ICCAUA Proceedings Journal
Proceedings of the international conference of contemporary affairs in architecture and urbanism-ICCAUA
Volume 9 (December 2026), 2620050

Journal homepage: https://journal.iccaua.com/

ICCAUA Proceedings Journal

DOI: https://doi.org/10.38027/ICCAUA2026TR0050

The Impact of Large Language Models on Functional Diagram Generation: An Extended Cognition-Based Investigation of First-Year Architecture Students

* [1] **Adil Yirmibeş** , [2] **Arzu Ispalar Çahantimur** , [3] **Gözde Kırlı Özer**
[1, 2, & 3] Department of Architecture, Faculty of Architecture, Bursa Uludag University, Bursa, Türkiye
[1] E-mail: adilyirmibes@uludag.edu.tr, [2] E-mail: arzucahan@uludag.edu.tr, [3] E-mail: gozdekrl@uludag.edu.tr

Received: 24.04.2026
Revised: 25.06.2026
Accepted: 01.07.2026
Available online: 10.07.2026

This article has been selected and peer-reviewed for publication in this journal as part of the 9th International Conference of Contemporary Affairs in Architecture and Urbanism, held on 7–8 May 2026 in Istanbul, Türkiye.

Abstract

This study investigates the impact of Large Language Models (LLMs) on the functional diagram generation process of novice architecture students within the framework of extended cognition theory. Through a quasi-experimental design involving 41 first-year students, the research compares topological networks generated through traditional sketching with those produced through GenAI interaction for a weekend house project. Quantitative analyses of spatial nodes and edges reveal a significant cognitive expansion: LLM assistance increased the number of spaces by 11.4% and connections by 17.9%. However, the findings also indicate a qualitative shift toward instrumental dominance. Students made statistically insignificant curatorial reductions to the AI-generated complex networks, acting more as passive editors than active co-creators. Furthermore, extended dialogue loops resulted in additive topological expansion rather than hierarchical refinement. The study concludes that while AI offers substantial capacity enhancement, architectural pedagogy must cultivate spatial critical thinking to filter the hyper-productive nature of AI.

Keywords: Architectural Design Education; Generative Artificial Intelligence; Extended Cognition; Functional Diagram; Human–AI Co-Creativity.

1. Introduction

The early conceptual stages of the architectural design process constitute an intensive cognitive problem-solving activity characterized by high ambiguity and the necessity to resolve complex spatial relationships (Goel, 1995). During this phase, functional (bubble) diagrams emerge as fundamental tools for designers to establish topological structures between spaces and visualize hierarchical connections (Do & Gross, 2001). Conducted with pen and paper in traditional studio culture, this action represents a classic practice of "distributed cognition," wherein the designer offloads cognitive load onto the physical environment (Hutchins, 1995). In distributed cognition theory, tools are passive; they assist in the storage and externalization of information. However, the contemporary integration of Generative Artificial Intelligence (Gen-AI) and Large Language Models (LLMs) into architectural design has altered the ontological status of design tools. This integration has transformed them from passive recording mediums into idea-generating, "co-creative" partners (Karadağ, 2025; Rezwana & Maher, 2023).

In this context, the incorporation of LLMs into design transcends the boundaries of distributed cognition, evolving toward the theory of "extended cognition" posited by Clark and Chalmers (1998). Within extended cognition, the boundaries of the mind integrate with external algorithms that actively process information and provide feedback to the designer. Recent studies emphasize that rather than displacing human creativity in design education, Artificial Intelligence (AI) functions as a cognitive prosthesis that stretches mental boundaries (Cila, 2022; Leach, 2022). Under this theoretical framework, the act of designing transitions from directly drawing lines into a "dialogue-based design" process that requires negotiating with AI through natural language. In particular, prompt engineering emerges as a novel mode of sketching to define spatial relationships (Tong et al., 2023). Rezwana and Maher (2023) state that co-creative AI is not merely a system that autonomously generates content, but rather a dynamic partner that engages in complex interaction with the user. Consequently, the generation of architectural functional diagrams ceases to be a mere prompt-output relationship, transforming instead into an act of co-creativity where AI negotiates with the human designer.

In the literature, the integration of AI into architectural studios has predominantly been discussed through the lens of visualization, image generation, and text-to-image design workflows. Recent studies have shown that generative AI can expand creative possibilities in architectural design education, particularly during conceptual visualization and representational exploration (Montenegro, 2024). Similarly, AI-assisted architectural programming and design courses have demonstrated the pedagogical potential of AI in improving work efficiency and supporting early-stage design analysis (Jin et al., 2024). However, these studies mainly address AI as a tool for representation, programming support, or general studio transformation. Empirical studies focusing on the structural impact of LLMs on topological networks and functional diagrams, which constitute the invisible infrastructure of architectural design, remain highly limited. Specifically, a significant gap remains in the literature regarding how the topological accuracy and connection density of resulting diagrams transform when novice designers, whose spatial mental schemas are not yet fully established, offload the cognitive burden of establishing spatial relationships to AI. In broader educational contexts, LLMs have been associated with both opportunities for personalized learning and risks related to overreliance, bias, and the need for critical human oversight (Kasneci et al., 2023). Therefore, there is a need to evaluate, through quantitative and qualitative metrics, whether human-AI interaction in foundational design education goes beyond a simple prompt-response dynamic and precipitates a crossing of cognitive boundaries within the framework of extended cognition. Iranmanesh and Lotfabadi (2025) highlight the lack of theoretical foundations and guidelines regarding the use of these emerging technologies in creative disciplines. The present study aims to bridge this theoretical gap within the framework of Extended Cognition.

1.1 Purpose, scope, and methodology of the study

The primary objective of this study is to evaluate the integration of LLM assistance in the functional diagram generation processes of first-year architecture students within the framework of extended cognition, and to quantitatively assess its impact on spatial topology. Furthermore, the research aims to examine how novice designers transition from their traditional roles as "producers" to acting as "curators" or "editors" when confronted with the complex networks generated by AI. To achieve these objectives, the study is structured around the following hypothesis (H) and research questions (RQs):

- *Hypothesis 1 (H1):* The utilization of LLM assistance in the generation of functional diagrams significantly expands the topological complexity, specifically the number of nodes and edges, of the spatial networks produced by novice designers.
- *Research Question 1 (RQ1):* How does offloading cognitive load to an LLM transform novice designers' graph management strategies (i.e., generating/adding versus pruning/selecting) and their curatorial behaviours?
- *Research Question 2 (RQ2):* Does initiating the design process with a traditional hand-drawn sketch (cognitive preparation) versus prompting the AI directly yield a statistically significant difference in the topological complexity of the final diagram?
- *Research Question 3 (RQ3):* Are the iterative prompting loops engaged with the AI primarily utilized to expand the diagram quantitatively (an additive approach) or to correct and synthesize it qualitatively (a refining approach)?
- *Research Question 4 (RQ4):* Do the distinct commercial LLMs (e.g., Gemini, ChatGPT) utilized by students exhibit inherent algorithmic differences or biases regarding their topological density and their default reflexes for constructing architectural spatial networks?

This empirical research was conducted with architecture students in their first year of study enrolled in the MIM1004 course at Bursa Uludag University. Quantitative data were collected on a voluntary basis during an academic application conducted in the classroom. In terms of scope and methodology, the participants were tasked with generating functional networks for a weekend house project. The study comprises a total of 41 students evaluated under two distinct methodological approaches. Participants in Group 1 initiated the process with a manual preliminary sketch, whereas participants in Group 2 started directly by prompting the LLM. The resulting architectural diagrams were standardized using graph theory metrics to quantify spatial nodes and topological connections. Finally, the collected data were subjected to comprehensive statistical analysis to measure the structural differences between the two methodologies.

2. Materials and Methods

This research employs a quantitative quasi-experimental design to investigate the dialogical interaction established between students in their first year of architecture and AI, as well as their topological space generation processes (Creswell & Creswell, 2017; Rezwana & Maher, 2023). The study consists of an experimental group based on a repeated measures logic to compare cognitive states before and after the AI intervention. Furthermore, to isolate the effects of the traditional manual method, the study incorporates a comparison or control group that begins the design process directly with AI without any preliminary cognitive preparation (Fraenkel & Wallen, 1990). The experimental setup of the study and the workflow indicating the specific data collection points are illustrated in Figure 1.

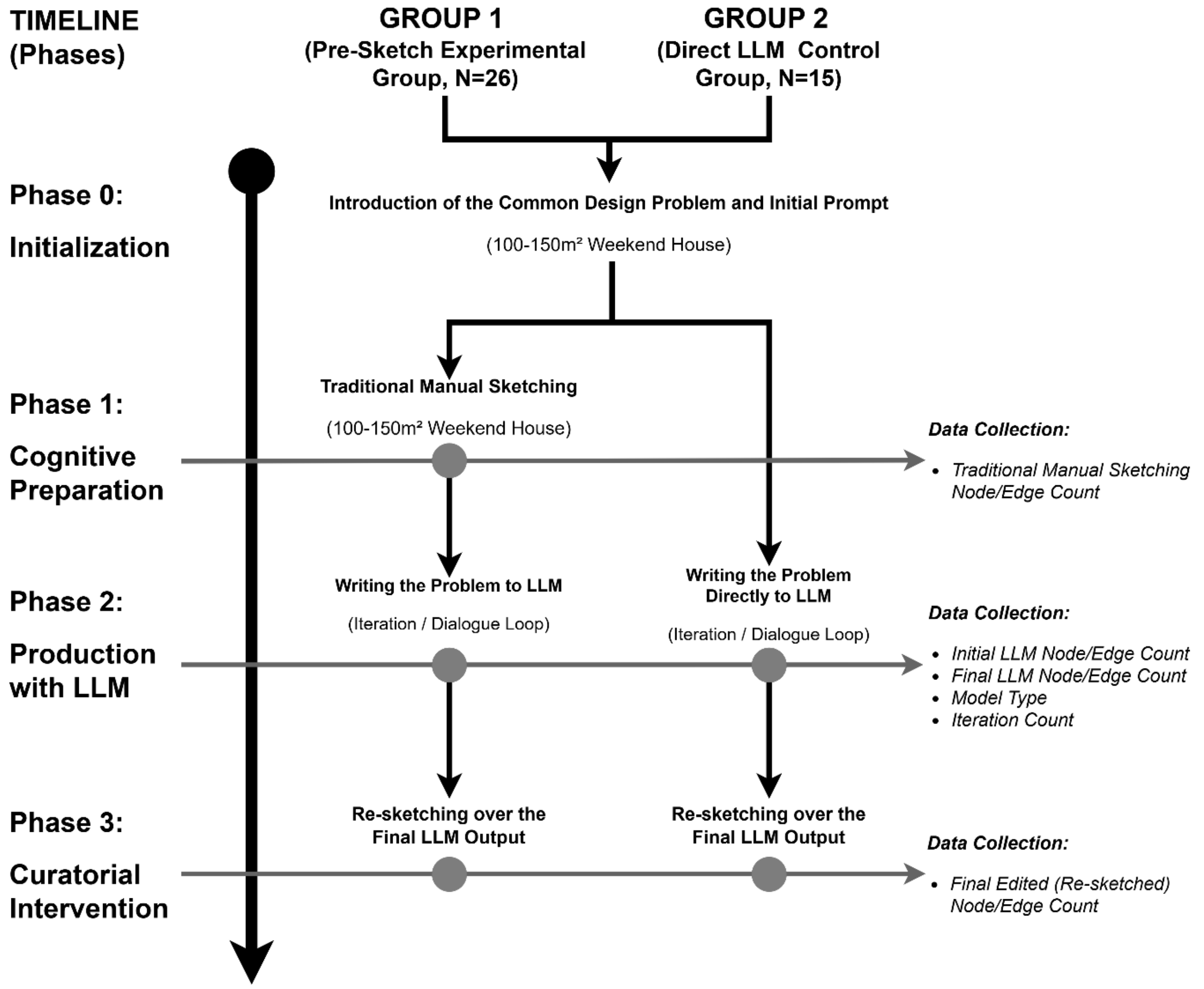

Figure 1. Methodological workflow of the experimental design, illustrating the interaction phases and data collection points across the two participant groups.

2.1 Experimental setting and study group

The study was conducted as an in-class application following two weeks of theoretical instruction on functional diagrams within the scope of the MIM1004 course at the Faculty of Architecture of Bursa Uludag University. The research study group consists of 41 first-year students enrolled during the 2025-2026 academic year who participated voluntarily. Informed consent was obtained from all participants, and ethical committee approvals were secured for the research. To increase the reliability of the study and to determine whether the traditional sketching stage influences the Artificial Intelligence (AI) process, students were divided into two independent study groups.

- *Group 1 (Pre-Sketch Experimental Group, N=26):* These students initiated the design process with a preliminary sketch representing cognitive preparation and subsequently engaged in a dialogue by transferring this mental schema to the Large Language Model (LLM).
- *Group 2 (Direct LLM Control Group, N=15):* These students began the design problem directly by writing an initial prompt on the LLM interface without any involvement in the preliminary sketching stage.

The same design problem was established for both groups. Furthermore, both groups utilized the same initial prompt when commencing their work with the LLM, as shown in Table 1. Consequently, it is possible to observe how the cognitive process undertaken prior to the task affects the final design productions for the defined design problem.

Table 1: Design problem and the initial prompt submitted to the LLM.

Design Problem	Formulate a spatial requirements program comprising open, semi-open, and enclosed spaces for the design of a two-story weekend house featuring an indoor area of 100 to 150 square meters. Furthermore, construct a functional diagram illustrating the relationships among these spaces.
Initial Prompt	Your task is to establish an optimal spatial requirements program comprising open, semi-open, and enclosed spaces for a two-story weekend house design with an indoor area of 100 to 150 square meters, and to code the fundamental relationship network among these spaces in Mermaid.js (graph TD) format. Do not provide any introductory or concluding remarks. Output only the raw Mermaid code block that can be directly copied and executed. Continue to code subsequent spatial arrangement instructions provided later in the conversation in the Mermaid.js (graph TD) format.

2.2. Data collection methods

Within the scope of the research, students were required to develop an optimal spatial requirements program comprising open, partially open, and enclosed spaces for a two-story weekend house project with an indoor area of 100 to 150 square meters. Furthermore, they were asked to construct the functional diagram, specifically the topological relationship network, among these spaces. The questions utilized for data collection are presented in Table 2. The data collection process consists of the following sequential phases:

- *Phase 1, Functional Diagram Generation via Traditional Sketching Methods (Group 1 Only, N=26):* Students drew their initial functional diagrams using traditional pen and paper methods relying entirely on their inherent cognitive capacities.
- *Phase 2, Functional Diagram Generation via Artificial Intelligence (Group 1 and Group 2, N=41):* Students entered their initial prompts into the system utilizing various LLM tools. Google Gemini was the predominantly preferred tool at 78 percent, accompanied by other models such as ChatGPT, Copilot, and DeepSeek. To revise the initial schema in accordance with their specific architectural intentions, the students continued the process by establishing a dialogue through iterative prompting with the model. The initial and final LLM outputs were collected in the format of Mermaid.js (graph TD), which is a text-based network definition language. Consequently, this standardization allowed the diagrams generated by the system to be visualized on platforms such as Draw.io.
- *Phase 3, Curatorial Arrangement via Traditional Sketching Methods (Group 1 and Group 2, N=41):* Students redrew the final output they obtained from the LLM using traditional methods. Through this action, they organized the spaces or relationships upon their own initiative to execute a curatorial intervention and ultimately formulated their definitive functional diagrams.

Table 2: Questions posed at the data collection points.

Study Group	Data Collection Phase	Questions Posed for Data Collection
Group 1	Phase 1 Generation via traditional sketching method	How many independent spaces in total did you define in the diagram you generated using the traditional sketching method?
Group 1	Phase 1 Generation via traditional sketching method	What is the total number of connections you established between spaces in the diagram you generated using the traditional sketching method?
Group 1 and Group 2	Phase 2 Generation via LLM (Initial)	How many independent spaces in total did you define in your initial functional diagram generated via the LLM?
Group 1 and Group 2	Phase 2 Generation via LLM (Initial)	What is the total number of connections you established between spaces in your initial functional diagram generated via the LLM?
Group 1 and Group 2	Phase 2 Generation via LLM (Final)	How many independent spaces in total did you define in your final functional diagram generated via the LLM?
Group 1 and Group 2	Phase 2 Generation via LLM (Final)	What is the total number of connections you established between spaces in your final functional diagram generated via the LLM?
Group 1 and Group 2	Phase 2 Generation via LLM (Employed LLM model)	Which LLM model did you use during the "Functional Diagram Study with the Assistance of Generative Artificial Intelligence" application?
Group 1 and Group 2	Phase 2 Generation via LLM (Iteration Depth)	How many iterations including the initial prompt did you conduct for the LLM to generate the final functional diagram during the production phase?
Group 1 and Group 2	Phase 3 Curatorial arrangement via traditional sketching	How many independent spaces in total did you define in your final functional diagram where you performed curatorial arrangement using the traditional sketching method?
Group 1 and Group 2	Phase 3 Curatorial arrangement via traditional sketching	What is the total number of connections you established between spaces in your final functional diagram where you performed curatorial arrangement using the traditional sketching method?

2.3. Measurement and evaluation criteria

This section establishes the methodological framework for quantifying the architectural outputs by defining the specific topological metrics grounded in graph theory. Furthermore, it details the comprehensive statistical procedures employed to rigorously analyse the experimental data and evaluate the cognitive transformation of the students.

2.3.1. Topological metrics and dependent variables

To objectively measure the complexity of the architectural functional diagrams generated by the students and to quantify the human-machine interaction, the calculations in this study were conducted utilizing two fundamental metrics grounded in graph theory (Hillier & Hanson, 1989).

- *Number of Nodes (Spaces):* This metric represents each independent functional space, whether enclosed, semi-open, or open areas, within the design and measures the quantitative volume of the spatial program.
- *Number of Edges (Spatial Relationships):* This metric represents the permeability, accessibility, and topological connections between spaces to reveal the relational complexity and network density of the diagram (Ostwald, 2011).

In addition to these two fundamental metrics, the "Curatorial Arrangement Amount" was defined as a dependent variable to measure the curatorial response students provided to the network proposed by the AI. This amount was derived mathematically by subtracting the number of elements in the final sketch, which the student redrew by taking initiative, from the number of elements in the final diagram proposed by the LLM (Final LLM Output minus Re-sketch).

2.3.2. Statistical analysis procedure

The statistical analyses of the obtained data were conducted using the Python programming language and the scientific computing library SciPy (Virtanen et al., 2020). The steps followed during the analysis process and the methodological justifications for the preferred tests are as follows:

- *Normality testing:* The Shapiro-Wilk normality test was initially applied to determine whether the data satisfied the prerequisites for parametric tests at each stage (Field, 2024). The analysis methods, whether parametric or nonparametric, were determined dynamically based on the results of this test.
- *Measuring cognitive transformation (dependent groups)***:** To detect the cognitive expansion and topological change between the internal stages of the same experimental group (Group 1), such as the difference between the Initial Sketch and the final output from the LLM, the paired-samples t-test or the Wilcoxon signed-rank test was utilized in accordance with the data distribution (Field, 2024; Siegel, 1957).
- *Testing methodological differences (independent groups):* The Mann-Whitney U Test for independent samples was applied to compare the differences in the final outputs of students who began with the traditional sketching process (Group 1) and those who started the process directly with the LLM (Group 2) (MacFarland & Yates, 2016).
- *Dialogue dynamics and the iteration effect:* Pearson and Spearman correlation analyses were conducted to reveal the relational patterns between the depth of dialogue, meaning the number of iterations, that students established with the artificial intelligence and both the topological complexity they achieved and the arrangement behaviours they performed (Field, 2024).
- *Effect sizes:* Beyond the obtained p-values ($p<0.05$), Cohen's d and r effect size coefficients were calculated and reported according to the test type to demonstrate the real-world magnitude and practical significance of the artificial intelligence on topological generation (Cohen, 2013).

2.4. Data reliability

To ensure the internal validity of the topological data declared by the participants in the online form, an Interrater Reliability test was applied (Shrout & Fleiss, 1979). Expert counts were conducted by the researchers on diagram images belonging to 8 randomly selected participants from the dataset, which constitutes approximately 20 percent of the study population, and these values were crosschecked with the self-declarations of the students. By examining the diagrams of each participant at different design stages, the analysis was performed across a total of 40 distinct data points. Following the Two-Way Mixed Effects Intraclass Correlation (ICC 2,1 Absolute Agreement) analysis, the ICC score for node (space) counts was determined to be 1.000 indicating perfect agreement, while the score for edge (relationship) counts was found to be 0.933 indicating excellent agreement (Koo & Li, 2016). Additionally, the Mean Absolute Error (MAE) in the edge counts remained at the 0.25 level. Ultimately, the collected data were deemed highly reliable for quantitative analyses.

3. Results

The experimental application of the research was conducted with a total of 41 first year architecture students comprising two independent groups. Group 1, which initiated the process with a traditional preliminary sketch, consists of 26 students, whereas Group 2, which commenced directly with an artificial intelligence prompt, includes 15 students. The findings of the study are presented sequentially under four main headings: the cognitive expansion effect of AI on topological complexity, the impact of the initial method on the overall process, the curatorial arrangement behaviours of the students, and the dynamics of iteration and tool preference.

3.1. The impact of LLMs on topological complexity

The initial phase of the research investigated the primary hypothesis of the study regarding how artificial intelligence tools affect the design cognition and topological network building capacities of first year architecture students. For this purpose, topological metric differences were analysed between the initial sketches of the Group 1 students (N=26) who drew manual functional diagrams using their own cognitive capacities at the beginning of the process and the final LLM outputs they developed by establishing a dialogue with the AI. The Shapiro-Wilk test results indicated that the differences between the node and edge counts exhibited a normal distribution (Table 3). Therefore, the parametric paired-samples t-test was utilized in the analysis. The results of the analysis are presented in Table 4.

Table 3: Testing the Normality Assumption for Group 1 (N=26) (Shapiro-Wilk).

Topological Metric Variables (Difference Scores)	**Shapiro-Wilk (W)**	**p-value**	**Distribution Type**
Number of Nodes (Spaces)	0.931	0.0804	Normal
Number of Edges (Relationships)	0.951	0.2470	Normal

Table 4: Topological change between the traditional sketch and the final LLM output for Group 1 (N=26)

Topological Metric	**Initial Sketch (Phase 1)**	**Final LLM Output (Phase 2)**	**t-value**	**p-value**	**Effect Size (Cohen's d)**
Number of Nodes (Spaces)	12.23	13.62	-3.28	0.0031	0.642
Number of Edges (Relationships)	12.62	14.88	-3.57	0.0015	0.699

Upon examination of Table 4, it is observed that while the average number of spaces (nodes) in the initial diagrams drawn independently by the students was 12.23, this number rose to 13.62, an approximate increase of 11.4%, in the final code

obtained following their interaction with the LLM, and this increase is statistically significant (t=-3.28, p=0.0031). Similarly, the number of relational connections (edges) established between spaces increased from an average of 12.62 to 14.88, representing a growth of approximately 17.9% after the LLM intervention (t=-3.57, p=0.0015).
The Cohen's d values calculated for both metrics (0.642 for the number of nodes and 0.699 for the number of edges) reside at the "medium-to-high" effect threshold within the literature (Cohen, 2013), indicating that the observed difference is not coincidental and corresponds to an effect that can be considered practically robust. In light of these findings, Hypothesis 1 is confirmed: Generative Artificial Intelligence has significantly expanded the fundamental architectural functional networks that first-year students could construct mentally, thereby extending the design process toward a more complex and richer topological structure. In this context, LLMs have assumed the role of a cognitive expander that enhances the student's capacity for spatial generation.

3.2. The impact of the initial method on topological generation

In the second phase of the research, it was examined whether the traditional approaches exhibited by designers at the beginning of the process (creating a preliminary sketch as a cognitive preparation) generated a significant difference on the final topology produced by the artificial intelligence. For this purpose, the final LLM outputs and the re-sketch (Curatorially Arranged) phases of Group 1 (N=26), who initiated the process by drawing a preliminary sketch, and Group 2 (N=15), who entered the process directly with a text prompt on the LLM interface without any preliminary preparation, were compared.
Following the normality analysis of the dataset, it was determined that although the data belonging to Group 2 exhibited a normal distribution, the dependent variable scores belonging to Group 1 did not satisfy the normal distribution assumption (Table 5). Due to the fact that at least one of the two compared groups did not meet the parametric prerequisites, and the group sizes (N=26 and N=15) differed from one another, the non-parametric Mann-Whitney U Test was applied for the intergroup comparisons. The results of the analysis are presented in Table 6.

Table 5: Testing the normality assumption for Group 1 and Group 2 (N=26 and N=15) (Shapiro-Wilk).

Topological Metric	Group 1 (p-value)	Group 1 Distribution	Group 2 (p-value)	Group 2 Distribution
Final LLM Output (Node)	0.0275	Not Normal	0.8293	Normal
Final LLM Output (Edge)	0.0001	Not Normal	0.2487	Normal
Re-sketch / Arranged (Node)	0.0086	Not Normal	0.2860	Normal
Re-sketch / Arranged (Edge)	0.0503	Normal	0.4960	Normal

Table 6: Comparison of topological outputs of Group 1 and Group 2 according to the initial method using the independent samples test (Mann-Whitney U).

Topological Metric	Group 1 Mean (N=26)	Group 2 Mean (N=15)	U-value	p-value	Effect Size (r)
Phase-2 Final LLM Output (Node)	13.62	14.67	146.0	0.1851	0.207
Phase-2 Final LLM Output (Edge)	14.88	15.33	155.0	0.2807	0.169
Phase-3 Re-sketch / Arranged (Node)	13.04	14.80	130.0	0.0783	0.275
Phase-3 Re-sketch / Arranged (Edge)	13.69	15.13	138.0	0.1239	0.241

Upon examining Table 6, it was determined that whether students initiated the process with a preliminary cognitive preparation (sketching) or via direct dialogue did not yield a statistically significant difference on the node (U = 146.0, p = 0.185) and edge (U = 155.0, p = 0.280) dimensions of the "Final LLM" diagrams they produced. Similarly, no topological difference was found between the two groups during the curatorial arrangement phase, wherein students, upon their own initiative, transformed these AI-generated outputs into a "Re-sketch" (p > 0.05). The fact that all calculated effect size (r) values remained between 0.16 and 0.27, thereby falling within the weak effect threshold established in the literature (Cohen, 2013), mathematically corroborates this absence of variance.
These findings demonstrate that the determining factor in the process is neither traditional human preparation nor methodological preference; rather, Gen-AI dominates the process according to its own "hyper-productive" character, irrespective of the input type. Regardless of the starting point (linear or textual) from which the first-year students embarked, the AI drove them toward a similar topological network density, specifically within the range of 14 to 15 nodes.

3.3. Management of networks and curatorial behavior

LLM models inherently possess a hyper-productive character and can frequently propose overly complex spatial relationships that are disconnected from the architectural context. In the third phase of the research, the manner in which the human designer reacted to this topological complexity achieved by the AI and whether they took initiative were investigated. For this purpose, the "Final LLM Output" of all students (N=41) and the "Re-sketch (Curatorially Arranged)" diagrams they achieved by passing them through their own architectural filters were compared. Prior to the analysis, the normality distributions of the difference scores were examined using the Shapiro-Wilk test; due to the data not exhibiting a normal distribution (Table 7), the analysis proceeded with the non-parametric Wilcoxon signed-rank test. The results of the analysis are presented in Table 8.

Table 7: Testing the Normality Assumption for Group 1 and Group 2 (N=41) (Shapiro-Wilk).

Topological Metric Variables (Difference Scores)	Shapiro-Wilk (W)	p-value	Distribution Type
Number of Nodes (Spaces)	0.573	0.0001	Not Normal
Number of Edges (Relationships)	0.547	0.0001	Not Normal

Table 8: Curatorial Arrangement Phase for Group 1 and Group 2 (N=41) (Wilcoxon)

Topological Metric	Final LLM Output (Phase 2)	Curatorial arrangement via re-sketch (Phase 3)	Z-value	p-value	Effect Size (r)
Number of Nodes (Spaces)	14.00	13.68	-1.079	0.2804	0.168
Number of Edges (Relationships)	15.05	14.21	-1.711	0.0871	0.267

Upon examination of Table 8, it is observed that while redrawing the complex networks inherited from the AI upon their own initiative, the students reduced the average number of spaces (nodes) from 14.00 to 13.68, representing a decrease of approximately 2.3%, and the number of relational connections (edges) from 15.05 to 14.22, representing a decrease of approximately 5.5%. However, the conducted Wilcoxon signed-rank test revealed that this very limited reduction between the complexity of the AI-generated network and the final network intervened by the human is not statistically significant ($Z=-1.079$, $p=0.280$ for Nodes; $Z=-1.711$, $p=0.087$ for Edges). The reductions made by the students did not reach a magnitude that would alter the overall topological density of the diagram.

3.4. Human-machine interaction: the impact of iteration and instrumental preferences on generation

In the final phase of the research, moving beyond the topological outputs, how the form of interaction students established with Generative Artificial Intelligence (iteration depth) and the instrumental preferences they utilized in the process (LLM Model) affected the generated spatial organization was examined. However, prior to proceeding with these relational and comparative analyses, the distribution characteristics of the relevant variables were examined using the Shapiro-Wilk test to determine the assumptions of the statistical tests to be employed (Table 9). It is observed that the variables "Number of Iterations," representing the repetition of prompts written by the students to the artificial intelligence, and "Number of Edges," representing the relationships within the diagram, did not satisfy the normal distribution assumption ($p<0.05$). Consequently, it was decided to utilize the Spearman correlation analyses and Mann-Whitney U tests for all analyses in this section.

Table 9: Normality Assumption Test for Iteration and Final LLM Topology Metrics for Group 1 and Group 2 (N=41) (Shapiro-Wilk).

Variables	Shapiro-Wilk (W)	p-value	Distribution Type
Number of Iterations	0.833	0.0001	Not Normal
Final LLM Number of Nodes (Spaces)	0.946	0.0507	Normal
Final LLM Number of Edges (Relationships)	0.893	0.0010	Not Normal

3.4.1. The iteration effect in dialogue-based design

To determine the relationship between the successive revision prompts (the number of iterations) that students wrote to the artificial intelligence and the final size reached by the diagram, a Spearman Correlation analysis was conducted (Table 10). As a result of the analysis, a statistically significant, positive, and moderate relationship was found between the number of iterations students performed and the number of spaces (nodes) in the final diagram they obtained ($r = 0.368$, $p = 0.018$). In other words, as students prolonged the dialogue they established with the artificial intelligence to develop or correct the diagram, the artificial intelligence exhibited a tendency to continuously add new spaces (nodes) to the design. This finding indicates that revisions made with generative artificial intelligences systems are directed toward adding information and elements, rather than refining them. The prolongation of the dialogue increases the topological volume.

Table 10: Correlation of the number of iterations between the initial LLM generation and the final LLM generation for Group 1 and Group 2 (N=41) (Spearman).

Relationship (Number of Iterations)	Correlation Coefficient (r)	p-value	Significance
Number of Nodes (Spaces)	0.3676	0.0181	Significant positive relationship
Number of Edges (Relationships)	0.2721	0.0852	Not Significant

3.4.2. The impact of the LLM model on topological generation

Whether there was a difference between the architectural topology building reflexes of the different Large Language Models (LLMs) chosen by the students according to their own preferences during the in-class application was tested using the Mann-Whitney U test (Table 11). For this purpose, the participants were divided into two groups: those using Google Gemini (N=32) and those using other models (predominantly ChatGPT) (N=9). The results of the analysis showed that the artificial intelligence model used did not create a significant difference in the total number of spaces (nodes) in either the initially generated (starting) diagram or the final diagram reached after iterations ($U=158.5$, $p=0.656$ for the Final

Node). However, it was determined that there was a statistically significant difference between the models concerning the number of connections (edges) representing spatial relationships from the very beginning of the process.

Table 11: The effects of the differences in the Gen-AI models used by Group 1 and Group 2 (N=41) on the initial generation and the final generation conducted with the LLM (Mann-Whitney U).

Topological Metric	Gemini Mean (N=32)	Other Mean (N=9)	U-value	p-value	Difference Status
Initial Number of Nodes (Spaces)	13.78	15.55	113.50	0.339	No difference
Initial Number of Edges (Relationships)	14.15	20.66	71.00	0.021	Significant difference
Final Number of Nodes (Spaces)	14.21	13.22	158.50	0.656	No difference
Final Number of Edges (Relationships)	14.59	16.66	73.50	0.026	Significant difference

The average number of spatial connections in the initial diagrams that students using other models received from the artificial intelligence (M=20.67) is significantly higher than the average of the students using Gemini (M=14.16) (U=71.0, p=0.021). This characteristic difference in the network-building capacity of the models was preserved throughout the iterations; in the attained "Final" diagrams, the connection average of the other models (M=16.67) continued to remain significantly higher than Gemini's average (M=14.59) (U=73.5, p=0.026). This situation demonstrates that the different selected commercial language models (Gemini vs. GPT) interpret architectural programs differently even when the exact same initial prompts are entered; while some models (Gemini) propose a more isolated space syntax, others generate a much more integrated and permeable spatial organization from the very beginning through to the final generation.

4. Discussion

This research investigated how the heavy cognitive load encountered by novice designers, first year students, during the early conceptual stages of architectural design (Goel, 1995) undergoes a structural transformation through the use of GenAI. The findings reveal two opposing yet complementary fundamental dynamics regarding the role of AI in design education: a quantitative cognitive expansion and a qualitative instrumental dominance. As detailed in the introduction, the practice of passive distributed cognition (Hutchins, 1995) conducted with pen and paper in architectural studios is evolving into a new dimension alongside the active and co-creative nature of LLMs (Rezwana & Maher, 2023). However, an empirical gap exists in the literature regarding the outcomes of the process when students, particularly those whose spatial mental schemas are not yet fully established, delegate this heavy task of topological network construction to the machine. Building upon this theoretical deficiency highlighted by Iranmanesh and Lotfabadi (2025), our study addresses human AI interaction not merely as a prompt output relationship, but rather through the lens of Extended Cognition theory (Clark & Chalmers, 1998), opening up the discussion of whether this interaction constitutes a genuine capacity enhancement for the student or a passivating boundary violation.

4.1. Cognitive Expansion or Qualitative Surrender?

The initial findings of the research demonstrate a significant growth of approximately 11.4% in the number of spaces and 17.9% in the number of edges in the final diagrams students achieved with AI compared to their manual generation. At first glance, this situation mathematically supports the argument put forward by Cila (2022) and Leach (2022) in the literature that AI is a prosthesis that stretches mental boundaries. AI taking over the cognitive load of novice designers presented them with more complex spatial networks and theoretically created a cognitive expansion.

However, when the stage in which students subjected the AI-generated network to curatorial arrangement is examined, serious doubts arise regarding the nature of this cognitive expansion. Students made only very limited reductions to the diagrams proposed by the system, decreasing the number of spaces by 2.3% and the number of edges by 5.5%, both of which were statistically insignificant. This finding suggests that the design act may shift from the intensive cognitive problem-solving activity described by Goel (1995) toward a more passive approval process shaped by the anchoring effect (Tversky & Kahneman, 1974). This interpretation is consistent with broader educational discussions warning that LLM-supported learning environments require critical thinking, human oversight, and explicit awareness of algorithmic limitations (Kasneci et al., 2023). Therefore, although a numerical expansion occurred, this expansion should be interpreted cautiously: it may represent the passive acceptance of AI-imposed complexity rather than an organic mastery achieved by students through their own spatial reasoning.

4.2. The Illusion of Dialogue Based Design and Instrumental Dominance

The concepts of dialogue-based design and co-creativity, framed in the introduction through Tong et al. (2023) and Rezwana and Maher (2023), are also problematized by the research findings. Recent architectural education studies have emphasized the need to structure GenAI-supported co-creative processes through pedagogical frameworks that combine creative cognition, algorithmic thinking, technical guidance, and ethical reflection (Hikmet & Ozay, 2026; Wang et al., 2025). In the present study, however, no statistically significant difference was found between the topological density achieved by students who began with a preliminary hand sketch and those who directly prompted the LLM. This suggests that, within this experimental setting, the LLM interaction may have exerted a stronger structuring effect on the final diagram than the students' preliminary cognitive preparation.

Furthermore, the correlation analysis regarding the number of iterations showed that the AI-supported dialogue tended to operate in an additive direction rather than as a process of refinement or simplification. As students continued to write prompts, the diagrams increased in topological volume instead of moving toward hierarchical resolution. This pattern suggests that LLM-mediated design dialogue may produce an illusion of co-creative negotiation while encouraging uncontrolled expansion if it is not accompanied by explicit spatial criteria and critical curatorial judgment.

4.3. Topological Bias of Models

Finally, the findings suggest that the AI models forming the invisible infrastructure of architectural design should not be treated as neutral tools. Despite using the same initial prompt, students who used other LLMs, predominantly ChatGPT, obtained more integrated and densely connected spatial networks from the beginning of the process compared with those who used Gemini. However, given the unequal group sizes and the heterogeneous composition of the "Other" model category, this result should be interpreted cautiously. Rather than proving a fixed model bias, the finding points to a possible model-specific tendency in the construction of architectural spatial networks. In this sense, extended cognition does not only imply the outward expansion of the human mind; it may also involve the incorporation of algorithmic tendencies into students' spatial reasoning and diagrammatic decision-making.

5. Conclusions

This study empirically confirms that GenAI carries a strong potential as a capacity-enhancing extended cognition tool in the architectural design process. The findings support H1 by showing statistically significant increases in both node and edge counts after LLM interaction. However, the limited curatorial reductions made by students and the positive association between iteration depth and node count indicate that this expansion was primarily additive rather than critically selective. In this respect, the study contributes to architectural design education literature by shifting the discussion of GenAI from visual representation and image-based production toward the measurable topology of functional diagrams. This contribution complements recent studies on AI-supported architectural programming, co-creative design pedagogy, and ethical AI integration in studio education (Jin et al., 2024; Wang et al., 2025; Hikmet & Ozay, 2026). Commercial AI models, which appear to structure the process through their hyper-productive character, may carry the risk of imposing continuously expanding topological networks lacking hierarchy instead of refining the design as iterations prolong. In order to truly benefit from the potential of these tools within ethical boundaries in architectural education, it is essential that students are equipped not only with prompt engineering skills but also with a strong spatial curatorial critical ability capable of filtering this authoritative and additive nature of AI.

This research has certain methodological limitations. The study is limited to 41 first year architecture students and a singular, relatively small-scale design problem such as a weekend house. Furthermore, the findings reflect the algorithmic characters of specific LLMs (predominantly Google Gemini and ChatGPT) accessible in the years 2025 and 2026. Future research repeating similar experimental setups comparatively with upper year experienced architecture students or professional architects will fill a critical literature gap in terms of observing whether human initiative increases with experience. Additionally, it is recommended that the transformation processes of text-based topological generations (functional diagrams) into three-dimensional spatial morphology (mass generation) are examined through longitudinal studio studies.

Acknowledgements

The authors would like to express their gratitude to the 41 first-year architecture students at the Faculty of Architecture, Bursa Uludağ University, who voluntarily participated in the study conducted during the MIM1004 course in the 2025–2026 spring semester. Their engagement and design iterations were essential to the completion of this research.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflicts of Interest

The authors declare that they have no competing interests.

Data Availability Statement

The datasets generated during the current study, including anonymised topological metrics, LLM interaction logs, and architectural diagrams, are available from the corresponding author upon reasonable request.

Institutional Review Board Statement

The study was conducted in accordance with the Declaration of Helsinki and approved by the Science and Engineering Research and Publication Ethics Committee of Bursa Uludağ University (Session No. 2026/06, Decision No. 2).

CRediT Author Statement

Conceptualisation: A.Y., A.I.Ç., G.K.Ö.; Methodology: A.Y., A.I.Ç., G.K.Ö.; Software: A.Y.; Formal analysis: A.Y.; Investigation: A.Y.; Data curation: A.Y.; Visualisation: A.Y.; Writing – original draft: A.Y.; Writing – review & editing: A.Y., A.I.Ç., G.K.Ö.; Supervision: A.I.Ç., G.K.Ö. All authors have read and approved the final version of the manuscript.

References

Cila, N. (2022). Designing human-agent collaborations: Commitment, responsiveness, and support. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (Article 304, pp. 1–18). Association for Computing Machinery. https://doi.org/10.1145/3491102.3517500

Clark, A., & Chalmers, D. (1998). The extended mind. Analysis, 58(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Cohen, J. (2013). Statistical power analysis for the behavioral sciences. Routledge. https://doi.org/10.4324/9780203771587

Creswell, J. W., & Creswell, J. D. (2017). Research design: Qualitative, quantitative, and mixed methods approaches. SAGE Publications.

Do, E. Y. L., & Gross, M. D. (2001). Thinking with diagrams in architectural design. Artificial Intelligence Review, 15(1), 135–149. https://doi.org/10.1023/A:1006661524497

Field, A. (2024). Discovering statistics using IBM SPSS statistics. SAGE Publications.

Fraenkel, J. R., & Wallen, N. E. (1990). How to design and evaluate research in education. McGraw-Hill.

Goel, V. (1995). Sketches of thought. MIT Press. https://doi.org/10.7551/mitpress/6270.001.0001

Hikmet, Ş., & Ozay, N. (2026). Human–AI collaboration in architectural design education: Towards a conceptual framework. Buildings, 16(6), Article 1097. https://doi.org/10.3390/buildings16061097

Hillier, B., & Hanson, J. (1989). The social logic of space. Cambridge University Press.

Hutchins, E. (1995). Cognition in the wild. MIT Press.

Iranmanesh, A., & Lotfabadi, P. (2025). Critical questions on the emergence of text-to-image artificial intelligence in architectural design pedagogy. AI & Society, 40(5), 3557–3571. https://doi.org/10.1007/s00146-024-02111-x

Jin, S., Tu, H., Li, J., Fang, Y., Qu, Z., Xu, F., Liu, K., & Lin, Y. (2024). Enhancing architectural education through artificial intelligence: A case study of an AI-assisted architectural programming and design course. Buildings, 14(6), Article 1613. https://doi.org/10.3390/buildings14061613

Karadağ, D. (2025). AI in architectural education: Rethinking studio culture. PLANARCH - Design and Planning Research, 9(2), 243–253. https://doi.org/10.54864/planarch.1749891

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of Chiropractic Medicine, 15(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Leach, N. (2022). Architecture in the age of artificial intelligence: An introduction to AI for architects. Bloomsbury Visual Arts.

MacFarland, T. W., & Yates, J. M. (2016). Introduction to nonparametric statistics for the biological sciences using R. Springer. https://doi.org/10.1007/978-3-319-30634-6

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. https://doi.org/10.3846/jau.2024.20870

Ostwald, M. J. (2011). The mathematics of spatial configuration: Revisiting, revising and critiquing justified plan graph theory. Nexus Network Journal, 13(2), 445–470. https://doi.org/10.1007/s00004-011-0075-3

Rezwana, J., & Maher, M. L. (2023). User perspectives on ethical challenges in human–AI co-creativity: A design fiction study. In Proceedings of the 15th Conference on Creativity and Cognition (pp. 62–74). Association for Computing Machinery. https://doi.org/10.1145/3591196.3593364

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Siegel, S. (1957). Nonparametric statistics. The American Statistician, 11(3), 13–19. https://doi.org/10.1080/00031305.1957.10501091

Tong, H., Türel, A., Şenkal, H., Yagci Ergun, S. F., Güzelci, O. Z., & Alaçam, S. (2023). Can AI function as a new mode of sketching? A teaching experiment with freshman. International Journal of Emerging Technologies in Learning (iJET), 18(18), 234–248. https://doi.org/10.3991/ijet.v18i18.42603

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods, 17(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wang, J., Shi, Y., Chen, X., Lan, Y., & Liu, S. (2025). Teaching with artificial intelligence in architecture: Embedding technical skills and ethical reflection in a core design studio. Buildings, 15(17), Article 3069. https://doi.org/10.3390/buildings15173069