

Research Article

Cite this article: Çelik T (2025). Evaluating daylight performance of AI-generated housing plans via diffusion models and climate-based simulation. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 39, e19, 1–22
<https://doi.org/10.1017/S0890060425100085>

Received: 27 May 2025
Revised: 30 June 2025
Accepted: 06 July 2025

Keywords:

artificial intelligence; text-to-image diffusion models; sustainable housing design; daylighting simulation; performance-based design

Corresponding author:

Tuğçe Çelik;
Email: tugce.celik@ostimteknik.edu.tr

Evaluating daylight performance of AI-generated housing plans via diffusion models and climate-based simulation

Tuğçe Çelik 

Faculty of Architecture and Design, Ostim Technical University, Ankara, Turkey

Abstract

This study investigates the applicability of generative artificial intelligence (AI) in early-stage architectural design by evaluating the daylight performance of AI-generated sustainable housing plans across five distinct climate zones. A three-phase methodology was implemented: (1) Plan generation using text-to-image diffusion models (ChatGPT, Copilot, and LookX); (2) digital reconstruction in AutoCAD; and (3) daylight simulation via Velux Daylight Visualizer. Climate-adaptive prompts were formulated to guide the AI tools in producing context-specific floor plans with passive strategies. Out of 31 initial plans, eight valid outputs (five from ChatGPT and three from Copilot) were reconstructed in AutoCAD and simulated. Quantitative simulations were conducted on equinox and solstice dates, and average illuminance values were analyzed for key interior spaces (living room, kitchen, and bedroom). ChatGPT-generated plans demonstrated higher spatial clarity and more balanced daylight performance, whereas Copilot outputs varied significantly, and LookX was excluded due to insufficient architectural legibility. Results revealed that none of the models consistently integrated solar orientation or seasonal lighting considerations, indicating a gap between generative representation and environmental logic. The research contributes a replicable workflow that bridges generative AI and performance-based evaluation, offering critical insight into the current limitations and future potential of AI-assisted architectural design. The findings underscore the need for next-generation AI systems capable of semantic, spatial, and climatic reasoning to support environmentally responsive design practices.

Introduction

As the architecture discipline embraces artificial intelligence (AI), designers are increasingly engaging with machine-generated visual content in the early phases of design (Tamke et al., 2018). Among recent breakthroughs, text-to-image (T2I) generation models, such as Stable Diffusion and DALL·E, have gained attention for their ability to produce architectural drawings, forms, and spatial ideas from simple verbal prompts (Turrin et al., 2023). However, their potential to produce environmentally responsive architecture, especially in relation to daylighting, remains an underexplored area.

Daylight plays a critical role in sustainable architecture, directly impacting energy efficiency, thermal comfort, and user well-being (Zhao and Tian, 2023; Nazari and Matusiak, 2024). Passive daylighting strategies differ significantly across climatic regions; thus, contextual responsiveness is essential for performance-oriented design. This inquiry aligns with the paradigm of performative design, wherein architectural intelligence is assessed not only through form but through environmental responsiveness embedded at the generative stage (Oxman, 2007). Yet, it is unclear whether generative AI tools incorporate such considerations when producing plan drawings or spatial arrangements. This study addresses this gap by simulating and analyzing the daylighting performance of AI-generated floor plans in cities with varying climatic conditions. Furthermore, this study contributes to the theoretical debate surrounding design intentionality in machine-generated architecture. It asks not only what AI can produce visually, but also whether those spatial constructs bear any environmental logic, especially in regions where climatic sensitivity is not optional but essential for habitability. By focusing on daylight – a non-negotiable aspect of sustainable housing – this work challenges the optimistic narrative of AI as a neutral design assistant, and instead poses a more fundamental question: Can machines think like architects, or do they merely draw like them?

The rapidly growing interest in AI-assisted generative tools – particularly among urban housing authorities and design professionals – renders the question of environmental adequacy not merely a theoretical concern but a critical issue for early-stage housing design practices. In this study, this inquiry is addressed through a hybrid methodological workflow that goes beyond prompt-based generation alone. By integrating prompt-driven image outputs with AutoCAD-based reconstruction and climate-informed daylight simulation using Velux Daylight Visualizer

(VDV), the research investigates whether AI systems function merely as stylistic generators or exhibit latent performative intelligence embedded within their spatial propositions.

Literature review

T2I generation in architecture

T2I AI tools – driven by diffusion models – enable designers to create visual architectural concepts based on natural language prompts. These models, trained on large datasets of images and captions, have been widely adopted for conceptualization, ideation, and rapid prototyping in architectural design (Hanafy, 2023; Horvath and Poulidou, 2024; Paananen et al., 2024). Research has demonstrated their usefulness in visual storytelling, early-stage façade studies, and atmospheric renderings (Çelik, 2024; Montenegro, 2024). However, critical investigations into their spatial logic, functional accuracy, or environmental responsiveness are limited.

Despite their potential, current T2I systems often prioritize aesthetic coherence over structural realism or performative intelligence (Iranmanesh and Lotfabadi, 2024). Some recent studies (e.g., Liao et al., 2022) suggest combining these outputs with CAD and BIM (Building Information Modeling) tools to evaluate architectural feasibility. This hybrid workflow – prompt to image to simulation – is emerging as a method to assess the generative output's alignment with real-world architectural standards.

Recent work has explored text-to-image generative models (e.g., Midjourney, DALL-E, and Stable Diffusion) as design tools for early-stage architectural ideation. These models allow architects to rapidly visualize multiple concept sketches from written prompts, effectively broadening the range of design options. For example, Paananen et al. (2024) found that using Midjourney, DALL-E, and Stable Diffusion in a classroom exercise supported “serendipitous discovery of ideas and an imaginative mindset,” enriching the concept phase of a cultural center design. Similarly, a recent study reports that T2I tools help architects “conceptualize new architectural ideas more clearly,” providing fresh perspectives and expanding creativity by tapping into large visual datasets (Thampanichwat et al., 2025). In practice, leading design firms already use these AIs for divergent-thinking exploration: Researchers note that models like Midjourney, DALL-E, and Stable Diffusion “promote rapid exploration and iteration through visualization, enabling designers to better express their design concepts” (Chen et al., 2025). These findings suggest that text-based image generators can become meaningful parts of the design workflow, especially when prompts and constraints are carefully managed (although attention must be paid to their limitations).

Daylight simulation in architectural design

Daylight simulation tools, including VDV, Radiance, and ClimateStudio, have long been utilized to predict and optimize natural light penetration within built environments (Kim and Chung, 2011; Sancho-Salas et al., 2023). These tools evaluate daylight autonomy, uniformity, and glare to guide decisions on window orientation, room depth, material reflectance, and spatial configuration (Cammarano et al., 2015; Gibson and Krarti, 2015). Among them, VDV provides a user-friendly interface and accurate simulation outputs based on geographic data and seasonal solar positions.

Computational modeling and simulation play a critical role in evaluating the performance of architectural designs. When these

tools are embedded within the design workflow, the process is referred to as performance-based design. Although rapid feedback loops are essential throughout design development, conducting daylight simulations remains a complex task due to their intensive computational demands and time-consuming nature (Queiroz et al., 2024). Performance-based design using such tools is standard practice in passive design strategies, especially in climates where lighting energy demand is critical (Oxman, 2007). However, their application to AI-generated content is novel. No known study has systematically simulated T2I-generated plans for daylighting performance across multiple climate zones – a gap this research seeks to address.

Daylight performance is now usually assessed with dynamic, climate-based simulation rather than static rules of thumb. In the past, simple metrics like the Daylight Factor under a single overcast sky were common; however, climate-based daylight modeling (CBDM) has gained prominence (Jiang, 2021). CBDM uses hourly weather data (latitude, climate, orientation, etc.) to predict interior luminance over the year, accounting for solar geometry and sky variability. This approach “has attracted considerable attention” and is gradually replacing fixed-condition methods (Jiang, 2021). Accordingly, researchers have adopted both advanced simulation engines (Radiance and Daysim) and industry tools like the VDV. VDV 3.0 (a free Radiance-based program) is widely used for building-scale analyses. For instance, Mandala et al. (2021) used VDV 3.0 to model a large-volume building with various skylight designs and computed daylight factors and uniformity across the space.

In practice, standards often call for checking daylight at equinox conditions to represent average seasonal performance. For example, LEED (Leadership in Energy and Environmental Design) lighting rules instruct teams to simulate illuminance at 9:00 and 15:00 h on March 21 (or September 21) to gauge daylight availability. In one study, a classroom testbed in India was modeled in VDV 3.0 and simulated on March 21 (equinox) from 8:00 to 18:00 h; comparing modeled results to on-site measurements showed that using an accurate sky model (based on prevailing weather data) improved prediction of work-plane illuminance by about 24% (Edwards and Torcellini, 2002; Boyce, 2004). Such studies illustrate that software like VDV can reliably compute time-resolved illuminance for design evaluation. Thus, the literature emphasizes climate-based, hourly simulation (e.g., Spatial Daylight Autonomy and Useful Daylight Illuminance) and the use of realistic sky models and metrics, rather than single-point calculations.

AI-driven climate-based housing design

Several recent projects have experimented with AI-generated architectural solutions tailored to climatic demands. Recent research has highlighted the potential of AI and data mining techniques in addressing the challenges of energy poverty (EP), particularly in warm climate zones. According to Bienvenido-Huertas et al. (2023), AI-based automated tools can effectively detect EP without the need for detailed energy performance analyses. This approach not only streamlines the workload of energy managers and social workers but also enhances the predictive capacity of EP diagnosis, paving the way for more efficient and scalable interventions.

In the context of smart environmental management, the integration of AI with Internet of Things technologies offers promising advancements in greenhouse monitoring systems. As highlighted by Riskiawan et al. (2024), traditional greenhouse environments still largely depend on manual regulation of temperature and humidity, which imposes labor-intensive demands. Their study

demonstrates that AI-enabled automated environmental control significantly enhances the precision and efficiency of indoor climate regulation. The system's capacity to autonomously predict and adjust environmental conditions marks a substantial shift toward intelligent and sustainable greenhouse practices.

The intersection of generative AI and environmental simulation presents new opportunities to evaluate AI's design "intelligence." This article builds upon emerging discourse by combining T2I models with simulation software to assess the daylight performance of climate-specific house plans. It also contributes to the methodological expansion of architectural research by incorporating computational workflows that link creativity and performance.

Researchers are beginning to combine AI generative design with climate data to create performance-aware housing layouts. Generative AI can produce novel floorplan alternatives, but climate adaptation requires coupling generation with environmental evaluation. Recent reviews highlight that many existing AI floorplan tools implicitly depend on local weather data and, therefore, yield solutions tailored to specific regions (Meselhy and Almalkawi, 2025). For true regional adaptation, climate must be explicitly integrated: Different climate zones impose different design priorities (e.g., maximizing shading and minimizing solar gain in hot-arid regions versus enhancing natural ventilation and passive heating in temperate zones) (Meselhy and Almalkawi, 2025). Meselhy and Almalkawi (2025) note that without incorporating these factors, a design optimized for one climate "may not be directly applicable elsewhere" unless adapted to new climate inputs (Meselhy and Almalkawi, 2025).

In response, some recent studies use AI pipelines that generate floor plans and simultaneously predict performance metrics from climate data. For example, Hu et al. (2024) developed a workflow where a diffusion model generates a variety of residential floor plan layouts, and a neural network (GAN, Generative Adversarial Network) rapidly predicts each layout's daylight performance (Hu et al., 2024). They fine-tuned the model on regionally realistic house plans and then used the GAN as a surrogate to evaluate daylight autonomy. Their AI workflow reproduced daylight simulation results with high fidelity (within ~5% error of ground-truth Radiance simulations) and ran over 200× faster than traditional iteration (Hu et al., 2024). This demonstrates that AI-generated designs can be quantitatively assessed for daylight even at an early stage. Other efforts similarly integrate climate: for example, surrogate modeling studies show that including detailed hourly weather features (temperature, solar angles, etc.) improves prediction accuracy across diverse climate zones (Manmatharasan et al., 2025) (suggesting future AI tools should ingest location-specific climate datasets). In sum, the literature indicates that AI methods for housing design are starting to account for environmental performance: Generative models propose geometry, then machine-learning predictors (or embedded simulations) assess metrics like daylight, allowing architects to iteratively refine climate-adapted designs.

Methodology

Research design

This study adopts a mixed-methods design combining generative AI modeling, digital drafting, and environmental simulation to evaluate the daylighting performance of AI-generated sustainable housing plans across diverse climate zones. The workflow is structured into three main phases: (1) Plan generation via T2I diffusion models, (2) plan reconstruction in AutoCAD for simulation

compatibility, and (3) daylight analysis using VDV based on equinox conditions. To ensure methodological clarity, the workflow is further divided into seven stages, including climate zone and city selection, prompt engineering and AI tool evaluation, two- (2D) and three-dimensional (3D) modeling processes, and multi-date daylight simulations using false color mapping. Evaluation criteria are based on average illuminance values (lux) in key interior spaces, benchmarked against recognized standards from the literature. The comparative analysis highlights the extent to which generative models consider passive lighting design principles during the early stages of architectural production (Table 1).

Table 1. Workflow stages of the AI-based climate-sensitive housing study

Step	Description
1	Climate zone and city selection Five cities were selected to represent distinct Köppen–Geiger climate zones (Peel et al., 2007), based on criteria such as climatic diversity, geographical distribution, and urban relevance.
2	Prompt engineering and AI tools Prompts included spatial program details (one bedroom, one kitchen, and one living room) and climate-adaptive strategies (ventilation, shading, and biophilia). AI tools used: • ChatGPT (OpenAI DALL-E 3) for both prompt engineering and rendering (OpenAI, 2025) • Microsoft Copilot Designer – general-purpose, accessible AI tool (Copilot, 2025) • LookX AI – architecture-specific model (Lookx, 2025) *LookX was excluded due to a lack of window/opening outputs. Copilot results for two cities were also excluded. Only valid ChatGPT and Copilot outputs were used.
3	2D CAD drafting (AutoCAD) • External wall thickness: 20 cm • Internal walls standardized to 10 or 15 cm based on AI output • Doors/windows placed according to AI-generated imagery • Plans scaled to 1:50, with north orientation preserved
4	3D modeling (AutoCAD) • 3D geometry developed for each plan • Separate layers assigned for floors, walls, and glazing • Layering ensured correct material assignment in Velux Daylight Visualizer
5	Daylight simulation (Velux Daylight Visualizer) • Realistic reflectance values assigned to materials • Simulations conducted on four key solar dates: Spring equinox (March 21), summer solstice (June 21), autumn equinox (September 23), and winter solstice (December 21) • Visual outputs are displayed using false color mapping
6	Evaluation criteria • Average illuminance (lux) measured for the living room, kitchen, and bedroom • Daylight performance evaluated based on established thresholds (Mardaljevic, 2000; Reinhart and LoVerso, 2010)
7	Comparative analysis • Performance results compared across different AI-generated floor plans *Key insight: Evaluated whether AI models inherently consider daylight in the generative design process

Climate zone and city selection

To ensure climatic diversity, five cities were selected from distinct Köppen-Geiger climate zones (Peel et al., 2007):

- **Jakarta, Indonesia (Af)** – Tropical rainforest climate
- **Alice Springs, Australia (BWh)** – Hot desert climate
- **Madrid, Spain (Csa)** – Warm-summer Mediterranean climate
- **Winnipeg, Canada (Dfb)** – Cold continental climate
- **Tromsø, Norway (ET)** – Polar tundra climate

The selection criteria included the following: (a) Representativeness of major global climatic typologies, (b) urban settings with housing demand, and (c) availability of geographic and solar data for simulation.

To ensure comparability across climate zones and isolate the influence of environmental context on model behavior, a standardized prompt was used for each location. This decision was informed by the need to minimize confounding variables arising from prompt variation, which can significantly affect generative outputs. The single-prompt approach enabled a more controlled assessment of model performance under varying climatic conditions.

Prompt design and AI model selection

Architectural floor plans were generated using leading open-source T2I diffusion models through high-resolution inference, incorporating architecture-specific fine-tuning. In this study, three distinct AI tools were selected to generate sustainable housing plans tailored to various climate zones: ChatGPT (OpenAI), Microsoft Copilot Image Creator, and LookX AI. Each tool offers unique advantages in architectural plan generation, and their selection was based on criteria such as accessibility, architectural specialization, and technical capability.

- **ChatGPT (OpenAI):** Developed by OpenAI, ChatGPT integrates the DALL·E 3 image generation model within the GPT-4o multimodal framework. This enables T2I conversion based on user-defined prompts. In this study, ChatGPT served a dual role as both a prompt engineering assistant and a generative model, producing visual outputs derived from its own textual descriptions. The multimodal capacity of GPT-4o allows seamless interaction between textual and visual content, facilitating the generation of high-quality, context-sensitive architectural images.
- **Microsoft Copilot Image Creator:** Offered through Microsoft's Copilot platform, this tool utilizes Bing Image Creator technology to transform textual prompts into photorealistic images. Distinguished by its free and open-access availability, user-friendly interface, and rapid rendering capability, Copilot efficiently processes natural language inputs to produce high-quality visuals, making it an accessible solution for early-stage generative design tasks.
- **LookX AI:** Specifically designed for architecture and interior design applications, LookX AI converts both text-based and sketch-based inputs into high-resolution renderings. The platform caters to design professionals by enabling custom model training and stylistic output generation. By streamlining the integration of AI into the architectural design workflow, LookX supports creative exploration and iterative visualization in professional contexts.

Each city was assigned a tailored prompt that included:

- climate-adaptive design features (e.g., elevated structures, thermal mass, insulation, and glazing orientation),

- material specifications (e.g., bamboo, adobe, timber, and concrete),
- spatial configuration (one bedroom, one kitchen, and one living room),
- bioclimatic and passive design principles (e.g., cross-ventilation, shading devices, and courtyard inclusion).

Prompt engineering focused on maximizing climatic specificity and functional clarity to encourage AI-generated layouts with architectural intent. Below are the final versions of the climate-based prompts used in this study:

• Jakarta, Indonesia (Tropical climate):

Tropical climate sustainable house plan in Jakarta, Indonesia. Elevated structure with natural ventilation. Large operable windows for cross-ventilation, positioned to maximize daylight. Open floor plan with semi-outdoor living spaces. Locally sourced materials such as bamboo, teak, and recycled wood. Biophilic design with lush greenery, internal courtyards, and shading devices. Floor plan must include one bedroom, one kitchen, one living room, and one toilet. The design must prioritize airflow, cooling efficiency, and sustainable urban living.

• Alice Springs, Australia (Desert climate):

Desert climate sustainable house plan in Alice Springs, Australia. Thick adobe or rammed earth walls for thermal mass. Small, strategically placed windows with shading devices to reduce heat gain. Courtyard layout to allow passive cooling and outdoor living. Minimal openings on west-facing walls. Roof insulation and solar panels integrated. The floor plan must include one bedroom, one kitchen, one living room, and one toilet. Locally sourced materials and efficient water usage features should be included.

• Madrid, Spain (Mediterranean climate):

Mediterranean climate sustainable house plan in Madrid, Spain. Thick masonry walls for thermal stability. Shaded verandas and balconies to manage solar gain. Cross-ventilation through aligned operable windows. Sloped tile roof for seasonal adaptability. Local stone and timber used. One-bedroom, one-kitchen, one-living room, one-toilet layout with semi-open spaces like patios or loggias.

• Winnipeg, Canada (Cold continental climate):

Cold climate sustainable house plan in Winnipeg, Canada. Compact form with high insulation levels. Triple-glazed south-facing windows to maximize solar heat gain. Minimal window openings on north walls. Airtight construction with passive solar principles. Locally sourced timber and insulated concrete. One-bedroom, one-kitchen, one-living room, and one-toilet layout. Enclosed vestibule to prevent heat loss.

• Tromsø, Norway (Polar climate):

Polar climate sustainable house plan in Tromsø, Norway. Elevated foundation to address snow buildup. Super-insulated envelope with minimal thermal bridging. South-facing windows with deep frames. Airtight design with mechanical ventilation and heat recovery. Compact one-bedroom, one-kitchen, one-living room, one-toilet layout. Use of sustainable local timber and snow-shedding roof form.

These prompts served as standardized instructions to assess whether AI-generated plans integrated region-specific climate strategies and spatial logic.

The floor plans generated by the three selected AI tools were developed based on previously defined climate-specific prompts. These outputs were subsequently transferred to the AutoCAD environment and subjected to daylight performance simulations using VDV. This workflow enabled a comparative assessment of the extent to which each AI model integrated climate-adaptive design strategies into spatial planning.

However, not all outputs produced by the AI tools were equally suitable for architectural drafting. The results generated by LookX AI were excluded from the CAD modeling phase due to the absence of windows or other fenestration elements, as well as poor legibility from a plan-reading perspective. Similarly, image outputs for two cities produced by Microsoft Copilot Image Creator were deemed unfit for AutoCAD translation, as they lacked architectural clarity and instead exhibited highly abstract forms that could not be interpreted as floor plans. Consequently, the CAD drafting and daylight simulation stages of the study were carried out exclusively using the plans generated by ChatGPT (OpenAI) and valid outputs from Microsoft Copilot. This selective process underscores the importance of evaluating the consistency and architectural usability of AI-generated floor plans, particularly in climate-responsive design contexts.

While this study focused on three AI models selected for their accessibility and architecture relevance (ChatGPT/DALL-E, Microsoft Copilot, and LookX AI), it must be acknowledged that the rapidly expanding ecosystem of generative tools, including Midjourney and Stable Diffusion, offers additional capabilities that may lead to different architectural outcomes. The limited model diversity is thus recognized as a methodological limitation.

Plan extraction and CAD translation

Since AI-generated images were raster-based and lacked precise metric information, all floor plans were manually redrawn and vectorized using AutoCAD 2025 to ensure simulation compatibility and dimensional consistency. The translation process followed these standard conventions:

- Wall thicknesses were standardized across all plans as follows: 20 cm for external walls, and 15 cm or 10 cm for internal partitions, depending on the wall typology suggested in the AI output.
- Window and door placements were preserved as visible in the AI-generated plans to maintain spatial intent and ventilation logic.
- Each plan was scaled to 1:50, and north orientation was maintained.

This CAD translation process enabled consistent geometric input for subsequent daylight simulation and comparative analysis across different climate-adaptive designs.

Following the drafting phase, a 3D model of each plan was constructed in AutoCAD to prepare for daylight simulation. In this stage:

- Architectural elements, such as floors, walls, ceilings, windows, and doors, were assigned to separate layers.
- This stratification allowed for accurate material assignment in the VDV environment, as the simulation software interprets geometry and materials based on imported CAD layer configurations.

This modeling step provided the necessary semantic and geometric clarity for translating AI-generated architectural intent into scientifically valid simulation scenarios.

Daylight simulation with VDV

After 3D modeling in AutoCAD, each plan was exported to VDV (version 3.0) to assess natural light performance under varied solar conditions. VDV is a validated daylighting analysis tool based on the Radiance rendering engine. While a range of daylight simulation engines exists, such as Radiance, DIVA, and Honeybee, VDV was selected due to its intuitive user interface, high compatibility with AutoCAD models, and suitability for residential-scale

analysis. Although not as customizable as Radiance-based engines, it enables efficient and replicable simulations for comparative purposes within the scope of this study.

Simulations were conducted using Parameters included:

- **Location settings:** City-specific geographic coordinates (latitude and longitude)
- **Time of year:** March 21 and September 21 (equinoxes), June 21 (summer solstice), and December 21 (winter solstice)
- **Sky model:** CIE overcast and intermediate sky types
- **Simulation outputs:** Illuminance levels (lux)
- **Material reflectance values:** Assigned based on realistic assumptions about locally sourced materials inferred from the AI prompt. For example:
 - **Walls:** Reflectance of 0.50–0.65 depending on surface finish (e.g., wood, adobe, and concrete)
 - **Floors:** Reflectance of 0.30–0.45
 - **Windows:** Transparent glazing with visible transmittance of 0.70
 - **Ceilings and shading elements:** Matte materials with reflectance below 0.25, where applicable

The 3D models were prepared with layered materials in AutoCAD to facilitate proper material mapping within Velux. Each building element (walls, windows, and floors) was placed on a distinct CAD layer and assigned corresponding material properties upon import.

Analysis parameters included:

- Rendered daylight scenes at 12:00 h (local solar time) for each simulation date,
- Generation of false-color illuminance maps for internal spaces.

This simulation framework enabled the objective evaluation of how well AI-generated plans performed under natural daylight scenarios, revealing whether the T2I model implicitly considered daylight access during design generation.

Evaluation criteria

The daylight performance of each AI-generated housing plan was evaluated based on the average indoor illuminance (lux) values obtained from simulations conducted on equinoxes (March 21 and September 21) and solstices (June 21 and December 21). The analysis focused on three primary functional spaces: the living room, the kitchen, and the bedroom.

According to commonly accepted daylighting performance standards (Illuminating Engineering Society, 2011; Mardaljevic, 2000), the following average illuminance thresholds were applied: 300–500 lux for living rooms, 500–750 lux for kitchens, and 200–300 lux for bedrooms. These benchmarks are widely recognized in lighting design literature and serve as a basis for evaluating the sufficiency of natural light in residential settings.








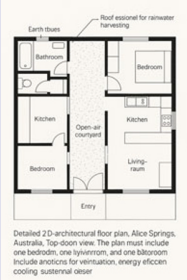





By comparing the simulated illuminance values with these standards, the study aimed to determine whether the AI-generated designs inherently considered adequate daylighting strategies suitable for their respective climatic contexts.

Results

AI-generated plan visualizations










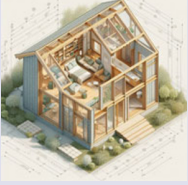


The initial phase of the study involved the generation of architectural floor plans using three different T2I AI models (Table 2): ChatGPT with integrated DALL-E, Microsoft Copilot Image

Table 2. AI-generated plan visualizations

	ChatGPT	Microsoft Copilot Image Creator	LookX AI
Jakarta, Indonesia (tropical climate)		   	 
Alice Springs, Australia (desert climate)	 <p>Detailed 2D architectural floor plan, Alice Springs, Australia, Top-down view. The plan must include one bedroom, one living room, and one bathroom. Include annotations for ventilation, energy efficient cooling, sustainable design.</p>	   	




(Continued)

Table 2. (Continued)

	ChatGPT	Microsoft Copilot Image Creator	LookX AI
Madrid, Spain (Mediterranean climate)		   	
Winnipeg, Canada (cold continental climate)		   	

(Continued)

Table 2. (Continued)

	ChatGPT	Microsoft Copilot Image Creator	LookX AI
Tromsø, Norway (Polar climate)			

Creator, and LookX AI. Each model was prompted with climate-specific design inputs tailored to selected cities representing five distinct Köppen–Geiger climate zones.

Among the tested models, ChatGPT served both as the prompt engineer and as the generator of architectural images via its integrated T2I engine, DALL·E. The model was capable of interpreting complex spatial requirements and producing top-down schematic layouts in 2D, with recognizable architectural components, such as doors, windows, and internal divisions.

Microsoft Copilot Image Creator, an openly accessible, diffusion-based generator, was selected for its transparency and accessibility. It yielded moderately structured architectural layouts for most prompts, although two city-specific outputs (for Winnipeg and Tromsø) lacked the spatial clarity required for plan reconstruction. For each prompt, Microsoft Copilot Image Creator generated four alternative plan images. The most suitable version was manually selected based on architectural legibility, spatial coherence, and the presence of identifiable functional zones, such as living areas, kitchen, bedroom, and sanitary facilities. This selection process aimed to ensure that the chosen output was representative of a plausible housing plan, enabling accurate vector-based reconstruction in AutoCAD and valid daylight simulation in VDV.

LookX AI, despite being marketed as an architecture-specific model, consistently failed to depict critical plan elements, such as fenestration or defined room boundaries. Although LookX AI was





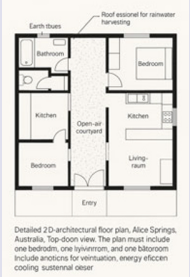
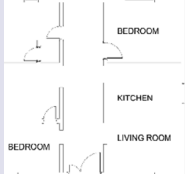
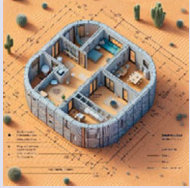

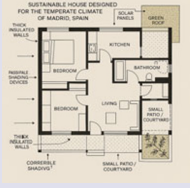







initially included in the experimental setup due to its domain-specific training for architectural visualization, the outputs generated across multiple prompt iterations lacked essential architectural features – particularly visible and distinguishable window openings. Since fenestration plays a vital role in spatial functionality, visual connectivity, and daylight performance (Tregenza and Wilson, 2013; Galasiu and Reinhart, 2008), the absence of such elements compromised the architectural usability of the outputs. According to Chaillou (2022), the geometric accuracy and semantic clarity of AI-generated floor plans are prerequisites for further computational analysis. Therefore, as the outputs from LookX did not meet these fundamental criteria, they were excluded from the AutoCAD reconstruction and simulation workflow.

AutoCAD plan reconstruction

All selected AI-generated plans were digitally reconstructed in AutoCAD (Table 3) to prepare them for daylight simulation. The reconstruction process followed a standardized protocol:

- Exterior walls were drawn with a thickness of 20 cm, while interior walls were set to 10 or 15 cm, depending on the detail level present in the AI outputs.
- Doors and windows were placed in the same positions as depicted in the AI-generated visualizations, preserving orientation and spatial organization.

Table 3. AutoCAD plan reconstruction

Jakarta, Indonesia (tropical climate)	ChatGPT		
	Microsoft Copilot Image Creator		
Alice Springs, Australia (desert climate)	ChatGPT		
	Microsoft Copilot Image Creator		
Madrid, Spain (Mediterranean climate)	ChatGPT		
	Microsoft Copilot Image Creator		
Winnipeg, Canada (cold continental climate)	ChatGPT		
	Microsoft Copilot Image Creator		

(Continued)

Table 3. (Continued)

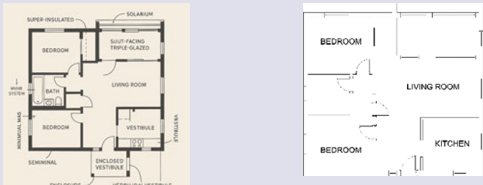








Tromsø, Norway (polar climate)	ChatGPT	
--------------------------------	---------	--

Table 4. 3D models of each AI-driven plan

Jakarta, Indonesia (tropical climate)	ChatGPT	
	Microsoft Copilot Image Creator	
Alice Springs, Australia (desert climate)	ChatGPT	
	Microsoft Copilot Image Creator	
Madrid, Spain (Mediterranean climate)	ChatGPT	
Winnipeg, Canada (cold continental climate)	ChatGPT	
	Microsoft Copilot Image Creator	
Tromsø, Norway (polar climate)	ChatGPT	

- Each drawing was scaled to 1:50 and aligned to true north according to the urban context of the target city.

After generating the 2D plans, a 3D model of each plan was created in AutoCAD using separate layers for each material type (walls, floors, glazing, and roofing) (Table 4). This separation was essential for assigning reflectance properties during the Velux daylight simulation phase.

Daylight simulation results

Using the VDV, daylight performance simulations were conducted for each of the five architectural plans under both equinox (March

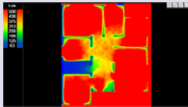
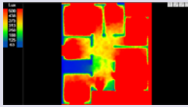
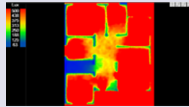
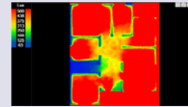
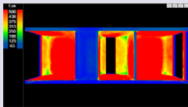
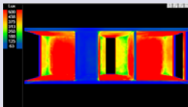
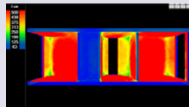
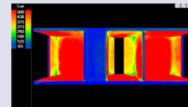
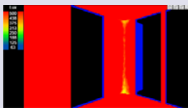
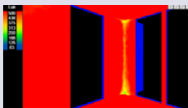
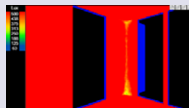
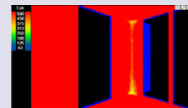
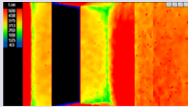
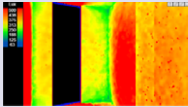
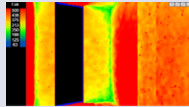
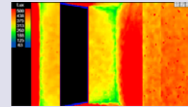
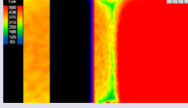
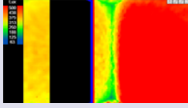
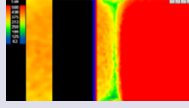
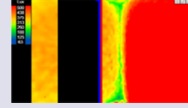
21 and September 21) and solstice (June 21 and December 21) conditions. Material reflectance values were input into the software based on typical surface properties (e.g., white painted wall = 0.85, clear glazing = 0.65, and wooden flooring = 0.35), enhancing the realism of the simulations.

The simulation focused on three primary spaces:

- Living room
- Kitchen
- Bedroom

To analyze the daylight performance of the AI-generated housing plans, all simulations conducted in VDV were visualized using false color mapping (Tables 5–12). This method provides a detailed

Table 5. Daylight simulation results for Jakarta, Indonesia (ChatGPT-generated plan) (Author, 2025)

		Equinox (March 21)	Solstice (June 21)	Equinox (September 21)	Solstice (December 21)
Jakarta, Indonesia (tropical climate) 6.2088° S 106.8456° E ChatGPT	Plan				
	Section				
	Living room				
	Average illuminance (lux)	936.7	823.3	903.1	881.2
	Kitchen				
	Average illuminance (lux)	423	370.2	412.5	403.7
	Bedroom				
	Average illuminance (lux)	567.7	498.4	557	543.1

representation of illuminance distribution across the interior spaces by assigning a gradient of colors to specific lux levels, thereby allowing clear identification of over- or underlit areas. False color rendering is a widely accepted technique in daylighting studies for its effectiveness in visually communicating quantitative lighting data (Ruck, 1986; Andersen et al., 2008).

Evaluation

Architectural evaluation of AI-generated plans

In the initial phase of the study, AI-generated architectural floor plans were subjected to a qualitative assessment within a professional architectural framework. The evaluation was structured around four main criteria. First, spatial organization was examined in terms of how effectively the plans addressed user needs, with particular attention to the placement of rooms, proportionality of spatial dimensions, and the logic of functional relationships (Li et al., 2024). Second, functionality and circulation were assessed by analyzing the adequacy of usable spaces, the clarity of circulation routes, and the overall legibility of the plan. The third criterion focused on climate responsiveness, evaluating whether the design allowed for natural ventilation through appropriate openings and whether climate-specific features, such as wide eaves in tropical regions or compact layouts in polar zones, were evident. Lastly, architectural coherence was

considered, emphasizing the presence of a consistent design logic and the architectural feasibility of the proposed floor plan. This multilayered evaluation framework (Table 13) provides a critical foundation for interrogating the architectural validity of AI-driven design outputs. Although the primary focus of this study was climate adaptation, a supplementary morphological reading was conducted to examine culturally resonant architectural cues. In this context, features such as the raised platform and shaded veranda in Jakarta, the inner courtyard in Alice Springs, and the use of timber-heavy construction in Tromsø can be interpreted as echoes of vernacular typologies, albeit emergent through AI-generated outputs. While these elements may not indicate conscious cultural encoding, they suggest latent representational patterns worthy of further exploration.

Spatial organization

The floor plans generated by ChatGPT demonstrated a relatively coherent level of spatial organization, with discernible functional zoning in cities like Jakarta, Alice Springs, and Madrid. In particular, the Jakarta plan for a tropical climate suggested a plausible spatial sequence between indoor and semi-outdoor areas, aligning with biophilic principles and local climatic conditions. However, proportional inconsistencies were observed in some layouts – most notably in the Tromsø plan – where spatial adjacencies compromised privacy and spatial logic, indicating a prioritization of formal expression over functional rigor.

Table 6. Daylight simulation results for Jakarta, Indonesia (Microsoft Copilot Image Creator-generated plan) (Author, 2025)

		Equinox (March 21)	Solstice (June 21)	Equinox (September 21)	Solstice (December 21)
Jakarta, Indonesia (tropical climate) 6.2088° S 106.8456° E Microsoft Copilot Image Creator	Plan				
	Section				
	Living room				
	Average illuminance (lux)	442.7	388.5	434	425.9
	Kitchen				
	Average illuminance (lux)	1347.8	1051	1182	1168.2
	Bedroom				
	Average illuminance (lux)	496.4	435.9	484.1	474.8

Conversely, plans produced by Microsoft Copilot displayed only moderate spatial intelligibility. While the selected outputs presented visual completeness, their internal spatial relationships often lacked the clarity and responsiveness necessary for residential use. In several instances, the delineation between public and private zones remained ambiguous, reflecting the tool's limited capacity to resolve complex spatial hierarchies.

Functionality and circulation

ChatGPT's outputs exhibited relatively well-defined circulation patterns, with effective transitions between entry zones and primary living spaces, particularly in the Alice Springs and Madrid examples. The logic of spatial sequencing – while not always optimal – was generally comprehensible. However, certain design choices, such as the direct adjacency between the kitchen and the living area in the Jakarta plan, suggested a lack of sensitivity to functional separation and user comfort.

In contrast, the circulation patterns in Microsoft Copilot's plans were less intelligible. For instance, the Alice Springs output lacked clear spatial transitions, making navigation within the layout ambiguous. The AI-generated plans often remained at a conceptual level, lacking the operational logic required for architectural implementation and spatial legibility.

Climate responsiveness (presimulation)

ChatGPT's ability to respond to climate-adaptive design prompts was partially successful. In the Winnipeg plan, for example, the

inclusion of large, south-facing glazed areas and minimal north-facing openings adhered to passive solar design principles. Similarly, in Alice Springs, the introduction of a central courtyard and thick masonry walls suggested a meaningful engagement with hot-arid climate strategies. Nonetheless, these responses remained inconsistent; in the Tromsø plan, the placement and scale of windows did not adequately address the thermal or daylighting challenges of a polar climate.

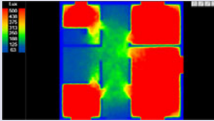
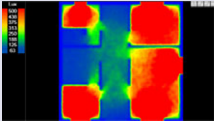
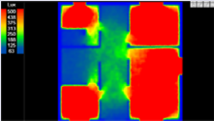
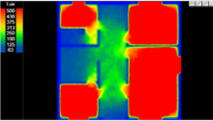
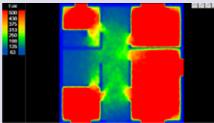
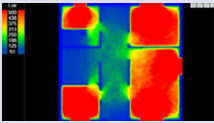
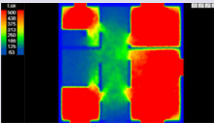
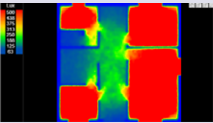
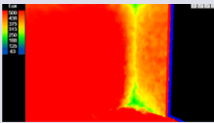
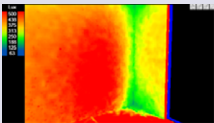
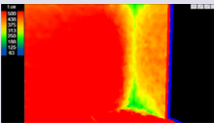
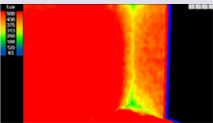
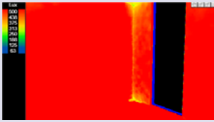
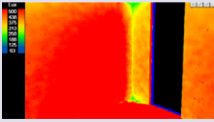
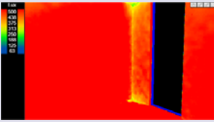
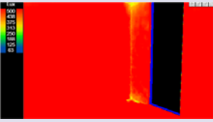
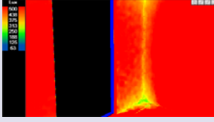
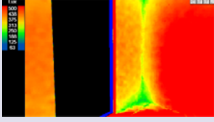
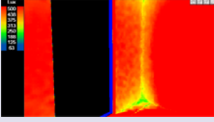
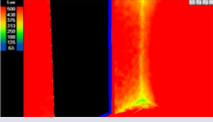
Microsoft Copilot's outputs demonstrated a more superficial engagement with the climate context. While the tropical plans featured large openings, critical passive design strategies, such as orientation control, shading devices, and compact volume manipulation, were inconsistently applied or entirely absent, raising concerns about the tool's comprehension of climatic specificity.

Architectural coherence

From a standpoint of architectural coherence, ChatGPT-generated plans exhibited a higher degree of internal consistency. The Madrid example, in particular, reflected a rational structural system, coherent material logic (e.g., masonry and timber use), and a buildable configuration grounded in local construction practices. These qualities suggest an implicit awareness of tectonic and spatial discipline within the AI's generative logic.

By comparison, Microsoft Copilot's plans, although aesthetically engaging, lacked structural legibility. In several cases, wall continuity and the implied load-bearing structure were either unresolved or implausible, undermining the technical

Table 7. Daylight simulation results for Alice Springs, Australia (ChatGPT-generated plan) (Author, 2025)

		Equinox (March 21)	Solstice (June 21)	Equinox (September 21)	Solstice (December 21)
Alice Springs, Australia (sesert climate) 23.6980° S 133.8807° E ChatGPT	Plan				
	Section				
	Living room				
	Average illuminance (lux)	567	411.5	521	593.2
	Kitchen				
	Average illuminance (lux)	741.7	537.7	685.5	782.3
	Bedroom				
	Average illuminance (lux)	625.4	455.9	573	635.2

feasibility of the architectural proposals. As such, the Copilot outputs appeared more representational than architecturally grounded.

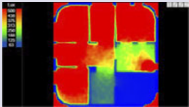
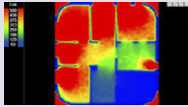
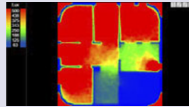
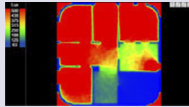
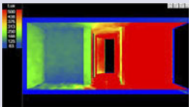
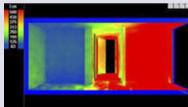
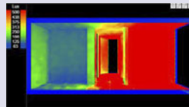
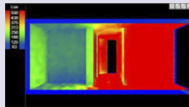
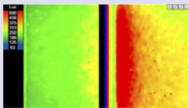
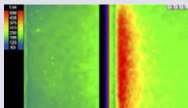
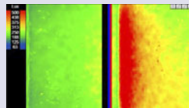
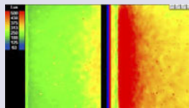
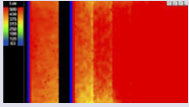
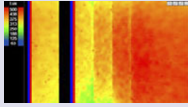
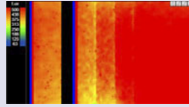
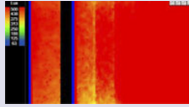

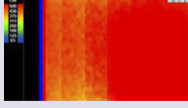
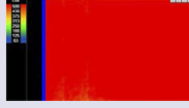

In light of the visualizations presented in Table 2, ChatGPT emerged as the more competent model in terms of spatial articulation, functional hierarchy, climate responsiveness, and architectural buildability. Although Microsoft Copilot succeeded in producing compositionally coherent images, its outputs lacked the structural clarity and environmental sensitivity required for viable architectural design. LookX AI, which was excluded from further analysis due to fundamental representational deficiencies (e.g., absence of fenestration), further underscored the necessity of rigorous architectural criteria when assessing AI-generated content. These findings collectively suggest that while T2I models can support schematic visualization, their capacity to internalize and apply architectural principles – particularly in the context of climate-responsive housing – remains limited and warrants further refinement.

The architectural evaluation was conducted by the lead researcher, who is a licensed architect and associate professor with expertise in design theory and AI-driven generative methods. The assessment followed a structured content analysis protocol based on established architectural criteria (Li et al.,

2024), ensuring consistency and interpretive depth across both AI models. Four core dimensions – spatial organization, functionality and circulation, climate responsiveness, and architectural coherence – were used to comparatively assess the plans produced by ChatGPT and Microsoft Copilot. Each criterion was rated on a five-point scale, where higher scores indicate greater alignment with architectural logic and climate-adaptive design principles. The resulting scores are presented in Table 14 as a synthesized comparison of the two models’ architectural performance. Although the architectural evaluation involved qualitative judgments, the use of predefined criteria, a calibrated scoring system, and a structured comparison framework ensured a high degree of methodological objectivity. The evaluation did not rely on subjective preferences but rather on professional architectural standards and simulation-informed assessments, allowing for replicable and critically grounded interpretations.

In addition to the unweighted qualitative scores (Table 14), a supplementary weighted scoring matrix (Table 15) was introduced to enhance transparency and quantifiability. Criterion-specific weights were assigned based on their relevance to spatial performance, and final scores were computed as weighted sums. This method offers a more structured comparative lens, reaffirming

Table 8. Daylight simulation results for Alice Springs, Australia (Microsoft Copilot Image Creator-generated plan) (Author, 2025)

		Equinox (March 21)	Solstice (June 21)	Equinox (September 21)	Solstice (December 21)
Alice Springs, Australia (desert climate) 23.6980° S 133.8807° E Microsoft Copilot Image Creator	Plan				
	Section				
	Living room				
	Average illuminance (lux)	306.1	223.7	281	322.8
	Kitchen				
	Average illuminance (lux)	547.6	399.6	504.9	576.3
	Bedroom				
	Average illuminance (lux)	708.1	518.4	652.9	747

ChatGPT’s relative strength in spatial reasoning while addressing reviewer concerns regarding subjectivity.

Daylight performance analysis

To assess the daylighting performance of AI-generated architectural plans, simulations were conducted using VDV on four key solar dates {spring equinox [March 21], summer solstice [June 21], autumn equinox [September 21], and winter solstice [December 21]} at 12:00 h local solar time. Each plan was evaluated across three primary functional zones: the living room, the kitchen, and the bedroom. Average illuminance values (lux) were extracted using false color mapping, providing a visual and quantitative understanding of spatial daylight distribution. The performance metric was average illuminance (lux) across these rooms during daytime hours (12:00 h). Results were compared to daylighting benchmarks derived from literature. According to Reinhart and Walkenhorst (2001) and IES *Lighting Handbook* (2011), the thresholds presented in Table 16 were applied.

In tropical and desert climates (Jakarta and Alice Springs), ChatGPT-generated plans generally met or exceeded the recommended daylight thresholds for all rooms, reflecting a relatively robust spatial openness and appropriate window placement. Particularly, the Jakarta plan exhibited an average illuminance of over 900 lux in the living area across equinox dates, suggesting potential

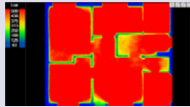
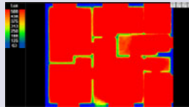
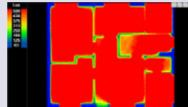
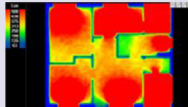
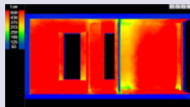
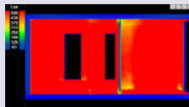
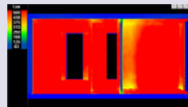
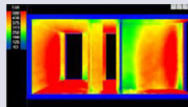
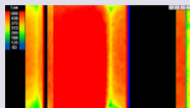
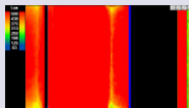
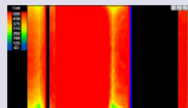
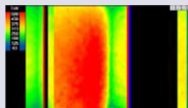

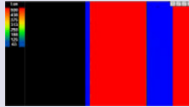

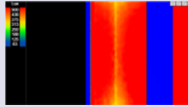



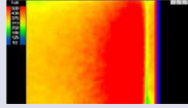
for glare and over-illumination, yet confirming ample daylight availability. In contrast, Copilot outputs in the same locations showed underperformance in the living room (e.g., 306.1 lux in Alice Springs), but surprisingly high kitchen values (e.g., 1347.8 lux in Jakarta), possibly due to exaggerated opening placements or unshaded orientations.

In temperate and cold continental climates (Madrid and Winnipeg), ChatGPT plans performed reliably within the daylight thresholds, although solstice data revealed seasonal drops, especially in the winter months (e.g., kitchen illuminance of 193.5 lux in Winnipeg). Microsoft Copilot’s outputs in Winnipeg revealed very high illuminance levels in summer (e.g., 1740.6 lux in the living room), indicating limited passive shading logic and poor modulation of seasonal exposure.

In Tromsø’s polar context, ChatGPT’s floor plan exhibited critically low daylight performance in winter (e.g., only 9.6 lux in the living room), aligning with extreme daylight limitations at high latitudes. Even during the equinox, results hovered below minimum standards, suggesting the need for active daylighting systems or architectural adaptations like larger glazed surfaces or light-reflecting interiors.

Comparative analysis across all plans highlights that while some AI-generated layouts coincidentally met daylighting standards, these successes were not consistently climate-responsive or seasonally calibrated. In particular, Copilot’s outputs showed

Table 9. Daylight simulation results for Madrid, Spain (ChatGPT-generated plan) (Author, 2025)

		Equinox (March 21)	Solstice (June 21)	Equinox (September 21)	Solstice (December 21)
Madrid, Spain (Mediterranean climate) 40.4168° N 3.7038° W ChatGPT	Plan				
	Section				
	Living room				
	Average illuminance (lux)	518.7	702	607.3	339.9
	Kitchen				
	Average illuminance (lux)	665.3	898.4	776.5	434.8
	Bedroom				
	Average illuminance (lux)	655	883.3	765.5	427.4

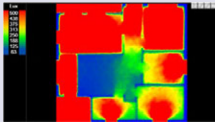
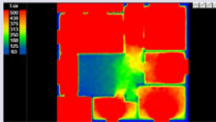
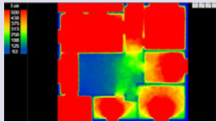
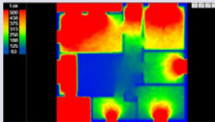
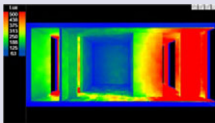
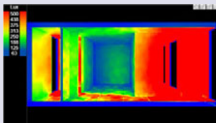
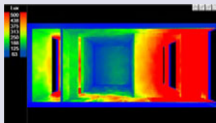
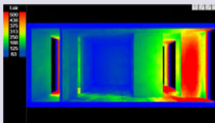
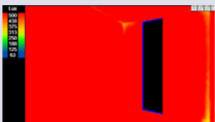


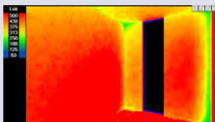
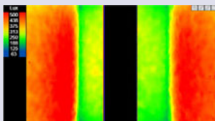
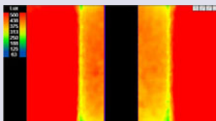
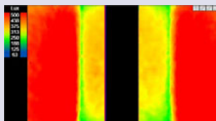
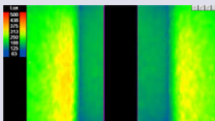
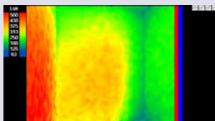
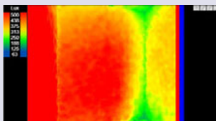
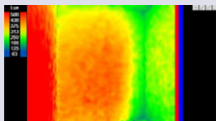
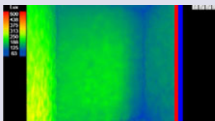
erratic performance patterns, with certain spaces far exceeding thresholds while others failed to meet minimum requirements. This inconsistency suggests that generative AI tools, despite interpreting prompt keywords like “natural light” or “cross-ventilation,” do not yet possess a functional understanding of solar geometry or daylight performance across seasons.

Figures 1–3 present a comparative analysis of average daylight illuminance levels across three critical residential spaces (living room, kitchen, and bedroom) based on AI-generated plans from ChatGPT and Microsoft Copilot. Overall, ChatGPT-generated layouts exhibited more consistent performance within the recommended daylight thresholds defined by Illuminating Engineering Society (2011), particularly in temperate and tropical climates. In kitchens, both AI tools produced sufficiently high illuminance values, often exceeding the 500 lux threshold; however, Copilot outputs occasionally demonstrated overexposure, suggesting a lack of refined daylight modulation. For bedrooms, where the target range is lower (200–300 lux), most AI outputs, especially from ChatGPT, surpassed minimum requirements but also risked excessive lighting in equatorial and desert contexts. The graphical comparisons underscore the variability in daylight response between models and highlight the need for postgenerative calibration when employing AI-driven plans in performance-sensitive design workflows.

Overall, ChatGPT-based plans demonstrated greater coherence in spatial logic and lighting performance, although even these required careful adjustment to mitigate overexposure or insufficient daylight during certain periods. The findings emphasize that current T2I AI tools can generate plausible architectural layouts, but their capacity to produce climate- and daylight-aware design solutions remains limited without postgenerative environmental refinement.

A synthesis of the findings (Table 17 and Figure 4) reveals that ChatGPT demonstrates a marked superiority in architectural criteria, particularly in terms of *Architectural Coherence* and *Spatial Organization*. However, its performance in daylight compliance remains limited, indicating a strong capacity for formal composition but a notable deficiency in meeting illumination thresholds. In contrast, Microsoft Copilot exhibits weaker architectural performance, yet achieves comparatively higher scores in *Daylight Compliance*, particularly in certain climate-space configurations. The radar chart (Figure 4) visually reinforces these observations, highlighting ChatGPT’s architectural advantage while simultaneously underscoring the inconsistency of both models in climate-responsive daylight performance. This suggests that while generative models excel in visual form-making, their output remains insufficiently informed by performance-oriented design logic.

Table 10. Daylight simulation results for Winnipeg, Canada (ChatGPT-generated plan) (Author, 2025)

		Equinox (March 21)	Solstice (June 21)	Equinox (September 21)	Solstice (December 21)
Winnipeg, Canada (cold continental climate) 49.8951° N 97.1384° W ChatGPT	Plan				
	Section				
	Living room				
	Average illuminance (lux)	824	1271.3	1036.2	444.1
	Kitchen				
	Average illuminance (lux)	359.6	551.4	445.2	193.5
	Bedroom				
	Average illuminance (lux)	293.6	449.2	362.3	157.8

Conclusion

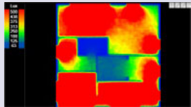
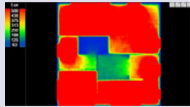
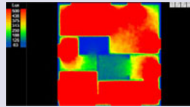
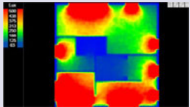
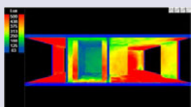
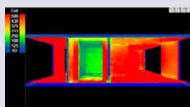
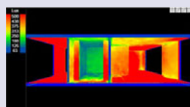
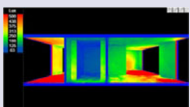


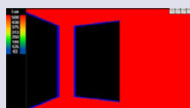
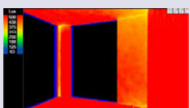
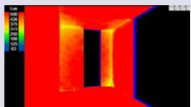


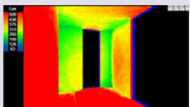
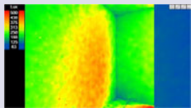
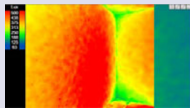
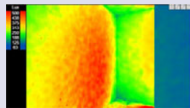
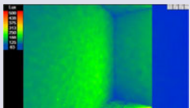
This study critically examined the integration of generative AI tools within the early-stage design process of sustainable housing in diverse climate zones, with a specific focus on daylight performance. By combining T2I diffusion models, CAD-based digital reconstruction, and validated daylight simulation methods, the research offers a novel mixed-methods approach that bridges speculative design generation and performance-based environmental assessment.

The findings reveal that while current AI-driven models, such as ChatGPT (OpenAI) and Microsoft Copilot, are capable of producing visually coherent floor plan representations, their capacity to embed climate-responsive logic, particularly in relation to solar orientation, fenestration, and seasonal daylight modulation, remains inconsistent and largely superficial. ChatGPT-generated plans displayed comparatively higher architectural legibility and more balanced illuminance values across different functional spaces. However, even these outputs required substantial postprocessing for standardization and

simulation compatibility. The exclusion of LookX AI due to its inability to render fundamental spatial features, such as window openings, further underscores the limitations of existing domain-specific AI models in producing technically operable architectural solutions.

From a methodological perspective, the study contributes to the growing discourse on performance-driven generative design by establishing a replicable workflow that spans prompt engineering, architectural visualization, CAD reconstruction, and dynamic simulation. The comparative evaluation across five distinct climate contexts revealed the sensitivity of daylight adequacy not only to spatial configuration but also to AI model behavior, suggesting that generative intent alone does not guarantee environmental suitability. While the use of a single standardized prompt per location enhanced experimental control, future research should explore the effects of prompt variation and repetition. Repeating the experiment with diverse prompts could yield more robust conclusions regarding the stability and generalizability of AI-generated design outcomes across different environmental contexts.

Table 11. Daylight simulation results for Winnipeg, Canada (Microsoft Copilot Image Creator-generated plan) (Author, 2025)

		Equinox (March 21)	Solstice (June 21)	Equinox (September 21)	Solstice (December 21)
Winnipeg, Canada (cold continental climate) 49.8951° N 97.1384° W Microsoft Copilot Image Creator	Plan				
	Section				
	Living room				
	Average illuminance (lux)	1128.8	1740.6	1396.2	614
Kitchen					
	Average illuminance (lux)	854	1311.5	1042.7	458.4
Bedroom					
	Average illuminance (lux)	207.3	317.3	255.4	112.4

It is important to note that the scope of this research was limited to a fixed spatial program and a single set of daylight performance metrics; therefore, future studies should expand the analysis to include more complex building typologies, additional environmental variables (e.g., thermal comfort, glare, and energy loads), and a broader range of AI models. Future research should consider expanding the range of AI models tested to include emerging platforms, such as Midjourney and Stable Diffusion, which exhibit distinct stylistic and spatial affordances. This would allow for a more comprehensive evaluation of generative model behavior in architectural contexts. Furthermore, the simulation process assumed static material reflectance values and did not account for context-specific obstructions or dynamic façade systems, which may affect real-world daylighting outcomes. Addressing these limitations through more comprehensive datasets and real-time AI feedback loops could significantly enhance the environmental fidelity and architectural applicability of generative design outputs. While this study focused exclusively on daylighting performance as a measurable environmental indicator, future research could be extended to include additional parameters, such as natural ventilation potential, thermal gain/

loss, and operational energy demands, particularly for climate-sensitive residential design scenarios.

Future research should explore how generative models might be refined not only through prompt engineering at the input stage but also through performance-based feedback loops that allow for iterative fine-tuning of existing design outputs. Rather than regenerating entire schemes when performance deficits are identified, such as inadequate daylight compliance or poor spatial organization, models could be steered incrementally using evaluative feedback to adjust form and layout. This would align with emerging research in reinforcement learning and adaptive conditioning in AI-assisted architectural design, paving the way for more intelligent, environmentally responsive generative systems.

Ultimately, this research calls for the development of next-generation AI systems that move beyond aesthetic speculation and toward semantically aware, performance-informed architectural reasoning. Such advancements will be essential if generative tools are to support architects meaningfully in addressing pressing challenges such as climate adaptation, daylight sufficiency, and energy equity in the built environment.

Table 12. Daylight simulation results for Tromsø, Norway (ChatGPT-generated plan) (Author, 2025)

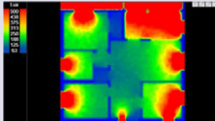
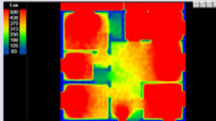
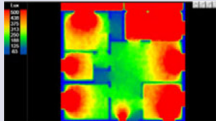

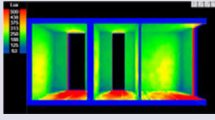
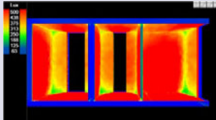
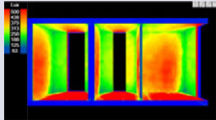

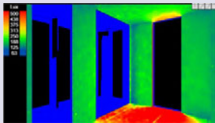
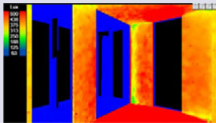
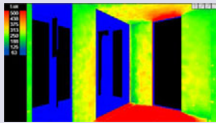
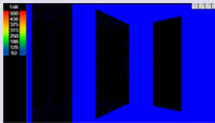
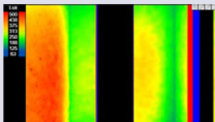

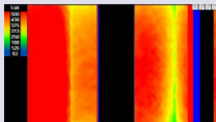

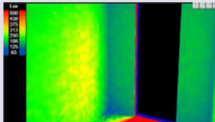
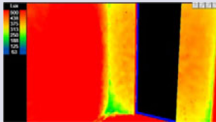
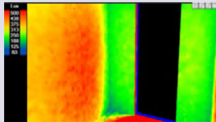

		Equinox (March 21)	Solstice (June 21)	Equinox (September 21)	Solstice (December 21)
Tromsø, Norway (polar climate) 69.6496° N 18.9560° E ChatGPT	Plan				
	Section				
	Living room				
	Average illuminance (lux)	145.3	357.6	233	9.6
	Kitchen				
	Average illuminance (lux)	290	733	472.6	21
	Bedroom				
	Average illuminance (lux)	190.3	482.2	307.8	14.3

Table 13. Multilayered evaluation framework

Criterion	Evaluation questions	Key considerations
a. Spatial organization	Do the spatial arrangements address user needs effectively?	Logical room layout, appropriate room sizes, and functional relationships (Li et al., 2024)
b. Functionality and circulation	Are the usable areas sufficient? Are circulation paths clear? Is the plan legible?	Clear zoning, unambiguous movement flow, and ease of understanding the plan
c. Climate responsiveness (presimulation)	Are openings positioned to enable natural ventilation? Are there climate-specific design features?	Wide eaves (tropical), compact forms (polar), orientation, and passive design strategies
d. Architectural coherence	Does the design exhibit a unified architectural concept? Is it technically buildable?	Consistency in design logic and realistic structural and spatial configuration

Table 14. Comparative scoring of AI models based on architectural evaluation criteria

Evaluation criteria	ChatGPT (OpenAI)	Microsoft copilot
A. Spatial organization	4.0/5	2.5/5
B. Functionality and circulation	3.5/5	2.0/5
C. Climate responsiveness	4.0/5	2.0/5
D. Architectural coherence	4.5/5	2.5/5
Total (average score)	4.0/5	2.25/5

Scoring legend (1–5 scale):

1 = Very poor: No architectural value or responsiveness.

2 = Poor: Inadequate spatial or climatic logic.

3 = Fair: Acceptable but limited architectural consistency.

4 = Good: Functionally and environmentally thoughtful.

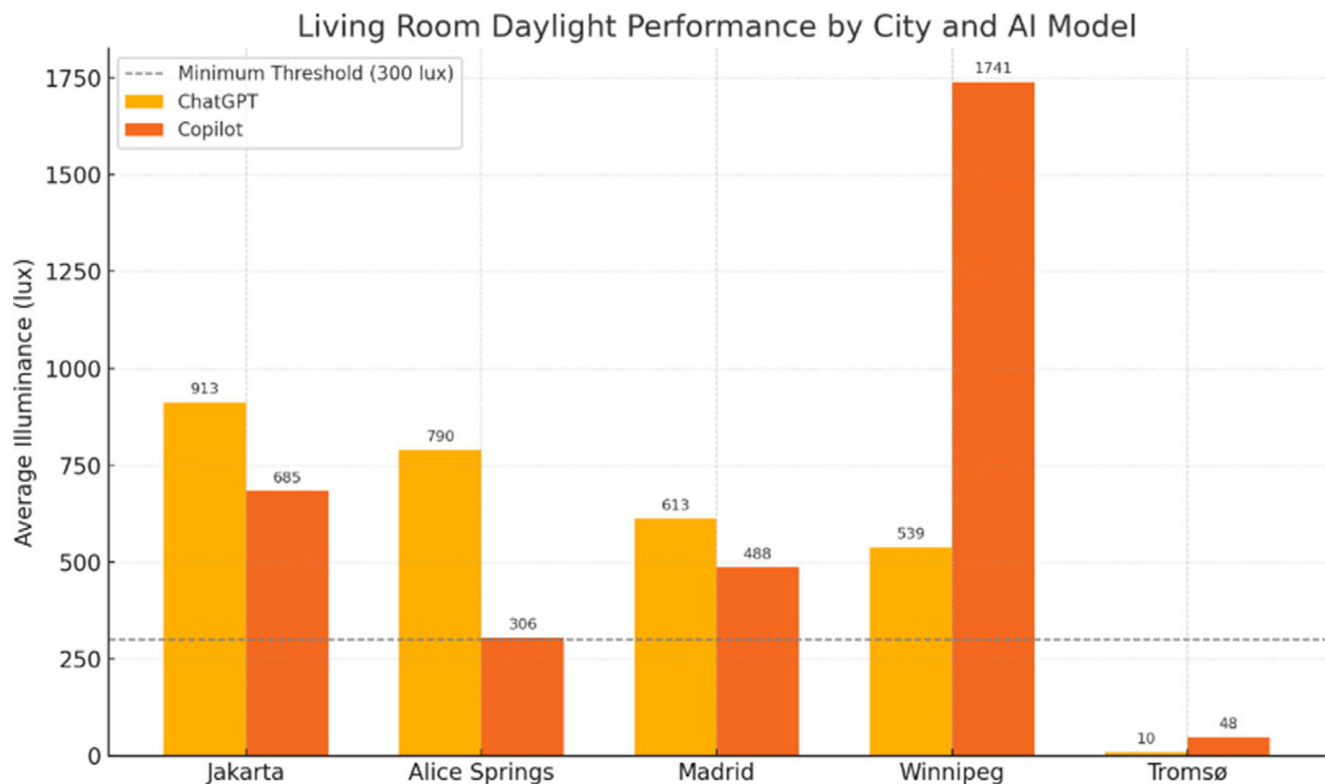
5 = Excellent: Highly coherent, climate-adaptive, and buildable.

Table 16. Daylight thresholds

Room	Recommended average illuminance
Living room	300–500 lux
Kitchen	500–750 lux
Bedroom	200–300 lux

Table 15. Weighted architectural evaluation matrix

Criteria	Weight	ChatGPT score	Copilot score	ChatGPT weighted	Copilot weighted
Spatial clarity	0.30	4	3	1.20	0.90
Functional hierarchy	0.25	3	2	0.75	0.50
Circulation logic	0.20	4	3	0.80	0.60
Environmental responsiveness	0.25	3	2	0.75	0.50
Final weighted score	1.00	–	–	3.50	2.50

**Figure 1.** Living room daylight performance by city and AI model.

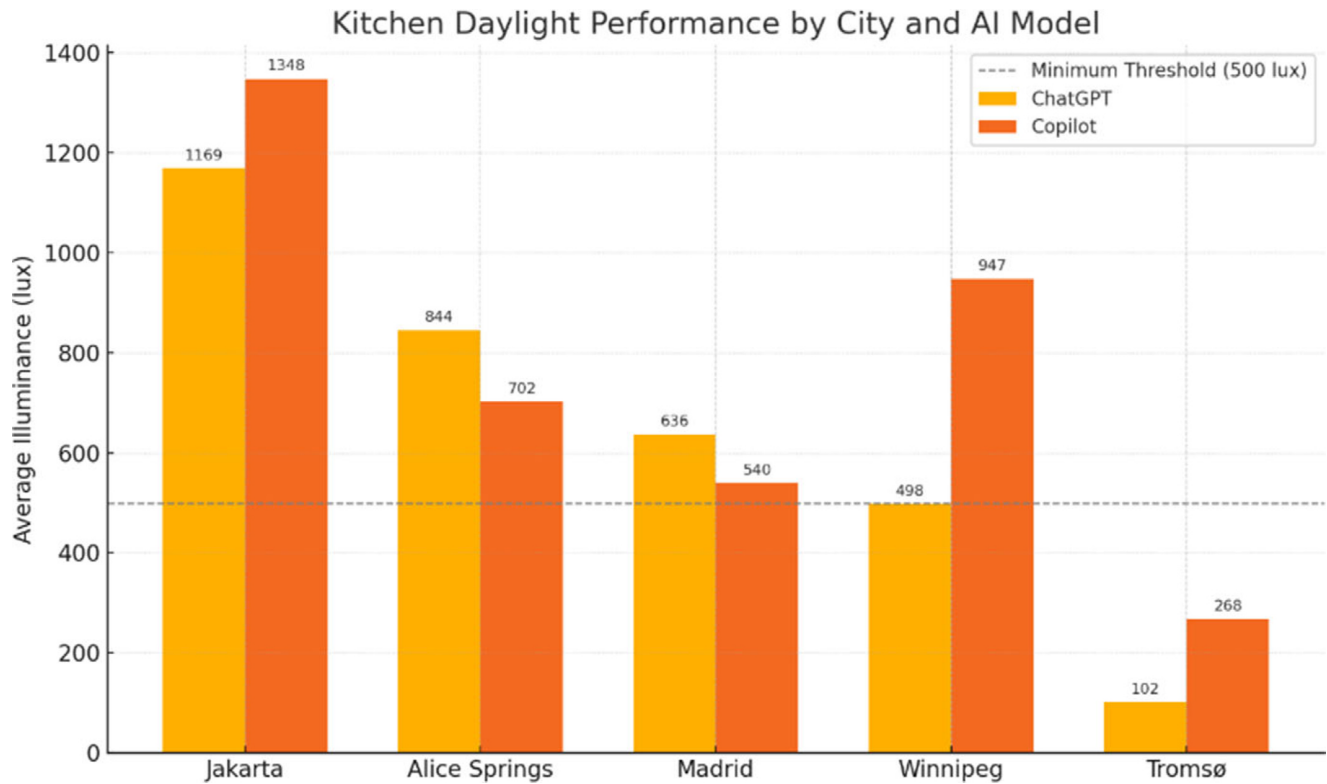


Figure 2. Kitchen daylight performance by city and AI model.

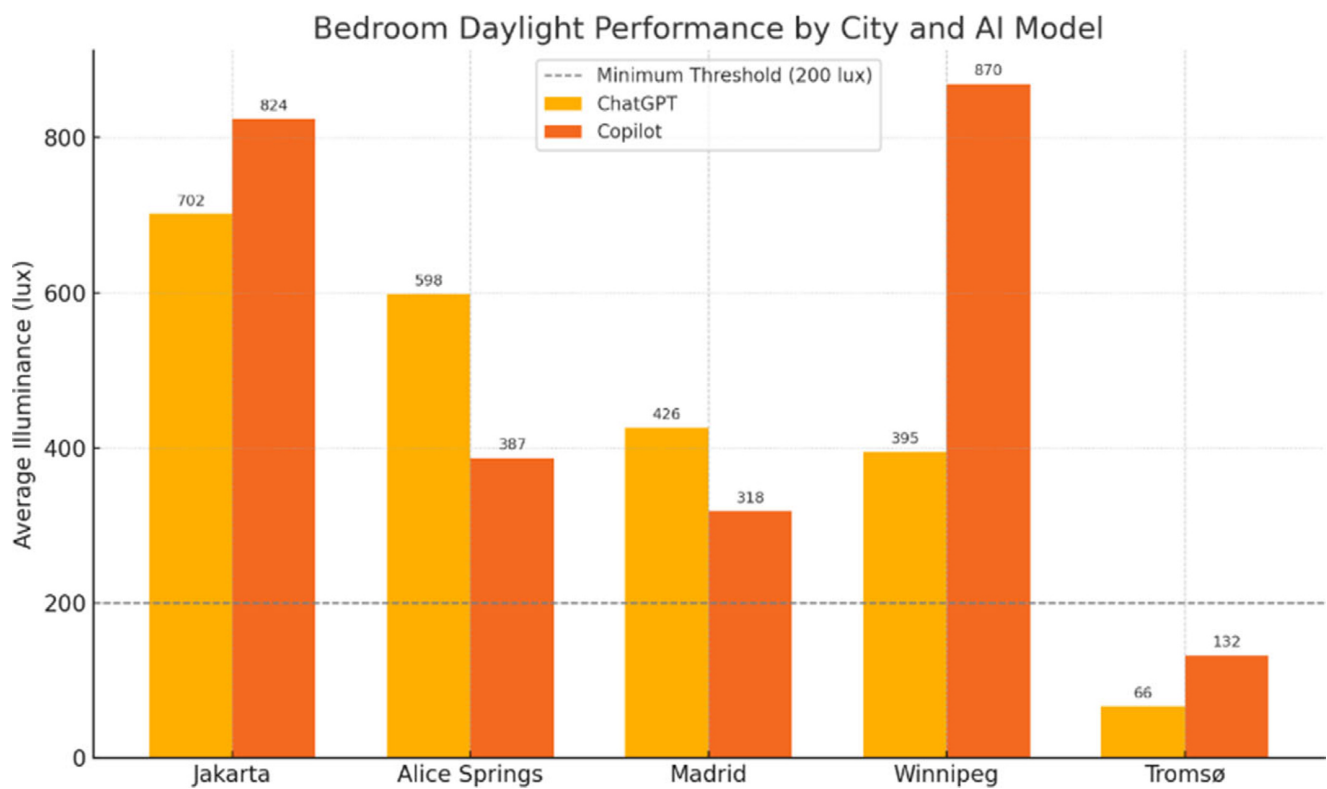


Figure 3. Bedroom daylight performance by city and AI model.

Table 17. Summary comparative chart of architectural and environmental performance

Axis	Scale	Source
Spatial organization	1–5	Table 14
Functionality and circulation	1–5	Table 14
Climate responsiveness	1–5	Table 14
Architectural coherence	1–5	Table 14
Daylight compliance	0–5 (normalized) ^a	Tables 5–12

^aThe number of spaces within the recommended daylight thresholds was proportionally scaled from 0 to 1, and then linearly converted to a 0–5 range.

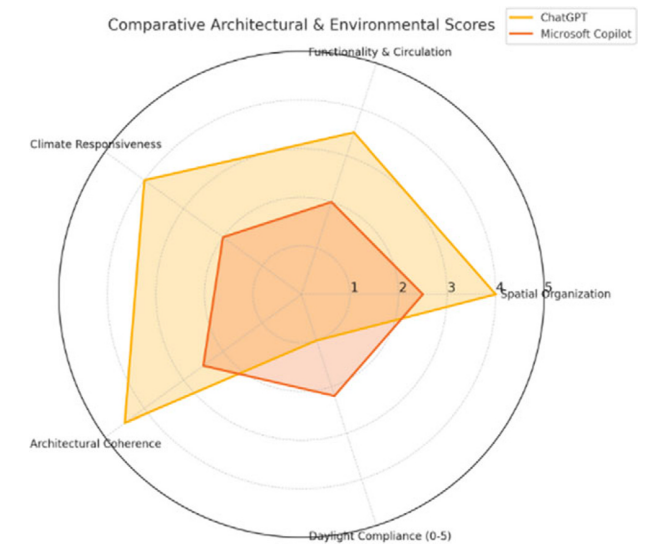


Figure 4. Summary comparative chart of architectural and environmental findings.

Acknowledgments. The author gratefully acknowledges the use of ChatGPT (OpenAI) for assistance in academic English editing during the preparation of this article.

References

Andersen M, Kleindienst S, Yi L, Lee J, Bodart M and Cutler B (2008) An intuitive daylighting performance analysis and optimization approach. *Building Research & Information* 36(6), 593–607.

Bienvenido-Huertas D, Sánchez-García D, Marín-García D and Rubio-Bellido C (2023) Analysing energy poverty in warm climate zones in Spain through artificial intelligence. *Journal of Building Engineering* 68, 106116.

Boyce P (2004) Reviews of technical reports on daylight and productivity. *Daylight Dividends Programs*, 14.

Cammarano S, Pellegrino A, Lo Verso VRM and Aghemo C (2015) Assessment of daylight in rooms with different architectural features. *Building Research & Information* 43(2), 222–237.

Çelik T (2024) Generative design experiments with artificial intelligence: Reinterpretation of shape grammar. *Open House International* 49(5), 822–842.

Chaillou S (2022) *Artificial Intelligence and Architecture: From Research to Practice*. Switzerland: Birkhäuser

Chen L, Song Y, Guo J, Sun L, Childs P and Yin Y (2025) How generative AI supports human in conceptual design. *arXiv preprint arXiv:2502.00283*.

Copilot (2025) Available at <https://copilot.microsoft.com/chats/sXUbaRTL TiXNqdYxTnRfD>

Edwards L and Torcellini P (2002) *Literature review of the effects of natural light on building occupants* (NREL/TP-550-30769). National Renewable Energy Laboratory, Golden, CO, United States.

Galasiu AD and Reinhart CF (2008) Current daylighting design practice: A survey. *Building Research & Information* 36(2), 159–174.

Gibson T and Krarti M (2015) Comparative analysis of prediction accuracy from daylighting simulation tools. *Leukos* 11(2), 49–60.

Hanafy NO (2023) Artificial intelligence’s effects on design process creativity: a study on used AI text-to-image in architecture. *Journal of Building Engineering* 80, 107999.

Horvath AS and Poulou P (2024) AI for conceptual architecture: Reflections on designing with text-to-text, text-to-image, and image-to-image generators. *Frontiers of Architectural Research* 13(3), 593–612.

Hu X, Zheng H and Lai D (2024) Performance prediction of AI-generated architectural layout design: Using daylight performance of residential Floorplans as an example, Computational Design Academic Forum and Annual Meeting of the Architectural Society of China’s Academic Committee on Computational Design. https://www.researchgate.net/profile/Hao-Zheng-20/publication/385863457_Performance_Prediction_of_AI-generated_Architectural_Layout_Design_Using_Daylight_Performance_of_Residential_Floorplans_as_an_Example/links/673809994a70511f07202b42/Performance-Prediction-of-AI-generated-Architectural-Layout-Design-Using-Daylight-Performance-of-Residential-Floorplans-as-an-Example.pdf.

Illuminating Engineering Society (2011) *IESNA Lighting Handbook: Reference and Application*, 10th Edn. New York: Illuminating Engineering Society of North America

Iranmanesh A and Lotfabadi P (2024) Critical questions on the emergence of text-to-image artificial intelligence in architectural design pedagogy. *AI & SOCIETY*, 40, 3557–3571.

Jiang K (2021) Use of climate-based daylight models for lighting simulation: A state of the art review of the literature published between 2010 and 2020, (Master’s thesis, Master in Product Development with a Specialization, Jönköping University). Jönköping, Sweden.

Kim CS and Chung SJ (2011) Daylighting simulation as an architectural design process in museums installed with toplights. *Building and Environment* 46(1), 210–222.

Li C, Zhang T, Du X, Zhang Y and Xie H (2024) Generative AI models for different steps in architectural design: A literature review. *Frontiers of Architectural Research*, 14(3), 759–783.

Liao W, Huang Y, Zheng Z and Lu X (2022) Intelligent generative structural design method for shear wall building based on “fused-text-image-to-image” generative adversarial networks. *Expert Systems with Applications* 210, 118530.

Lookx AI (2025) Available at <https://www.lookx.ai/pc/gc/t2i>

Mandala A, Sutanto EH and Santoso AR (2021) The effectiveness of daylighting through the toplighting design in large-volume building models. *ARTEKS: Jurnal Teknik Arsitektur* 6(2), 223–234.

Manmatharasan P, Bitsuamlak G and Grolinger K (2025) AI-driven design optimization for sustainable buildings: A systematic review. *Energy and Buildings*, 332, 115440.

Mardaljevic J (2000) Simulation of annual daylighting profiles for internal illuminance. *International Journal of Lighting Research and Technology* 32(3), 111–118.

Meselhy A and Almalkawi A (2025) A review of artificial intelligence methodologies in computational automated generation of high performance floorplans. *npi Clean Energy* 1(1), 2.

Montenegro N (2024) Integrative analysis of text-to-image AI systems in architectural design education: Pedagogical innovations and creative design implications. *Journal of Architecture and Urbanism* 48(2), 109–124.

Nazari M and Matusiak B (2024) Daylighting simulation and visualisation: Navigating challenges in accuracy and validation. *Energy and Buildings*, 312, 114188.

OpenAI (2025) Available at <https://chatgpt.com/>

Oxman R (2007) A performance-based model in digital design: PERFORMATIVE—Design beyond aesthetic. *Architectural Engineering and Design Management* 3(3), 169–180.

- Paananen V, Oppenlaender J and Visuri A** (2024) Using text-to-image generation for architectural design ideation. *International Journal of Architectural Computing* **22**(3), 458–474.
- Peel MC, Finlayson BL and McMahon TA** (2007) Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences* **11**(5), 1633–1644.
- Queiroz N, Fernandes L and Pereira FOR** (2024) Comparison of fast daylighting climate-based simulation methods for parametric design: Two-phase, three-phase method and path tracing. *Architectural Engineering and Design Management* **20**(3), 559–577.
- Reinhart CF and LoVerso VRM** (2010) A rules of thumb-based design sequence for diffuse daylight. *Lighting Research & Technology* **42**(1), 7–31.
- Reinhart CF and Walkenhorst O** (2001) Validation of dynamic RADIANCE-based daylight simulations for a test office with external blinds. *Energy and Buildings* **33**(7), 683–697.
- Riskiawan HY, Gupta N, Setyohadi DPS, Anwar S, Kurniasari AA, Hariono B, et al.** (2024) Artificial intelligence enabled smart monitoring and controlling of IoT-green house. *Arabian Journal for Science and Engineering* **49**(3), 3043–3061.
- Ruck NC** (1986) Letting in the daylight. *Building Research and Practice* **14**(5), 294–300.
- Sancho-Salas A, Flor JF and Muñoz SO** (2023) Predicting reality within the virtual: Experimental validation of daylight simulation tools for architectural spaces in the tropics. *Journal of Engineering Science and Technology* **18**, 72–81.
- Tamke M, Nicholas P and Zwierzycki M** (2018) Machine learning for architectural design: Practices and infrastructure. *International Journal of Architectural Computing* **16**(2), 123–143.
- Thampanichwat C, Wongvorachan T, Sirisakdi L, Chunhajinda P, Bunyaratitkit S and Wongmahasiri R** (2025) Mindful architecture from text-to-image AI perspectives: A case study of DALL-E, Midjourney, and stable diffusion. *Buildings* **15**(6), 972.
- Tregenza P and Wilson M** (2013) *Daylighting: Architecture and Lighting Design*. Routledge, London, United Kingdom.
- Turrin M, Andriotis C and Rafiee A** (eds) (2023) In *Computer-Aided Architectural Design. Interconnections: Co-Computing beyond Boundaries: 20th International Conference, CAAD Futures 2023, Delft, the Netherlands, July 5–7, 2023, Selected Papers*. Springer Nature, Cham, Switzerland.
- Zhao Y and Tian Z** (2023) Modified climate-based daylight modeling methods for buildings. *Building and Environment* **242**, 110598.