Human-AI Co-Creation: Evaluating the Impact of Large-scale Text-to-Image Generative Models on the Creative Process

Tommaso Turchi[1[0000-0001-6826-9688]], Silvio Carta[2[0000-0002-7586-3121]], Luciano Ambrosini[3[0000-0003-1529-2694]], Alessio Malizia[1, 4[0000-0002-2601-7009]]

[1] Department of Computer Science, Università di Pisa, Pisa, Italy
{name.surname}@unipi.it
[2] Department of Architecture and Design, University of Hertfordshire, Hatfield, United Kingdom
s.carta@herts.ac.uk
[3] Independent researcher
luciano.ambrosini@outlook.com
[4] Molde University College, Molde, Norway
alessio.malizia@himolde.no

Abstract. Large-scale Text-to-image Generative Models (LTGMs) are a cutting-edge class of Artificial Intelligence (AI) algorithms specifically designed to generate images from natural language descriptions (prompts). These models have demonstrated impressive capabilities in creating high-quality images from a wide range of inputs, making them powerful tools for non-technical users to tap into their creativity. The field is advancing rapidly and we are witnessing the emergence of an increasing number of tools, such as DALL-E, MidJourney and StableDiffusion, that are leveraging LTGMs to support creative work across various domains. However, there is a lack of research on how the interaction with these tools might affect the users' creativity and their ability to control the generated outputs. In this paper, we investigate how the interaction with LTGMs-based tools might impact creativity by analyzing the feedback provided by groups of design students developing an architectural project with the help of LTGMs tools.

Keywords: Generative AI, Creativity, Human-AI, AI-driven design process.

1 Introduction

In the past year we have witnessed the rise of impressive AI-based tools capable of generating images from textual descriptions, holding coherent conversations, providing writing suggestions for creative writers, and even writing code alongside a human programmer. All these examples share a common characteristic: the AI does not simply categorize data or interpret text based on predetermined models, but instead it generates something entirely new such as images or designs. This type of work pushes the potential of AI systems beyond problem-solving and towards problem-finding, which often results in the AI functioning as a creative human collaborator and supporter rather than a decision-maker.

These tools are often based on Large-scale Text-to-Image Generative Models (LTGMs), a rapidly evolving class of Artificial Intelligence (AI) algorithms with the ability to generate images based on natural language descriptions, called prompts. These models have shown remarkable capabilities in creating high-quality images from a wide range of inputs, making them powerful tools for non-technical users to tap into their imagination. As the field of LTGMs continues to advance, we are witnessing the growing adoption of a number of tools, such as DALL-E, MidJourney, and Playground, that are leveraging the power of these models to support work in various creative domains.

However, despite their increasing popularity, the impact of LTGMs-based tools on creativity remains largely understudied. The users' ability to effectively direct and control a creative support tool to fit their needs is an essential component of the creation process and plays a crucial role in determining the successful outcome of a project. Therefore, it is important to understand how the interaction and collaboration between humans and AI through these tools might affect creative processes.

In this paper, we address this gap by investigating the impact of LTGMs-based tools on creativity. Our study analyzes post-hoc feedback provided by different groups of design students as they work on an architectural project with the support of some of these tools. We aim to gain insights into how the interaction with LTGMs affects the students' creative process, focusing on the ability to effectively control them to generate new ideas.

Therefore, the research question tackled in this work is: *how does the interaction with LTGMs affect users' creativity?*

This research provides a valuable contribution to the field of End-User development by exploring the impact of Human-AI Co-Creation on users. Our findings will inform the development of future tools and investigate their use in creative work.

2 Related Works

Generative AI and Text-to-Image Generative Models

Generative AI refers to a new class of Artificial Intelligence models that create new content, as opposed to simply analyzing existing data like Expert Systems do. These Generative Models consist of a discriminator (or transformer) and a generator, trained on a dataset and can map input information into a high-dimensional space, producing novel content on each new trial, even from the same input. Thus, unlike predictive Machine Learning systems, Generative Models can both discriminate information and generate new content [1]. Within the domain of architecture and spatial design, the automated generation of spatial configuration has a long tradition, starting with the seminal work of Shape Grammars [2] in the 1980s, developing with Spatial Synthesis [3], with more recent developments with more sophisticated graph-based models [4, 5]. A detailed account of such developments in architecture can be found in [6].

The recent growth of Generative AI is due to the availability of large datasets and the latest advancements in computing power. Such models can map any input format, like text, to any output format, like video or images, allowing the generation of new media from prompts-like text inputs, or a set of relevant images. The taxonomy of

existing systems mapping the different input formats to different outputs is growing day by day, as new models are introduced to jumpstart new domain-specific applications [1].

The availability of massive datasets and a wide range of use cases enabled by their widespread has contributed to the rapid development of Text-to-Image Generative Models, with new tools emerging on a daily basis. These models can be exploited by different disciplines such as architecture or product design, and throughout many phases of the creative process: ideation, sketches, variants building, texture creation, … [7]. They can be used to spark new ideas and inspire innovative designs.

Developing such large-scale Generative AI Models proved to be a challenging task, as the estimation of their parameters requires enormous computational power and a highly skilled and experienced team in data science and engineering [1]. Thus, only a handful of companies have been successful in deploying Generative Models.

Among the firsts, StabilityAI introduced Stable Diffusion in 2022 and its main purpose is to generate highly detailed images based on textual descriptions [8]. Additionally, it can be utilized for other tasks like image editing and image translation. The model is trained on 512x512 images from "2b English language label subset of LAION 5b, a general crawl of the internet created by the German charity LAION"[1]. The model incorporates a fixed CLIP [9] ViT-L/14 text encoder to influence the model's output based on text inputs.

Most notably, OpenAI created DALL-E 2, which is an improvement over its predecessor DALL-E. It generates more lifelike images at higher resolutions and has the ability to blend together concepts, attributes, and styles [8]. DALL-E 2 was trained using approximately 650 million image-text pairs obtained from the Internet. A clear comparison[2] where salient points are summarized in table 1 below.

DALL-E 2 (OpenAI)	**Stable Diffusion (Stability.ai)**
code is not open-source	code is open-source
Training data are not disclosed and publicly available	Training data are disclosed and publicly available
The model uses heavily curated data. This results in strict control of the outputs.	Training data are generally non-curated. The model can generate uncontrolled images.
The model uses GPT-3 and its large number of parameters (over 175 billion machine learning parameters). This allows for a high capability of generation of unseen visuals.	The model uses a diffusion technique based on existing data. Outputs are restricted to training images (limited capability of generating unseen visuals).

Table 1. Comparison between DALL-E 2 and Stable Diffusion models.

These two Generative Models were selected and tested in our study with the help of a purposefully-designed tool integrating them into the participants' workflows, as reported in Section 3.

[1] https://stability.ai/blog/stable-diffusion-public-release
[2] https://nimblebox.ai/blog/stable-diffusion-ai

Human-AI Co-Creation

The idea of humans and AI agents collaborating to achieve creative endeavors is becoming increasingly common, stemming from a long tradition of Computer-Supported Cooperative Work and creativity support systems, thanks also to the recent popularity of Generative AI Models. This contributed to the rise of a new research area called Human-AI Co-Creation [10], involving both the human and the AI contributing to the creative process and sharing responsibility for the resulting artefact.

Nonetheless, while powerful Generative AI Models are now commonly available to designers, artists [11], and knowledge workers, there is still much to learn about how to make these tools interactive and design effective user experiences around them. Additionally, little is known about the long-term effects of this technology on creative practices, the overarching role of Generative AI in society as a whole, and the regulations that will govern this area of design [12].

A recent literature survey [13] showed how fostering productive use in Human-AI Co-Creation systems is still a challenge. Researchers found that many such systems failed to achieve positive synergy, which refers to the ability of a Human-AI team to produce superior outcomes compared to either party working alone. In fact, some studies have even found the opposite effect, with Human-AI teams producing inferior results compared to a Human or AI working alone [14].

Furthermore, fostering the safe use of Generative AI is also a challenge due to the potential risks and harms associated with these systems. These risks can stem from how the model was trained [15] or how it is applied [16].

Several theoretical frameworks [17, 18, 19] have been proposed to guide the design of these systems and to make sure that the collaboration between Humans and AI is fruitful. However, there is still a limited amount of on-the-field studies investigating how this synergy impacts creative outcomes and affects existing design processes, now more than ever with the rising popularity of new tools. With this study, we aim to collect user feedback on these models and analyze how these tools can impact creative works in a real-world scenario.

3 User Study

This section presents the goals, hypotheses, and description of the user study we carried out, following the guidelines of Wohlin et al. [20].

Goals

The goal of this study was to collect user feedback on how LTGMs-based tools can affect creative works. The purpose is to evaluate their impact on creativity.

Research Questions

Large-scale Text-to-Image Generative Models (LTGMs) can generate high-quality images from natural language descriptions, and they are increasingly used to support work in creative domains. However, their impact on creativity remains understudied,

and it is important to understand how users can effectively direct and control these tools to generate new ideas.

The main research question derived from this context is: *how does the interaction with LTGMs affect users' creativity?*

Participants

Overall, 22 students took part in the workshop in mixed groups representing the spectrum of genders, age, discipline (architecture, urban design and interior architecture) and level of study (undergraduate and postgraduate) of the student cohorts in the Department of Architecture and Design at the University of Hertfordshire, UK. Students were divided into 8 groups of 3-4 students. Generally speaking, none of them had prior experience of Generative AI Tools for design, and all had experience in developing architectural projects in both academic and professional settings.

The students were asked to develop design solutions responding to a given design brief called the Art of Bathing. This brief required to create a public repository of water, namely a building serving as a sanctuary for individuals to contemplate, meditate, replenish, and heal, from the daily pressure of life. The students were tasked to design a structure, with a maximum building envelope of 10x10x10m in Stanborough Park, Welwyn Garden City (UK), characterized by a rich sensory experience and spatial configuration. Students were asked to create a building concerned not simply with style, image or beautiful materiality, but resonant with memories of volumetric weight, contiguity and enclosure of space, as well as sound and light effects related to water.

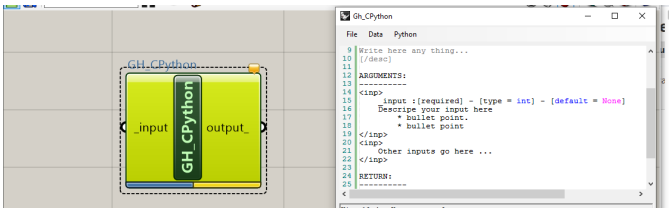

Figure 1. Example of scripting feature on Grasshopper.

The workshop took place in one of the computer labs at the University of Hertfordshire, UK. In order to facilitate the workshop, we developed a suite of Grasshopper (GH) Components to allow students to experiment with the different options provided by OpenAI and Stability.ai. Grasshopper[3] is a node-based visual programming environment working within McNeel's Rhinoceros 3D software widely used in architecture and design industry and research. Rhino and GH allow for great tool customization by including scripting capability through Visual Basic (VB), C# and Python for developers.

Grasshopper has been used for this workshop since it allowed us to generate ad hoc scripts to introduce diffusion models into the design process, and it also represents a familiar design environment for the students.

[3] https://www.grasshopper3d.com/

The tools used in the Grasshopper (GH) definitions distributed to students belongs to the "AI" subcategory included within the "Ambrosinus-Toolkit" plugin[4]. These services are underpinned by two models: DALL-E 2 [21] and Stable Diffusion [22] respectively. Both the OpenAI and StabilityAI platforms allow students to perform three different types of image generation: Creative mode, Variation mode and Edit mode. The study presented therein will comply with the first two methods which are performed differently by the two aforementioned platforms.

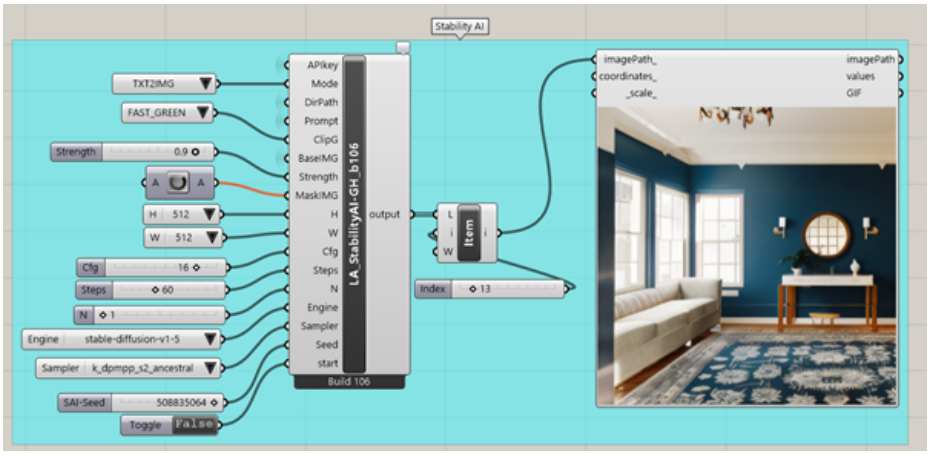

Figure 2. Grasshopper with Stable Diffusion model.

There are two GH tools that perform these operations[5]: "OpenAI-GHadv" and "StabilityAI-GHadv". The first tool processes images through the neural model called DALL-E (v.2), while the second tool processes images through the neural model called Stable Diffusion. Before quickly illustrating some of the most significant parameters that will be used in this experiment by the students, it is important to underline that the substantial difference between the two neural models is that DALL-E can not discretize the generative process in the current version, while the Stable Diffusion model can and this translates into the possibility of making the outputs, and therefore all the parameters used, completely identifiable and recallable with the same settings.

OpenAI-GHadv main parameters:

- *Mode*: execution mode selector (Create, Variation and Edit);
- *BaseIMG* is the source image path, it is required in order to run the Variation and Edit modes;

[4] The author of the plugin started to develop and share the Toolkit in November 2022 and the current version is v1.1.6 (2023/02/06). The Main AI components are available from this GitHub page: https://github.com/lucianoambrosini.

[5] The current version of the tool "LA_OpenAI-GHadv" is the build 111 and that one of the tool "LA_StabilityAI-GHadv" is the build 107.

- *MaskIMG* is the source image-mask path (it will be ignored in this experimentation);
- *DirPath* is the target path where the tool will store the generated image by DALL-E;
- *Prompt* is the textual description passed as input;
- *S* is the image size. The pre-trained model used for the training process of the DALL-E allows only these sizes: 256, 512 and 1024 pixels (it is allowed only squared pictures);
- *N* is the number of images to generate;

StabilityAI-GHadv main parameters[6]:

- *Mode* is the execution mode selector (TXTtoIMG, IMGtoIMG and IMGtoIMG Masking);
- *DirPath* is the target path where the tool will store the generated image by Stable Diffusion;
- *Prompt* is the textual description passed as input;
- *ClipG* is the CLIP guidance mode. It is a tricky procedure executed by the neural network encoder to increase the consistency of the image with the text given as input;
- *BaseIMG* is the source image path, it is required in order to run the IMGtoIMG mode;
- *Strength* is how much "weight" has the text prompt in relation to the initial image (admitted values from 0.0 to 1);
- *MaskIMG* is the source image-mask path (it will be ignored in this experimentation);
- *H* and *W* are the height and width of the output image. Only the size in the range 256 to 1024 pixels with 64px as increment value is admitted;
- *Cfg* scale dictates how closely the engine attempts to match a generation to the provided prompt; v2-x models respond well to lower CFG (eg: 4-8), whereas v1-x models respond well to a higher range (eg: 7-14);
- *Steps* affect the number of diffusion steps performed on the requested generation.
- *N* is the number of images to generate;
- *Engine* current version includes eight engines[7], all selectable by the user;
- *Sampler* is the sampling engine to use. Currently have been implemented in the toolkit nine samplers. They are a sort of statistical samplers that are used in the diffusion model prediction process (especially for the denoising process);
- *Seed* is an integer number useful to discretize all parameters used in the generative process. The toolkit assigns a random value if no slider is connected to the *Seed* input.

[6] This workshop focused only on text-to-image and image-to-image procedures, so all "masking" mode parameters have been skipped in this description.

[7] The engines are different neural models that are developed by the use of a specific pre-trained set of images with different sizes. More info here: https://stability.ai/blog/stable-diffusion-v2-release

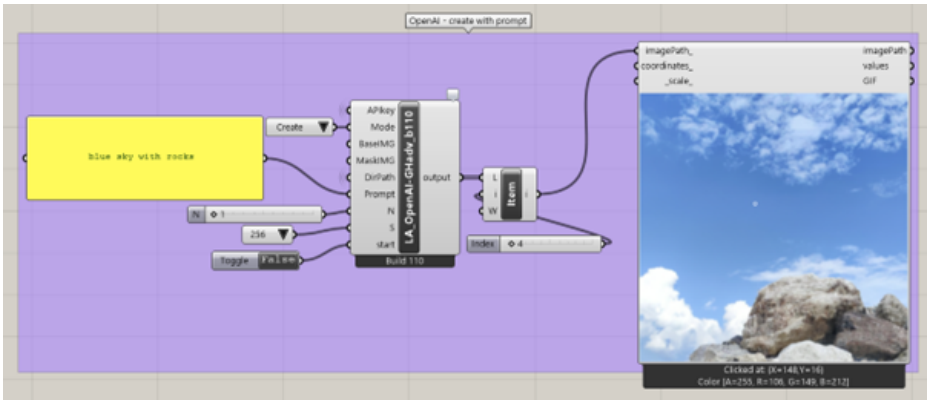

Figure 3. Grasshopper script with OpenAI model (DALL-E) manually inputted by a prompt.

Another aspect to take into account concerns the image-to-image mode, i.e. the mode that allows a user to get output variations starting from a source image. Basically, the source image can be generated previously and alternatively by OpenAI or by StabilityAI tools, using DALL-E allows only the source image to be processed as input. On the other hand, using Stable Diffusion lets the user have more control over the image-to-image process (IMGtoIMG) due to the possibility of adding a second text prompt as input besides the source image. Finally, thanks to the parameters *Strength* and/or *ClipG*[8] will possibly shift the weight of the generative process towards text or the source image.

Both tools mentioned above generate three output formats: a PNG image stored in a subfolder called "IMGs", a TXT text file stored in a subfolder called "TXTs" and finally a Log file in CSV format located in the folder specified in the "DirPath" parameter. This way of managing the output files allows students to keep track of all their design exploration iteratively by archiving the input and output parameters used during the investigative stage, but also to access each metadata stored in the TXTs subfolder.
The script we created for the workshop offered the students multiple options for the task. Students were able to run the 2 models (Stable Diffusion and DALL-E) using different parts of the prepared script. The components allowed for three modes of image generation: creation, variation and prompt editing. Input can be images (generated in previous iterations) or prompts (manually modified by the students as they progress with their tasks).

As students generated different images (examples shown in Figure 4) manipulating the prompts to achieve a satisfying solution to address the tasks set in the project brief, our model automatically generated a CSV log that records data that helped us to track

[8] Clip guidance mode (ClipG) works only with the "Ancestral Sampler" models, according to StabilityAI's API documentation. Source, https://platform.stability.ai/docs/getting-started/python-sdk

the entire process. The log includes a timestamp, the prompt used, the mode of creation, along with other metadata like the name and size of the image file saved, and the base and image-mask used.

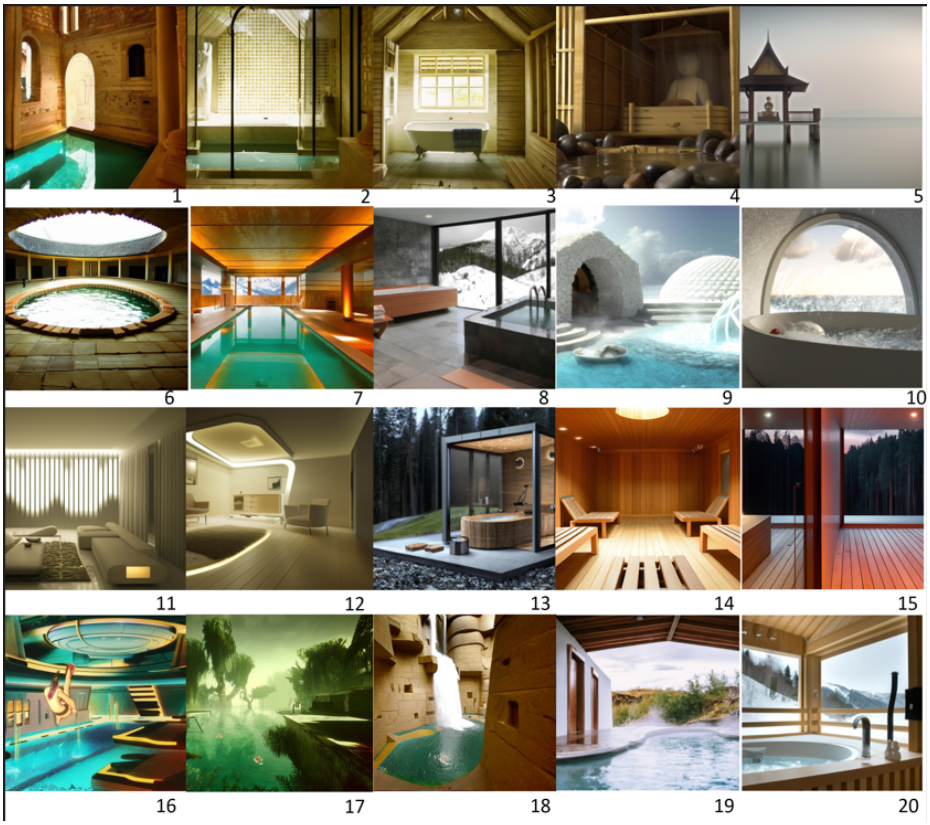

Figure 4 Example of the work produced by the students. 1-3 Group 7 Stability.ai, 4-5 Group 7 DALL-E; 6-7 Group 2 Stability.ai, 8-10 Group 7 DALL-E; 11-12 Group 3 Stability.ai, 13-15 Group 7 DALL-E; 16-18 Group 5 Stability.ai, 19-20 Group 5 DALL-E.

Tasks and Procedure

The workshop required students to design a public repository of water that serves as a sanctuary for individuals through the exploration and use of DALL-E-2 and Stable Diffusion models. The aim of the workshop was to introduce design students to the use of diffusion models and allow them to explore the ‘design process’ and the ‘digital representation’ of this novel method. Throughout the course of the workshop, students were asked to reflect on the application of diffusion models in architecture and how they could be used to assist in the design process and enhance their visual presentations.

The workshop ran for 4 hours and was divided into: Workshop Introduction (30 minutes), Phase 1: Groups to develop projects with diffusion models (duration 1.5 hours), Phase 2: Groups to fill in the questionnaire template (duration 1.5 hours) and finally, the group presentations (duration 30 minutes).

Phase 1: Once all groups had a solid concept for their building, they started using DALL-E-2 (for the first 45 minutes) and Stable Diffusion (for the second 45 minutes) to generate images of their repository of water, by providing models with text-based descriptions. This gave the students a better idea of what their building would look like and allowed them to make any necessary adjustments to their concept. Groups documented and commented on each iteration, explaining their own thinking process per each image generated.

Phase 2: In phase two of the workshop, the students were asked to reflect on their experience using DALL-E-2 and Stable Diffusion within the design process. This phase was an important part of the workshop as it allowed the students to think critically about the tools they have used and how they might be able to apply them in the future. The students combined snippets and comments describing Phase 1 providing an overall reflection on the process.

The students were asked to discuss the following topics:

1. The strengths and weaknesses of using DALL-E-2 and Stable Diffusion in the design process: What worked well and what didn't? What were the limitations of the tools and how did they affect the students' designs?
2. The impact of AI on the design process: How did using DALL-E-2 and Stable Diffusion change the way the students approached the design process? What were the advantages and disadvantages of using these tools compared to traditional design methods?
3. Potential future applications: How might the students use DALL-E-2 and Stable Diffusion in future projects? Are there other industries or fields where these tools could be applied?

Participants were then asked to fill in a questionnaire, divided into two parts. In the first part of the questionnaire we included the following questions:

- Can you briefly describe your experience with using the diffusion models?
- What is the aspect/activity you found more challenging?
- What is the aspect you found more interesting?
- On the basis of your experience today, what are the potentials you see in these models?
- What are the weaknesses/pitfalls?
- How about your learning experience? What did you learn (new) today?
- How would you compare your design activity today with the more conventional design methods (e.g. using CAD, 3D modeling etc. to produce architectural images/concepts)?
- Is there anything in particular that you think you are learning more or differently using diffusion models?
- Is there anything you think you are missing out by using these models?

Students were asked to comment on open-ended questions describing their experience through brief commentaries.

In the second part of the questionnaire, we asked the students to respond to the following questions:

Q1. Can you score (1-5) your experience today with diffusion models within GH?
Q2. How easy or difficult was it to experiment with these tools (1-5)?
Q3. How much the images that were generated differed from each other (1-5)?
Q4. How easy was it to instruct the AI to produce the solutions you had in mind (1-5)?
Q5. How many adjustments did you have to do to each prompt in order to produce a satisfying solution (1-5)?
Q6. To what extent do you feel you had an agency in the entire process? (How much of you as a designer do you think there is in the final results?) (1-5)

The discussion has been guided by the tutors who encouraged the students to share their thoughts and ideas, and provided feedback on their reflections. Overall, this phase of the workshop is an opportunity for the students to think critically about the role of AI in the design process, and how they can use these tools to enhance their creativity and improve their designs in the future.

The workshop was open-ended, meaning there was no specific design requirement, but students were encouraged to explore different design elements, typologies and styles. The workshop was, in fact, process-driven rather than finalized to the design outcome. The project was a great opportunity for students to explore the potential of AI in the design process, come up with creative solutions, and think critically about how the use of AI might shape the future of architecture.

Results

The data collected in the workshop were mainly a log of the prompts used by each group, the images generated with different models and prompts, and the replies of each student to the questionnaire. We use the latter to run a thematic analysis with an a priori coding [20], stemming from comments of the students in relation to two main topics: (i) usability of the pipeline; and (ii) relationship between automated process and the designer. Two of the authors coded independently the questionnaire's answers, using the following nodes (or codes) with the relative Intercoder Reliability scores [24] (Cohen's Kappa coefficients, with a fair-to-good strength of agreement between 0.41 and 0.75, very good between 0.75 and 1 [25]) as per below:

- Challenges: 0.7445
- Enhancement of Design: 0.8066
- General Opinion (interest surprise): 0.5202
- Limitation in Design: 0.8661
- Representation of ideas: 0.4955
- Usability
    - Awareness: 0.7705
    - Negative: 0.5379
    - Positive: 0.6648

With the relationship between the tool tested and the designer, we wanted to explore the extent to which the students (in their role as designers) felt that the diffusion models were enhancing or hindering their creative process in responding to a given design brief.

The first theme that emerged was the ability of the tools to support the students in **representing** their **design ideas visually**. Overall respondents commented positively on this point, yet they highlight a certain level of discrepancy between what they expected as the outcome and what the tools produced. This was considered positively and in many cases as a surprising product of the process: "[the tools] *gave a different perspective that I wouldn't have taught about in the first place*", yet with the awareness of some difference in the expected final results "*it was great, although the design did not represent what we wanted*". This aspect is also supported by the other emergent theme about the **interest and surprise** generated by the experience. Students seemed to appreciate the short time needed to generate strong visuals to describe and support their design: "*we learn a new way to generate idea in short amount of time*", "*it's good for getting a quick answer*" or "*I learnt how some changes when describing can give off big changes in the design*". Another aspect that has been emphasized in the responses is the **variety of images** that can be easily generated "*how AI creates different images by simply changing one or two words*". However, students realized the importance of words in generating prompts: "*How many vastly different versions it can create from the same prompt*" or "*it makes me focus more on the vocabulary I use to create something I can see in my mind*". Some students felt the need to be able to **manipulate prompts in a more granular way**, perhaps replicating the level of accuracy to which they are used in generating design with traditional tools (pencil while sketching or drafting, or CAD and 3D modeling): "*It would be nice to change one little thing In particular in each image as some were close to what we wanted to create but not exact*".

In our analysis, we noticed two most significant themes that emerged in a very similar measure: the tools as **enhancement of design** and as **hindrance of design**. These two aspects are reflected in two of the codes with highest value of agreement: Enhancement of Design (0.8066) and Limitation in Design: (0.8661). Students appreciated the capabilities of using diffusion models to generate images as a part of their creative process: "*it is great with creating something crazy*", "[useful] *for future use of quicker design and productivity*", "*exploring the potential of my thoughts*" and "*an easy way to communicate our ideas instead of only description. If there was a way of transmitting our sketches as well and from that it could generate a more realistic image*". Students emphasized the power of the used tools in creating unexpected images in a quantity and speed that is appealing and considered a strong addition to their designer toolkit. This is particularly true for early-stage design and representation or investigation of initial concepts.

In the same way, students highlighted the limits of the tools in helping them to produce the expected results: "Unexpected random outcomes, frustrating to communicate, no consistency", "*it was frustrating to* [be able to] *get a final model close to our ideas*", or "*the Dall-e is quite tricky because you have to choose proper words and play with it for a long time to get results you want*". The problems highlighted by the students in answering the brief through the tools proposed can be attributed to the fact that all participants were using diffusion models for the first time during the workshop. We recognise that some of the comments can be associated with any other

generative design approach where the designer needs to design a method to produce an outcome. This is in contrast with more traditional What-You-See-Is-What-You-Get (WYSIWYG) approaches (e.g. in 3D and CAD modeling). It would be interesting at this point to run a similar workshop with designers who are used to generative models (like for example, genetic algorithms or other form-finding approaches) to compare the results. This would allow us to isolate the comments about not being able to intervene directly into the final results and appreciate whether they are related to generative processes in general, or to the diffusion models used in our particular case.

More speculatively, we realized that students felt a sense of indirectness in their design process due to the complexity that underpins the AI models used. Students explained that there was a third agent in the creative process (besides them as the designers and the medium as the computer or pen and paper): "*we feel more following the 'machine minds' to interpret things, while in the conventional method we visualize ideas almost purely from our own mind*". Designers are used to impact directly on their design representations, very often through a lengthy iterative process. The process the students underwent in this workshop forced them to have a mediated approach to the development of their design, where the mediation was represented by an AI agent. While some students embraced this new element in the creative process as an enhancer "*When the images come out right, they are very good and extremely realistic. Producing such an image could have taken weeks*", others found it somewhat hindering "*it cannot replicate the image you might have in your head and can take time to get the perfect image that the description depicts*".

	Q1	*Q2*	*Q3*	*Q4*	*Q5*	*Q6*
Mean	3.83	3.72	3.61	3.06	**4.00**	**2.67**
Median	4	4	3.5	3	4	2
Max	5	5	5	5	5	5
Min	2	1	1	1	2	1
Range	3	4	4	4	3	4
Standard dev	0.90	0.99	1.01	1.18	0.94	1.25

Table 2. Results for the second part of the questionnaire.

The findings from the second part of the questionnaire reinforce the concept that AI models alter the notion of agency in the design process. In this part of the questionnaire, students confirm this idea by focusing on the unforeseen and spontaneous (generated/different/interesting) results during the process.

We obtained 18 valid responses from participants, which are summarized in table 2 above.

The overall findings suggest that there were no extreme attitudes towards the utilization of AI tools during the workshop. The mean values for all 6 questions varied between 4 (highest in Q4) and 2.67 (Q6). While students had a moderate viewpoint on the general experience of using the tools, with values around 3 for Q1-Q4, they expressed a strong opinion on the number of adjustments required to attain the desired outcomes, as indicated by a value of 4 in Q5. The most intriguing outcome was from the question about the students' agency in the design process (Q6), which suggested a low level of feeling in control for the students as designers during the process (mean = 2.67 for Q6).

4 Discussion

The results of our study point out several interesting things about the use of Generative AI tools in the context of creative work to support users.

First, as participants pointed out, the images produced are surprising and interesting, which, together with novelty and utility or value form three basic criteria for evaluating creativity [12]. However, the control over the outputs proved to be more challenging than initially expected, as the importance of the chosen words for each prompt became immediately clear to participants. Controlling the way in which outputs are produced through prompts by tweaking them slightly or completely, as it is used in traditional creative work can be a much-needed improvement. Enabling users to better adjust outputs by means of fixing prompts is a topic that can definitely fit within the End-User Development research area's point of view, as it would enable users to customize these tools to fit their intended use and interact with them more naturally, in turn fostering their widespread use [26]. Also, students felt the need to see a direct relation (or mapping) input (prompt/words) to output (image generated), almost mimicking the node structure that characterizes Grasshopper-like design and programming environments.

The contrast between some participants' feedback in relation to the effects of Generative AI on their design is quite striking. The reported unpredictability of results represents an added value for the initial stages of design when blue-sky concepts are welcome, but can be limiting in later stages when a convergence over an expected solution is sought. This is still an open problem, but it could be mitigated with tools implementing masking actions: for instance, some conditioning procedures acting on the Stable Diffusion's decoder have been implemented using a Neural Network named ControlNET. The latter enables conditional inputs like edge maps, segmentation maps, keypoints to enrich the methods to control Large Diffusion Models and further facilitate related applications. This technology showing promising results [27].

Finally, the key point arising from our results is closely related to the nature of these tools: the sense of indirectness sensed by participants over the results, together with the

lack of control of them in order to replicate what they had in mind is rooted into the black-box nature of Artificial Intelligence models. Generative AI tools will have to find ways to properly open up their inner models and allow users to properly direct them if they want to succeed in becoming the next companion of digital creators. Some proposed frameworks [17, 18, 19] already point out features needed in terms of Explainability of the models and characteristics of the outputs, but more research is needed in order to investigate proper solutions to these issues.

Finally, the inability to replicate what participants had in mind may cause a significant hindrance to the design process, resulting in frustration and lower creativity. These results have broader implications for the development and implementation of Generative AI Tools for Architecture, calling for a better understanding of the underlying algorithms and the need for greater transparency in the design process. By addressing these challenges, we can create more effective and accessible tools that really support and enhance the creativity and control of architects and designers, rather than be perceived as a tool to replace them, reflecting the true meaning of "Co-Creation".

Limitations

Although participants provided their feedback autonomously, it's important to recognize the significant role that facilitators played in the workshop. In this study, one of the authors moderated the workshop, but we plan to conduct future ethnographic research studies to increase the results' validity.

Our study aims to examine the impact of Generative AI tools on creativity. However, achieving this goal requires meticulous analysis and testing to ensure the results can be applied generally.

The group formation in the current study may have led to internal biases, which may have skewed certain groups, thereby limiting the results' overall validity. Question formulation could also have impacted the reliability and validity of the collected data. More studies are needed to cross-validate our results with different quantitative data coming from other sources of measure.

Finally, since participants were architecture students, additional research is necessary to evaluate the tools with participants from diverse backgrounds and levels of expertise.

5 Conclusions and Future Work

The rise of AI-based tools capable of generating new content has pushed the potential of AI systems beyond problem-solving and towards problem-finding, where the AI functions as a creative collaborator and supporter. Large-scale Text-to-Image Generative Models (LTGMs) are a rapidly evolving class of AI algorithms that have shown remarkable capabilities in creating high-quality images based on natural language descriptions, making them powerful tools for non-technical users. While the adoption of LTGMs-based tools is growing in various creative domains, their impact on creativity remains understudied.

This paper addressed this gap by investigating the impact of LTGMs-based tools on creativity through post-hoc feedback provided by design students. Our findings provide valuable insights into how the interaction with LTGMs affects the users' creative process, focusing on their ability to effectively control them to generate new ideas. This research contributes to the development of future tools and investigate their use in creative work. As LTGMs continue to advance, it is important to understand how the collaboration between humans and AI through these tools might affect creative processes.

Future works include further analysis of data in relation to the prompts issued by users to highlight, for instance, how many times a single prompt has to be refined in order to obtain a suitable result, as well as comparing the final outcomes with previous work that haven't made use of Generative AI tools. Moreover, involving participants without an architectural background would provide an interesting way of comparing our results outside of this domain and make them more generally valid.

Acknowledgements

The authors would like to thank Ian W. Owen for helping organize the workshop and the other tutors of the Architecture department at Hertfordshire who helped run the activities. We would also like to thank the students who proactively took part in the workshop and provided useful feedback for this study. The work has been organized as follows: TT coordination and leadership, LA developments of the tools, SC workshop, prep of the GH scripts, testing and troubleshooting, data analysis, AM feedback and advice throughout the study.

Research partly funded by PNRR - M4C2 - Investimento 1.3, Partenariato Esteso PE00000013 - "FAIR - Future Artificial Intelligence Research" - Spoke 1 "Human-centered AI", funded by the European Commission under the NextGeneration EU programme.

References

1. Gozalo-Brizuela, R., Garrido-Merchan, E.C.: ChatGPT is not all you need. A State of the Art Review of large Generative AI models, http://arxiv.org/abs/2301.04655, (2023).
2. Stiny, G.: Introduction to shape and shape grammars. Environ. Plann. B. 7, 343–351 (1980). https://doi.org/10.1068/b070343.
3. Jo, J.H., Gero, J.S.: Space layout planning using an evolutionary approach. Artificial Intelligence in Engineering. 12, 149–162 (1998). https://doi.org/10.1016/S0954-1810(97)00037-X.
4. Park, H., Suh, H., Kim, J., Choo, S.: Floor plan recommendation system using graph neural network with spatial relationship dataset. Journal of Building Engineering. 106378 (2023). https://doi.org/10.1016/j.jobe.2023.106378.
5. Jabi, W., Chatzivasileiadi, A.: Topologic: Exploring Spatial Reasoning Through Geometry, Topology, and Semantics. In: Eloy, S., Leite Viana, D., Morais, F., and Vieira Vaz, J. (eds.) Formal Methods in Architecture. pp. 277–285. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-57509-0_25.
6. Carta, S.: Self-Organizing Floor Plans. Harvard Data Science Review. (2021). https://doi.org/10.1162/99608f92.e5f9a0c7.
7. Ploennigs, J., Berger, M.: AI Art in Architecture, http://arxiv.org/abs/2212.09399, (2022).

8. Borji, A.: Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2. (2022). https://doi.org/10.48550/ARXIV.2210.00586.
9. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning Transferable Visual Models From Natural Language Supervision. (2021). https://doi.org/10.48550/ARXIV.2103.00020.
10. Wu, Z., Ji, D., Yu, K., Zeng, X., Wu, D., Shidujaman, M.: AI Creativity and the Human-AI Co-creation Model. In: Kurosu, M. (ed.) Human-Computer Interaction. Theory, Methods and Tools. pp. 171–190. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-78462-1_13.
11. Lyu, Y., Wang, X., Lin, R., Wu, J.: Communication in Human–AI Co-Creation: Perceptual Analysis of Paintings Generated by Text-to-Image System. Applied Sciences. 12, 11312 (2022). https://doi.org/10.3390/app122211312.
12. Muller, M., Chilton, L.B., Kantosalo, A., Martin, C.P., Walsh, G.: GenAICHI: Generative AI and HCI. In: CHI Conference on Human Factors in Computing Systems Extended Abstracts. pp. 1–7. ACM, New Orleans LA USA (2022). https://doi.org/10.1145/3491101.3503719.
13. Campero, A., Vaccaro, M., Song, J., Wen, H., Almaatouq, A., Malone, T.W.: A Test for Evaluating Performance in Human-Computer Systems, http://arxiv.org/abs/2206.12390, (2022).
14. Clark, E., Ross, A.S., Tan, C., Ji, Y., Smith, N.A.: Creative Writing with a Machine in the Loop: Case Studies on Slogans and Stories. In: 23rd International Conference on Intelligent User Interfaces. pp. 329–340. ACM, Tokyo Japan (2018). https://doi.org/10.1145/3172944.3172983.
15. Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L.A., Isaac, W., Legassick, S., Irving, G., Gabriel, I.: Ethical and social risks of harm from Language Models. (2021). https://doi.org/10.48550/ARXIV.2112.04359.
16. Houde, S., Liao, V., Martino, J., Muller, M., Piorkowski, D., Richards, J., Weisz, J., Zhang, Y.: Business (mis)Use Cases of Generative AI. (2020). https://doi.org/10.48550/ARXIV.2003.07679.
17. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P.N., Inkpen, K., Teevan, J., Kikin-Gil, R., Horvitz, E.: Guidelines for Human-AI Interaction. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. pp. 1–13. ACM, Glasgow Scotland Uk (2019). https://doi.org/10.1145/3290605.3300233.
18. Abedin, B., Meske, C., Junglas, I., Rabhi, F., Motahari-Nezhad, H.R.: Designing and Managing Human-AI Interactions. Inf Syst Front. 24, 691–697 (2022). https://doi.org/10.1007/s10796-022-10313-1.
19. Weisz, J.D., Muller, M., He, J., Houde, S.: Toward General Design Principles for Generative AI Applications, http://arxiv.org/abs/2301.05578, (2023).
20. Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B., Wesslén, A.: Experimentation in Software Engineering. Springer US, Boston, MA (2000). https://doi.org/10.1007/978-1-4615-4625-2.
21. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., Chen, M.: Hierarchical Text-Conditional Image Generation with CLIP Latents, http://arxiv.org/abs/2204.06125, (2022).
22. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-Resolution Image Synthesis with Latent Diffusion Models. (2021). https://doi.org/10.48550/ARXIV.2112.10752.
23. Saldaña, J.: The coding manual for qualitative researchers. SAGE Publishing Inc, Thousand Oaks, California (2021).

24. O'Connor, C., Joffe, H.: Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. International Journal of Qualitative Methods. 19, 160940691989922 (2020). https://doi.org/10.1177/1609406919899220.
25. Fleiss, J.L., Levin, B., Paik, M.C.: Statistical Methods for Rates and Proportions. Wiley (2003). https://doi.org/10.1002/0471445428.
26. Fogli, D., Tetteroo, D.: End-user development for democratising artificial intelligence. Behaviour & Information Technology. 41, 1809–1810 (2022). https://doi.org/10.1080/0144929X.2022.2100974.
27. Zhang, L., Agrawala, M.: Adding Conditional Control to Text-to-Image Diffusion Models, http://arxiv.org/abs/2302.05543, (2023).