Artificial Intelligence as a pedagogical tool for architectural education: What does the empirical evidence tell us?

Mohamed Sadek\textsuperscript{1,a}, and Nermine Abdel Gelil Mohamed\textsuperscript{2,b,*}

\textsuperscript{1,2} MSA University, 6th of October City, Giza, Egypt.
E-mail: \textsuperscript{a} mrafik@msa.edu.eg (Corresponding author), \textsuperscript{b} ngelil@msa.edu.eg

Abstract

Artificial Intelligence (AI) applications in arts and design have been growing in popularity in recent years. Numerous AI program applications are publicly available, both commercial and open source, offering a chance to convert the textual description of a design or context into original content images. While many research projects tackled AI use in architecture since 2015, there seems to be a lack of empirical evidence of its use in architecture education as a tool for concept generation. An experimental architecture design course took place in Fall 2022 where 34 students utilised AI programs in the conceptual design phase to produce original content images from narratives to help students generate concepts for their buildings. The designs of the experimental student groups are compared to another control group of 50 students which use more traditional conceptual design methods on a similar project in the same semester. This paper aims to highlight the opportunities and challenges of using AI tools in the concept generation phase in architecture design education by measuring their effectiveness to increase creativity in form finding. With the use of recent literature in architecture learning theories and pedagogies, this experimental approach will be critiqued to provide further insights into AI’s role in architecture education.

Keywords: Artificial Intelligence, Architecture pedagogy, Concept generation, Architecture design, Form finding.
1. Introduction

1.1. Research Background

Architecture education helps students be prepared for their professional design careers with a continuous effort to catch up with the technological advancements available to the vocation [1], [2]. One of the most critical and stressful tasks for an architecture student is to come up with an idea for their building form. These ideas or ‘concepts’ are developed through several one-to-one sessions where students get professional and academic feedback [3]. Students use traditional concept generation techniques which are widely used, such as model making, site visits, conceptual sketching, and physical or digital modelling. Oxman [4] has studied new practices and relations between the architecture student and the conceptual design process while integrating novel techniques, such as parametric formation, etc. This provided novel paths for concept exploration. However, the recent developments of Artificial Intelligence (AI) applications in conceptual design have been rarely used in design studios. A wholistic definition of AI is stated as “the activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment” [5, p. 905]. The slow progress in integrating AI in architecture curriculums, in some instances, has been attributed to a strong rejection by the studio’s professors to use digital tools and Computer Aided Design (CAD) programs in conceptual design, thinking that these tools would diminish the manual drafting and design skillset of architecture students [6]. Yet, Salman et al. [7] argue that while the digital conceptual design process has become a valid alternative, this does not exclude all other media that can aid the students to reach a suitable concept and form for their design.

1.2. Research rationale and significance

Considering the dizzying range of digital, traditional, and mixed tools now available to future architects, the following questions emerged: Can the combination of narratives and storytelling with AI image-producing applications produce creative concepts? Can AI aid students reach a mature architectural design in a time-efficient manner? This gap of knowledge inspired an experimental approach in an architectural design course, where students were asked to use narratives and AI-generated imagery as primary concept generators for their building form-finding. In addition to providing insights on the effectiveness of including AI image generation in the design studio, this research project adds to the discourse of figuring out the role of the architect in a future where AI takes on an increasing role in the arts. Thus, the following hypotheses are tested:

- Student projects that used the combination of narratives and AI imagery obtained better results than their traditional counterparts.
- Concepts generated from combining narratives with AI imagery were more creative than their traditional counterparts.
2. Literature review

2.1. AI in art and humanities: Is AI a true art form?

Goodfellow et al. [8] proposed a framework that uses a machine learning (ML) algorithm known as a generative adversarial network (or GAN) to generate images. This method produced a portrait entitled ‘Edmond De Belamy’, which was sold for the same amount as a collection of Adolf Hitler’s original artwork painted before his horrendous war crimes [9]. This was mainly because the auction house justifiably advertised the portrait as the first painting generated by an algorithm sold in an auction. The question that poses itself is, can AI-generated art be judged as ‘true’ art? To answer this question this section refers to three main criteria that define true art, namely, sociability, creativity, and incentive. First, historians refer to sociability to clarify the bonds that are consciously or unconsciously initiated to form social relationships between people [10]. This intersubjective interpretation of reality is the main drive behind many works of Art, for example, Picasso’s painting Guernica was a highly appreciated work of art, despite that it was nonconformist with his early work (see Figure 1), mainly because of the association of the Spanish Civil War and the bombing of the Basque town. AI images still lack this element as it only relies upon a set of data and an algorithm to create an acceptable form of art.

![Guernica and Lola portrait](https://example.com/guernica.png and https://example.com/lola.png)

Fig. 1 (Left) Guernica © Moleskine, [CC BY-NC-SA 4.0](https://example.com/guernica.png) (Right) (Picasso’s sister) Lola portrait © Gandalf’s Gallery, [CC BY-NC-SA 2.0](https://example.com/lola.png)

Second, creativity is defined by Boden [11] as everyday designs and innovations that are uniquely different from the individual artist or social perspective. Yet the leaps taken by artists that start a new creative art that has never been seen or attempted is yet to be mimicked by AI. For example, while he was not appreciated and fully understood in his own time, Claude Monet paved the path towards modernism by developing a unique style, impressionism. This style aimed to capture the human act of perceiving natural landscapes. Unfortunately, at this stage in their development, AI and ML cannot surprise us with an entirely new artistic expression. However, Wingström, et al. [12] evaluated the creativity of AI based on the five definitions: actor, process, outcome, domain, and space. In their study, the artists recognised their research before undertaking an artistic project to be classified as co-creativity. Hence, the authors argue that while creativity remains a human endeavour, creativity in the era of AI must be amended to co-creativity.
Third, what drove Vincent Van Gogh to draw the starry night the way that he did? The artist mentioned in the correspondence he shared with his brother “It is good for me to work hard. But that does not keep me from having a terrible need of—shall I say the word—yes, of religion. Then I go out at night to paint the stars” [13, p. 596]. If we inspect the painting with a keen eye for details, we can see the church in the middle of the painting is unlit (see Figure 2). From a psychiatric standpoint, Blumer [14] connects this artistic expression to the artist’s marked depression following the rejection by the church of his radical charitable actions. At that time, these acts were incompatible with his role as a member of the clergy. However, AI art has not reached this step of self-expression, of using intrinsic emotions to inspire their art. In the far future, perhaps, more conscious versions of AI would express their mistreatment by human beings in original content images portraying their emotions. Moreover, Van Gogh underlined the role of perseverance, persistence, and learning in the creative endeavour, recognising due diligence and devotion in the creative fields as prerequisites to create a masterful painting [15]. All this hard work is driven by the artist’s intrinsic motivation to express his mind and emotions. Presently, AI is driven by a human-made algorithm, not an intrinsic incentive. For example, an AI that is programmed to make art cannot make the choice not to create an art piece as a protest to a social problem [16].

To close this section, it is noteworthy to emphasise that AI is not an independent entity detached from human minds. The definition of “Art” is the application of imagination and skill by human beings to produce designs that have emotional power or beauty [17]. ‘Edmond de Belamy’ followed an algorithm that is based on the hard work and creativity of human beings who judged it to be an artistic contribution [18]. "Architecture is an art, yet we rarely concentrate our attention on buildings as we do on plays, books, and paintings." [19]. Hence, the artistic nature of architecture demands the investigation of AI use in the concept generation phase just as it is required for the arts in general.

![Starry Night by Vincent Van Gogh](https://msaeng.journals.ekb.eg//)

Fig. 2 Starry Night by Vincent Van Gogh © Ismoon, [CC BY-NC-SA 4.0](https://msaeng.journals.ekb.eg//)
2.2. Narratives in architecture: How can narratives be used in concept generation?

While architects are captivated by narratives, novelists are captivated by architecture. Since Dedalus wrote the ‘labyrinth’, all the way to gothic places illustrated by Edgar Alan Poe [20]. Whether or not it is written explicitly for architecture, stories create a cognitive spatial experience for the recipient [21]. Architecture form and storytelling have a sort of symbiotic relationship, narratives have the potential to inspire the architectural design and vice-versa. For example, Frank Lloyd Wright’s famous masterpiece Falling Water inspired Kelcey Parker’s Liliane’s Balcony, where he narrates the intriguing story of new residents haunted by the lives of the previous occupiers [22]. On the other hand, there are many architecturally significant buildings inspired by literary work, for example, an apartment block by Ricardo Bofill built in 1968 in Barcelona, is inspired by Franz Kafka’s The Castle. The cubic forms and confusing gateways highlight the influence of Kafka on Bofill [23]. When architects speak about their concept in a building, they usually refer to it as a narrative using a hypothetical space user and a journey through time and space [20]. While architecture education sometimes uses narration in theoretical lectures, it is not common to find it in the design studio. Nazidizaji et al [21] conducted a study to show the difference between narration and lecturing and found that narration produced better results for architecture students than its counterpart. While listening to or telling a story, our imagination produces a unique cognitive spatial element [24]. Thus, while an architect narrates their ideas in a story, a sort of internal hypothetical conversation happens between the space user and the designer that produces a description of the space.

2.3. AI in education: What can previous experiences teach us?

The fast pace of technologies and AI-related software in design pedagogies has caught many architecture schools unprepared to integrate them into their curriculums at the same pace [25]. However, AI tools are already employed in the educational process including curriculum development, methods of teaching, and assessment of student work [26]. Similarly, before GAN generated imagery, AI took many forms in education. For example, Ceylan [2] explains how AI is integrated with theoretical, technical, and representational architecture modules.

In theoretical courses, AI is useful in collecting, storing, and analysing textual data. This offers intelligent solutions to reduce the amount of time and labour required from students to collect and interpret vast amounts of data [27]. For example, AI in learning foreign languages led to the creation of ICALL (intelligent computer-assisted language learning), which supports students but also prevents unnecessary workload and useless monotonous tasks, e.g. via writing assistants and correction systems [28]. There seems to be a gap in the literature when it comes to theoretical architecture education empowered by AI tools, this can be explored in future research.

In technical modules, AI-friendly software like Building Information Modelling (BIM) and parametric design packages can generate 3D spatial data for design optimisation and construction purposes. Ceccon and Villa [29] explain that BIM and AI as two core technologies uniting with one another into an entirely new way of modelling reality. The present problem with BIM, according to a study conducted by Holzer [30], is the academic
focus on the tool while neglecting the fact that students should be taught how it is related to the actual construction in real-world contexts. Moreover, Guney [25] lists some disadvantages to CAD, specifically BIM, including decreasing the students’ imagination and creativity, technology addiction, visuals that outweigh design quality, and less contact between students and advisors. However, BIM software packages have started to embrace parametric design philosophy and AI applications to increase the level of flexibility early on in the design process [30]. This can be seen in the efforts of Park [31] to create a parametric model of traditional Korean houses using BIM software which can potentially preserve traditional Korean architecture by giving an economic advantage via a smart automated design process.

AI also impacts architecture representation through Virtual reality (VR), Augmented Reality (AR), and 3d printing technologies. These technologies contribute to spatial recognition, especially in terms of physical appearance [1]. However, it is mistaken to fully replace social interaction with AI. It should complement the conventional learning process [26]. Finally, while Ceylan [2] states that AI creates fast, efficient, and unconventional methods for architecture visualisation, there was no mention in the literature of GANs or any of their applications being implemented in a design studio. Thus, this research can present significant insights into using this tool in architecture education.

3. Methodology

3.1. Sampling

To answer the research questions mentioned and statistically test the effect of using narration and AI image generation in the concept generation phase, a quasi-experimental method of research was applied. This study consisted of 84 undergraduate students at the department of architecture engineering at MSA University, Giza, Egypt. The subjects are separated into two groups. One group of 34 students used narratives and AI imagery in their concept generation phase, while the control group consisted of 50 students who used traditional forms of concept generation. Both groups have a similar project of mixed-use high-rise structures. Out of the 34 students in AI group (A), 22 agreed to participate in the study, used the proposed method to a significant extent, and had all the required grades, and similarly 47 out of 50 students in the control group (B).

3.2. Methods and AI platforms

The design process of the AI test group (A) followed a similar design workflow as the control group (B) as seen in Figure 3, except for using the concept narratives and AI imagery on two occasions. First, the novel concept generator was introduced in an introductory lecture then the students were asked to write a 250 words story of a new occupier of the tower complex, and then they fed the AI platforms, such as Craiyon (formerly known as DALL-E Mini) and Midjourney, with prompts driven from the narratives (see Figure 4). These platforms have been selected based upon their ease of access, user-friendly interface, and their pattern recognition [32]. Second, a round of one-to-one feedback with each student, in addition to carrying out similar case study research and site analysis, led to an iterative loop between narratives and AI imagery till they settled on a form. Afterwards, group (A) followed a more-or-less similar path to group (B) till the final project submission.
Fig. 3. (Top) AI Group (A) design studio workflow (Bottom) Control group (B) design studio workflow
3.3. Data collection and analysis

To test the effect of the process on the final design output and the creativity of the form generated, data was collected from students via online surveys with e-signed consent forms, and the university student database after the ethical approvals were obtained. The data obtained from the student university database helped to create a mean for the design studio grade for each participating student, which is then compared to the final result of the design studio (A) and (B) to check whether they improved or worsened. Furthermore, to measure the creativity of the form, the grade given to the concept sketch for both groups was compared in the following section. It is important to note that only the external juror’s grade was taken as the final grade in group (A) to avoid any partial grading from the researcher. Both tests were undertaken in SPSS software using independent two-tailed t-tests to compare the means, then Levene’s test was used to examine the equality of variance from the data’s absolute centre [33]. Finally, the survey included several statements with Likert scale, a psychometric scale widely used scale in survey studies [34], to measure the creativity of the form from the student’s perspective and the effectiveness of using the proposed method in different stages in the project workflow.
4. Results and discussion

The following results were compared through an independent t-test using SPSS. The mean average of the group (A) students showed they improved by (0.8636), while group (B) students' mean worsened by (-5.3191) (see Table 1).

Table 1. Descriptive statistics for the grade difference in each group

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade Diff.</td>
<td>A</td>
<td>22</td>
<td>.8636</td>
<td>6.79142</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>47</td>
<td>-5.3191</td>
<td>9.13987</td>
</tr>
</tbody>
</table>

Since the two groups are independent of each other, the means of both were measured mainly to test the following hypothesis:

- Final design outputs that used the combination of narratives and AI imagery obtained better results than their traditional counterparts.

The hypothesis is then devised as such:

H0: (Null hypothesis): \( \mu_A \leq \mu_B \)

H1: (Contrastive hypothesis): \( \mu_A > \mu_B \)

To test the significance of the results, an appropriate statistical test was carried out. The equality of variances was tested the data listed in Table 2 can be examined as follows: Since the p-value of Levene's test is 0.350, which is bigger than 0.05, we can fairly assume that the variance of the two groups is the same. With equal variance assumed, since the p-value of the two-tailed t-test is 0.006, which is smaller than 0.05. Thus, we reject the null hypothesis and conclude that there is a difference between the mean grade of groups (A) and (B) at 5% significance level.

Table 2. Equality of means for final output grade difference mean

<table>
<thead>
<tr>
<th>Levene's Test</th>
<th>t</th>
<th>df</th>
<th>Sig.(2-tailed)</th>
<th>Mean Diff.</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade Diff.</td>
<td>.350</td>
<td>2.824</td>
<td>.006</td>
<td>6.18279</td>
<td>1.8134</td>
<td>10.5521</td>
</tr>
<tr>
<td>Equal var. assumed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal var unassumed</td>
<td>3.141</td>
<td>53.98</td>
<td>.003</td>
<td>6.18279</td>
<td>2.2367</td>
<td>10.1288</td>
</tr>
</tbody>
</table>

95% Confidence Interval
After testing the difference of grades (out of 100) in the concept sketch for both groups, the mean average of group (A) students is (76.5909), which is higher than that of group (B) (73.9894) (see Table 3).

Table 3. Descriptive statistics for the concept sketch grade in each group

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept Grade A</td>
<td>22</td>
<td>76.5909</td>
<td>13.92038</td>
<td>2.96784</td>
</tr>
<tr>
<td>Concept Grade B</td>
<td>47</td>
<td>73.9894</td>
<td>9.67768</td>
<td>1.41163</td>
</tr>
</tbody>
</table>

These means were measured mainly to test the following hypothesis:

- Concepts generated from combing narratives with AI imagery were more creative than their traditional counterparts.

The hypothesis is then devised as such:
H0: (Null hypothesis): μA ≤ μB
H1: (Contrastive hypothesis): μA > μB

The equality of variances was tested the data listed in Table 4 can be examined as follows: Since the p-value of Levene’s test is 0.450, which is bigger than 0.05, we can fairly assume that the variance of the two groups is the same. With the equal variance assumed, the p-value of the two-tailed t-test is 0.371. Since the p-value here is greater than 0.05, we fail to reject the null hypothesis and cannot conclude that there is a difference between the mean concept sketch grade of groups (A) and (B) at 5% significance level.

Table 4. Equality of means for concept sketch grade mean

<table>
<thead>
<tr>
<th>Grade</th>
<th>Diff.</th>
<th>Levene's Test</th>
<th>t</th>
<th>df</th>
<th>Sig.(2-tailed)</th>
<th>Mean Diff.</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>.450</td>
<td>.901</td>
<td>67</td>
<td>.371</td>
<td>2.601</td>
<td>-</td>
<td>8.36719</td>
</tr>
<tr>
<td></td>
<td>equal var assumed</td>
<td>55</td>
<td>3.16410</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.792</td>
<td>30.8</td>
<td>.435</td>
<td>2.601</td>
<td>-</td>
<td>9.30558</td>
<td></td>
</tr>
<tr>
<td></td>
<td>equal var unassumed</td>
<td>56</td>
<td>4.10248</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Finally, to measure the creativity of the concepts generated and the final design output from the students' perspectives in the test group (A), the answer to the statements in the survey were analysed as seen in Table 5.
Table 5. Descriptive statistics to measure the students' perspective on using AI in concept generation

<table>
<thead>
<tr>
<th>Statement</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI was useful in your form finding to produce a creative design morphology.</td>
<td>22</td>
<td>3.00</td>
<td>5.00</td>
<td>4.0909</td>
<td>.86790</td>
</tr>
<tr>
<td>The AI images and Narrative I used were helpful in my design project even after the form-finding phase.</td>
<td>22</td>
<td>2.00</td>
<td>5.00</td>
<td>3.5000</td>
<td>.91287</td>
</tr>
<tr>
<td>I will use AI imagery in different designs in the future.</td>
<td>22</td>
<td>2.00</td>
<td>5.00</td>
<td>4.0000</td>
<td>.81650</td>
</tr>
</tbody>
</table>

Note. 5 strongly agree, 4 agree, 3 neutral, 2 disagree, 1 strongly disagree

According to Pimentel [34], the five-point Likert scale is deemed an interval scale. ‘Strongly disagree’ is measured from 1 to 1.8, and ‘disagree’ from 1.81 to 2.60. ‘Neutral’ is measured in the interval from 2.61 to 3.40. ‘Agree’ falls between 3.41 to 4.20, and ‘strongly agree’ from 4.21 to 5. In the first statement, the mean is 4.0909. Thus, it means that the majority agree that narratives and AI imagery produced creative conceptual forms. The mean of the second statement is 3.5, thus most of the students agree that even after the concept generation phase was over, the narrative and AI were still useful in their design workflow. The third statement about whether or not, the students will use this technique in the future, the mean measured was 4.0. Accordingly, the majority agrees on using this technique in future designs.

5. Conclusions

The purpose of this study was to measure the effectiveness of using architectural narratives combined with the powerful tools of AI-generated imagery in increasing the creativity of the form generated to eventually end up with a decent architecture design project. From an external assessment perspective, this study statistically concluded that this method enhanced the final design outputs for architecture students. However, when its effect in the concept generation phase was measured, the results were complimentary to the hypothesis, yet statistically insignificant. A possible explanation can be the unorthodox nature of this method to the students. It required more time and effort to get accustomed to narration and AI image generation before reaching a mature concept.

For using this method in the design studio, the recommendations are as follows:

- Introduce the method as early as possible in the design studio to give sufficient time for the students to learn how to best employ the tools.
• This method is best used when the focus of the design studio project is towards creative form generation. It is not known if the proposed method will produce similar results in a more pragmatic or functional design project.

• The students should be exposed to architectural narratives in their early stages in the architecture program as it is essential for them to know appropriate literary terms to produce efficient text prompts for AI software.

• AI concept imagery cannot be independent of human beings, there needs to be a conversation between both parties to produce a mature concept and architectural form.

While this study has some limitations, namely, the small student population of the narrative-AI test group and the possible discrepancy in the grading criteria for the assessors, the experimental approach contributed to existing knowledge by providing insights into how AI can be added to future architects’ skillset.

Acknowledgement

The authors would like to thank the staff in the architecture department in the faculty of engineering at MSA University for their cooperation in the data collection. First and foremost, we wish to thank the students who outdone themselves and made a huge effort to experiment with novel methods in their design projects.

References


