

Article

Generative Artificial Intelligence Image Tools among Future Designers: A Usability, User Experience, and Emotional Analysis

Joana Casteleiro-Pitrez 

Department of Arts, University of Beira Interior, 6200-001 Covilhã, Portugal; joana.casteleiro.ferreira@ubi.pt

Abstract: Generative Artificial Intelligence (GenAI) image tools hold the promise of revolutionizing a designer's creative process. The increasing supply of this type of tool leads us to consider whether they suit future design professionals. This study aims to unveil if three GenAI image tools—Midjourney 5.2, DreamStudio beta, and Adobe Firefly 2—meet future designers' expectations. Do these tools have good Usability, show sufficient User Experience (UX), induce positive emotions, and provide satisfactory results? A literature review was performed, and a quantitative empirical study based on a multidimensional analysis was executed to answer the research questions. Sixty users used the GenAI image tools and then responded to a holistic evaluation framework. The results showed that while the GenAI image tools received favorable ratings for Usability, they fell short in achieving high scores, indicating room for improvement. None of the platforms received a positive evaluation in all UX scales, highlighting areas for enhancement. The benchmark comparison revealed that all platforms, except for Adobe Firefly's Efficiency scale, require enhancements in pragmatic and hedonic qualities. Despite inducing neutral to above-average positive emotions and minimal negative emotions, the overall satisfaction was moderate, with Midjourney aligning more closely with user expectations. This study emphasizes the need for significant improvements in Usability, positive emotional resonance, and result satisfaction, even more so in UX, so that GenAI image tools can meet future designers' expectations.

Keywords: generative artificial intelligence image tools; usability; user experience; emotional induction; future designers



Citation: Casteleiro-Pitrez, J. Generative Artificial Intelligence Image Tools among Future Designers: A Usability, User Experience, and Emotional Analysis. *Digital* 2024, 4, 316–332. <https://doi.org/10.3390/digital4020016>

Academic Editors: Yannis Manolopoulos, Mariusz Duplaga and Nicos Komninos

Received: 23 March 2024

Revised: 11 April 2024

Accepted: 15 April 2024

Published: 17 April 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the contemporary design landscape, the use of GenAI image tools in the summer of 2022, driven by machine learning algorithms and large language models, rose among the design community. Design practitioners have shown a growing curiosity about these tools, which has increased their availability on the market. Tools like DALL-E 2, Midjourney, Stable Diffusion, and Adobe Firefly, among many others, promise to revolutionize the paradigms of image production. These and other tools allow designers to redefine the creative process, improve methodologies, result in faster ideation and prototype development, faster creation, manipulation, and the exploration of visual and multimedia content [1]. However, the extent to which these conversational interface tools cater to the needs and preferences of future designers remains a subject of inquiry. Some authors advise that despite the potential benefits of AI image tools, there is a lack of research on their effectiveness and Usability [2,3], which prompted us to delve into this research topic. Nielsen [3] indicates that this new user interface paradigm, the Intent-Based Outcome Specification, has deep-rooted Usability problems that require knowledge of prompt engineering, such as the need to use prose text that obliges the user to tell the computer what outcome they want rather than tells the computer what to do. The user tells the computer what they want but not how to accomplish it, reverting the locus of control: all these new specifications prove the need to analyze these user interfaces. Yu, Dong, and Wu's [4] study also warns about the UX problems of GenAI tools, highlighting their high learning costs, limited effectiveness,

and trust issues. These gaps in understanding the Usability and effectiveness of GenAI tools underscore the importance of investigating and analyzing their user interfaces.

Given the circumstances described above, it is imperative to ask if GenAI image tools are made for future designers. Do these tools have good Usability and UX, induce positive emotions, and provide satisfactory results? This is the question we want to solve with this research. To answer this question, we conducted a literature review in the domain of GenAI tools for creative areas. We also undertook a quantitative study into the suitability of three GenAI image tools (Midjourney, DreamStudio, and Adobe Firefly) for the evolving needs and aspirations of 60 future designers. Through a quantitative analysis encompassing evaluations of the Usability, UX, Emotional Induction, and Generated Results Satisfaction, we aim to understand the assessments of future designers for each of these dimensions for the three chosen GenAI image tools, and from there, determine if these tools are ready for integration into the design workflow. The holistic research instrument joined together the USE Questionnaire: Usefulness, Satisfaction, and Ease of Use [5]; the User Experience Questionnaire (UEQ) [6]; the Positive Affect Negative Affect Scale (PANAS) [7–9]; and a question related to satisfaction regarding the results obtained.

By synthesizing insights from the Usability, UX, and emotional response analyses from three GenAI image tools, we endeavor to understand the perspectives and expectations of future designers about these conversational interfaces. This understanding can facilitate the improvement and refinement of GenAI image tools, which are crucial for promoting the effective utilization and adoption of these tools by design practitioners.

The findings from this study highlight the necessity for substantial enhancements in the Usability, Emotional Induction, Satisfaction with Generated Results, and particularly in the User Experience.

2. Literature Review

Researchers in creative areas have started to try to unveil the potential of GenAI in creative practice [10] and how these tools can be used in the creative process [11], like in the ideation phase [12,13]. Also, the creative power [14], practices, and risks of the co-creative artificial generative ecosystem have been analyzed [15]. Specific design areas also try to provide us with an understanding of the professional's perceived challenges and opportunities, like the visualization [16] or the communication [17] domain. Likewise, in web design, researchers try to understand how to integrate AI tools into the design process [18]. The first impact of GenAI image tools also raises profound ethical, legislative, and societal considerations [19–21] that interest researchers.

GenAI models typically evaluate the capabilities of generative image algorithms using metrics such as the CLIP score [22], FID [23,24], Inception Score [25], or Benchmark [26]. Nevertheless, because GenAI tools operate in a technologic-centric reality, the UX and user interface design have a low priority and still challenge these tools' creators. Human-computer interaction [27] and UX design are areas of research in which designers are trying to improve general AI-infused systems, as is the case of the work of Amer-shi [28], which reveals 18 guidelines created through the experimentation of AI-infused products. Concerning GenAI tools, some studies [4] warn about UX problems and the need to highlight their high learning costs, limited effectiveness, and trust issues. Some authors underline that these interfaces also have Usability problems that require the user to understand how to write prompts, be capable of writing prose text, and say what he/she wants instead of indicating to the computer how to accomplish the desired result [3]. Because of these specificities, some authors [29] have tried to develop a set of Usability and UX assessment scales for an in-depth evaluation of potentially essential characteristics of platforms like ChatGPT, Bing Chat, and Bard. Recognizing that these conversational interfaces have unique design requirements leads some researchers to investigate the fundamental UX design principles of conversational interface design [30]. Some researchers [31] have also tried to investigate the Usability of specific GenAI image tools like Midjourney. The results indicated a robust positive evaluation of the Perceived Ease of Use, Perceived Usefulness, Atti-

tude, and Behavioral Intention. In addition, the results suggest that Midjourney positively influences creativity.

Our brief literature review shows that GenAI text and image tools and their relationship with design and the creative industry have been interesting and have produced cutting-edge topics for research, covering areas like visualization design, communication design, web design, design methodologies and processes, ethics, UX, and user interface. There is still a need to understand if distinct GenAI image tools align with future designers' expectations; this understanding could help improve the development of these types of tools.

3. Method, Procedure, and Materials

3.1. Quantitative Experimental Design

Our research endeavors to explore whether three GenAI image tools respond to the requirements of the future generation of designers, proposing a good UX, Usability, inducing positive emotions, and providing satisfactory results. Therefore, the research questions driving this study are as follows:

RQ1. Do GenAI image tools (Midjourney 5.2, DreamStudio beta, and Adobe Firefly 2) have Usability? Moreover, do they have a positive evaluation of Usefulness, Ease of Use, Ease of Learning, and user Satisfaction?

RQ2. Do GenAI image tools (Midjourney 5.2, DreamStudio beta, and Adobe Firefly 2) have sufficient UX for design students? Additionally, do they receive favorable assessments for their Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty? Furthermore, do they align with the standards of the "Good" category when compared to the benchmark values?

RQ3. Do GenAI image tools (Midjourney 5.2, DreamStudio beta, and Adobe Firefly 2) induce positive/negative emotions in design students when they use them?

RQ4. Do GenAI image tools (Midjourney 5.2, DreamStudio beta, and Adobe Firefly 2) provide satisfactory results for design students?

We conducted an experiment involving three user groups, each utilizing and assessing three designated tools to address these research questions. The three groups provided responses to a holistic questionnaire that contained the USE Questionnaire: Usefulness, Satisfaction, and Ease of Use (USEQ); User Experience Questionnaire (UEQ); the Positive Affect Negative Affect Scale (PANAS); and a question related to the satisfaction with the generated results.

3.2. Procedure

We assigned each of the three GenAI image tools (Midjourney, DreamStudio, and Adobe Firefly) to a group of users. Then, between 4 and 19 December 2023, we requested users to visit a laboratory at the University of Beira Interior, where they followed a briefing: create a poster of a circus in Paul Rand style. They used and experimented with the GenAI tool assigned to them for the time they considered necessary, and following the experience, the students completed an online holistic questionnaire. Access to the Google Forms questionnaire was facilitated through a link provided on the University of Beira Interior Moodle platform.

3.3. Participants of the Study

We enlisted sixty volunteers (N = 60). The study participants were undergraduates pursuing a Bachelor's Degree in Multimedia Design within the Art Department at the University of Beira Interior, Portugal. These 60 participants were then distributed randomly into groups of 20 users each. The sample size range between 20 and 30 respondents guaranteed that we could attain trustworthy results [32].

3.4. Data Gathering

Information was gathered through the Portuguese version of the USEQ [5,33], the PANAS [7,8], the Portuguese version of the UEQ [6], and a question related to satisfaction with the results. Some studies recommend employing a variety of questionnaires to obtain a complete measurement of UX in conversational interfaces or enhance the assessment of a particular UX dimension [34]. Three instruments and one query were chosen to respond to the research questions:

- (a) The USEQ was chosen because it is a valid and reliable [35] survey instrument with 30 items that examine four dimensions of Usability: Usefulness, Ease of Use, Ease of Learning, and Satisfaction. Each item was rated on a seven-point Likert scale ranging from 1 = Strongly Disagree to 7 = Strongly Agree. This questionnaire evaluated the self-perceived Usability of GenAI image tools.
- (b) The UEQ was used because it facilitated the analysis of the complete UX beyond mere Usability. The questionnaire scales encompass Usability elements such as Perspicuity, Efficiency, and Dependability as well as UX factors like Novelty and Stimulation. This comprehensive approach provides a holistic understanding of the UX across product/system touchpoints [32]. The UEQ consists of 26 items spread across six seven-point Likert-type scales including Perspicuity, Attractiveness, Stimulation, Dependability, Novelty, and Efficiency. Each item in the UEQ is structured as a semantic differential, with two opposing terms representing each item. The reliability and validity of the UEQ have undergone thorough scrutiny in numerous studies [32,36].
- (c) The PANAS is among the most frequently utilized scales for assessing moods or emotions. This brief scale comprises 20 items, with half measuring positive effects (e.g., inspired and excited) and the other half measuring negative effects (e.g., afraid and upset). Each item employs a five-point Likert scale, from 1 = Very Slightly to 5 = Extremely, to measure the degree of experienced emotions within a defined timeframe [7]. This scale can measure emotional responses to events, such as the experience with GenAI image tools. Emotions are an increasingly important factor in human–computer interaction. Nevertheless, traditional Usability has mostly ignored the affective factors of the user and the user interface [37]. This scale can complement Usability testing by adding an information layer about the user’s affective state during testing.
- (d) A question about the satisfaction with the results obtained with GenAI image tools was included, and was rated on a seven-point Likert scale.

3.5. Materials

(1) Independent variables

The variables manipulated independently in this study were the three GenAI image tools. The chosen tools were Midjourney (see Figure 1), DreamStudio (see Figure 2), and Adobe Firefly (see Figure 3). We chose Midjourney because this tool is a leading generative AI tool, we chose DreamStudio because it is an easy-to-use interface using the latest version of the Stable Diffusion model, and is a leading generative AI tool as well, and we chose Adobe Firefly because Adobe created it. This company creates software specifically for designers and creatives.

Midjourney runs on the Discord Interface. Although the company is creating a new interface, it is not fully functional in Portugal. The image generation platform first entered open beta in July 2022. Midjourney has a text prompt box that admits using commands and parameters. The results presented allow the user to scale images, create slight image variations, and rerun the original prompt. Midjourney also accepts that terms are interpreted separately, allows for images to be uploaded, has a negative prompt, and has the creative tag prompt for unconventional outputs.

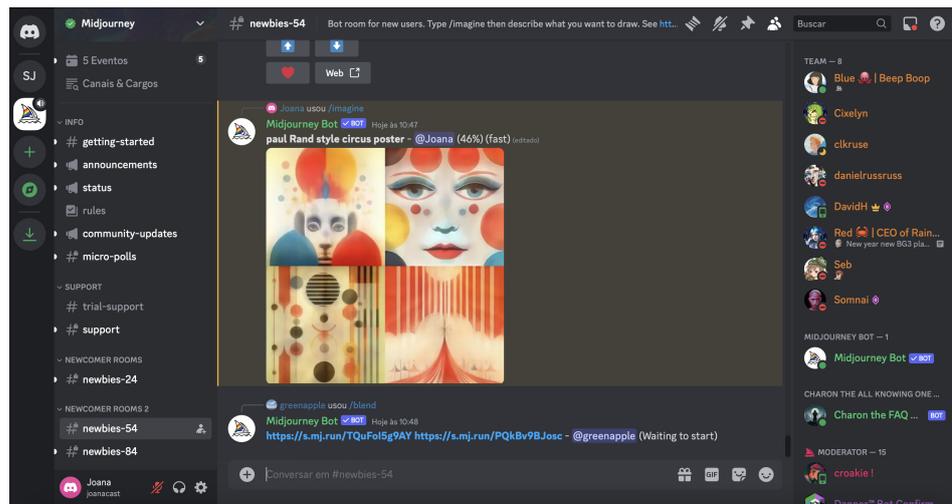


Figure 1. Midjourney 5.2 user interface.

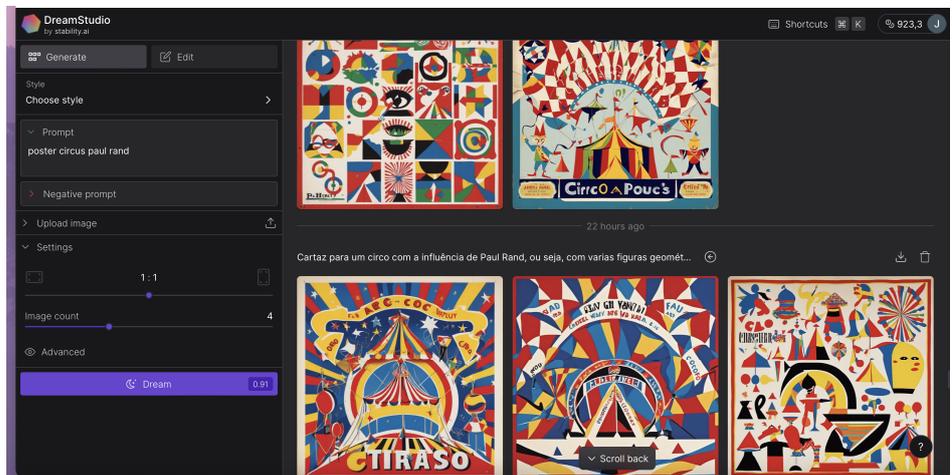


Figure 2. DreamStudio beta user interface.

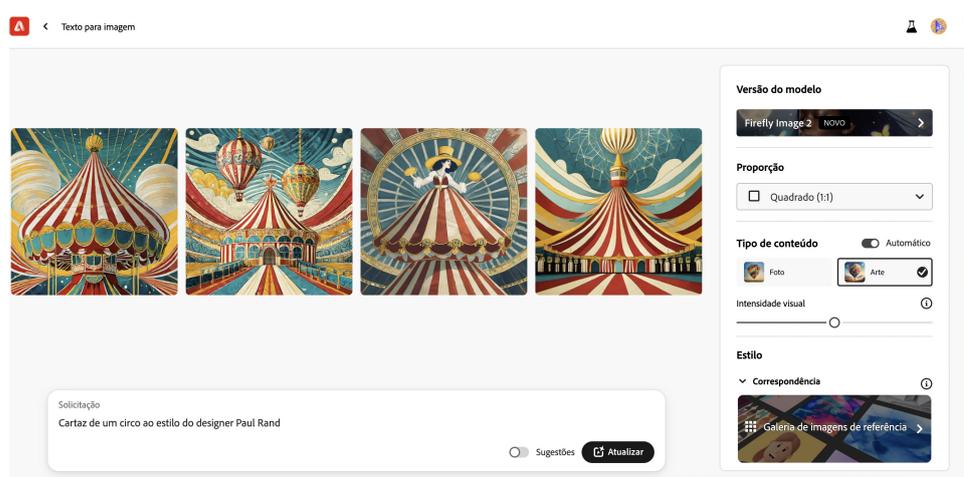


Figure 3. Adobe Firefly 2 user interface.

DreamStudio is a user-friendly interface designed for image creation, leveraging the latest Stable Diffusion image generation model iteration. It was released in August 2022. The DreamStudio interface features style options, a text prompt box, a negative prompt, image upload, image settings, advanced image settings, and the dream button.

Adobe Firefly was released in June 2023. Its features include a text prompt box, negative prompt, image upload, image proportions, style controls, and buttons for image effects, light, color, composition, and style intensity. It also has different possibilities, not only text to image, but also generative refill, text effects, generative recolor, and text to vectors. Adobe Firefly developers are now working on the possibilities of 3D to image and sketch to image.

(2) *Dependent variables*

The dependent variables used in this study were as follows:

Usability, UX, Emotional Induction, and Generated Results Satisfaction.

- (a) **Usability:** This variable was explored to ascertain the extent to which participants feel the system is easy to use and provides benefits in helping users obtain information or create a result. Four dimensions were used to evaluate this variable: Usefulness, Ease of Use, Ease of Learning, and Satisfaction. Given this aim, thirty statements and two open questions were used. As important as for other interfaces, the described Usability components are not sufficient concerning the evaluation of GenAI image tools, because they are not a traditional interactive device; their process is highly random and unexplainable, and the results may interfere with the Usability evaluation; because of this, other variables were considered in this study.
- (b) **UX:** UX focuses on the perceptions and behaviors of the user during their interactions with technical systems or products (ISO 9241-210:2019) [38]. Measuring UX means measuring hedonic (non-goal-oriented aspects) and pragmatic dimensions (goal-oriented aspects). Six dimensions were used to evaluate this variable: Efficiency (shows whether users can accomplish their tasks without undue effort), Attractiveness (indicates the overall impression of the product and gauges users' preferences towards it), Perspicuity (shows whether users can quickly become acquainted with the product or grasp its usage), Stimulation (indicates whether using the product is stimulating and engaging), Dependability (demonstrates whether users perceive a sense of control during interaction), and Novelty (reflects the product's level of innovation and creativity and its ability to capture the user's interest). Attractiveness represents a liability dimension. Efficiency, Perspicuity, and Dependability pertain to pragmatic quality, while Stimulation and Novelty relate to hedonic quality [39].
- (c) **Emotional Induction:** This variable was studied to determine the emotional impact exerted by the independent variable on participants. Participant motivation, ease of memorization, and ability to solve problems can be influenced by positive emotions [40]. Also, some studies consider the existence of carry-over effects of affective states on Usability appraisals [41]. We used twenty items to assess the emotions experienced during the experiment, and participants were queried about the extent to which they felt each emotion. The positive emotion items were attentive, active, enthusiastic, alert, determined, excited, interested, inspired, strong, and proud. The negative emotion items were scared, afraid, jittery, nervous, irritable, hostile, ashamed, guilty, distressed, and upset.
- (d) **Generated Results Satisfaction:** Because users' evaluations of GenAI image tools may tend to the results generated and confuse the ability to understand the other measurable scales, we opted to measure the users' satisfaction regarding the results generated.

3.6. *Data Analysis*

For the data analysis, we proceeded differently for each variable. For Usability (USEQ), we analyzed the average of the scales Usefulness, Ease of Use, Ease of Learning, and Satisfaction. Then, we measured the total average of all scales to obtain the total score of the Usability variable. The values were converted into percentages to facilitate the interpretation of the results. We also translated the results of the open questions related to positive and negative items and transcribed the most named arguments.

We analyzed each UEQ item for the UX and then used a benchmark to compare the results. Values for the single scale items between -0.8 and 0.8 represent a neutral evaluation, values >0.8 represent a positive evaluation, and values <-0.8 represent a negative evaluation [32]. Upon obtaining scores for each scale, the data were analyzed through a benchmark graph to evaluate the quality of the GenAI image tool in comparison to other products within the UEQ Analysis Data Tool dataset. Subsequently, we reference the benchmark intervals for the UEQ as outlined by Schrepp, Hinderks, and Thomaschewski [42]. The feedback is categorized into five levels:

- Excellent: the evaluated product is among the best 10% of results.
- Good: 10% of the benchmark results are better than the evaluated product, and 75% of the results are worse than the evaluated product.
- Above Average: 25% of the benchmark results are better than the evaluated product, so 50% of the results are worse.
- Below Average: 50% of the results in the benchmark are better than the evaluated product, and 25% of the results are worse.
- Bad: the evaluated product is among the worst 25% of results.

Additionally, Cronbach's Alpha was utilized to assess the reliability of the UEQ scales. A Cronbach's Alpha value between $1-0.90$ indicates an excellent internal consistency; between $0.70-0.90$, a good internal consistency; between $0.60-0.70$, an acceptable consistency; between $0.50-0.60$, a poor consistency; and less than 0.50 , an unacceptable consistency.

For the PANAS, we calculated the total score by finding the sum of the ten positive and ten negative items in a range of $10-50$.

We measure the total average for each GenAI image tool for the Generated Results Satisfaction variable. Again, values were converted into percentages to facilitate the interpretation of the results.

4. Results

Considering the questionnaires completed by 60 participants, we determined that the average age of our participants is 21 years old, with 25 (41.7%) male respondents, 34 (56.7%) female respondents, and one (1.7%) respondent identified as another gender. A total of 32 (53.3%) participants had never used GenAI image tools, and 28 (46.7%) had already experimented with this tool. Most of the 28 participants experimented with several GenAI tools; between the reported platforms, we can find Dall-E, for which 12 (20%) participants reported having used it, 8 (13.3%) participants indicated that had used Adobe Firefly, 7 (11.7%) participants had been experimenting with Midjourney, 4 (6.7%) participants used Stable Diffusion, and 14 (23.3%) also indicated that had used other GenAI image tools. From the analysis of each group, we can say that they are homogenous regarding the participants' previous use of GenAI image tools. In the DreamStudio group, 11 participants never used GenAI tools, and 9 did; in the Midjourney group, 9 participants never used GenAI tools, and 11 did; and in the Adobe Firefly group, 8 participants never used GenAI tools, and 12 persons did.

4.1. Impact upon Usability

The overall value of Usability for Midjourney, on a scale of $1-7$, is 4.72 , corresponding to 67.43% . DreamStudio reaches a 58.86% (4.12) score, and Adobe Firefly has a 62.29% (4.36) score (Table 1). These values are very little above the average, revealing the need to improve aspects of the user interface and the system itself to increase Usability. Regarding the dimensions of the Usability variable, Usefulness (61.71%), Ease of Use (65.86%), Satisfaction (66.14%), and Ease of Learning (76.29%) are higher for the Midjourney platform. The dimension Ease of Learning is the one that reaches the highest values. Almost all the dimensions reach positive values for all the platforms; the exception is the dimension Usefulness for the platform DreamStudio (45.57%), which is below the average.

Table 1. Total results of the USE Questionnaire.

	Midjourney		DreamStudio		Adobe Firefly	
	Mean	%	Mean	%	Mean	%
Total Usability	4.72	67.43%	4.12	58.86%	4.36	62.29%
Usefulness	4.32	61.71%	3.19	45.57%	3.82	54.57%
Ease of Use	4.61	65.86%	4.41	63.00%	4.17	59.57%
Ease of Learning	5.34	76.29%	5.26	75.14%	5.29	75.57%
Satisfaction	4.62	66.00%	3.63	51.86%	4.16	59.43%

Scale from: 1–7.

The results of the questions related to positive and negative points were essential for us to gather qualitative feedback to improve the GenAI image tool. The positive points mentioned most often for the Midjourney platform were related to the very quick generation of results and the possibility to inspire and improve creativity. Negative points included complaints about a confusing interface, difficulties related to image editing, results that did not meet users’ expectations, concerns over ethical and copyright issues, and difficulties in creating prompts.

For the DreamStudio GenAI image tool, the qualitative positive points mentioned by the participants highlight the platform’s ease of use, being fun to use, and its quick creation of images. The negative points emphasized by the participants indicated difficulties in editing the image results, results that did not change when altering the prompts, results that are too repetitive, an interface that should be simpler and organized, and copyright concerns.

For Adobe Firefly, students highlight the positive points of being easy to use, having an intuitive interface, creating exciting results, and the quick creation of results. The negative points accentuated were difficulties in putting together referential images, concerns over copyright, and the need for more editing tools.

4.2. Impact on UX

4.2.1. Midjourney GenAI Tool

The reliability assessment of the scale Attractiveness ($\alpha = 0.96$) indicated an excellent internal consistency; the analysis of the scales Perspicuity ($\alpha = 0.81$), Efficiency ($\alpha = 0.87$), Stimulation ($\alpha = 0.88$), and Novelty ($\alpha = 0.75$) indicated a good internal consistency; the scale Dependability ($\alpha = 0.50$) indicated a poor internal consistency. The experience with Midjourney created two types of results. Initially, we examined the significance of each UEQ item (see Figure 4), where the average indicated a positive evaluation of the UX for the scales Efficiency (0.950) and Stimulation (0.858). The average reveals a neutral evaluation for the scales Attractiveness (0.608); Perspicuity (0.675); Dependability (0.238); and Novelty (0.600).

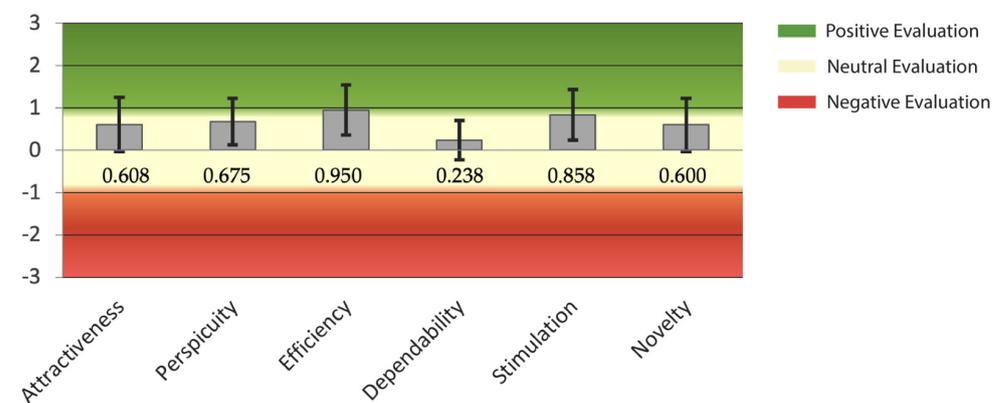


Figure 4. Average UEQ scale values of Midjourney.

Secondly, we acquired additional results based on the UEQ benchmark (see Figure 5). We analyzed the UX of the Midjourney GenAI tool compared to other digital products. The diagram shows that the scale values of Efficiency, Stimulation, and Novelty are in the “Below the Average” category, which indicates that 50% of the benchmark products have a better UX than this GenAI tool. The scenario is worse when we measure the scales of Attractiveness, Perspicuity, and Dependability, which are in the “Bad” category, meaning that 75% of the benchmark products have a better UX than this GenAI tool.

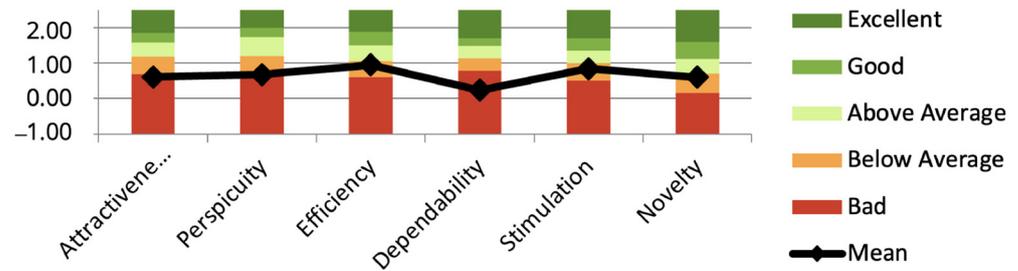


Figure 5. UEQ benchmark diagram on the Midjourney GenAI tool.

4.2.2. DreamStudio GenAI Tool

The reliability examination of the scales Attractiveness ($\alpha = 0.83$), Efficiency (0.76), Stimulation (0.73), and Novelty (0.80) indicated “Good” internal consistency. The analysis of the scales Perspicuity (0.50) and Dependability (0.55) indicated a “Poor” internal consistency. With the DreamStudio experience, we can perceive the value of each UEQ item (see Figure 6). The average analysis reveals a positive evaluation of the UX for the scales Perspicuity (0.888) and Efficiency (1.038). It also indicates a neutral evaluation for the following scales: Attractiveness (0.458), Dependability (0.238), Stimulation (0.138), and Novelty (0.188).

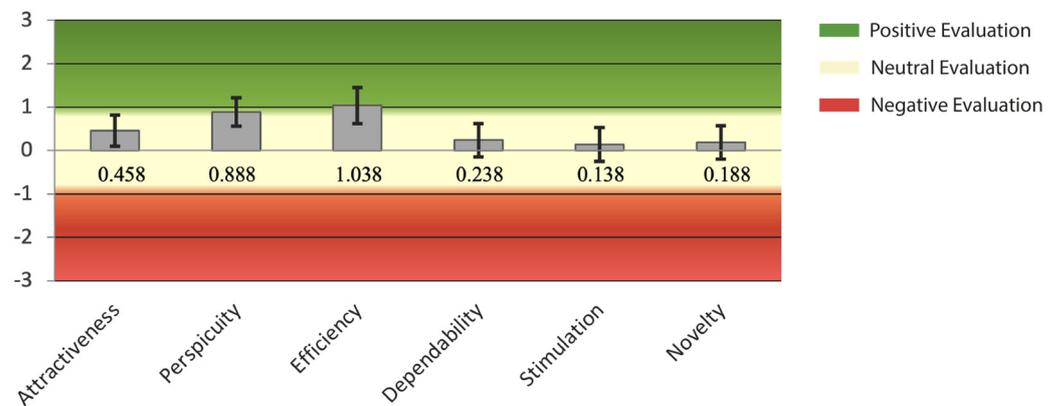


Figure 6. Average UEQ scale values of DreamStudio GenAI tool.

The outcomes of the UEQ benchmark (see Figure 7) for this GenAI tool reveal that the scale value for Perspicuity, Efficiency, and Novelty fall within the “Below the Average” category; this suggests that 50% of the benchmarked products outperform this GenAI tool in these aspects. On the other hand, the scales Attractiveness, Dependability, and Stimulation are in the “Bad” category, meaning that this product is in the 25% worst UX results.

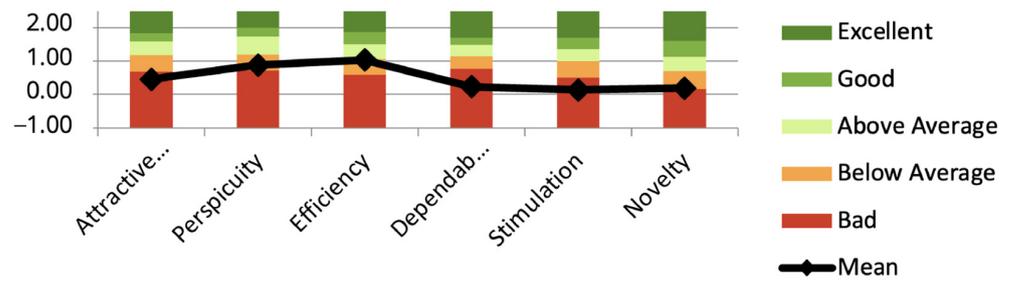


Figure 7. UEQ benchmark diagram on DreamStudio GenAI tool.

4.2.3. Adobe Firefly GenAI Tool

The reliability analysis of the scale Attractiveness ($\alpha = 0.91$) demonstrated “Excellent” internal consistency. The analysis of the Perspicuity scale (0.80), Efficiency scale (0.78), Stimulation scale (0.79), and Novelty scale (=0.70) indicated “Good” internal consistency. However, the Dependability scale exhibited poor internal consistency.

The user experience with the GenAI Tool Adobe Firefly also produced two types of results. First, we can observe the value of each UEQ item (see Figure 8), where the average indicates a positive evaluation of the UX for the following scales: Attractiveness (1.473); Perspicuity (1.642); Efficiency (1.607); Dependability (1.243); Stimulation (1.135). The average of the Novelty (0.774) scale indicates a neutral evaluation.

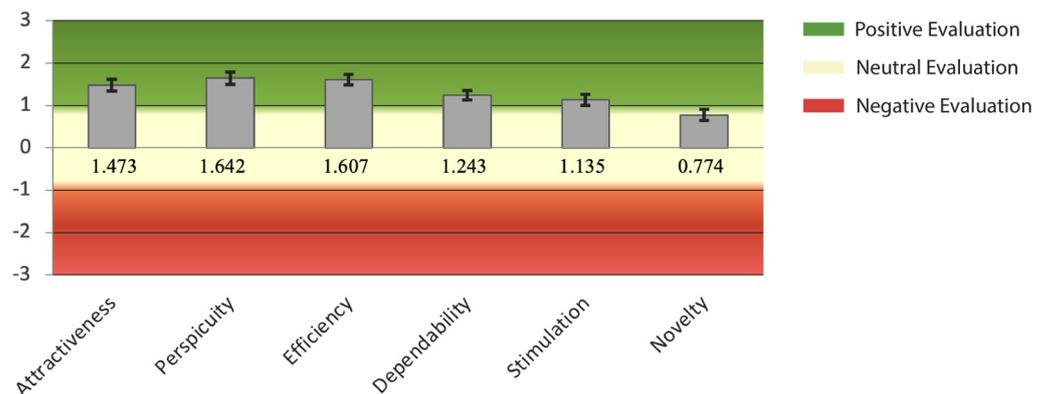


Figure 8. Average UEQ scale values of Adobe Firefly.

For this GenAI tool, the UEQ benchmark results (see Figure 9) indicate that the scale values for Attractiveness, Perspicuity, Dependability, Stimulation, and Novelty fall within the “Above Average” category. These results suggest that 25% of the benchmark products outperform this GenAI tool in these properties. Conversely, the Efficiency scale falls within the “Good” category, ranking among the top 25% results.

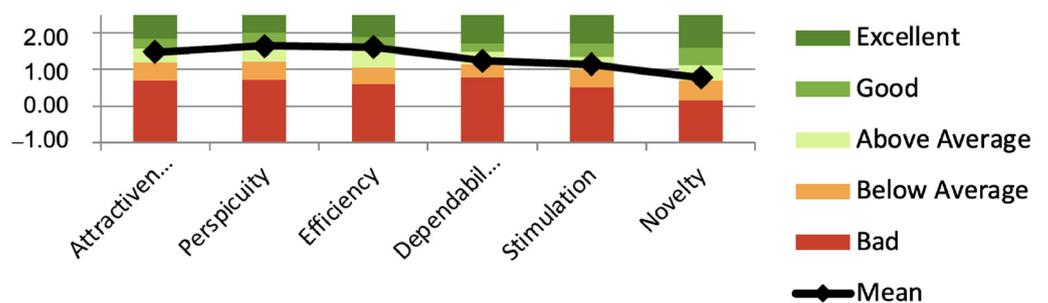


Figure 9. UEQ benchmark diagram on Adobe Firefly GenAI tool.

4.2.4. UX Comparison between the Three GenAI Tools

In the graphs presented below (see Figures 10 and 11), it is possible to compare the UEQ results of the three GenAI tools.

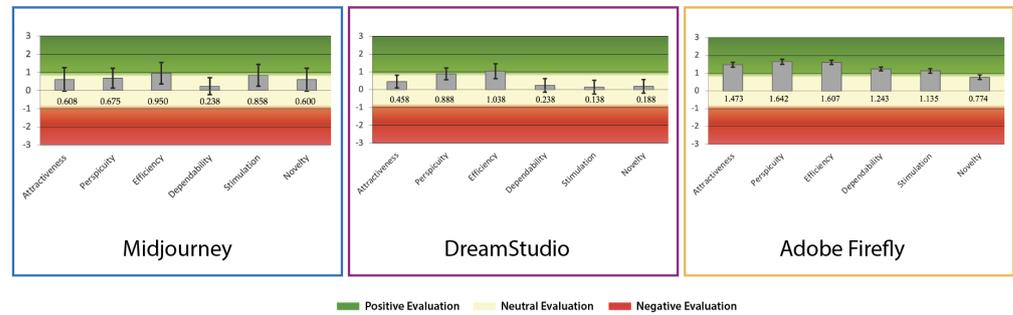


Figure 10. Comparison of average UEQ scales' values.

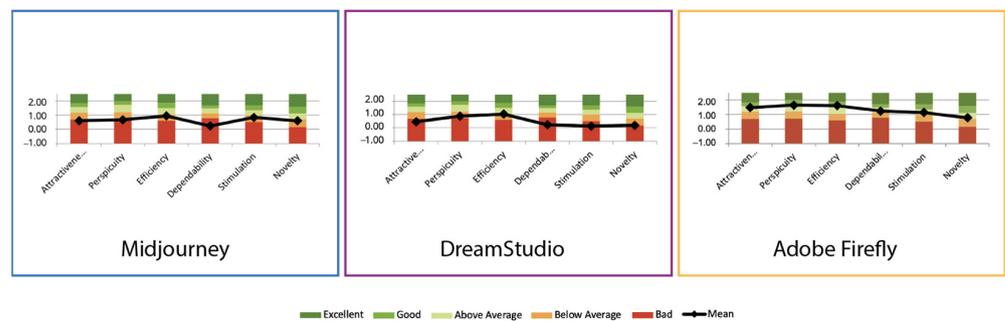


Figure 11. UEQ benchmark diagram comparison.

4.3. Impact on Emotional Induction

Table 2 shows the positive and negative average scores from the experience with Midjourney, DreamStudio, and Adobe Firefly.

Table 2. Average scores resulting from PANAS.

	Midjourney	DreamStudio	Adobe Firefly
Positive	29.455	19.600	25.800
Negative	15.954	16.200	18.100

Scale from: 10–50.

The results reveal that Midjourney (29,455) achieved a higher average of positive affects. Adobe Firefly reveals a neutral level of students' positive emotions, while DreamStudio reveals a low level of positive emotions. Midjourney (3.09, on a scale of 1–5) and Adobe Firefly (3.15, on a scale of 1–5) environments induce higher enthusiasm, while DreamStudio (3.05, on a scale of 1–5) induces a higher interest in students.

All the negative affect scores obtained with the PANAS questionnaire reveal a low level of students' negative emotions. Low levels of negative emotional induction indicate a lack of negative engagement, reflecting calmness and serenity [8]. Midjourney (2.14, on a scale of 1–5) and Adobe Firefly (2.15, on a scale of 1–5) induce higher levels of jittery behavior, while DreamStudio (2.5, on a scale of 1–5) induces a higher level of distress.

4.4. Generated Results Satisfaction

The analysis of the student's perceptions (Table 3) about the results indicates that students are more satisfied with the results obtained with Midjourney (67.14%).

Table 3. Evaluation of the results obtained by the students.

	Midjourney		DreamStudio		Adobe Firefly	
	Mean	%	Mean	%	Mean	%
Satisfaction with the results	4.70	67.14%	3.80	54.29%	4.59	65.57%

Scale from: 1–7.

Below, it is possible to see some of the results (Table 4) created by the students during the experience.

Table 4. Results created by the participants, with respective GenAI tool and prompt used.

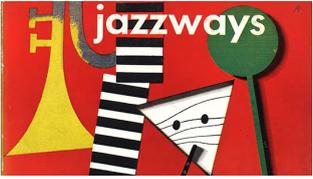
GenAI Tool	Prompt/Image Upload/ID, Seed	Results
Midjourney	<p>/imagine prompt create a poster of a circus based on the work of the designer Paul Rand—v 5.2 Id: 933008ae-8519-4c48-8043-628c79c9191b</p>	
Midjourney	<p>/imagine prompt contortionists and jugglers for a circus poster in the designer paul rand style—v 5.2 Id: 7d3a0416-1c5e-4368-869a-2bdc4324c57e</p>	
DreamStudio	<p>Prompt: Poster for a circus with the influence of Paul Rand, that is, with various geometric figures and color. The poster needs to contain objective elements of the circus. Seed: 554,335 Image:</p> 	
DreamStudio	<p>Prompt: Colorful poster for a circus, white background, geometrical objects in primary colors, different texture, minimalistic design. Seed: 93,581</p>	

Table 4. Cont.

GenAI Tool	Prompt/Image Upload/ID, Seed	Results
Adobe Firefly	<p>Prompt: Inside circus tent vector old retro vintage style of Paul Rand.</p> 	
Adobe Firefly	<p>Prompt: Vintage circus host with big pointed black hat, two women seated and two circus tents on the background vector poster Paul Rand style.</p> 	

5. Discussion

We developed this fieldwork study to understand if GenAI image tools, namely Midjourney, DreamStudio, and Adobe Firefly, are suitable for future designers, demonstrating that these individuals recognize that these tools have Usability, sufficient UX, provoke positive emotions, and provide satisfactory results in line with their expectations. The research questions were answered throughout the investigation. We consider the first research question: *RQ1. Do GenAI image tools have Usability? Moreover, do they have a positive evaluation of Usefulness, Ease of Use, Ease of Learning, and user Satisfaction?* The results indicate that future designers positively evaluate Usability in all GenAI tools but reach very little above-average values. The Midjourney tool shows a higher Usability level. This tool also achieves higher levels in all four Usability parameters: Usefulness, Ease of Use, Ease of Learning, and Satisfaction. Curiously, the dimension that reached the highest level for all the platforms was learnability; users considered it quick and easy to learn, without much effort to use these tools, and that once learned, it was easy to remember how to use them. All the tools reach positive values for almost all the dimensions of Usability; the exception is the dimension of Usefulness in the DreamStudio tool, which has a negative evaluation.

USEQ has two open questions that are very important for gathering qualitative feedback. For all tools, students refer to positive points, such as the readiness to create results, and negative points, such as difficulties in editing images and concerns over ethical and copyright issues. For Midjourney, students indicated that the platform could be a vehicle of inspiration that improves creativity; this is consistent with the results of other studies [31]. Students also indicated that when using Midjourney, it was challenging to create prompts, the interface could be confusing initially, and the results might not meet users' expectations. For DreamStudio, students alleged that the platform was fun and easy to use, the results were too repetitive and did not change when altering prompts, and the interface could be more organized and straightforward. For Adobe Firefly, the students underline that the interface is intuitive and easy to use, and the results are interesting; they also indicate that it is difficult to upload referential images.

Next, we regard the second research question: *RQ2 Do GenAI image tools have sufficient UX for design students? Additionally, do they receive favorable assessments for their Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty?*

Furthermore, do they align with the standards of the “Good” category compared to the benchmark values? The analysis of each GenAI tool’s UX suggests that the scale Efficiency is the only one that is positive for all platforms, meaning that users agree that with all platforms, it is possible to solve tasks without unnecessary effort, which corresponds to the result of the USEQ. The scale that is evaluated as neutral in all the platforms was Novelty, indicating these platforms do not catch the interest of our students and are not understood as innovative. For Midjourney, the other scale that reaches a positive level is Stimulation, meaning the users find it exciting and motivating to use it. The scale Perspicuity is the other one that reaches a positive score for the DreamStudio tool, which indicates that users easily became familiar with the platform and learned how to use it. Adobe Firefly is the platform that has more scales (Attractiveness, Perspicuity, Dependability, and Stimulation) reaching a positive level, which suggests that users like the tool overall, believe that it is easy to learn how to use the tool, feel control over the interactions, and feel excited and motivated to use it. For the analysis of the UX benchmark, it is essential to remember that “the general UX expectations have grown over time. Since the benchmark also contains data from established products, a new product should reach at least the Good category on all scales” [42] (p. 43). None of the analyzed platforms reached this goal. Only one platform, Adobe Firefly, reached a “Good” category in only one scale, which was Efficiency, a pragmatic goal-oriented quality, which means users feel they can solve their tasks without much effort. In all the platforms, the scales Attractiveness, Perspicuity, Dependability, Stimulation, and Novelty are below the category of “Good”. Midjourney (Attractiveness, Perspicuity, and Dependability) and DreamStudio (Attractiveness, Dependability, and Stimulation) have scales in the “Bad” category, meaning these products are in the 25% worst results of the benchmark.

Despite our group of design students recognizing that these tools could solve tasks easily, without much effort, they need significant improvements in their pragmatic and hedonic qualities. The improvements should concern the total impression of the GenAI tools, making users feel control over the interactions, promoting easy learning on how to use the tools and become familiar with them, making these tools more exciting and motivating to use, and transforming these tools so that they catch the user’s interest and be understood as an innovative product. These GenAI image tools do not have sufficient UX.

Next, we look at the third research question: RQ3. *Do GenAI image tools induce positive/negative emotions in design students when they use them?* The results reveal that these tools induce low negative emotions, provoking calmness and serenity [8]. The positive emotions reveal low levels for DreamStudio, neutral levels for Adobe Firefly, and a slightly above-the-average levels for Midjourney.

We consider the fourth research question: RQ4. *Do GenAI image tools provide satisfactory results for design students?* The results indicate that students are more satisfied with the Midjourney results, although the values are not much above the average. This scenario can reveal a possibility of a connection between the achieved image results, the Usability, and the emotional induction scores.

This study contributes to the growing body of quantitative research whose aim is to evaluate the UX/Usability/Emotions of GenAI tools [29,43,44], specifically GenAI image tools in the design domain [31,45]. This study can assess that our design students consider these platforms to have slightly above-the-average positive Usability levels, with insufficient UX scores, even more so when compared to other products. The positive emotions they induce in the students also range between low and just above average. The negative emotions induced are low. The satisfaction with the results obtained could be higher. The obtained results from this experience prove that GenAI image tools need improvements related to their Usability, Induced Emotions, and Satisfaction with the generated results. However, more efforts need to be made to improve the UX.

Because these tools have specific characteristics and the results obtained with them are not controllable, the users’ responses may tend to the results. In future studies, we recognize that it would be important to propose holistic measuring instruments used specifically for

Generative AI products, with a benchmark related to AI-infused products. This future instrument should evaluate not only the hedonic and pragmatic qualities, obtained image results, and emotions but also trustworthiness and reliability, concerns that our student participants revealed. Also, qualitative studies could obtain more data on what could be changed in these tools to achieve higher levels of Usability and UX. As in any study, there were limitations to this research. The sample size could be higher, with future designers from different universities and domain experts. The study could be completed using a Usability test to obtain more qualitative data, and we could also evaluate more GenAI image tools.

6. Conclusions

This study assessed the Usability, UX, Emotional Induction, and Results Satisfaction of three GenAI image tools, namely, Midjourney, DreamStudio, and Adobe Firefly, with 60 future designers as participants. While all platforms received favorable ratings for Usability, they fell short of achieving high scores, indicating room for improvement. Midjourney was perceived as having the highest Usability, particularly in Usefulness, Ease of Use, Ease of Learning, and Satisfaction. The students agreed that it is easy and quick to learn how to work with all the tools; the Ease of learning dimension is the one that reached the highest values. With respect to UX, none of the platforms reached a positive evaluation in all the scales, with only Adobe Firefly achieving positive ratings in five scales. Students positively evaluated all GenAI tools regarding their Efficiency, agreeing that they can solve tasks without unnecessary effort. The analyzed tools failed to meet the UX goal requirements concerning Novelty, meaning students do not understand these tools to be innovative products, and they do not catch their interest. The comparison with the benchmark shows that all GenAI image tools need improvements in the pragmatic and hedonic qualities; the exception is the scale Efficiency in Adobe Firefly. In fact, some Midjourney and DreamStudio scales are in the worst 25% benchmark results. Despite inducing above-the-average positive emotions, particularly for Midjourney, and low negative emotions for all tools, the overall result satisfaction was moderate, with Midjourney meeting expectations more closely. The results obtained with this study underscore the need for significant Usability, Emotional Induction, and Generated Results Satisfaction improvements, but further improvement is needed for the UX variable, so that these tools can meet the expectations of future designers.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to legal and ethical restrictions.

Acknowledgments: The researcher would like to extend her gratitude to all the participants who took part in this study, acknowledging their time, cooperation, and invaluable perspectives offered, which allowed her to obtain essential information for the realization of this work.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Lessard, J. Generative AI in Content Creation: Revolutionizing the Creative Process with Innovative Solutions. Medium. Available online: <https://medium.com/@jacobylessard/generative-ai-in-content-creation-revolutionizing-the-creative-process-with-innovative-solutions-e5049f9ed292> (accessed on 29 January 2024).
2. Suryadevara, C. Generating free images with openai's generative models. *Int. J. Innov. Eng. Res. Technol.* **2020**, *7*, 49–56.
3. Nielsen, J. AI: First New UI Paradigm in 60 Years. Nielsen Norman Group. Available online: <https://www.nngroup.com/articles/ai-paradigm/> (accessed on 29 January 2024).
4. Yu, H.; Dong, Y.; Wu, Q. User-centric AIGC products: Explainable Artificial Intelligence and AIGC products. In Proceedings of the 1st International Workshop on Explainable AI for the Arts (XAIxArts), ACM Creativity and Cognition (C&C), Online, 19 June 2023; ACM: New York, NY, USA, 2023.
5. Lund, A. Measuring Usability with the USE Questionnaire. STC Usability SIG. *Newsletter* **2001**, *8*, 3–6.

6. Cota, M.; Thomaschewski, J.; Schrepp, M.; Goncalves, R. Efficient Measurement of the User Experience. A Portuguese Version. *Procedia Comput. Sci.* **2014**, *27*, 491–498. [CrossRef]
7. Tran, V. Positive Affect Negative Affect Scale (PANAS). In *Encyclopedia of Behavioral Medicine*; Gellman, M., Turner, J., Eds.; Springer: New York, NY, USA, 2013; pp. 1508–1509. [CrossRef]
8. Watson, D.; Clark, L.; Tellegen, A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Personal. Soc. Psychol.* **1988**, *54*, 1063–1070. [CrossRef] [PubMed]
9. Casais, M. Emotions as an Inspiration for Design. In *Advances in Industrial Design*; Shin, C., Di Bucchianico, G., Fukuda, S., Ghim, Y., Montagna, G., Carvalho, C., Eds.; AHFE 2021. Lecture Notes in Networks and Systems; Springer: Cham, Switzerland, 2021; Volume 260, pp. 924–932. [CrossRef]
10. Huang, J.; Chen, Y.; Yip, D. Crossing of the Dream Fantasy: AI Technique Application for Visualizing a Fictional Character’s Dream. In Proceedings of the IEEE International Conference on Multimedia and Expo Workshops (ICMEW), Brisbane, Australia, 10–14 July 2023; pp. 338–342. [CrossRef]
11. Liu, V.; Vermeulen, J.; Fitzmaurice, G.; Justin Matejka, J. 3DALL-E: Integrating Text-to-Image AI in 3D Design Workflows. In Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS ’23), Pittsburgh, PA, USA, 10–14 July 2023; Association for Computing Machinery: New York, NY, USA, 2023; pp. 1955–1977. [CrossRef]
12. Brisco, R.; Hay, L.; Dhimi, S. Exploring the role of text-to-image ai in concept generation. *Proc. Des. Soc.* **2023**, *3*, 1835–1844. [CrossRef]
13. Paananen, V.; Oppenlaender, J.; Visuri, A. Using text-to-image generation for architectural design ideation. *arXiv* **2023**, arXiv:2304.10182. [CrossRef]
14. Oppenlaender, J. The Creativity of Text-to-Image Generation. In Proceedings of the 25th International Academic Mindtrek Conference (Academic Mindtrek ’22), Tampere, Finland, 16–18 November 2022; Association for Computing Machinery: New York, NY, USA, 2022; pp. 192–202. [CrossRef]
15. Oppenlaender, J. The Cultivated Practices of Text-to-Image Generation. *arXiv* **2023**, arXiv:2306.11393.
16. Schetinger, V.; Di Bartolomeo, S.; El-Assady, M.; McNutt, A.; Miller, M.; Passos, J.; Adams, J. Doom or Deliciousness: Challenges and Opportunities for Visualization in the Age of Generative Models. *Comput. Graph. Forum* **2023**, *42*, 423–435. [CrossRef]
17. Ferreira, Â.; Casteleiro-Pitrez, J. Inteligência Artificial no Design de Comunicação em Portugal Estudo de Caso sobre as Perspetivas de 10 Designers Profissionais de Pequenas e Médias Empresas. *ROTURA—Rev. Comun. Cult. Artes* **2023**, *3*, 114–133.
18. Lively, J.; Hutson, J.; Melick, E. Integrating AI-Generative Tools in Web Design Education: Enhancing Student Aesthetic and Creative Copy Capabilities Using Image and Text-Based AI Generators. *J. Artif. Intell. Robot.* **2023**, *1*, 23–33. Available online: <https://digitalcommons.lindenwood.edu/faculty-research-papers/482> (accessed on 29 January 2024).
19. Amer, S. AI Imagery and the Overton Window. *arXiv* **2023**, arXiv:2306.00080.
20. Martínez, G.; Watson, L.; Reviriego, P.; Hernández, J.; Juárez, M.; Sarkar, R. Towards Understanding the Interplay of Generative Artificial Intelligence and the Internet. *arXiv* **2023**, arXiv:2306.06130.
21. Samuelson, P. Generative AI meets copyright. *Science* **2023**, *381*, 158–161. [CrossRef]
22. Radford, A.; Kim, J.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J. Learning transferable visual models from natural language supervision. In Proceedings of the International Conference on Machine Learning, PMLR, Virtual, 18–24 July 2021; pp. 8748–8763.
23. Borji, A. Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2. *arXiv* **2023**, arXiv:2210.00586.
24. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. Gans trained by a two time-scale up- date rule converge to a local nash equilibrium. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6626–6637.
25. Betzalel, E.; Penso, C.; Navon, A.; Fetaya, E. A Study on the Evaluation of Generative Models. *arXiv* **2022**, arXiv:2206.10935.
26. Achterberg, J.; Arel, R.; Grinberg, T.; Chaibi, A.; Bach, J.; Tzagkarakis, N. Generative Image Model Benchmark for Reasoning and Representation (GIMBRR). In Proceedings of the AAAI 2023 Spring Symposium Series EDGeS, San Mateo, CA, USA, 27–29 March 2023.
27. Shneiderman, B. *Human Centered AI*; Oxford University Press: Glasgow, UK, 2022.
28. Amershi, S.; Weld, D.; Vorvoreanu, M.; Fournay, A.; Nushi, B.; Collisson, P.; Shamsilqbal, J.; Bennett, P.; Inkpen, K. Guidelines for human-AI interaction. In Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; pp. 1–13.
29. Bubaš, G.; Čizmešija, A.; Kovačić, A. Development of an Assessment Scale for Measurement of Usability and User Experience Characteristics of Bing Chat Conversational AI. *Future Internet* **2024**, *16*, 4. [CrossRef]
30. Rossouw, A.; Smuts, H. Key Principles Pertinent to User Experience Design for Conversational User Interfaces: A Conceptual Learning Model. In *Innovative Technologies and Learning*; Huang, Y., Rocha, T., Eds.; ICITL 2023; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2023; Volume 14099, pp. 174–186. [CrossRef]
31. Shen, S.; Chen, Y.; Hua, M.; Ye, M. Measuring designers use of Midjourney on the Technology Acceptance Model. In *Life-Changing Design*; De Sainz Molestina, D., Galluzzo, L., Rizzo, F., Spallazzo, D., Eds.; IASDR 2023; IASDR: Milan, Italy, 2023; pp. 1–8. [CrossRef]

32. Schrepp, M.; Hinderks, A.; Thomaschewski, J. Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios. Lecture Notes. In *Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2014; Volume 8517 LNCS, pp. 383–392. [[CrossRef](#)]
33. Dantas, C.; Jegundo, A.; Quintas, J.; Martins, A.; Queirós, A.; Rocha, N. European Portuguese Validation of Usefulness, Satisfaction and Ease of Use Questionnaire (USE). In *Recent Advances in Information Systems and Technologies*; Rocha, Á., Correia, A., Adeli, H., Reis, L., Costanzo, S., Eds.; Springer: Cham, Switzerland, 2017; Volume 570, pp. 561–570. [[CrossRef](#)]
34. Kocaballi, A.; Laranjo, L.; Coiera, E. Understanding and Measuring User Experience in Conversational Interfaces. *Interact. Comput.* **2019**, *31*, 192–207. [[CrossRef](#)]
35. Gao, M.; Kortum, P.; Oswald, F. Psychometric Evaluation of the USE (Usefulness, Satisfaction, and Ease of use) Questionnaire for Reliability and Validity. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Philadelphia, PA, USA, 1–5 October 2018; pp. 1414–1418.
36. Schrepp, M.; Hinderks, A.; Thomaschewski, J. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *Int. J. Interact. Multimed. Artif. Intell.* **2017**, *4*, 103–108. [[CrossRef](#)]
37. Zimmermann, P.; Gomez, P.; Danuser, B.; Schär, S. Extending usability: Putting affect into the user-experience. In Proceedings of the 4th Nordic Conference on Human-Computer Interaction, Oslo, Norway, 14–18 October 2006; pp. 27–32.
38. *ISO 9241-210:2019; Ergonomics of Human-System Interaction Part 210: Human-Centred Design for Interactive Systems*. ISO: Geneva, Switzerland, 2019. Available online: <https://www.iso.org/standard/77520.html> (accessed on 17 March 2024).
39. Laugwitz, B.; Schrepp, M.; Held, T. Construction and evaluation of a user experience questionnaire. In *HCI and Usability for Education and Work*; Holzinger, A., Ed.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 63–76.
40. Isen, A.; Reeve, J. The influence of positive affect on intrinsic and extrinsic motivation: Facilitating enjoyment of play, responsible work behavior, and self-control. *Motiv. Emot.* **2005**, *29*, 297–325. [[CrossRef](#)]
41. Velazquez, M. Understanding the Effects of Positive and Negative Affect on Perceived Usability. Ph.D. Thesis, PennState University, State College, PA, USA, 2010.
42. Schrepp, M.; Hinderks, A.; Thomaschewski, J. Construction of a benchmark for the User Experience Questionnaire (UEQ). *Int. J. Interact. Multimed. Artif. Intell.* **2017**, *4*, 40–44. [[CrossRef](#)]
43. Skjuve, M.; Følstad, A.; Brandtzaeg, P. The User Experience of ChatGPT: Findings from a Questionnaire Study of Early Users. In Proceedings of the 5th International Conference on Conversational User Interfaces (CUI '23), Eindhoven, The Netherlands, 19–21 July 2023; Association for Computing Machinery: New York, NY, USA, 2023; pp. 1–10. [[CrossRef](#)]
44. Baek, T.; Kim, M. Is ChatGPT scary good? How user motivations affect creepiness and trust in generative artificial intelligence. *Telemat. Inform.* **2023**, *83*, 102030. [[CrossRef](#)]
45. Mortazavi, A. Enhancing User Experience Design Workflow with Artificial Intelligence Tools. Master's Thesis, Linköping University, Linköping, Sweden, 2023.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.