstances.



Output: Iterations
Author, Midjourney, Nano Banana 2



Output: Region Edit
Author, Gendo AI

Final Result
Author, Nano Banana 2, Gendo AI

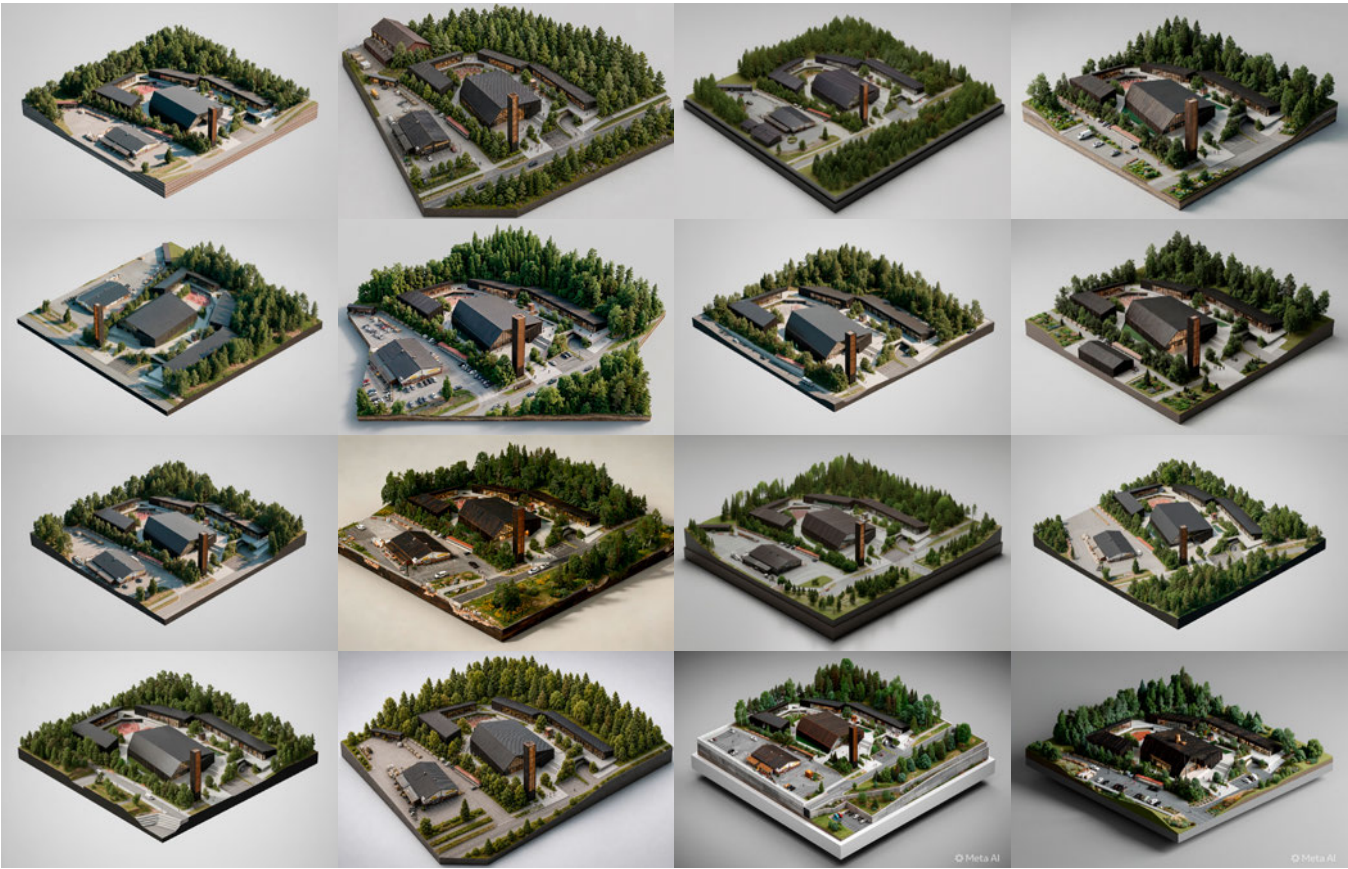
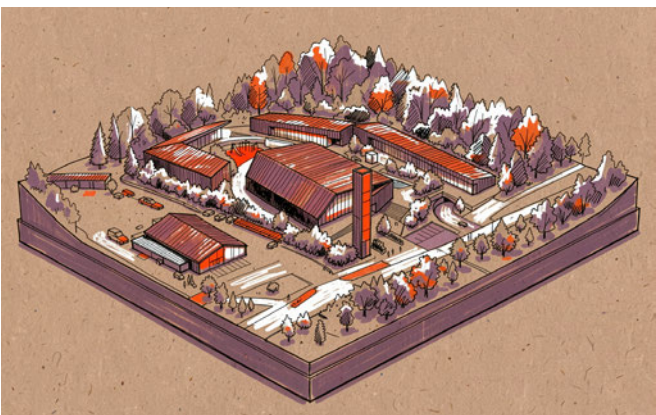
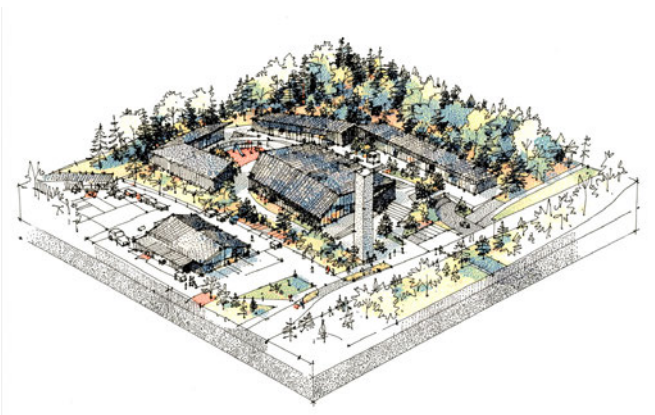
Image 09

Input 1: Text Prompt
Author, Claude

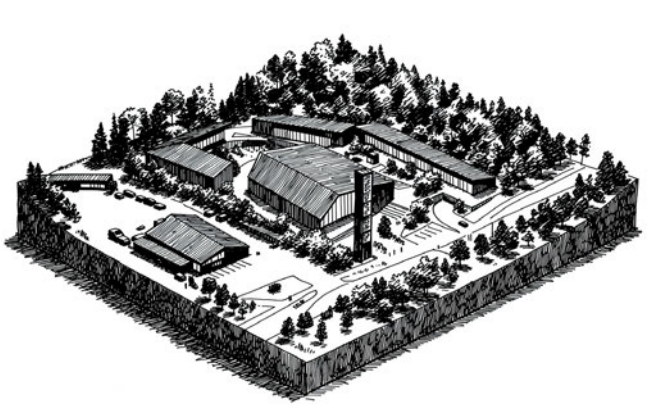
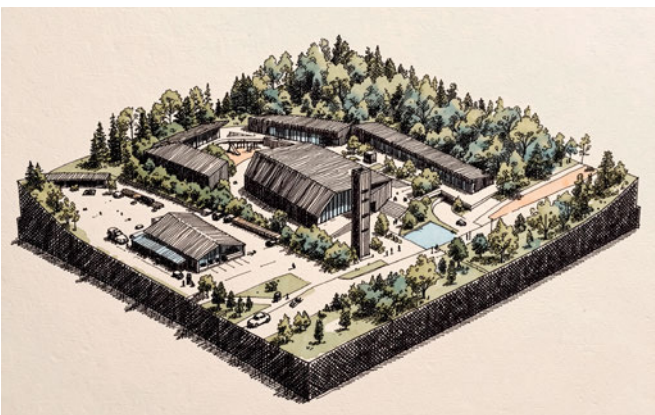
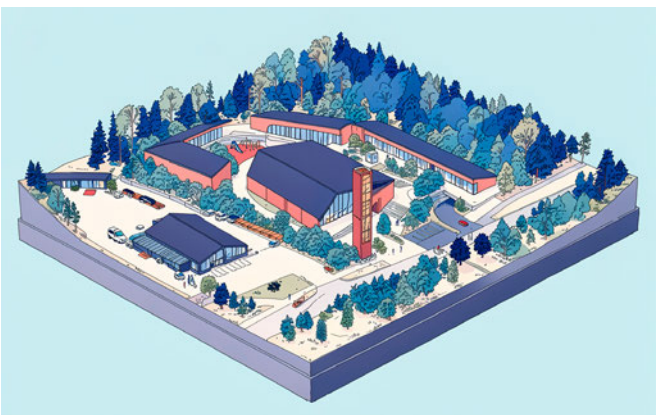
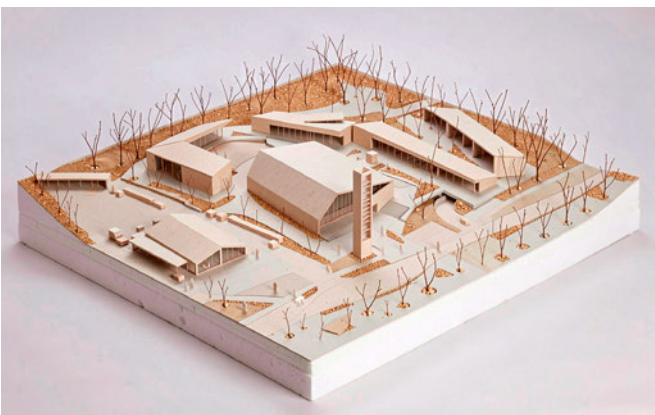
Create a highly detailed isometric architectural tile based on the attached aerial photograph of the Nordic Islamic center. Transform the entire site into a premium collectible architectural diorama tile, maintaining the recognizable layout and proportions of the original complex while presenting it in a clean, orthographic isometric view (approximately 30°/30° axonometric projection). The tile should contain the complete site composition: the central prayer hall with its steep gabled roof, the surrounding curved and linear building wings, the minaret-like vertical tower element, adjacent streets, sidewalks, parking areas, service roads, and pedestrian paths, dense Nordic forest surrounding the complex, site landscaping, courtyards, plazas, retaining walls, planting beds, and site furniture, vehicles and tiny human figures for scale, and terrain subtly modeled as a self-contained display piece. Material treatment should include all buildings clad in deep black-brown charred timber (Shou Sugi Ban) with rich wood grain and subtle variations, warm bronze-toned metal accents on the tower and selected façade details, large contemporary glazed openings with realistic reflections, light grey stone paving and concrete site elements, and natural Nordic vegetation with varied greens and realistic textures. For tile presentation, the entire scene should sit on a square floating display base with clean cut edges around the terrain, like a museum-quality architectural model or premium tabletop city-building game tile, slightly raised topography and layered landscaping, every object fully visible with no cropping, and balanced composition centered on the Islamic center. Lighting should feature professional studio lighting rather than outdoor sunlight, large softbox illumination from multiple directions, soft and controlled shadows, neutral light grey seamless studio background, subtle ambient occlusion emphasizing architectural depth, crisp material definition and texture readability, and no atmospheric haze, fog, lens flare, or sun glare. Rendering style should be ultra-high-end architectural visualization with a hyper-realistic miniature world aesthetic, combination of architectural model photography and photoreal CGI, extremely sharp details, global illumination, physically based materials, and 8K quality, presenting a clean composition suitable for an architecture competition board, design exhibition, or premium concept-art showcase. Preserve the overall site organization, building relationships, circulation routes, and forest context from the source aerial image, but reinterpret the scene as a self-contained isometric architectural display tile viewed in a studio environment.



Input 2: Architecture Reference
Author



Output: Iterations
Author, GPT Image 2, Nano Banana 2, Nano Banana Pro, Qwen Edit 2511, Flux 2 Max



Output: Style Reference
Author, Gendo AI



Final Result
Author, Qwen Edit 2511, Gendo AI, Topaz Generative

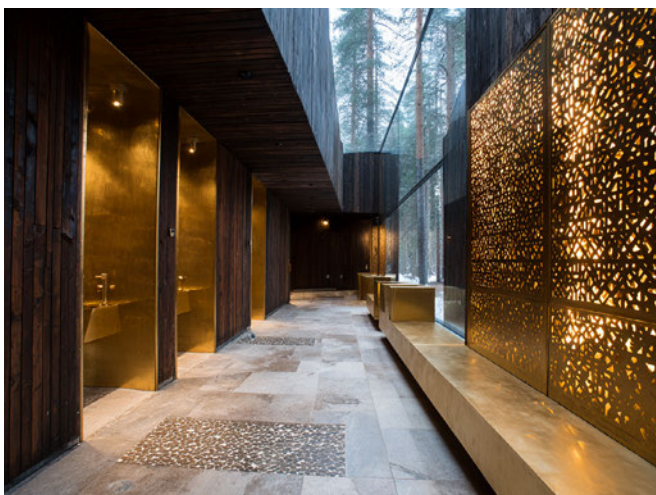
4.4.4. Productive Hallucinations

Most of the model's departures from intent were noise, discarded in the filtering already described. A number, however, proved to be inventions worth keeping rather than errors to be corrected. Telling the two apart was itself a form of design work: the curatorial role that Section 1.2.1 saw the technology imposing on the architect, reclaimed here as a deliberate and generative skill. The instances below are the most consequential, the hallucinations that entered, and in several cases shaped, the final building.

The first came from one of the concept generations produced under the "Mother and Her Children" direction, an image that proved unusually influential on the project's design language. The prompt had described the tall, dominant mass as the mosque itself, yet the model rendered it as a slender minaret rather than a prayer hall. The substitution was unasked for, but it read more convincingly than the intended massing, and the tall perforated tower it proposed became a defining element of the final scheme.



The second appeared among the spatial prompt images of Section 4.3.4, generated to study the character of the ablation space. The model added two features that had not been requested: brass accents lining the room and a long skylight running along its upper edge. Both were adopted in the final design, where the narrow band of daylight and the warm metal became central to the room's atmosphere.



The third also belongs to the Section 4.3.4 images, generated to test the prayer hall interior. As a literal proposal it failed completely, bearing no resemblance to a mosque, but its folded and deformed timber ceiling was compelling in its own right, and that ceiling was carried into the final design of the hall.



The fourth arose during the Stage 2 generations of Section 4.4.2, and was subtler. Neither the recessed glazed wall at the mosque entrance nor the triangular canopy over the entrance stair had appeared in the input sketch, yet the model introduced both, and both were kept in the resolved design.



The fifth likewise emerged in the Stage 2 generations. The model tilted the mihrab wall so that it met the pitched roof in a manner the sketch had not specified, creating a relationship between wall and roof that was retained in the final design.



Across these cases the deviation was curated rather than corrected: judged, kept, and developed. It bears out in practice the argument of Section 2.2.2, that filtered through architectural judgment, hallucination can work as estrangement rather than error, supplying form the designer would not have reached alone. These five are the deviations that judgment retained. They show that hallucination can be put to work when it is filtered with care; they do not show that the model reliably supplies form worth keeping. Far more deviations were discarded than were developed, and the value lay in the selection.

4.4.5. The Projection Gap

The preceding workflow relied on aerial and perspectival imagery throughout, and this was not a matter of preference. Repeated attempts to have the model produce orthographic drawings, the plans and sections on which architectural documentation depends, failed in instructive ways.

The first attempts asked the model to generate a floor plan, or a site layout, directly from a text prompt supplied with the spatial-logic matrix and the bubble diagram. The results bore almost no relation to the project. Feeding the model a rough but already-developed floor-plan sketch improved matters only slightly: the output was less incoherent, yet still unusable, and beyond rescue even by editing. Sections fared better. Conditioned on an accurate section sketch together with several renders of the project, the model came markedly closer, a clear improvement on the plans but still short of a usable drawing.

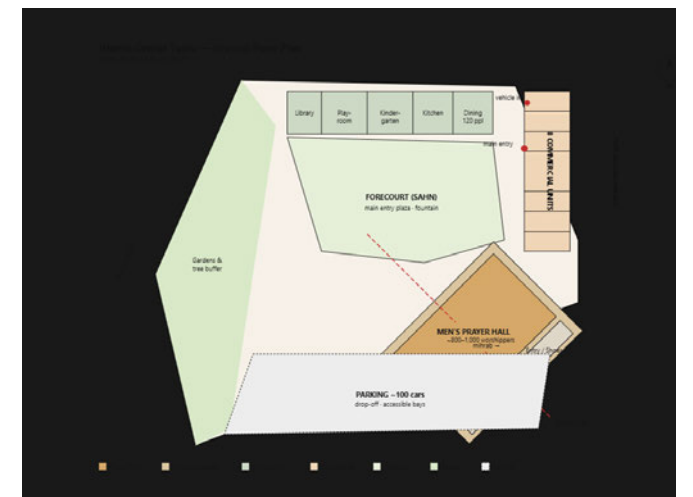


Generative Floor Plan 01
Author, GPT Image 2

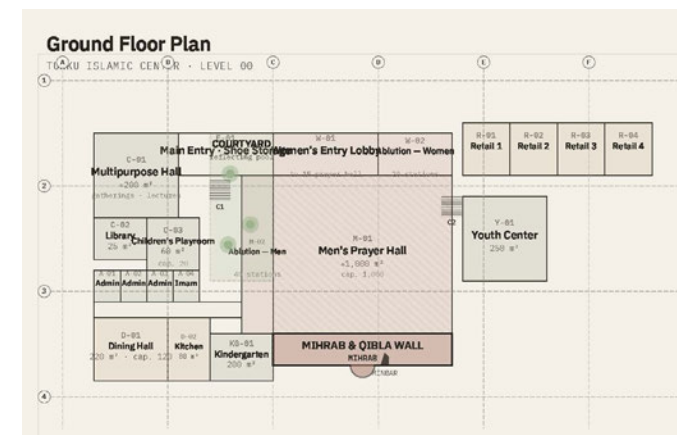


Generative Floor Plan 03
Author, Nano Banana Pro

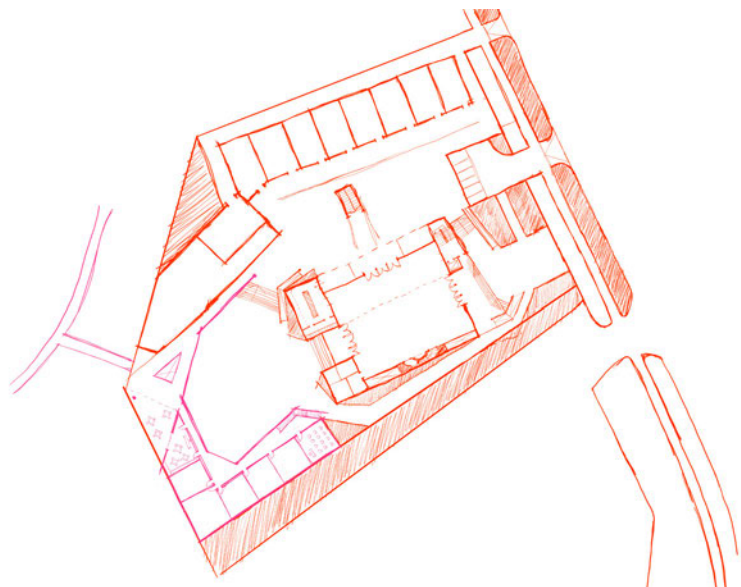
The pattern is the Raster-Vector Gap of Section 1.2.2 observed at first hand. The model performs well on the aerial and perspectival images for which its training data is abundant, and poorly on orthographic projection, where a plan or section is a dimensioned, topological construction rather than a statistically likely picture, and where the rarity of the typology removes even the statistical crutch. Plans, the most abstract and convention-bound of the projections, failed worst; sections, nearer to a sectioned perspective, came closest without arriving. This is why the projections in the project were authored in conventional drafting and modeling software, and the model confined to what it does reliably: the generation of images locked to that authored geometry.



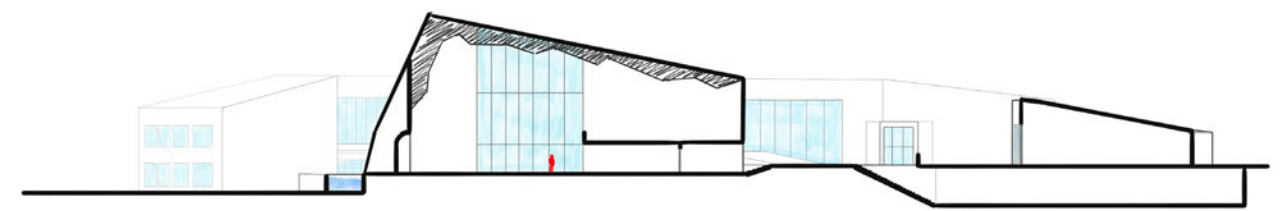
Generative Floor Plan 02
Author, Claude



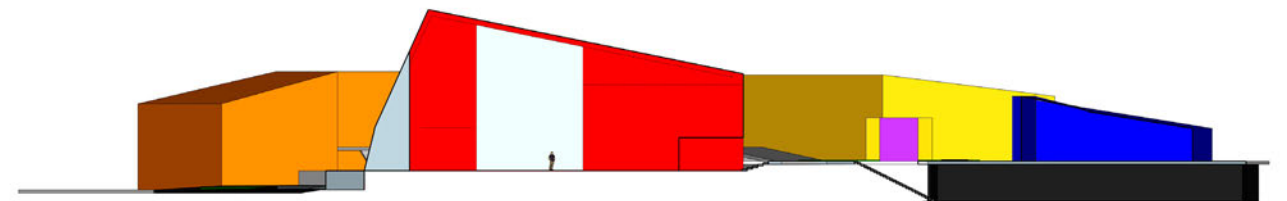
Generative Floor Plan 04
Author, Claude Design



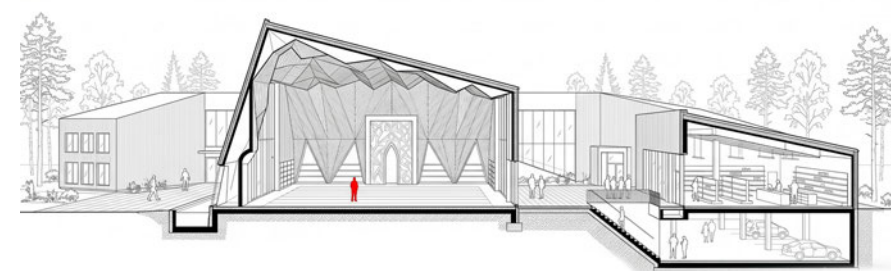
Reference Floor Plan Sketch
Author



Reference Cross-Section Sketch
Author

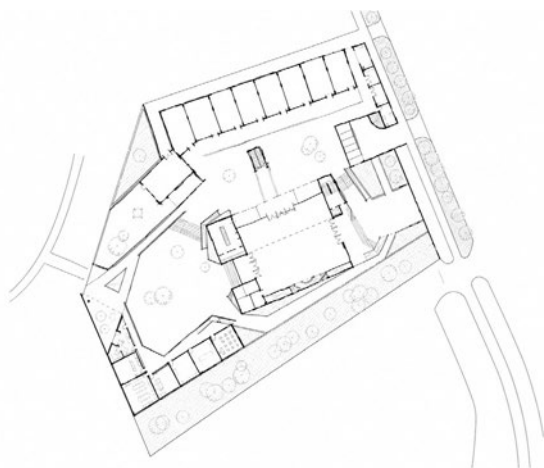


Reference Cross-Section Screenshot
Author

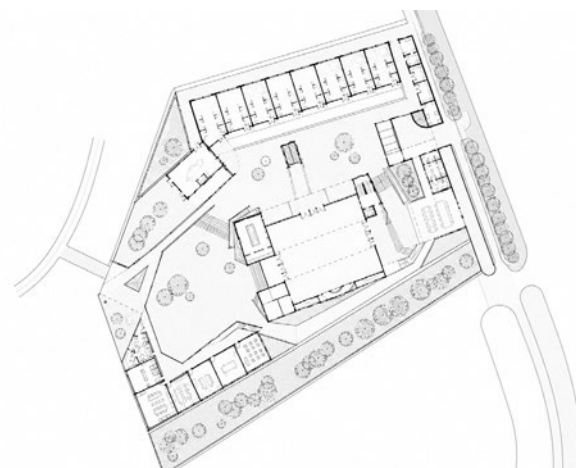


Generative Cross-Section 01
Author, Nano Banana Pro

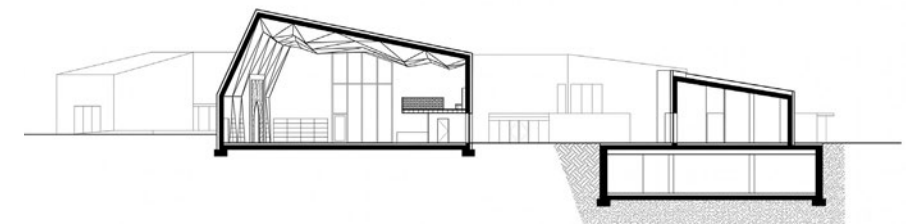
Generative Floor Plan 05
Author, Nano Banana Pro



Generative Floor Plan 06
Author, Claude Design



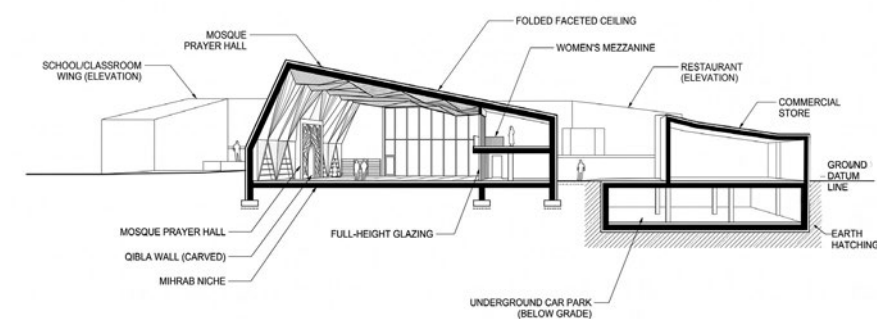
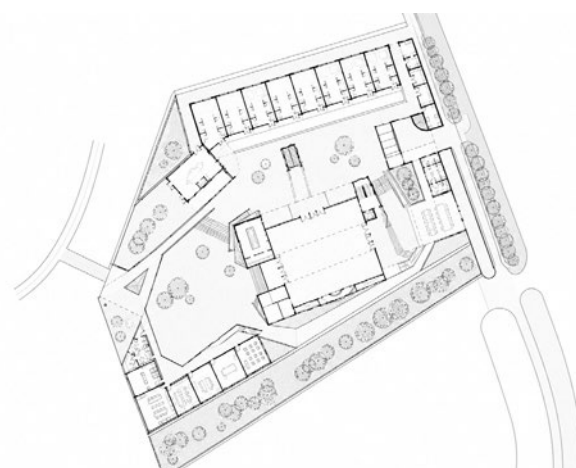
Generative Cross-Section 02
Author, Claude Design



Generative Floor Plan 07
Author, Nano Banana Pro



Generative Floor Plan 08
Author, Claude Design



Generative Cross-Section 03
Author, Nano Banana Pro

CONCLUSION





[The Olive]
Author, Midjourney

5.1. Practical Outcome

The practical results of the research fall into three parts, each meant to leave an architect with knowledge that can be carried straight into a studio. The first records the diffusion models used in the study and what each proved able and unable to do, since no single model serves every purpose. The second turns to the interfaces through which those models are run, from the node-based environment of ComfyUI to dedicated websites and editing plug-ins, each suited to a different kind of work. The third consolidates both into a workflow organized by design stage, from concept generation through to final upscaling, and presents it as a roadmap that can be followed in order. The aim across all three is plainness. Elsewhere the thesis builds its argument and reads the design process through theory, but here the findings are reported as directly as possible, without twisting the observations into something they are not or weighting them with more concept than the work itself supports.

5.1.1. Diffusion Models in Practice

No model performs well across the board. A few come close, but each carries its own strengths and weaknesses, and competent practice is largely a matter of matching the right model to the task at hand. The notes that follow record the strengths and weaknesses observed for each model used in the study.

Nano Banana, Google's image model (officially Gemini Flash Image, spanning the Nano Banana, Nano Banana 2, and Nano Banana Pro versions), is marketed specifically as an editing model, with Imagen positioned as a separate and weaker generation model. Its strengths are considerable. It reasons well about complex problems and understands what an image represents rather than only what its pixels are, to the point of constructing a genuinely new view of a subject from several reference images. It is unusually good at completing missing information, extending a partial facade into a full elevation, section, or isometric, and it is among the few models that produce diagrams with legible text. These qualities make it one of the most versatile models available. Its weaknesses follow from the same training. Its style is highly distinctive and resists change: outputs are consistent within a category, with diagrams sharing diagrammatic qualities and renders sharing render qualities, yet the categories bear little resemblance to one another, and the diagram style in particular appears locked, most likely because few diagrams featured in its training. As a closed, remote model it also depends on its provider's servers; when those servers are overwhelmed the model is reportedly throttled, and output quality drops sharply. The power behind the model is not the user's to control.

GPT Image 2, the current OpenAI model and the successor to GPT Image 1, GPT Image 1.5, and the largely unusable DALL-E, is marketed for both generation and editing. It is versatile and context-aware in much the same way as Nano Banana, though weaker across several areas, and on occasion, depending on the seed, it surpasses Nano Banana in realism. Its limitations are a more pronounced version of Nano Banana's: an even more distinct and near-unchangeable style, marked by a recurring color cast and a dark, moody register; an inability to produce usable diagrams or drawings; and a weaker grasp of context.

Both Nano Banana and GPT Image share two traits worth isolating. Each follows long, detailed prompts almost to the letter, which is a genuine strength. Yet both are markedly stronger at editing than at generation, and their generated images tend toward sameness, repeating similar angles, palettes, and overall look, as though every project had passed through the hands of the same visualization artist and little remained distinctive.

Midjourney, whose underlying model is undisclosed and may be either proprietary or a trained derivative of Flux or Stable Diffusion, was the principal concept-generation model throughout this study. It is the strongest model tested for generative creativity and conceptual imagery, with controls for weirdness, variety, and adherence to its house style, and it produces strikingly inventive, occasionally unusable results. Its weaknesses are equally clear: it follows specific prompts poorly,

and it is among the weakest models for both editing and image prompting. In an architectural context its profile is sharp, best in class for concepts and weakest at almost everything else.

Flux, an open-source family spanning the Flux 1 and Flux 2 generations along with variants such as Flux Max, a Flux Context model, and a Flux 2 editing variant, produces interesting images. It is less conceptually wild than Midjourney but adheres more closely to prompts, which is notable for a comparatively light open-source model, and Flux 2 is exceptional at environmental edits such as weather changes, where Nano Banana is weaker. It is less capable at object substitution than Nano Banana or GPT Image, struggles with very long and complex prompts, and is inconsistent, sometimes spectacular and sometimes poor. For an open-source model that can be iterated freely, that inconsistency is a fair trade.

Two further models merit brief mention. Grok, from X, proved surprisingly capable, particularly at editing, short of Nano Banana but solid. Meta's model is a more curious case: apparently closed-source, yet with no API nodes or partnerships discoverable on any playground or within ComfyUI, it nonetheless produced decent and occasionally surprising results. Qwen 2511, an open-source editing model rather than a generation model, handles both simple and complex tasks well; as with Flux, complex prompts require iteration, but its open-source nature allows that iteration to continue indefinitely without loss of quality.

5.1.2. Interfaces in Practice

These models are reached through three kinds of interface, and the choice of interface shapes what can be done with them almost as much as the choice of model. The first is the node-based local environment, of which ComfyUI is the principal example. The second is the plug-in embedded in existing software, such as Photoshop. The third is the website or playground, such as Krea and Gendo.

For breadth of access and control, ComfyUI is the recommended primary environment. It is the only platform that reaches most models at once, open weights directly and closed models through paid API tokens, the cost of those tokens being its main drawback. It also does something no other interface allows: running several models simultaneously against the same prompt and images, which is how the parallel generations described in Section 4.4.2 were produced. Its disadvantage is a steep learning curve.

The simpler alternative is to move between dedicated websites, using the Gemini app directly for Nano Banana, Krea for Flux, and Midjourney for concept generation. This is far easier to learn but inconvenient in practice, since it scatters the workflow across separate tools. For editing specifically, the Photoshop plug-in is worth singling out, because its region selection and manual retouching sidestep the whole-image degradation that repeated edits otherwise cause. It should not, however, be used to generate images from scratch.

Two playgrounds build or customize their own models and deserve closer attention. Krea is a general playground rather than an architectural tool, run on a token subscription. It offers access to many models through their APIs alongside its own, which began as a trained Flux variant and, with Krea 2, became a model built from the ground up, with strong results and good LoRA support, capable in generation but not in editing. More usefully, Krea wraps individual models in customized interfaces: its Nano Banana editing tab, for instance, adds reference-image categorization, style images, keyword highlighting, camera shifting, color coding, and annotation, none of which are available in the Gemini app directly. Its tablet application generates in real time from an Apple Pencil sketch and a short spoken or written prompt; the real-time model is necessarily weaker, but the feature points to a promising direction for architects as models grow faster.

Gendo, by contrast, is built specifically for architects. It arranges work on a board in the manner of a mood board, where images are generated, edited, enhanced, and upscaled in one place. Its underlying model is undisclosed and is most likely an API or a trained variant rather than an in-house system. What distinguishes it is fluency in architectural terms: it rerenders with style transfer, renders from sketches or 3D screenshots, applies drawing styles, and includes a dedicated function for populating empty architectural images with people whose age, clothing, gender, and seasonal dress can be specified. It does not match Nano Banana on the most complex tasks, but it covers the large majority of an architect's needs.

5.1.3. Roadmap

The practical findings above consolidate into a recommended workflow of six steps. Four of these form the core of the process and run in sequence: analysis, concept, schematic, and realization. An initial input precedes them and a final output follows, so that the workflow carries through from the material the architect already holds to the assembled presentation of what the tools produce. The text below sets out each step in turn; the accompanying full-page diagram maps the same sequence together with the alternative routes available at each step.

Input. The workflow begins with whatever the architect already knows of the project: the brief, the program, the site, the governing constraints, the references, and the design vision. None of this is generated. It is the predetermined material the architect supplies, and the richer it is, the stronger every later step becomes. EMBER's brief, with its program schedule, site coordinates, and stated intentions, is the example carried through this thesis. A thin input yields a thin result, so the effort spent assembling this material is repaid downstream.

Analysis. A large language model such as Claude turns the raw input into structured research. From little more than a set of coordinates it assembles a site analysis spanning climate, sun, ground conditions, and the regulatory envelope, alongside relevant precedent and the cultural and architectural context the project must enter. This is the "Logic Synthesis" use of large language models from Section 1.3.2, and the procedure built into the first half of the concept-generation skill of Section 4.3.1, demonstrated at length in the EMBER site analysis of Section 4.2. The model gathers and cross-references in minutes what would otherwise take days, leaving the architect to verify, interpret, and select.

Concept. The structured analysis feeds concept generation. Midjourney remains the strongest tool for generating and visualizing concepts, run alongside a large language model that researches references and structures the proposals into design logics rooted in the analysis rather than in stylistic association. The custom concept-generation skill of Section 4.3.1 governs this step, turning unstructured prompting into a repeatable procedure and producing a field of typology-specific directions of the kind catalogued for EMBER in Section 4.3.2.

Schematic. One or more of the concepts are then developed through an iterative generation loop, the most demanding step in the process. The work calls for reasonably accurate visualization and the study of layout, plan, form, and mass. The most efficient route that avoids 3D modeling is to sketch and then render those sketches with Nano Banana or GPT Image. A Google Earth screenshot serves well as a base to sketch over, as does sketching directly in Krea. Model choice is flexible here, since most perform adequately; Nano Banana is the strongest, though a local setup makes Flux the better option. As the design firms up, the input tightens, moving from loose sketches to accurate sketches of the relevant projections or screenshots of a 3D massing model, rendered with Nano Banana or with an architecturally developed tool such as Gendo. The loop tightens constraint as it proceeds, in the manner traced through Section 4.4.

Realization. Once the iteration reaches a satisfactory result, a final pass of enhancement and upscaling is applied to all the generated data. Regional editing, handled cleanly in the Photoshop plug-in, corrects local defects without degrading the whole image. Upscaling then adds detail and texture while removing the artifacts that generation leaves behind, such as disfigured figures; Topaz Generative and Magnific both serve. One caveat emerged in the EMBER project and is worth recording: Topaz appears under-trained on Islamic imagery, and during upscaling it rendered figures in hijab and Islamic dress as Caucasian figures with uncovered hair, an alteration Magnific did not make. Training-data blind spots persist even into post-processing.

Output. The workflow closes by assembling all the generated data, the concepts, the resolved views, and the enhanced final images, into a single coherent presentation layout, composing the scattered products of the preceding steps into a legible whole.

A qualification applies to every model named above. Each is the strongest available for its task at the time of writing, and the recommendations reflect that moment alone. The field moves quickly, as Section 5.3.1 sets out, and the specific tools are certain to be overtaken; what should outlast them is the sequence of the six steps and the kind of tool each calls for, which hold regardless of which model occupies any given slot.

Roadmap

the sequence with the alternative routes at each step, initial input and a final output. The diagram maps concept, schematic, realization — framed by anFour core steps run in sequence — analysis,

core sequence —————
alternative route / iteration - - - - -

PREDETERMINED MATERIAL
Brief · Program · Site
Constraints · References
Design vision
e.g. EMBER brief

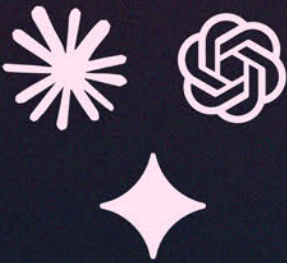
Input

The workflow begins with what the architect already holds; brief, program, site, constraints, references, and the design vision. None of it is generated. The richer this material, the stronger every step that follows.

LLM · LOGIC SYNTHESIS
Claude
climate · sun · ground · regulatory
precedent · cultural context

Analysis

A large language model turns raw input into structured research. A site analysis spanning climate, sun, ground and the regulatory envelope, alongside precedent and context. It cross-references in minutes what would take days; the architect verifies and selects.



GENERATE + VISUALISE
Midjourney
+ LLM → design logics
typology-specific directions

Concept

The structured analysis feeds concept generation. Midjourney remains strongest for generating and visualising, run alongside an LLM that structures proposals into design logics rooted in the analysis, a field of typology-specific directions.



ITERATIVE LOOP
Sketch → Render
base — Google Earth · Krea
Nano Banana · GPT Image
alt — Flux (local) · Gendo

Schematic

One or more concepts are developed through an iterative loop, the most demanding step. Sketch, then render, over a Google Earth base or directly in Krea. As the design firms up the input tightens, from loose sketches to accurate projections and 3D massing.



ITERATE

ENHANCE + UPSCALE
Regional edit — Photoshop
upscale — Topaz · Magnific
caveat — Bias

Realization

A final pass of enhancement and upscaling is applied to all generated data. Regional editing in the Photoshop plug-in corrects local defects; upscaling adds detail and texture while removing the artifacts that generation leaves behind.



ASSEMBLE
Presentation layout
concepts · views · final images

Output

The workflow closes by assembling all the generated data, concepts, resolved views and enhanced final images, into a single coherent presentation layout, composing the scattered products of the preceding steps into a legible whole.

Each tool named is the strongest available for its task at the time of writing; the field moves quickly and specific tools are certain to be overtaken. What outlasts them is the sequence of the six steps, and the kind of tool each calls for.

[Roadmap]
Author

5.2. Theoretical Realizations

This Section explores the philosophical implications of generative AI on architecture, arguing that while AI is currently an assistive tool rather than a replacement for architects, it fundamentally alters the design process. It addresses the "hype gap" between AI's theoretical potential and its current state as a largely speculative, non-autonomous technology. The section further posits that because AI lacks cultural and contextual understanding, it cannot make subjective design decisions and therefore will not replace human architects. Instead, the research highlights a shift in design hierarchy, suggesting that AI's fluency in experiential imagery may invert traditional practices that prioritize orthographic projections, while emphasizing that successful human-AI collaboration requires the architect to actively manage the "flow of constraint" to ensure design intent remains in human hands.

5.2.1. The Hype Gap

The potential of AI to reshape architectural practice is real, and this thesis does not doubt it. The industry, however, discusses that integration far more than it practices it. In current use the technology is largely assistive and, in many firms, still theoretical. Nothing central to architectural work has in fact been handed over to it; it helps with small and repetitive tasks rather than replacing the core of the discipline, because it is not yet reliable enough to replace anything an architect actually does. The potential remains, across schematic design, concept generation, documentation, and fact-checking, but potential is not present practice.

Part of the distortion is structural. The profession, like many others, rides each wave of new technology, and there is reputational advantage in being seen aboard early. The result is misrepresentation rather than outright falsehood. Articles and videos claim that a given model can draw sections and elevations at complete accuracy, when in practice it falters on unusual typologies, or on anything it has not seen many times before, because its training data does not cover everything. A wide gap separates the researchers and genuine innovators from the influencers and content creators, and it exists because the technology has not yet settled. Reliable insight requires testing the tools directly and at depth, and the great majority of the material promising to build a whole building with AI does not survive that scrutiny, most of it coming from outside research. As the technology matures and workflows stabilize, that gap should narrow.

5.2.2. AI vs. Architect

Among the claims that circulate about AI in architecture, the most far-reaching is that it will replace architects. It is also among the least convincing. The debate echoes the older question of whether computers are smarter than people, and like that question it can be argued both ways. AI plainly excels at what computers have long done better than humans: computation, repetition, calculation, simulation, and analysis. The claim that it can therefore make a better building, however, collapses on the meaning of "better." To call one building better than another is to have already defined the best building, and that definition is irreducibly subjective. The most beautiful, the most functional, the grandest, and the most quietly improving of daily life are not the same target. AI can be made to excel at the measurable ends, optimal daylight or the most efficient structure, but not at the contested matter of beauty.

The deeper problem is context. A diffusion model carries no cultural background and is, in effect, the same everywhere, much as a given computer is the same machine in North America, the Middle East, or East Asia. It will not register the sequence and privacy that a particular culture asks of a dwelling, the Japanese sensitivity to light and shadow, or the Nordic hunger for sun. Its sense of the best building is therefore speculative. Two refinements sharpen the point. First, even were all of a culture's knowledge available to it, the model would not know that knowledge in any meaningful sense: one may read the whole of a country's literature and still not know the country without having lived in it, and the model is precisely such a scholar, informed of everything and acquainted with nothing. Second, modern architecture has already travelled some

distance down this road. The early-twentieth-century cult of the individual architect, the International Style, and the ambition to unify architecture across the globe all de-prioritized context, a direction now widely regarded as a mistake, and AI left to its own logic would press further the same way, maximizing efficiency and treating buildings as well-functioning machines without a soul.

The familiar analogy to the calculator and the mathematician does not quite hold and is set aside here. A calculator good enough might indeed displace some mathematical work, but architecture is unlike mathematics in that it changes across cultures, countries, and ways of living, and it is exactly that variability the model cannot inhabit.

5.2.3. Architect and Machine

Whether the tool is best understood as an assistant or as a fellow architect is unsettled. It is not obvious who leads, who makes the final decision, or who influences whom, and a case can be built either way. What is not in doubt is that working with the tool influences the architect's decisions, and that its output therefore has to be treated with care rather than accepted at face value.

The most workable relationship found in this study treats the model's tendency to fill gaps as something to be managed deliberately, and managed differently at different stages. Early in a project the model is given less information, so that the room left for it to fill becomes room to innovate, and its hallucinations are accepted as creative noise worth sifting. Late in a project it is given a great deal more, so that it returns something close to exactly what is wanted. This is the same constraint gradient enacted in the case study, read now as a way of conducting the relationship itself: the architect decides, at each moment, how much authorship to retain and how much to lend.

5.2.4. Shift in Thinking

Sustained work with these tools changed the way the design task itself was approached. Instructing a generative model is less like operating a piece of software than like addressing someone who lacks the background knowledge ordinarily taken for granted in any collaborator. Asked simply to hold a cup, such an interlocutor might take it sideways and spill its contents, less from incapacity than from a missing assumption no one thought to state. Communicating intent therefore demands invention, because the model frequently misreads plain instruction.

The same faculty that makes the model useful complicates this further. It is adept at filling gaps, but given too little to work with it fills them with implausible things, having no context for what was meant. Practice settles into a balance between two failures: instruction explicit enough to forestall misreading, and context generous enough to forestall the stranger hallucinations. Learning to hold that balance, rather than any single technique, is the lasting change in method this research produced.

5.2.5. Design Beyond the Drawing

The case study returned one result that bears on the discipline more widely than on the project itself. The generative tools were at their most fluent in the experiential image, the aerial and the perspectival view, and at their weakest in the orthographic projection, the plan, the section, and the elevation. Section 4.4.5 recorded this as the Projection Gap, and the Raster-Vector Gap of Section 1.2.2 explains its mechanism. Read narrowly, it is a limitation to be patched. Read more widely, it unsettles an assumption the discipline has held for the better part of a century.

That assumption is the primacy of the drawing. Architecture has long thought through projection. The plan was held to generate the building, while the section and elevation were the instruments through which form was reasoned, dimensioned, and communicated. The perspectival image, by contrast, was treated with suspicion, a seductive presentation made after the real work was done and distrusted for the very persuasiveness that made it useful. Generative AI inverts this order. It is fluent in the mode the discipline held at arm's length, and it falters in the mode the discipline placed at its center.

The provocation worth taking seriously is what follows if that inversion is accepted rather than resisted. A building is encountered in three dimensions, in light, over time, at the scale of the body. No one has ever inhabited a section. In the built world the drawing carries no weight; it is an instrument for arriving at the building, never the building itself. If the tool generates most powerfully in the mode closest to how architecture is actually experienced, then the origin of a design could migrate there, beginning in the simulated experience of a space and treating projections as documentation drawn from it afterward, by conventional means. The drawing would survive as a record and a set of instructions, produced once the experiential intent had been settled, rather than as the place where that intent is first formed.

The counterweight is substantial and should not be understated. Projections are more than representations; they are the instruments by which a building is measured, coordinated, costed, regulated, and constructed. A section carries the information a render conceals: structure, assembly, tolerance, the thickness of things. To design from the image alone is to risk privileging the photogenic over the constructible, deferring the structural and dimensional discipline that the section enforces, and the long suspicion of the seductive view has exactly this danger in mind. A practice that allowed the experiential image to lead would have to import that discipline deliberately at the point of documentation, rather than receiving it for free from a process that began in the drawing.

The realization is narrower than abolition. The drawing does not disappear, though its position in the sequence is no longer fixed. Where generative tools are strong, the experiential image can precede the projection rather than follow it, and the architect is left to decide, project by project, which mode should lead. That decision is itself an exercise of the agency this thesis argues for: to let the tool's fluency reorganize the order of design where it serves the work, and to hold the projection's authority in place where construction demands it.

5.3. Limitations

5.3.1. Pace and Standardization

The most immediate limitation of any work in this area is that the field will not hold still. Over the period during which this thesis was written, the tools it examines changed repeatedly, and the work had to be revised to keep pace: GPT Image moved through three versions, Nano Banana through three, Flux through two generations, and Krea rebuilt its own model as Krea 2. Section 3.3.4 shows the effect directly, recording a state of the field already partly overtaken by the time the conclusions were drafted, which is why a model assessed there as GPT Image 1.5 reappears in Section 5.1.1 as GPT Image 2. Any finding tied to a specific model is therefore perishable, and the durable contribution lies one level up, in the workflow and the architect-tool relationship, which outlast the models that occasioned them.

The pace is compounded by an absence of standards. Where the mature toolset is interoperable by convention, exchanging a drawing as DWG or a model as IFC, generative tools share no common interface, no agreed language of prompting, and little interoperability between platforms, so skill earned in one does not carry cleanly into another. For a profession this is a real barrier: a practice cannot easily build durable workflows on tools that change monthly and refuse to interoperate, nor justify training staff on software that may be obsolete before the training repays its cost. The friction should ease as the technology matures and a few interfaces consolidate, much as it once did for computer-aided drawing, but until then the pace of change and the want of standards remain among the more practical limits on what generative AI can contribute.

5.3.2. Hardware

These models are computationally heavy, and that weight has to be borne somewhere. With open-source models it falls on local hardware: the architect, or the practice, must own and run a machine capable of the work. The compensation is consistency and independence, since a local model performs at whatever level the setup allows and answers to no external scheduler. With closed-source models the computation is the provider's, reached remotely, which removes the hardware barrier but introduces a dependency. When a provider's servers are overwhelmed, the model can be throttled without notice; Nano Banana has reportedly been reduced in capability at such moments, returning markedly poorer results. The power behind a closed model is, in the end, not the user's to govern. For professional adoption the choice is therefore between a real capital cost in hardware and a standing reliance on infrastructure the practice neither sees nor controls.

5.3.3. Training Data and Bias

Much of what these models can and cannot do traces directly to the data they were trained on. The clearest deficits appear where that data is thin. None reliably produces architectural projections, and all struggle with uncommon typologies, the contemporary Nordic mosque of the case study among them, because such buildings appear rarely, if at all, in the training set. The same mechanism produces bias of a more troubling kind. During the EMBER project, the upscaling tool Topaz, evidently under-trained on Islamic imagery, altered figures in hijab and Islamic dress into Caucasian figures with uncovered hair, a substitution another tool did not make. What a model has seldom seen, it renders poorly or wrongly, and at times it overwrites what it was given.

A subtler limitation is stylistic. Because each model carries a recognizable signature, heavy reliance on any one of them tends toward homogenization, with a practice's output drifting toward a single machine's defaults. For closed models this is difficult to escape. For open-source models it is far less constraining, since LoRA training and related customization allow the default style and its biases to be retrained away, which is among the strongest arguments for the open-source approach in a professional setting.

5.3.4. Open or Closed Source

The field divides into open-source and closed-source models, and at present neither offers everything. Closed-source models are more powerful out of the box, but they are opaque and fixed: their signatures cannot be removed, their behavior cannot be inspected, and their availability depends on a provider. Open-source models can be customized without limit, through further training, LoRAs, inpainting, and noise-level control, and they run under the user's own command, but they remain, for now, the weaker performers. The honest summary is a trade-off, power on one side and control on the other, and at the time of writing the customization available to open-source models is not yet strong enough to outweigh the raw capability of the best closed ones.

That balance seems unlikely to hold. Two futures are plausible, and they are not exclusive. Open-source models may simply close the capability gap and prevail on the strength of their flexibility, or their rise may force the closed providers to open their systems further to remain competitive. Either way, the present settlement, in which the most capable tools are also the least controllable, reads more like a phase than a destination.

5.4. Conclusion

The fear that a generative model would turn the architect into a spectator of their own project turns out to be conditional rather than inevitable. An architect can work with a probabilistic tool and still author the result. What decides the matter is the flow of constraint. Give the model little and it returns invention no one asked for; give it the architect's own geometry and it returns something close to what was intended. Authorship survives in the hand that turns that dial, and the single most useful thing to take from this work is the knowledge that the dial exists and can be operated on purpose.

From this follows a less comfortable conclusion. Hallucination is better treated as a material than as a malfunction. The reflex to correct the model's errors is premature, because those errors are worth most at the beginning, when a design is still open and a strange suggestion can carry it somewhere unreached by ordinary means, and they turn into liabilities only later, when precision is owed. EMBER kept five of them. A mass the model misread as a minaret, a skylight it invented over an ablution space, a ceiling it folded where none had been drawn: these were absorbed into the design rather than corrected back toward intention. For a discipline raised on control this is hard to accept. It is nonetheless true that part of the building was authored by keeping what the machine got wrong.

The architect is never a neutral party to this exchange. The tool shapes the choices of whoever directs it even as they constrain the tool, and to deny the influence is only to stop noticing it.

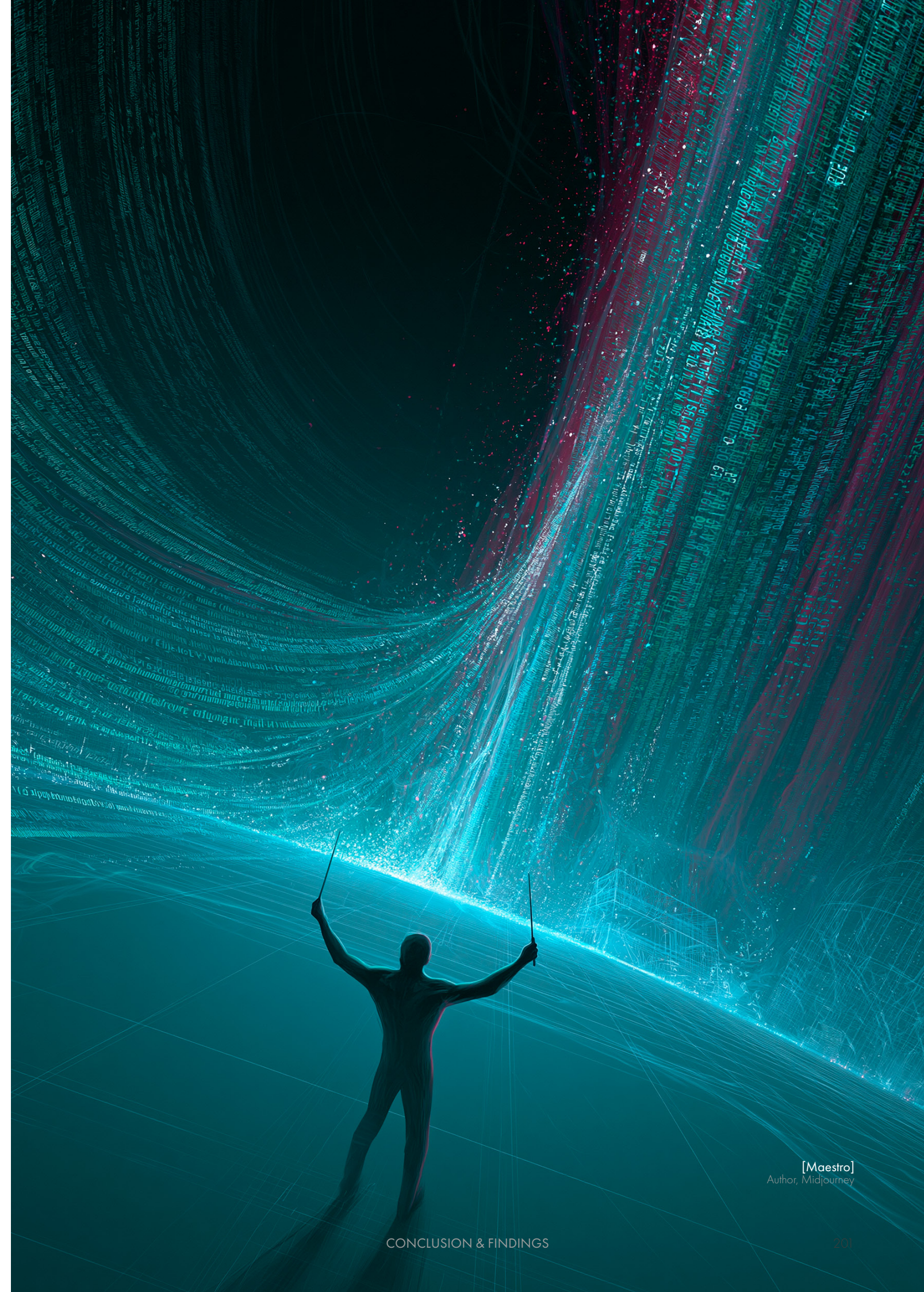
A wider conclusion outruns the project entirely. These tools are fluent in exactly the register architecture has long dismissed as decoration, the rendered and experiential view, and they are clumsy with the orthographic projection the discipline has treated as the real site of design. For the better part of a century the plan was held to generate the building and the perspective merely to sell it. That order is now reversible. A practice could begin in the experiential image, where the tools are strong, and draw its plans and sections afterward, unsettling a hierarchy old enough to pass for nature. Whether the profession takes that road is uncertain; that the road has opened is not.

On the loudest question, whether the architect will be replaced, the answer is for now definite. It will not happen soon, and the reason is precise rather than sentimental. A model can optimize whatever is measurable, daylight and structure and movement, but the claim that it makes a better building falls apart the instant anyone asks what better means, since the answer is cultural and particular and lived, and the model holds all the information about such things while having lived none of it. It is a scholar with no country. The measurable will keep migrating to the machine. The judgment that weighs a building against a way of living will not, and that judgment is the part of architecture least like the work computers were always good at.

None of this arrives without heavy qualification. The tools are shaped by their training, sure with the common and unreliable with the rare, and able to erase the specificity a project rests on, as the upscaling that turned figures in Islamic dress into

something else demonstrated plainly. They go stale within months. They still cannot produce a section worth building from, and they remain split between closed systems that are powerful and sealed and open ones that are free and weaker, leaving the architect to choose which compromise to live with. These limits are real. They fence the conclusions above without undoing them.

What is left, once the forecasts and the cautions are set down, is one finding stated without ornament. The machine did not design the building. It enlarged what could be attempted and pressed its collaborator to think harder about intention, yet the building was authored in the decisions it could not make, the choice of what to keep and what to refuse, and the harder question of what a space in this place and for these people ought to be. That is where the discipline still lives. The lasting contribution here is neither a tool nor a model, since both will date within the year; it is a way of keeping that judgment in the architect's hands while the instruments around it grow stronger. Of the two paths, augmentation was always the more demanding. It asks the architect to stay present at every turn, which is precisely how it keeps the architect in the work.



Bibliography

-Alammar, J., *The Illustrated Stable Diffusion*, [jalammar.github.io](https://github.com/jalammar/stable-diffusion), 2022.

-American Institute of Architects (AIA), *Artificial Intelligence Adoption in Architecture Firms*, 2025.

-Bernstein, P., *Architecture | Design | Data: Practice Competency in the Era of Computation*, Birkhäuser, 2018.

-Bolojan, D., *Creative AI: Deep Learning for Architecture*, 2022.

-Carpo, M., *A Very Short Story of the Digital Turn in Architecture*, 2013.

-Carpo, M., *Beyond Digital: Design and Automation at the End of Modernity*, MIT Press, 2023.

-Castelvecchi, D., *Can we open the black box of AI?*, Nature, 2016.

-Chaillou, S., *AI + Architecture: Towards a New Approach*, Harvard GSD Thesis, 2019.

-del Campo, M., *Neural Architecture: Design and Artificial Intelligence*, ORO Editions, 2022.

-Deutsch, R., *Superusers: Design Technology Specialists and the Future of Practice*, Routledge, 2019.

-Dimension Market Research, *Generative AI in Architecture Market Report*, 2024.

-The Double Diamond: A Universally Accepted Depiction of the Design Process, Design Council.

-Dürr, O., Sick, B., & Murina, E., *Probabilistic Deep Learning with Python*, Manning Publications, 2020.

-The Evolution of Architectural Practice: From Hand Drawings to AI Integration, Parametric Architecture.

-The Evolution of Modelling in Architecture, eCAADe, 2023.

-Frayling, C., *Research in Art and Design*, Royal College of Art Research Papers, 1993.

-Gartner, *Hype Cycle for Emerging Technologies*, 2023.

-Géron, A., *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.)*, O'Reilly Media, 2022.

-Goodfellow, I., et al., *Generative Adversarial Networks*, Advances in Neural Information Processing Systems, 2014.

-Goodfellow, I., Bengio, Y., & Courville, A., *Deep Learning*, MIT Press, 2016.

-Ho, J., Jain, A., & Abbeel, P., *Denoising Diffusion Probabilistic Models*, Advances in Neural Information Processing Systems, 2020.

-Ho, J., & Salimans, T., *Classifier-Free Diffusion Guidance*, arXiv:2207.12598, 2022.

-Kaplan, J., et al., *Scaling Laws for Neural Language Models*, arXiv:2001.08361, 2020.

-Karras, T., et al., *Elucidating the Design Space of Diffusion-Based Generative Models*, Advances in Neural Information Processing Systems, 2022.

-Kingma, D. P. & Welling, M., *Auto-Encoding Variational Bayes*, arXiv:1312.6114, 2013.

-Kiviniemi et al., *BIM adoption and implementation*, 2008.

-Kvan, K., *The effects of technology on Architecture*, 2008.

-Leach, N., *Architecture in the Age of Artificial Intelligence*, Bloomsbury, 2022.

-Leach, N., *AI and the Death of the Architect*, Dezeen, 2023.

-Licklider, J. C. R., *Man-Computer Symbiosis*, IRE Transactions on Human Factors in Electronics, 1960.

-Manovich, L., *AI Aesthetics*, Strelka Press, 2018.

-MDPI, *Enhancing Efficiency and Creativity in Mechanical Drafting*, 2024.

-Mitchell, T., *Machine Learning*, McGraw-Hill, 1997.

-Nawari, N. O., *Building Information Modeling and Artificial Intelligence*, Routledge, 2024.

-NBS, *National BIM Report*, 2017.

-Negroponte, N., *The Architecture Machine*, MIT Press, 1970.

-Ontology of BIM in a Construction Project Life Cycle, 2020.

-Radford, A., et al., *Learning Transferable Visual Models From Natural Language Supervision*, arXiv:2103.00020, 2021.

-RIBA Plan of Work 2020 Overview, Royal Institute of British Architects.

-Rombach, R., et al., *High-Resolution Image Synthesis with Latent Diffusion Models*, CVPR, 2022.

-Ronneberger, O., et al., *U-Net: Convolutional Networks for Biomedical Image Segmentation*, MICCAI, 2015.

-Russell, S., & Norvig, P., *Artificial Intelligence: A Modern Approach (4th ed.)*, Pearson, 2020.

-Schön, D. A., *The Reflective Practitioner: How Professionals Think in Action*, Basic Books, 1983.

-Schumacher, P., *Parametricism 2.0: Gearing Up to Impact the Global Built Environment*, Architectural Design, 2016.

-Shklovsky, V., *Art as Technique*, 1917.

-Sociotechnical Systems Approach to BIM Implementation, 2015.

-Steyerl, H., *Mean Images*, New Left Review, 2023.

-Taraszkiewicz, A., *Architectural Drawing Throughout History*, 2021.

-Vaswani, A., et al., *Attention Is All You Need*, Advances in Neural Information Processing Systems, 2017.

-Vidalakis et al., *BIM Adoption in Small Architectural Practices*, 2019.

-Zhang, L., & Agrawala, M., *Adding Conditional Control to Text-to-Image Diffusion Models*, arXiv:2302.05543, 2023.

-Zimmerman, J., Forlizzi, J., & Evenson, S., *Research Through Design as a Method for Interaction Design Research in HCI*, CHI Conference on Human Factors in Computing Systems, 2007.

