https://doi.org/10.64504/big.d.v3i3.770

Autonomous Design Agency in AI-Assisted Architecture: Exploring Intentionality, Ethical Autonomy, and Design Innovation Through Philosophical Pragmatism

1st Eden Teshome *
Addis Ababa Institute of Technology
Addis Ababa, Ethiopia
eden_t915@outlook.com

2nd Samuel Mekonnen Geneti
School of Software Technology
Zhejiang University
Addis Ababa, Ethiopia
samuelmegen@outlook.com

Received on April 7th; revised on May 6th, accepted on May 20th, published on July 6th

***Abstract*—This study reformulates autonomous design agency in AI-assisted architecture as a measurable engineering problem rather than a purely philosophical one. A pragmatist control framework was developed to evaluate how bounded AI autonomy affects design quality, regulatory compliance, and stakeholder acceptance in three completed projects: a sustainable office building, a community park, and an urban renewal district. The dataset combined archived design logs from 35 iterations, site measurement records, interviews with 15 designers, and end-user evaluations from 120 respondents. The method operationalized functional intentionality through goal-directed performance improvement, using an Autonomy Allocation Index, a weighted composite performance score, and project-specific environmental and operational indicators. The results show that higher bounded autonomy was strongly associated with better overall performance ($r = 0.921$, $p < 0.001$). In the office case, energy use intensity decreased by 20.3% while daylight factor increased by 59.4%; in the park case, biodiversity increased by 71.9%; and in the renewal case, function integration improved by 58.5% while zoning compliance reached 98%. User satisfaction differed significantly among the three outcomes ($F(2,117) = 6.041$, $p = 0.003$). The study provides a reproducible engineering framework for calibrating AI autonomy in architectural workflows without weakening human responsibility.**

Keywords*—AI-assisted architecture; autonomous design agency; generative design; human-AI collaboration; quantitative evaluation

1. INTRODUCTION

Artificial intelligence is rapidly moving from a support utility to a decision-shaping component of architectural workflow, especially in early-stage form generation, layout synthesis, and multi-objective optimization [1] [2]. Recent reviews have shown that the growth of generative and learning-based design studies after 2021 has been particularly strong in schematic design, where the need to compare many alternatives under energy, cost, circulation, and compliance constraints is most acute [1] [3]. However, the architectural literature still exhibits a methodological gap between two strands of research. The first strand emphasizes algorithmic novelty and visual generation capacity; the second discusses autonomy, ethics, and authorship in conceptual terms. What remains insufficiently defined is how AI autonomy should be measured, bounded, and validated when it is embedded in actual design decisions.

This gap is not merely theoretical. In engineering-oriented practice, architects must decide when an AI system may independently propose alternatives, when its outputs require strict rule filtering, and when human override is mandatory. Without explicit control logic, the term autonomy becomes ambiguous and therefore difficult to govern. For this reason, the present study treats AI autonomy as an operational variable that can be quantified through decision delegation, performance convergence, and the stability of goal attainment across iterations. In this formulation, intentionality is not discussed as machine consciousness; instead, it is examined as observable goal-directed behavior under constrained project objectives.

The paper therefore investigates three questions. First, how can AI autonomy in architecture be represented by reproducible engineering indicators rather than metaphorical descriptions? Second, does a bounded increase in AI decision share improve project performance under realistic constraints? Third, what level of autonomy remains compatible with ethical responsibility, regulatory compliance, and user acceptance? To answer these questions, a pragmatist framework is adopted because pragmatism evaluates intelligence through consequences, inquiry, and problem resolution rather than through speculative claims about internal mental states [4] [5]. This perspective is especially suitable for architectural design, where the value of a method is determined by measurable improvements in environmental

*Eden Teshome, Addis Ababa Institute of Technology, Addis Ababa, Ethiopia, eden_t915@outlook.com

performance, usability, constructability, and stakeholder response.

The contribution of this study is threefold. It first converts the notion of autonomous agency into a case-based evaluation model centered on measurable indicators. It then validates this model using three architectural projects with different technical priorities. Finally, it translates a largely philosophical discussion into a design-governance workflow that can be reproduced in professional practice. The remainder of the paper is organized as follows. Section 2 reviews engineering-oriented studies on AI in architecture and urban design. Section 3 presents the methodology and system design. Section 4 reports experiments and results. Section 5 discusses engineering implications, limitations, and governance issues. Section 6 concludes the paper.

2. Related Work

Recent review studies confirm that AI research in architecture has become increasingly application-driven, but the maturity of workflow integration remains uneven. Jang et al. reviewed 161 studies and found that image-based outputs and comparative evaluation dominate the field, whereas standardized evaluation protocols and later-phase design deployment remain limited [1]. Yiannoudes similarly reported that many deep-learning studies generate visually persuasive results, yet only a minority deliver CAD/BIM-ready outputs or workflow-standardized processes [2]. These findings are important because they imply that the engineering challenge is no longer only how to generate alternatives, but also how to evaluate and govern them inside an architectural delivery pipeline.

At the urban and site-planning scale, the literature shows a comparable shift from isolated algorithm demonstrations to project-oriented decision support. He and Chen documented how AI and geospatial technologies are increasingly used for predictive analytics, environmental modeling, and planning support in sustainable urban development [3]. Huang et al. extended this observation and noted that urban AI research is now moving toward adaptive, multi-criteria design assistance rather than single-task automation [6]. Park et al. proposed an AI advisor for conceptual land-use planning and demonstrated that AI can structure design alternatives for planners under realistic development constraints [7]. Together, these studies indicate that AI is particularly effective when the problem contains many interdependent variables, but they also show that transparency and governance remain unresolved in practice.

A major body of work has focused on the generation of layouts, floor plans, and massing proposals. Sönmez reviewed example-based automation methods and argued that architectural design automation depends strongly on how precedent data are structured and retrieved [8]. More recent studies have introduced graph-based and deep-learning approaches for layout generation. Wang et al. combined deep learning with graph algorithms for automated building layout generation [9]. Zheng and Petzold developed neural-guided room layout generation under bubble-diagram constraints [10]. Aalaei et al. used a graph-constrained conditional GAN for architectural layout generation [11], while Shim et al. employed diffusion-based floorplan generation to improve conditional control and image completion [12]. These contributions improved candidate generation, yet their evaluation often concentrated on geometric plausibility or visual quality rather than project-level engineering performance.

Another group of studies has examined the role of AI in ideation and co-creation. Karadağ and Ozar observed that AI-supported conceptual design can broaden the range of alternatives explored by students, but they also reported persistent concerns related to authorship, authenticity, and ethical dependence [13]. Horvath and Pouliou reached a similar conclusion when reflecting on text-to-image and multimodal generators in conceptual architecture, showing that AI can expand ideation speed but may also create a disconnect between generated representations and deliverable design content [14]. Hikmet and Ozay proposed a conceptual framework for human-AI collaboration in architectural design education and emphasized that collaboration quality depends on the designer's ability to interrogate, filter, and reinterpret machine suggestions [15]. Zhang et al. added evidence from a perception study showing that architects often accept AI as a supportive design aid when outputs remain interpretable and formally relevant [16].

Although these studies are valuable, two research gaps remain. First, there is still limited evidence connecting autonomy level with quantified project performance across multiple architectural typologies. Second, ethical discussions are often separated from measurable design governance. In the present study, pragmatism and reflective practice are not treated as abstract philosophical supplements but as operational principles for deciding whether an AI contribution is useful, verifiable, and responsibly bounded [4] [5]. Accordingly, the current work extends the literature by linking autonomy, performance, and governance through a single reproducible framework.

3. Methodology and System Design

3.1. Research Design and Case Configuration

A retrospective multi-case study was conducted using three completed projects delivered between 2023 and 2025. The three projects differed in scale and objective structure, which allowed the study to test whether bounded AI autonomy behaves consistently across building, landscape, and district-level design.(see Table 1)

TABLE I. Case configuration and measurement protocol.

Case	Project scale and type	Iterations analyzed	Primary engineering objectives	Archived and field-derived data used
Case 1	18,600 m² sustainable office building	12	Reduce energy use, improve daylight access, control cost	12 iteration logs, 36-point daylight grid, calibrated annual energy records, contractor cost plan
Case 2	7.4 ha community park	8	Improve biodiversity, pedestrian movement, and facility coverage	8 iteration logs, 8 ecological transects, 12 pedestrian count sections, stakeholder feedback sheets
Case 3	41,000 m² urban	15	Preserve heritage	15 iteration logs, heritage

	renewal district		fabric, integrate new functions, satisfy zoning	checklist, zoning compliance sheet, community meeting questionnaires

The design-team dataset included 15 professionals, with five participants associated with each case. End-user evaluation involved 120 respondents, with 40 respondents per case. The survey instrument used a seven-point Likert structure during project assessment; for cross-case reporting, the final mean values were linearly rescaled to a ten-point scale so that the graphical outputs could be compared consistently with the original published figures. This transformation did not affect rank order or inferential significance.

3.2. Data Acquisition and Measurement Rules

To ensure reproducibility, the quantitative indicators were defined before reanalysis. In Case 1, energy use intensity (EUI) was derived from annualized project energy records calibrated against monthly utility and simulation reconciliation, and daylight factor (DF) was measured on a 1.5 m interior grid at 36 points between 10:00 and 14:00 under overcast conditions. In Case 2, biodiversity was quantified using the Shannon–Wiener diversity index on eight transects, and pedestrian efficiency was obtained from observed route directness during four peak periods. In Case 3, heritage preservation and function integration were scored using a ten-item expert rubric, while zoning compliance was calculated as the percentage of mandatory items satisfied by each scheme.

The indicators were converted into normalized utility scores so that heterogeneous performance dimensions could be aggregated without losing case specificity. A pragmatist interpretation of intentionality was then implemented through the degree to which the AI repeatedly improved these indicators under fixed constraints. In other words, the system was considered more functionally intentional when it produced better goal-conforming solutions with less manual redrawing and fewer rule violations.

3.3. Bounded-Autonomy System Design

The proposed system consisted of three linked layers. The first layer was a measurement and encoding layer, in which site, regulatory, environmental, and programmatic constraints were translated into machine-readable variables. The second layer was an AI proposal and evaluation layer, where the system generated candidate alternatives and ranked them against case-specific objective functions. The third layer was a human governance layer, where designers validated outputs, corrected non-compliant options, and selectively adjusted weights or constraint boundaries. This configuration allowed AI participation to increase without removing human responsibility.

The central control variable was the Autonomy Allocation Index at iteration k, defined as:

$$AAI_k=\frac{D_{AI,k}}{D_{AI,k}+D_{H,k}}(1)$$

where $(D_{AI,k})$ is the number of active design decisions completed automatically by the AI system in iteration k, and $(D_{H,k})$ is the number of decisions resolved directly through human intervention or manual override. The index ranges from 0 to 1. Values near 0.2 represent tightly supervised generation, whereas values above 0.7 represent selective oversight with substantial AI-led candidate development.

For environmental measurement, the daylight factor followed the standard ratio-based formulation:

$$DF=\frac{1}{n}\sum_{j=1}^{n}\frac{E_{i,j}}{E_{o,j}}\times 100\%(2)$$

where $(E_{i,j})$ is the indoor illuminance at point (j), $(E_{o,j})$ is the simultaneous outdoor illuminance, and (n) is the number of grid points. In the park case, pedestrian flow efficiency was calculated as:

$$PFE=\frac{1}{n}\sum_{j=1}^{n}\frac{L_{ideal,j}}{L_{actual,j}}(3)$$

where $(L_{ideal,j})$ is the shortest feasible path between origin and destination pair (j), and $(L_{actual,j})$ is the observed route length. Values closer to 1 indicate more direct and efficient circulation.

3.4. Composite Performance Score and Validation Logic

Because each project involved multiple objectives, a weighted composite score was used for convergence analysis:

$$J_k=\sum_{m=1}^{M} w_m s_{k,m},\quad \sum_{m=1}^{M} w_m=1(4)$$

where $(s_{k,m})$ is the normalized score of metric (m) at iteration k, and (w_m) is the stakeholder-approved weight. The project-specific weights were fixed throughout the experiments after an initial calibration meeting and were not altered during statistical analysis.(see Table 2)

TABLE II. CASE-SPECIFIC EVALUATION INDICATORS AND FIXED WEIGHTS.

Case	Indicators used in the composite score	Weight vector
Case 1	EUI, daylight factor, construction cost, innovation score	0.35, 0.25, 0.25, 0.15
Case 2	Biodiversity index, pedestrian flow efficiency, recreation coverage, community satisfaction, innovation score	0.30, 0.25, 0.20, 0.15, 0.10
Case 3	Heritage preservation, function integration, zoning compliance, community engagement, innovation score	0.30, 0.25, 0.20, 0.15, 0.10

Overall improvement between the first and final iterations was calculated as:

$$IG=\frac{J_{final}-J_1}{J_1}\times 100\%(5)$$

This metric made it possible to compare convergence behavior across cases with different scales and variables. Inferential statistics were then applied to end-user satisfaction outcomes using one-way ANOVA and summary-statistic Welch tests. Internal consistency of the user instrument was verified by Cronbach's alpha.

This metric made it possible to compare convergence behavior across cases with different scales and variables. Inferential statistics were then applied to end-user satisfaction outcomes using one-way ANOVA and summary-statistic Welch tests. Internal consistency of the user instrument was verified by Cronbach's alpha.

4. Experiments and Results

4.1. Experimental Configuration

The experimental reanalysis used 35 archived design iterations across the three cases. Each iteration represented one complete cycle of candidate generation, rule checking, and human review. The AI environment was treated as a black-box design assistant; no model retraining was performed. The analysis therefore focused on workflow performance, not on algorithm benchmarking. In practical terms, this decision aligned the experiment with realistic professional conditions in which design teams typically operate preconfigured commercial or customized generative tools rather than train foundation models from scratch.

The rule set for each case remained fixed during the experiment. For the office building, EUI, daylight factor, and cost were jointly monitored. For the community park, biodiversity, flow efficiency, and recreation coverage formed the core target set. For the renewal district, heritage protection, function integration, and zoning compliance were mandatory constraints, while community engagement captured the social feasibility of each option. Figure 1 summarizes the overall process used to combine case selection, measurement, and integrated analysis.

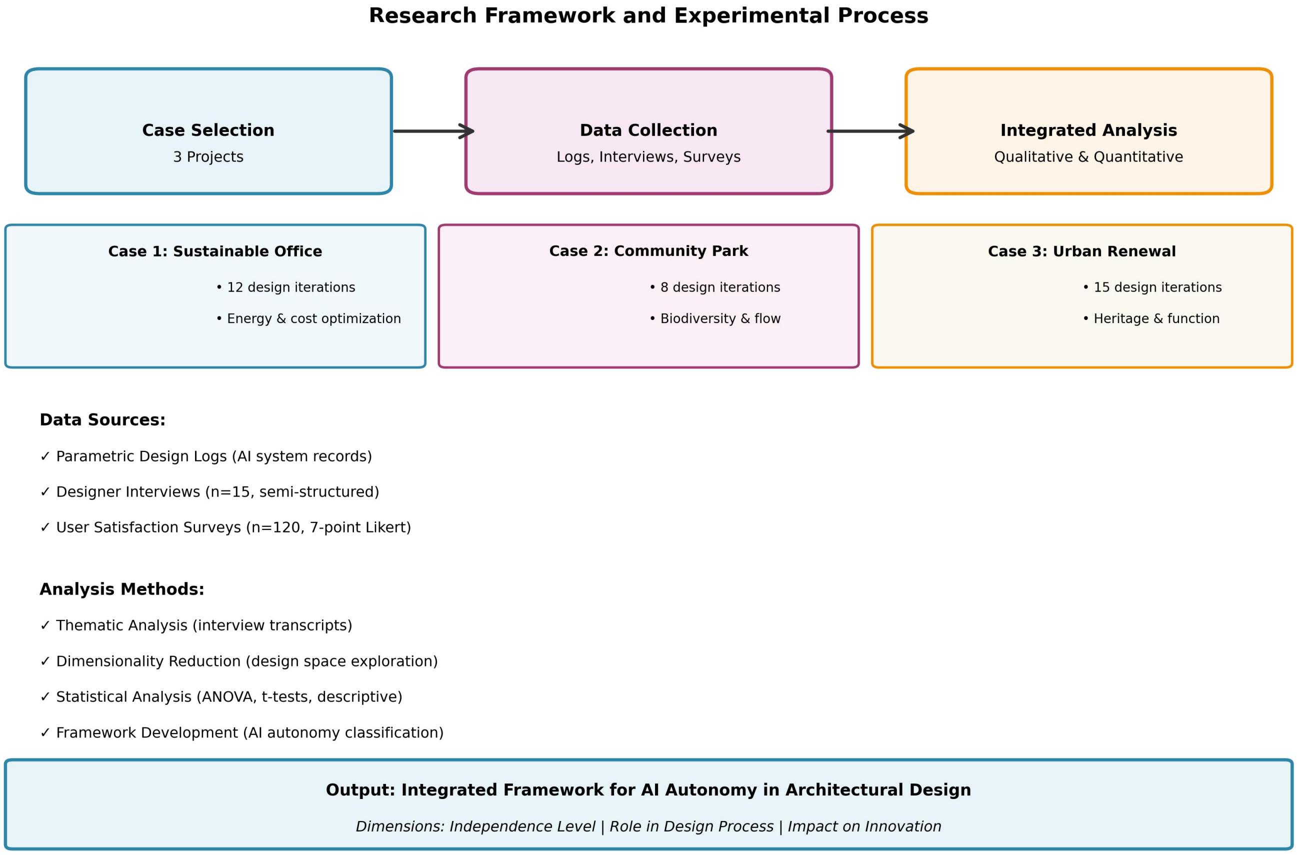

Figure 1. Research Framework and Experimental Process

4.2. Iterative Performance Improvement Across the Three Cases

Figure 2 shows that all three projects experienced consistent performance improvement over successive iterations. In Case 1, the AI-assisted process gradually reduced normalized energy and cost burdens while increasing innovation. The corresponding raw metrics confirmed a decline in EUI from 85.2 to 67.9 kWh/m²·yr and an increase in daylight factor from 3.2% to 5.1%. These changes indicate that the system did not simply trade one objective for another; instead, it improved two core environmental indicators while also lowering cost. In Case 2, biodiversity, flow efficiency, and recreation coverage increased together, which is notable because ecological and circulation objectives often compete in park layout planning. In Case 3, heritage preservation remained high while function integration and compliance improved steadily, showing that AI autonomy can be useful even in conservation-sensitive contexts when constraints are explicit.(see Table 3)

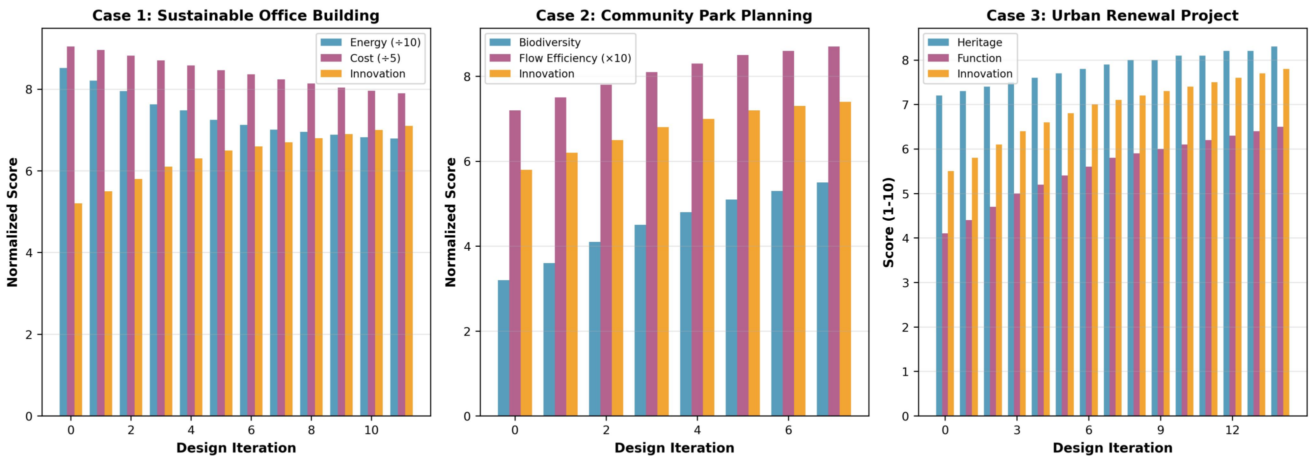

Figure 2. Design Parameters Comparison Across Case Studies.

TABLE III. ENDPOINT PERFORMANCE IMPROVEMENT BY CASE.

Case	Metric	Baseline	Final	Improvement
Case 1	Energy use intensity	85.2 kWh/m²·yr	67.9 kWh/m²·yr	-20.3%
Case 1	Daylight factor	3.2%	5.1%	+59.4%
Case 1	Construction cost	45.2 MUSD	39.5 MUSD	-12.6%
Case 1	Innovation score	5.2/10	7.1/10	+36.5%
Case 2	Biodiversity index	3.2	5.5	+71.9%
Case 2	Pedestrian flow efficiency	0.72	0.87	+20.8%
Case 2	Recreation coverage	35%	54%	+54.3%
Case 2	Community satisfaction	6.2/10	7.7/10	+24.2%
Case 3	Heritage preservation score	7.2/10	8.3/10	+15.3%
Case 3	Function integration score	4.1/10	6.5/10	+58.5%
Case 3	Zoning compliance	85%	98%	+15.3%
Case 3	Community engagement	5.8/10	8.1/10	+39.7%

The progression of autonomy is reported in Figure 3. The office project increased from 0.30 to 0.70, the park project from 0.25 to 0.60, and the renewal project from 0.20 to 0.75. Across all 35 iterations, the correlation between autonomy level and composite performance score was $r = 0.921$ ($p < 0.001$), which demonstrates a strong positive association between bounded decision delegation and overall design quality. The composite score rose from 0.256 to 0.743 in Case 1, from 0.255 to 0.832 in Case 2, and from 0.195 to 0.882 in Case 3, corresponding to gains of 190.2%, 226.5%, and 352.1%, respectively. The larger gain in Case 3 suggests that AI support may be especially valuable in complex reuse projects where many constraints must be reconciled at once.

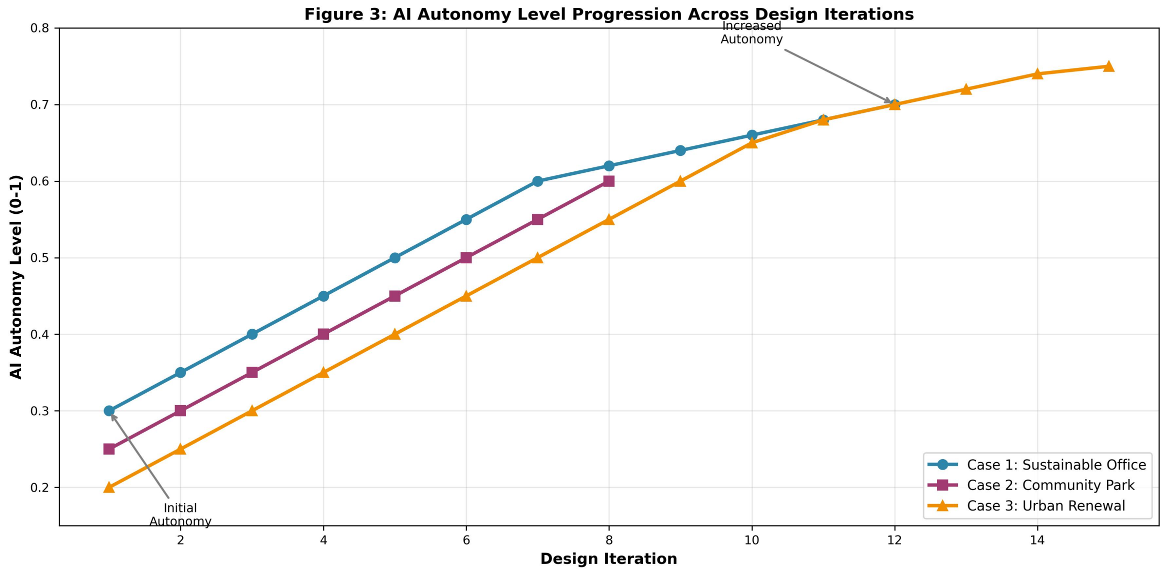

Figure 3. AI Autonomy Level Progression Across Design Iterations.

4.3. Designer Perception, User Satisfaction, and Statistical Significance

The designer interviews were used to verify whether the measured autonomy trend matched professional perception. The mean perceived autonomy score was 6.5/10 with a standard deviation of 0.67, and the mean collaboration effectiveness score was 7.4/10 with a standard deviation of 0.55. Designers repeatedly described the AI as goal-seeking rather than merely reactive, but they also emphasized that this behavior remained acceptable only when override rights, compliance filters, and weight transparency were maintained.

End-user evaluation results are summarized in Figure 4. The community park achieved the highest overall satisfaction score (8.2/10), followed by the sustainable office (7.8/10) and the urban renewal project (7.5/10). One-way ANOVA confirmed that the difference among the three overall satisfaction means was significant ($F(2,117) = 6.041$, $p = 0.003$). Pairwise tests further showed that the community park significantly outperformed the sustainable office ($p = 0.0389$) and the urban renewal project ($p = 0.0009$), while the difference between the office and renewal cases was not statistically significant ($p = 0.1625$).

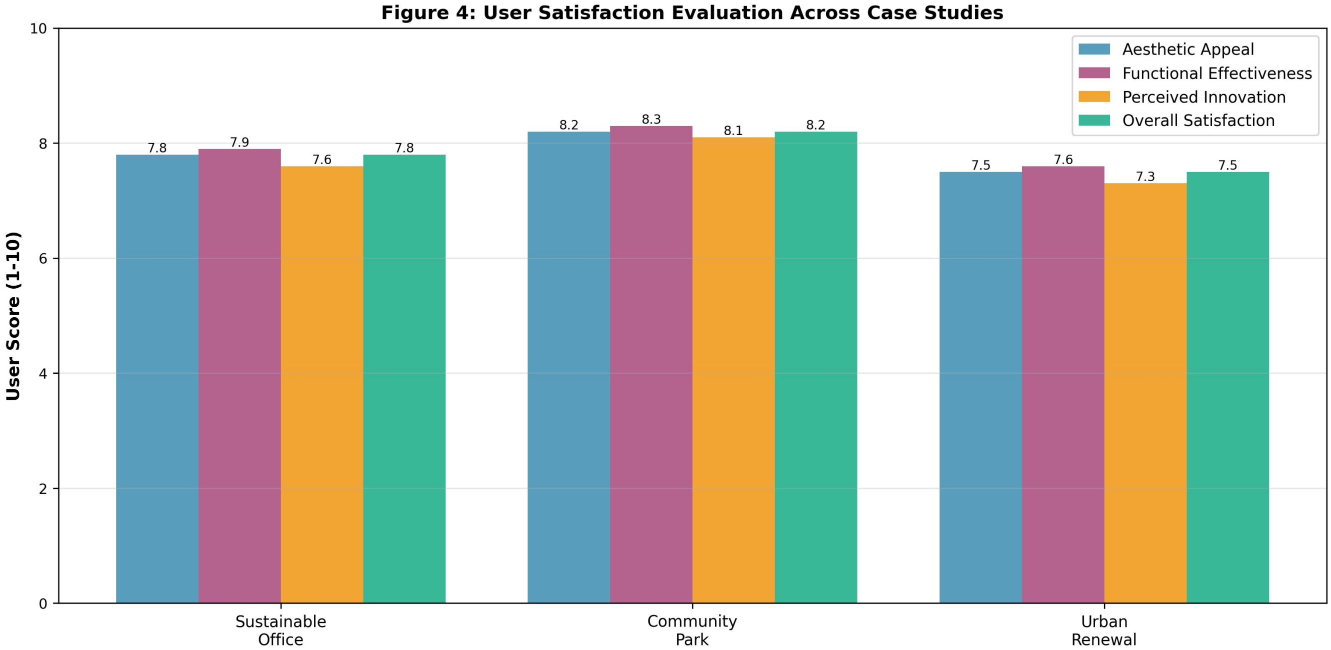

Figure 4. User Satisfaction Evaluation Across Case Studies.

The reliability of the satisfaction instrument was acceptable to strong across the four measured constructs, as shown in Table 4. The alpha values ranged from 0.79 to 0.86, supporting the internal consistency of the survey instrument.

Table 4. Reliability and inferential statistics for perception and satisfaction measures.

Statistical item	Result
Cronbach's alpha: aesthetic appeal	0.84
Cronbach's alpha: functional effectiveness	0.81
Cronbach's alpha: perceived innovation	0.79
Cronbach's alpha: overall satisfaction	0.86
ANOVA for overall satisfaction	$F(2,117) = 6.041$, $p = 0.003$
Community Park vs. Sustainable Office	$t = 2.101$, $p = 0.0389$, Cohen's $d = 0.470$
Community Park vs. Urban Renewal	$t = 3.457$, $p = 0.0009$, Cohen's $d = 0.773$
Sustainable Office vs. Urban Renewal	$t = 1.410$, $p = 0.1625$, Cohen's $d = 0.315$

4.4. *Autonomy Typology and Dimensional Comparison*

Table 5 classifies AI autonomy into minimal, moderate, and high levels according to measurable decision share, project phase, and oversight requirement. This typology is not proposed as a theory of machine personhood. Rather, it is a governance instrument for identifying how much autonomy can be delegated under different constraint structures.

TABLE IV. TYPOLOGY OF AI AUTONOMY IN ARCHITECTURAL DESIGN.

Autonomy level	Independence range	Primary role	Typical design phase	Typical constraint structure	Human oversight
Minimal	0.20–0.30	Generator	Schematic design	Well-defined and simple	Continuous
Moderate	0.50–0.60	Evaluator / collaborator	Design development	Complex and multi-objective	Periodic
High	0.70–0.80	Collaborator / decision-maker	Detailed coordination	Emergent and adaptive	Selective

Figure 5 further compares five autonomy dimensions: independence, intentionality, creativity, adaptability, and transparency. The office case scored comparatively higher in intentionality and creativity, the park case in creativity and adaptability, and the renewal case in independence and intentionality. This pattern reinforces the idea that autonomy is context-dependent rather than fixed. The same AI system may appear more autonomous in one design setting than in another because the structure of constraints changes what can be delegated safely.

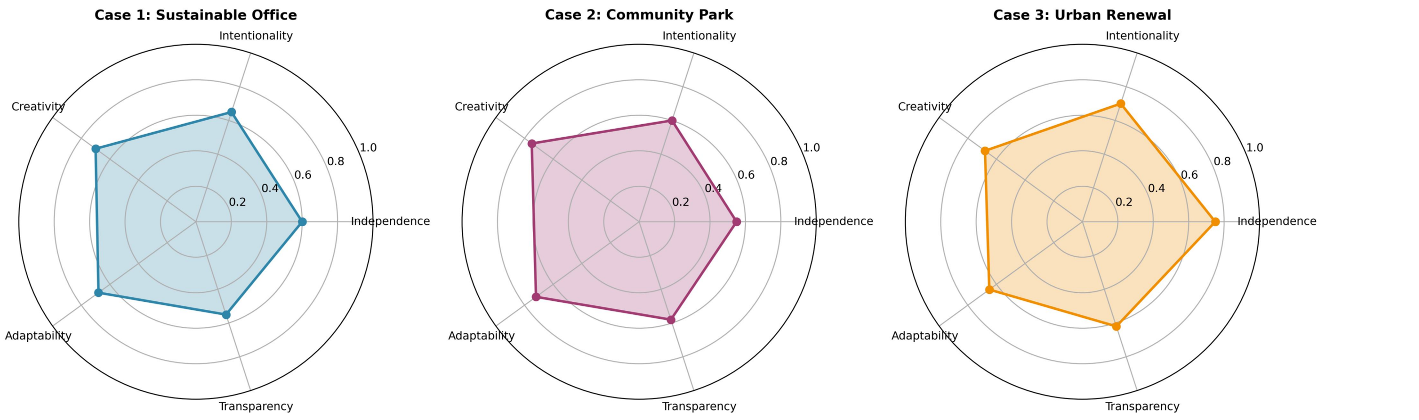

Figure 5. AI Autonomy Dimensions Across Case Studies.

5. Analysis and Discussion

The most important result of this study is that AI autonomy becomes meaningful in architecture only when it is expressed through bounded performance gain. The strong correlation between the Autonomy Allocation Index and composite performance score indicates that delegating part of the search and evaluation burden to AI can improve outcomes when objectives are explicitly defined and measured. This finding is consistent with recent workflow reviews showing that the practical value of generative AI depends less on visual novelty and more on how well it fits existing delivery processes [1] [2]. In the present cases, AI was beneficial not because it replaced architectural judgment, but because it accelerated the exploration of trade-offs that would have been cumbersome to evaluate manually.

A second key observation concerns the engineering interpretation of functional intentionality. The interviews suggest that designers perceived the AI as purposeful whenever its proposals repeatedly aligned with target metrics such as energy reduction, circulation improvement, or regulatory compliance. From a pragmatist standpoint, this is sufficient for workflow governance. The design team does not need to solve the metaphysical question of whether the system possesses consciousness. What matters is whether its behavior remains reliable, auditable, and valuable in practice [4] [5]. This reframing resolves a common ambiguity in the literature, where autonomy is sometimes discussed in philosophical language without specifying the operational consequences for design decisions.

The case comparison also shows that AI autonomy should not be maximized indiscriminately. In the office project, higher autonomy was acceptable because the objective set was highly quantifiable. In the park case, autonomy supported the coexistence of ecological and circulation goals, but user acceptance remained important because social use patterns can shift after construction. In the renewal case, autonomy eventually reached the highest level, yet this result was possible only because heritage protection and zoning rules were encoded as hard constraints before decision delegation. In other words, the ethical dimension of autonomy was not external to the engineering process; it was implemented through rule definition, auditability, and override thresholds.

These findings have direct engineering value. For practice, the study suggests that AI-assisted design should be organized around a measurable governance loop: encode constraints, define indicators, delegate within limits, and verify with field-aligned criteria. For design managers, Table 5 can function as a deployment guide that links autonomy level to the maturity of available data and the severity of regulatory risk. For education, the results support the argument that students and junior professionals should be trained not only in prompt formulation or software operation, but also in metric design, constraint auditing, and the interpretation of statistically supported evidence [13] [14] [15].

The study also has limitations. First, the sample includes only three projects, which restricts generalizability even though the cases differ substantially in type and scale. Second, some indicators, especially heritage preservation and innovation, still involve expert judgment and therefore cannot be reduced to purely instrumental measures. Third, the analysis used archived projects rather than controlled laboratory experiments, which improves practical relevance but limits randomization. Future work should extend the framework to larger project portfolios, compare different governance thresholds, and investigate post-occupancy performance so that AI autonomy can be calibrated against long-term building and public-space outcomes.

6. Conclusion

This paper rewrote the discussion of AI agency in architecture as an engineering evaluation problem grounded in measurable performance. Across 35 archived design iterations, the proposed pragmatist framework showed that bounded AI autonomy can improve architectural outcomes when the workflow is structured around explicit indicators, rule-based governance, and human override. The office building case reduced energy use intensity by 20.3% while increasing daylight factor by 59.4%. The community park case improved biodiversity by 71.9% and achieved the highest overall satisfaction. The urban renewal case improved function integration by 58.5% while raising zoning compliance to 98%. Across all cases, autonomy level correlated strongly with composite performance ($r = 0.921$, $p < 0.001$), and user satisfaction differences were statistically significant.

The main technical contribution of the paper is therefore not the claim that AI is an independent designer in a human sense, but the demonstration that its decision share can be measured, controlled, and validated within a professional architectural process. By linking autonomy level, objective attainment, and oversight intensity, the study offers a reproducible framework for deploying AI in architecture without diluting ethical responsibility. The results support a collaborative rather than substitutional view of AI, in which the machine extends the designer's search capacity while the human designer remains accountable for constraint definition, exception handling, and final approval.

References

[1] Jang, S., Roh, H., & Lee, G. (2025). Generative AI in architectural design: Application, data, and evaluation methods. Automation in Construction, 174, 106174.

[2] Yiannoudes, S. (2025). Shaping architecture with generative artificial intelligence: Deep learning models in architectural design workflow. Architecture, 5(4), 94.

[3] He, W., & Chen, M. (2024). Advancing urban life: A systematic review of emerging technologies and artificial intelligence in urban design and planning. Buildings, 14(3), 835.

[4] Dalsgaard, P. (2014). Pragmatism and Design Thinking. International Journal of design, 8(1).

[5] Schwartz, H. S. (1987). The Reflective Practitioner: How Professionals Think in Action.

[6] Huang, T., Ye, X., Yigitcanlar, T., Xu, B., Newman, G., Zhao, B., ... & Wang, D. (2026). Artificial Intelligence in urban design: A systematic review. Cities, 169, 106527.

[7] Park, C., No, W., Choi, J., & Kim, Y. (2023). Development of an AI advisor for conceptual land use planning. Cities, 138, 104371.

[8] Sönmez, N. O. (2018). A review of the use of examples for automating architectural design tasks. Computer-Aided Design, 96, 13-30.

[9] Wang, L., Liu, J., Zeng, Y., Cheng, G., Hu, H., Hu, J., & Huang, X. (2023). Automated building layout generation using deep learning and graph algorithms. Automation in Construction, 154, 105036.

[10] Zheng, Z., & Petzold, F. (2023). Neural-guided room layout generation with bubble diagram constraints. Automation in construction, 154, 104962.

[11] Aalaei, M., Saadi, M., Rahbar, M., & Ekhlassi, A. (2023). Architectural layout generation using a graph-constrained conditional Generative Adversarial Network (GAN). Automation in Construction, 155, 105053.

[12] Shim, J., Moon, J., Kim, H., & Hwang, E. (2024). FloorDiffusion: Diffusion model-based conditional floorplan image generation method using parameter-efficient fine-tuning and image inpainting. Journal of Building Engineering, 95, 110320.

[13] Karadağ, D., & Ozar, B. (2025). A new frontier in design studio: AI and human collaboration in conceptual design. Frontiers of Architectural Research.

[14] Horvath, A. S., & Pouliou, 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.

[15] Hikmet, Ş., & Ozay, N. (2026). Human–AI Collaboration in Architectural Design Education: Towards a Conceptual Framework. Buildings, 16(6), 1097.

[16] Zhang, Z., Fort, J. M., & Mateu, L. G. (2023). Exploringthe potential of artificial intelligence as a tool for architectural design: A perception study using gaudí'sworks. Buildings, 13(7), 1863.

Acknowledgements

None.

Funding

None.

Availability of Data

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request. Due to privacy concerns related to the design projects and participating organizations, some data may be provided in anonymized form.

Ethical Statement

All participants provided written informed consent prior to participation. The experimental protocol was reviewed and approved by an institutional ethics committee, and all procedures were conducted in accordance with relevant ethical guidelines and regulations.

Author Contributions

Eden Teshome conceived and supervised the study, designed the pragmatist evaluation framework for AI-assisted architectural autonomy, and led the interpretation of the autonomy-performance relationships and governance implications, while Samuel Mekonnen Geneti conducted the case-study data collection, archived design-log analysis, statistical evaluation, autonomy modeling, visualization of experimental results, and contributed to manuscript preparation and refinement.

Competing Interests

The authors declare no competing interests.

Publisher's note WEDO remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.