

Temes de Disseny #40

The Role of Designers in Society: Ethical, Theoretical and Practical Perspectives.

EDITED by
BORI FEHÉR, JONATHAN VENTURA
and ARIEL GUERSENZVAIG

CONTRIBUTIONS by
VALENTINA AURICCHIO, FRANCESCO BURLANDO,
PETER BUWERT, FLAVIANO CELASCHI,
ERIK CIRAVEGNA, ALEXANDRA COUTSOUCOS,
ANNALINDA DE ROSA, MARCO ANDREA FINARDI,
ÁGNES JEKLI, OXANA JEOUNG-RAKOVA,
MASSIMO MENICHINELLI, SEVI MERTER,
VANESSA MONNA, VERONICA PASINI,
ALEJANDRO IVÁN PAZ ORTIZ, LLUÍS SALLÉS DIEGO,
MATT SINCLAIR, LAURA SUCCINI
and ANNAPAOLA VACANTI.

FULL VERSION in
ENGLISH - CATALÀ - CASTELLANO



B. Fehér, J. Ventura & A. Guersenzvaig

Reflecting on the Role of Designers in Society

Reflexions sobre el rol dels dissenyadors a la societat / Reflexiones sobre el rol de los diseñadores en la sociedad

Editorial (p. 10 - 23)

V. Monna & V. Auricchio

*Designing the In-Between:
Outlining the Roles of the Designer
in the Civic Arena*

Dissenyant l'intermedi: Descripció de les funcions del dissenyador en l'àmbit cívic. / Diseñando lo intermedio: Descripción de las funciones del diseñador en el ámbito cívico.

Original Paper (p. 24 - 43)

L. Sallés Diego

*Design as a Political Object:
The Necessary Positioning of the Designer
as an Agent of Change.*

El disseny com a objecte polític: El reclamat posicionament del dissenyador com a agent de canvi. / El diseño como objeto político: El requerido posicionamiento del diseñador como agente de cambio.

Provocation (p. 44 - 53)

P. Buwert & M. Sinclair

Thought Experiments in Design Ethics

Experiments mentals en l'ètica del disseny / Experimentos mentales en la ética del diseño

Original Paper (p. 54 - 73)

S. Merter

*Co-Learning Scenarios with the Community:
Democratic Practices In and Out of the
Design Studio*

Escenaris de coaprenentatge amb la comunitat: Pràctiques democràtiques dins i fora de l'estudi de disseny / Escenarios de coaprendizaje con la comunidad: Prácticas democráticas dentro y fuera del estudio de diseño

Original Paper (p. 74 - 91)

L. Succini, E. Ciravegna, F. Celaschi
& V. Pasini

*Responsible Advanced Design:
Achieving Sustainability through
Collaborative Processes*

Disseny avançat responsable: Aconseguir la sostenibilitat gràcies a processos col·laboratius / Diseño avanzado responsable: Lograr la sostenibilidad valiéndose de procesos colaborativos

Original Paper (p. 92 - 111)

M. A. Finardi, A. Coutsoucos & A. De Rosa

*Co-designing with and for Small
and Remote Communities:
Reflections on Transformative Innovation,
Servitisation, and Agonism.*

Codissenyar amb i per a comunitats petites i remotes: Reflexions sobre la innovació transformadora, la servitització i l'agonisme / Codiseñar con y para comunidades pequeñas y remotas: Reflexiones sobre la innovación transformadora, la servitización y el agonismo

Case Study (p. 112 - 135)

O. Jeoung-Rakova

*Digitising the Physical:
The Evolving Role of Collaboration Tools
in Citizen Participatory Design Amidst the
COVID-19 Pandemic in Seoul*

Digitalitzar el que és físic: El paper evolutiu de les eines de col·laboració en el disseny participatiu ciutadà enmig de la pandèmia de COVID-19 a Seül / Digitalizar lo físico: El papel evolutivo de las herramientas de colaboración en el diseño participativo ciudadano en mitad de la pandemia de COVID-19 en Seül

Case Study (p. 136 - 155)

A. Vacanti, F. Burlando, A. I. Paz Ortiz
& M. Menichinelli

*Challenges and Responsibilities in the
UX Design of Text-to-Image AI Models:
A Discussion Through a Comparative
Heuristics Evaluation*

Reptes i responsabilitats del disseny d'UX en models d'IA de text a imatge: Discussió a través d'una avaluació heurística comparativa / Retos y responsabilidades del diseño de UX en modelos de IA de texto a imagen: Discusión a través de una evaluación heurística comparativa

Case Study (p. 156 - 175)

Á. Jekli

*Visual Communication
Bridging Intercultural Barriers:
Collaborative Methods Supporting the Social
Inclusion of Young Refugees*

Comunicació visual per superar les barreres interculturals: Mètodes de col·laboració en suport de la inclusió social dels joves refugiats / Comunicación visual para superar las barreras interculturales: Métodos de colaboración en apoyo de la inclusión social de los jóvenes refugiados

Original Paper (p. 176 - 200)

Challenges and Responsibilities in the UX Design of Text-to-Image AI Models: A Discussion Through a Comparative Heuristics Evaluation

The rapid advancement and adoption of generative AI in creative and professional domains mark a significant evolution in technological interaction, raising pivotal discussions on its integration, usability and broader societal impact. This paper delves into the multifaceted challenges and opportunities presented by generative AI, focusing in particular on text-to-image tools and their integration into daily work practices of artists, designers, architects and researchers. As these creators increasingly incorporate AI's inherent unpredictability and ambiguity into their work, a critical examination of the ethical, legislative and environmental implications becomes imperative. We highlight how the User Experience (UX) and User Interface (UI) aspects of these groundbreaking technologies remain mostly underexplored. UX pioneer Jakob Nielsen has recently warned that professionals in the field must adapt to the fast-paced advancements in AI or become obsolete for an industry that is becoming pivotal in the contemporary creative field. He draws a parallel between the current AI era and the dot-com boom of the 1990s, highlighting how UX professionals were slow to adapt to the Internet, which led to missed opportunities and a lack of user-centred design. Consequently, the usability of generative AI tools is a paramount concern, especially as non-expert users increasingly engage with these systems. This paper explores the usability challenges, specifically the articulation barrier where users struggle to effectively communicate prompts to AI, and the consequent need for more intuitive and accessible interfaces.

Through comparative analysis and a heuristic evaluation of four leading models, i.e. Midjourney, Dall.E, Stable Diffusion and Adobe Firefly, which presents a unique case study for comparison due to their divergent UI and UX elements, we aim to provide insights into the current state of AI tools and propose pathways for their responsible and effective use.

1 INTRODUCTION

Generative AI has been both the buzzword and the main subject of criticism in the year 2023, as Artificial Intelligence – about seventy years after the term was coined – has started to be introduced into the daily work practices of creators, in the broader meaning of the term. Generative AI involves creating media objects with artificial neural networks, including text, voice, music, 3D models, datasets and code. Terms like generative image, AI Image or visual

AI specifically refer to visual objects, encompassing static images that mimic various media forms, from photographs to drawings, and dynamic images resembling animations and videos (Manovich 2024). Artists, designers, architects and researchers are now weaving AI into their work, deliberately using its inherent unpredictability and ambiguity as a form of expression (Caramiaux and Fdili Alaoui 2022). In some contexts, AI has been defined as a novel design material, offering vast creative potential (Dove et al. 2017), as well as a medium ripe for experimentation (Holmquist 2017).

However, we should acknowledge that designing with AI means working within complex systems that deter-

ANNAPAOLA VACANTI
Università Iuav di Venezia;
Elisava, Barcelona School of Design and Engineering
(UVic-UCC)

FRANCESCO BURLANDO
Università di Genova;
Elisava, Barcelona School of Design and Engineering
(UVic-UCC)

ALEJANDRO IVAN PAZ ORTIZ
Elisava, Barcelona School of Design and Engineering
(UVic-UCC)

MASSIMO MENICHINELLI
Elisava, Barcelona School of Design and Engineering
(UVic-UCC)

KEYWORDS
AI Politics, Generative AI, Text-to-Image,
User Experience, Graphical User Interface,
Heuristic Evaluation.

LICENSE
CC BY

HOW TO CITE
Vacanti, Annapaola, Francesco Burlando Alejandro Ivan Paz Ortiz and Massimo Menichinelli. 2024. "Challenges and responsibilities in the UX design of text-to-image AI models: A discussion through a comparative heuristics evaluation". *Temes de Disseny* 40: 156-175.
<https://doi.org/10.46467/TdD40.2024.156-175>

mine outcomes which are inherently less predictable than those associated with traditional, passive materials and tools. As pointed out by Cattabriga and Joler (2023), AI’s autonomy, characterised by its ability to perform tasks in ways that might not strictly follow given instructions, necessitates a broader understanding and engagement from designers beyond simply viewing it as a tool or application. Designers need to integrate their practice with a comprehensive understanding of the political, ethical and social ramifications of AI systems to ensure its responsible utilisation and to uncover new avenues for professional growth.

The continuous evolution and refinement of generative AI tools, particularly text-to-image tools such as Midjourney, DALL-E and Stable Diffusion, are driving the rapid emergence of new issues at the intersection of human interaction and technological systems. These issues predominantly revolve around three main areas: the legislative and ethical challenges of AI integration, the environmental impact of deploying these technologies and the usability considerations that ensure these tools are accessible and effective for their intended users.

1.1. The legislative and social dilemma

As the production of AI-generated content escalates, the social role of digital visual information is becoming increasingly evident. Such content is being ingrained in the practices, skills, norms and institutions that govern the creation and dissemination of information, adding layers of complexity to how information is contextualised, interpreted and utilised across various societal activities and constructs. The interaction of AI systems with society leads to several and diverse implications, like the extraction of knowledge and exploitation of creative labour (Pasquinelli and Joler 2020).

At the heart of text-to-image generative AI systems is an extensive collection of raw data in the form of images, which are essential for learning and generating novel content. These images, devoid of inherent meaning, are assigned significance through the process of labelling, a central aspect of AI’s performance and output. The categorisation and labelling of images endow them with cultural and social significance, embedding societal assumptions and stereotypes within these massive datasets (Crawford and Paglen 2021). The biases, stereotypes or inaccuracies inherent in the labelling process can be amplified and perpetuated across various applications, leading to significant real-world consequences and perpetrating misconceptions across visual representation. The size and opacity of these datasets pose significant challenges to regulatory bodies and researchers dedicated to ensuring AI’s fair and responsible use. Recently, for example, LAION-5B – a huge training dataset containing 5 billion images scraped from the Internet – has been found to contain links to child abuse imagery, potentially allowing popular AI models such as Stable Diffusion to create harmful content (Thiel 2023). In the meantime, we are already facing the so-called synthetic loop problem, causing AI systems to be trained on AI generated content (Shumailov et al. 2023). The mix of synthetic data and human-made data appears to cause models to fail, producing inaccurate outputs and an inability to detect what is generated and what is not.

The uncontrolled proliferation of freely accessible AI web applications further complicates the task of regulation, with copyright laws facing enormous challenges. Training datasets make extensive use of artistic works across various disciplines, which is generally regarded as fair use at present, although this concept is evidently being stretched beyond its intended original scope of enabling creators to build upon existing content. Moreover, it should be noted that under current legal frameworks, non-human entities cannot hold a copyright, a principle now extended to generative AI creations. As a result, the myriad of works generated by AI systems are essentially unclaimed in the legal sense, regardless of the human prompts or instructions provided. Billions of pieces are being generated with no clear authorship, challenging the very concept of artistic ownership (Crawford 2021).

1.2. The environmental dilemma

Like all tools pertaining to the digital realm, AI is often perceived as an intangible entity. This consideration by the public results in its inordinate and uninformed use. However, AI requires an ecosystem to function, and that ecosystem consists of physical infrastructure such as the user’s device, server and network, which have a certain energy expenditure (Crawford 2018; Paoletti 2021). Based on the latest data, worldwide electricity usage by data centres has increased annually by 20-40% in the past few years, amounting to 1-1.3% of total global electricity consumption and accounting for 1% of energy-related greenhouse gas emissions as of 2022 (Hintemann and Hinterholzer 2022).

It is widely acknowledged that the process of training AI through gigantic databases proves to be very energy intensive. However, it is less known how much energy is consumed in the process of generating content through inference. Although processing a single instance during inference demands considerably less computation compared to training the same model, inference occurs much more often – potentially up to billions of times daily for models underpinning widely-used services. The least energy-efficient model for generating images consumes the equivalent amount of energy as 950 smartphone charges (11.49 kWh), roughly equating to nearly one charge for each image produced. However, energy usage varies significantly among different image generation models, largely depending on the size of the images they produce (Luccioni et al. 2024). In a society that rightly pays more and more attention to environmental issues and conscious consumption, the exponential growth in the immoderate use of digital solutions, among which generative AI tools have appeared in the past year, stands in a paradoxical counter-trend.

1.3. The usability dilemma

Each AI system possesses a certain degree of autonomy and decision-making capacity, a characteristic integral to its design (Wachter et al. 2017). This autonomy is essential, as it enables AI to perform complex tasks independently. However, the utility of generative AI systems is contingent upon the user’s ability to comprehend and effectively navigate these technologies. A lack of understanding in operating these autonomous systems paradoxically under-

mines their intended efficiency and creative augmentation. Therefore, as generative AI continues to evolve, emphasis must also be placed on interaction issues through design that enhances transparency and interpretability. The popularity that generative AI tools have been able to achieve has led to a great number of users, often non-experts, approaching these tools in a short time, revealing the need to make them user-friendly even more urgent. This stands in contrast to the complexity of the possibilities provided by generative AI, with the risk of falling within the paradoxical situation of having a very powerful technology within everyone’s reach, yet not being able to use it fully. We do not refer to the quality or nobility of use, but to the barrier that is likely to be created between the user’s expectations and the actual result. It would be like being in front of a genie from a lamp who does not speak the same language as you.

UX pioneer Jakob Nielsen (2023a) warns that professionals in the field must adapt to the fast-paced advancements in AI or become obsolete for an industry that is becoming pivotal in the contemporary creative field. He draws a parallel between the current AI era and the dot-com boom of the 1990s, highlighting how UX professionals were slow to adapt to the Internet, which led to missed opportunities and a lack of user-centred design. He predicted a future scarcity of UX experts with AI usability experience, emphasising the career implications for those in the field. Generative AI signifies the advent of the third user-interface paradigm in computing history, shifting from command-based interaction (first user-interface paradigm) to intent-based outcome specification (second user-interface paradigm) (Nielsen, 2023a). This paradigm shift revolutionises how users interact with technology, moving from telling the computer how to perform tasks to specifying the desired outcome (Zamfirescu-Pereira et al. 2023). Unlike previous paradigms, users of intent-based systems communicate their goals, and the system determines the necessary steps to achieve those results. Although extremely time-efficient, the abstraction from specific commands to broader intentions might obscure the process, making it difficult for users to

correct or refine outputs. Secondly, a noteworthy challenge emerges in the form of the articulation barrier (Mugunthan 2023). To generate high-quality images, users need to write complex and articulate prompts, requiring the ability to precisely communicate in written form. Users are tasked with not only describing the image’s subject matter but also articulating abstract concepts related to visual aesthetics. This requirement represents a sophisticated task, particularly challenging for individuals who are not acquainted with specific artistic styles or who lack the necessary technical vocabulary to express their vision effectively.

Although novel features and models appear on the market at an incredibly quick rate, text-to-image AI models are now accessible through three main approaches that we briefly describe in the following section, providing a map of the main features available at the time of writing and offering real-life examples for each one of these approaches.

2.1. Dedicated app and/or website

Text-to-image generators that run on dedicated web pages usually present an extremely simplified interface that allows users to write their prompt and often add a few more specifications through a graphical user interface (GUI) that varies widely from tool to tool. In the most basic scenario, the user is presented with nothing more than an input box where the prompt can be written. Sometimes, the input box offers writing suggestions and examples to help inexperienced users. However, most recent interfaces usually include GUI elements such as negative prompt boxes, buttons, toggles, checkboxes, dropdown menus and slider controls. These elements are typically used to enhance control for users without burdening them with the necessity to write long and redundant prompts (redundancy is used in prompting to force the model to generate a desired characteristic of the image).

At the moment, there is a great variety and lack of standardisation in GUIs for AI tools. The free generator Craiyon¹ only offers a limited number of buttons to choose the artistic style of the image and an input box for negative prompting. However, it supports the user by searching for relevant inspirations in its database of generated images by searching for keywords in the prompt (see Fig. 1, where we wrote “an avocado with a smiley face” and several anthropomorphised avocados appeared before clicking the draw button).

Freepik AI Image Editor² instead aims to reduce the length and complexity of the prompt by using a detailed lateral menu with several choices to be made by the user for generating the image concerning image ratio, style, colour, lighting and framing (Fig. 2). Selected choices appear as flags in the input box right before the prompt.

Stable Diffusion³, probably the most famous open source text-to-image tool, seems to be targeted at professional users due to the higher level of control offered by its

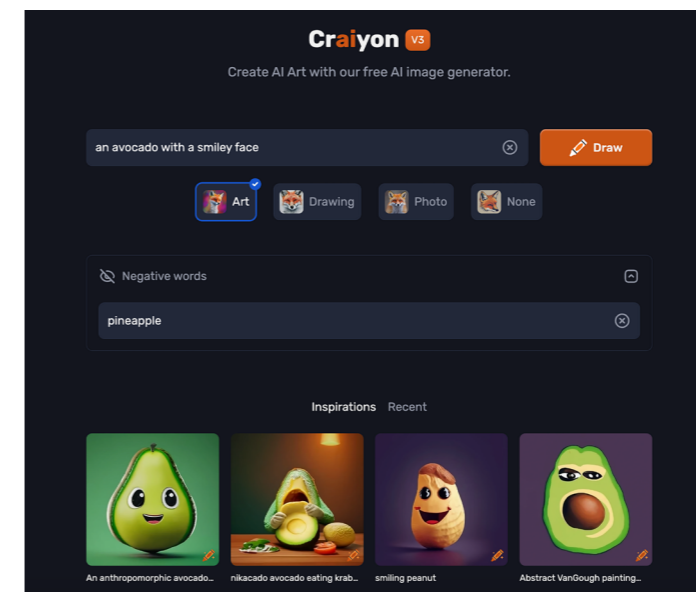


Fig. 1. The interface of Craiyon [screenshot].

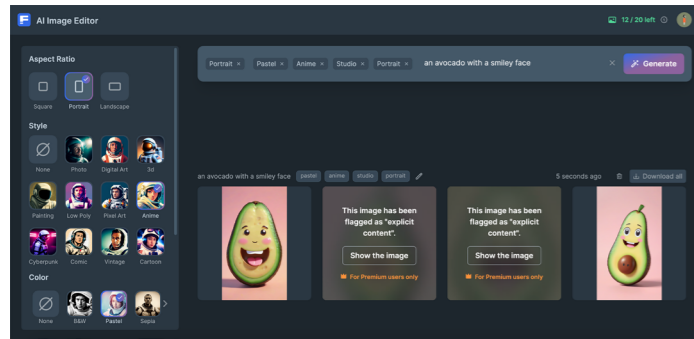


Fig. 2. The interface of Freepik AI Image Editor [screenshot].

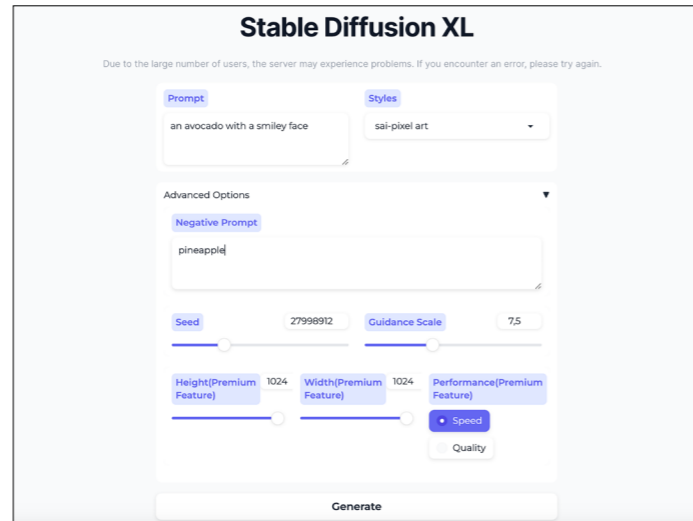


Fig. 3. The interface of Stable Diffusion [screenshot].

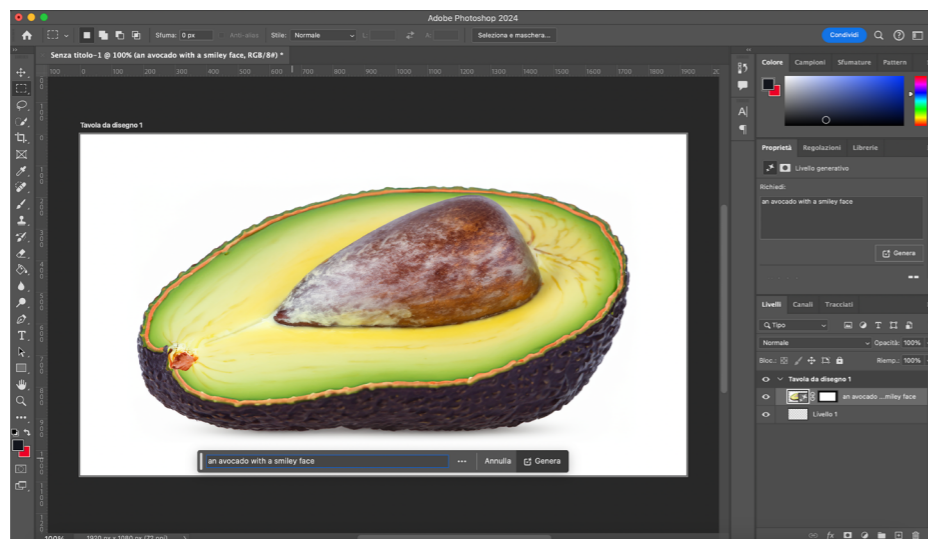


Fig. 4. Adobe Firefly integrated in Photoshop [screenshot].

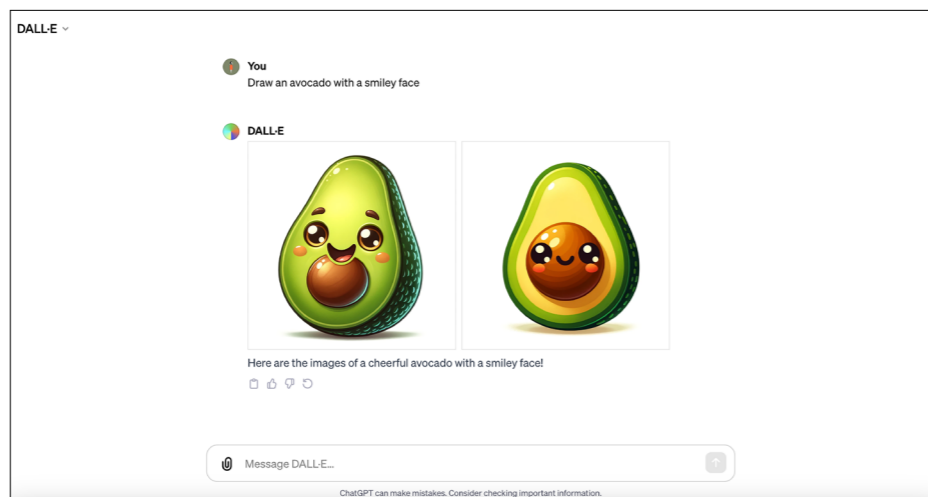


Fig. 5. The interface of DALL-E 3 integrated in ChatGPT [screenshot].

interface which, in turn, does not offer visual or written suggestions. It uses several GUI elements to control different aspects of the generation such as: input boxes for a prompt and negative prompt, dropdown menu for styles, slider controls for advanced settings such as seed, guidance scale and height/width as well as a button to choose speed over quality or vice versa (Fig. 3).

2.2. Integrated in software

Integrating generative text-to-image AI tools into professional software is intended as an augmentation to the creative process, introducing a dynamic new method for content creation. In a typical software ecosystem, such tools are often integrated as either a dedicated module or a feature accessible via the main toolbar or menus, ensuring easy discovery without disrupting the user's workflow. The interface involves a text input field, often accompanied by helpful cues or examples and possibly features like autocomplete or syntax suggestions to aid in crafting effective prompts. Given that in this case image generation is a part of a larger creative process, the UX flow is focused on allowing users to further refine or incorporate the images into the project using manual editing tools. Therefore, the generative tool must seamlessly integrate within the broader toolkit of the software. Also, due to the novel nature of generative AI, integrated help features, tutorials or tips are often present, taking the form of inline help, tutorial videos or example galleries.

Adobe has developed its own generative model, Firefly, which is accessible both from a web interface and within its most used software such as Photoshop and Illustrator. In Photoshop (Fig. 4), the GUI is integrated within the native environment, likely aiming to reduce the learning curve for existing users. The tool merely uses a simple input box for prompting. The model always generates three images, prominently displayed within the main canvas area, and the user can keep on switching between the options or generate further versions. Generated images can be easily manipulated in a dedicated layer, maintaining the software's core functionality.

It is clear that access to the use of this type of AI is mainly reserved for those who already use the software. Indeed, Photoshop is already complex in itself, and it is unlikely that a user will approach it merely to take advantage of this new feature.

2.3. Integration in other platforms

Integrating generative text-to-image tools into platforms represents a nuanced approach, focused on enhancing user interaction without overcomplicating the existing service structure. These integrations should be designed to feel like natural extensions of the original platforms, allowing users to engage with complex AI functionalities through familiar interfaces and interactions. However, this choice could prove to be a double-edged sword. The decision to rely on pre-existing platforms inevitably results in relying on the UX and UI of the latter as well. Designers do not have full control over the user experience, and therefore the success of the generative AI will depend largely on factors over which they have negligible control.

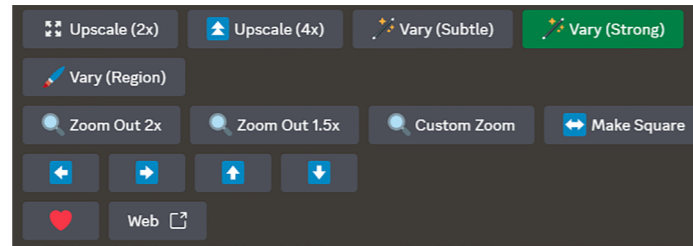
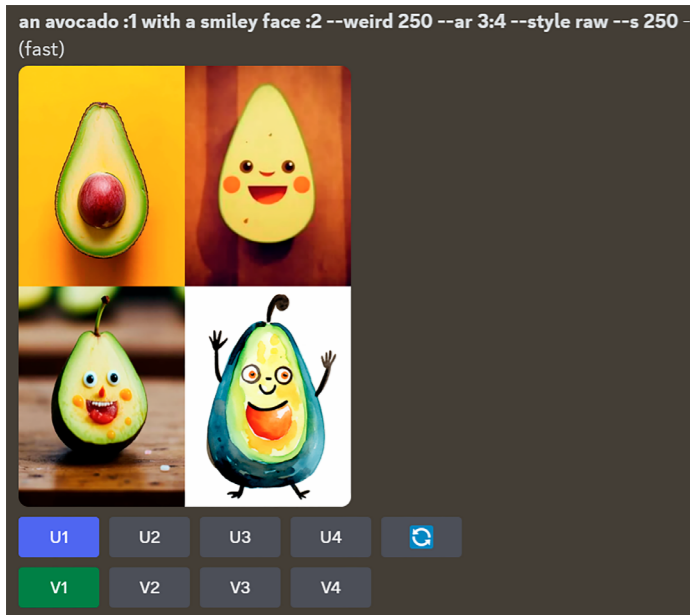
In its latest release, DALL-E 3 has been integrated within ChatGPT, the successful text-to-text AI. In this case,

the choice turns out to be particularly appropriate since the native platform within which the tool is integrated is already designed for that specific use, with the only variant that until then it could only produce textual output. DALL-E 3 relies on the UX and UI of ChatGPT (Fig. 5), which provide a minimalist interface in which interaction with the software is essentially accomplished through simple conversational prompts in a chat-like environment. Users are allowed to get images generated as if they were simply conversing with someone. Thanks to this choice, users who were already using ChatGPT can take advantage of a new potential, while it is very easy for those who were using DALL-E on the dedicated interface to migrate to the new platform.

As mentioned, the minimalistic interface allows users to write all the specifications for the image generation in conversational language through an input box. It is also possible to attach an image to the prompt, via a paperclip button. Once the image is generated, it is possible to read the actual prompt that DALL-E has used by clicking on the output, which is an extension of the concept expressed by the user. For example, we instructed the tool to "draw an avocado with a smiley face", and ChatGPT extended the request as follows: "a cheerful avocado with a big smiley face. The avocado should be green and ripe, with its brown pit visible. It has large, round eyes and a wide, happy smile. The avocado is cut in half, standing upright, and its skin has a textured, bumpy surface typical of a Hass avocado". Further editing is simply achieved by continuing the conversation.

An entirely different trend is present in Midjourney's approach, which is integrated in the social messaging application Discord through dedicated channels (Fig. 6). This generates community interaction that makes the use of AI a complex social process. In fact, it is not uncommon to see members using another user's production as a starting point to try to generate something new, modifying it through different prompts. In addition to these dynamics, Midjourney allows users a great deal of freedom to operate within the generation of images.

In fact, it offers six different commands with which new images can be generated, 13 parameters that can be used to direct the production toward certain required features and six additional tools through which the generated image can be further modified. However, unlike other generative AI tools, most of these functions are not shown explicitly to the user. Most commands are activated by adding some specific wording to the end of the prompt, as can be seen at the top of Fig. 6. In addition, these commands often follow a specific coding language. This feature makes Midjourney a tool suitable for experienced users who have the time and skill to inquire about how it works. However, even without knowing how to use the AI to its full potential, it is possible to produce images simply by entering the prompt thanks to some additional possibilities that are provided through buttons (Fig. 7) that allow, for example, varying the generated image or just one part of it or making a new image with a wider view. It can be assumed that Midjourney is among the most complex AI systems to use as only the most experienced users know how to master all the features and can control how the AI will react to the combined use of several commands together.



Figs. 6 and 7. The interface of Midjourney integrated in Discord [screenshots].

3 COMPARATIVE HEURISTICS EVALUATION

3.1. Methodology

To deepen the understanding of the usability of generative AI tools, a comparative analysis has been conducted on four leading models: Stable Diffusion 2.0, Adobe Firefly (integrated within Photoshop 2024 software), Dall.E 3 and Midjourney V6. We have used the renowned inspection method of the heuristics evaluation, introduced by Jakob Nielsen (1994), to support UX design teams in individuating usability problems in an interface. The selected heuristics include the eight golden rules of interface as defined by Shneiderman et al. (2016). Such principles have been adapted and applied to the specifics of AI tools to support our analysis:

- I. *Strive for consistency*: the interface for inputting prompts, adjusting parameters and viewing results is consistent throughout the tool. Terminology, layout and design elements are uniform.
- II. *Seek universal usability*: the tool accommodates a wide range of users, from visual artists to beginners, and considers their various technical proficiencies and creative needs, including adjustable settings or modes, and ensures accessibility to individuals with disabilities.
- III. *Offer informative feedback*: the interface gives immediate and clear feedback when users input prompts or adjust settings. When the AI is processing an image, it shows progress indicators. Upon completion, the results are displayed clearly and any errors or limitations in the generation process are communicated effectively.
- IV. *Design dialogs to yield closure*: the process of creating images is organised into clear, logical steps. Users feel a sense of completion after each action, whether it involves finalising a prompt or generating an image. It is clear when a task is completed

- V. *Prevent errors*: the prompt input and parameter adjustment interfaces are designed to minimise user errors. This could involve validating prompts, suggesting corrections or guiding users away from known issues that might lead to poor results. Clear guidance is given on the types of prompts that work well and any limitations of the AI model.
 - VI. *Permit easy reversal of actions*: users can easily undo actions or revise prompts without having to start from scratch. If a user isn't satisfied with a generated image, they are able to tweak the prompt or parameters and regenerate the image quickly and efficiently.
 - VII. *Keep users in control*: users feel they are guiding the creative process, with the AI serving as a tool to realise their vision. Users control as many parameters as possible, including style, content and format, while keeping the interface intuitive and not overwhelming.
 - VIII. *Reduce short-term memory load*: users don't need to continuously remember information across different parts of the tool. Essential information is visible or easily retrievable, especially when dealing with complex prompts or multiple parameters.
- Every AI tool is evaluated on a scale from 1 (very poor) to 5 (very good) for each heuristic.
- A heuristic evaluation involves experts reviewing a system against a set of criteria to identify usability issues. In the context of our paper, the four authors engaged with the four selected generative AI tools, leveraging their expertise. The exploration conducted in December 2023 involved navigating the tools and using them freely to generate and further edit some images and try different features. After giving individual grades, through active discussion, we shared our individual impressions and experiences with these tools to reach a consensus on a common evaluation framework, ensuring our findings were comprehensive and agreed upon by all authors. In the following section, we

present a synthesis of our evaluation without mentioning the quality of the visual outputs as they are not the subject of this research.

3.2. Results and discussion

At a glance, the results show that DALL-E is the most usable model, while Firefly proves to be the most complex and least approachable. The first performs very well thanks to the simplicity of the chat-like interface and the possibility of further editing the images by asking to modify the model in natural language. ChatGPT's minimalist approach, which hides the complexity of AI behind simple conversations, although it makes the process user friendly, diminishes the user's freedom to operate. While such user friendliness certainly proves profitable in generating more traffic on the platform, limiting the user's freedom and thus leaving more freedom to the generative AI can lead to a variety of problems in the images, such as excessive recurrence to certain biases and difficulty in obtaining very precise requests. Also, the tool sometimes generates only one image at a time, sometimes more, which causes confusion as users are not aware of the reason.

DALL.E design choices are radically opposed to those of Midjourney, which clearly has the highest learning curve, and requires expert knowledge of its commands and prompting style. However, Midjourney is by far the best model in terms of user control. In fact, it is possible to instruct the software with great precision as to how it should operate. Thus, in this case, the complexity of the production is largely managed by the users, which leads to a lower risk of falling back on a standardisation of generated images, but a higher risk that they will somehow manage to bypass the ethical limits imposed by AI by generating inappropriate content. Moreover, while for an experienced and passionate audience the ability to have a big influence on production can certainly be a winning feature, Midjourney certainly does not prove to be the most user friendly among generative AI tools. Also, it lacks consistency both in the parameter ranges for image generation and in the layout of buttons in relation to the images (Nielsen 2023b).

When one considers the fact that repeating an identical prompt leads to the generation of potentially infinite different images, the addition of numerous variables can create a sense of disorientation in the user. Paradoxically, the exponential growth in the use of Midjourney was accompanied by rapid development of new features, and even the most experienced users may have found themselves lost in the speed with which new releases of the software are made available, with the associated changes in controlling newer versions. It seems clear that Midjourney ranks among the AI tools best suited for experienced users but, nevertheless, this allows them to play a leading role in the production of new visual content without completely devolving the work to the software. In essence, in this case designers can have a non-marginal role within the creative process.

The integration of generative AI models into Photoshop was perceived as remaining a work in progress, making its usability somewhat limited. Although not relevant in terms of usability, it should be noted that Firefly is the least effective tool in producing quality results that effectively respond to user requests as it often generates distorted or surreal images. Such problems are not directly related to UX issues, but the tool may benefit from the addition of more advanced controls of the prompt such as negative prompting.

However, this integration may represent a significant development for many professionals, offering substantial potential despite the need for significant enhancements in feedback clarity, error prevention and overall consistency to improve the user experience. For those who are already familiar with the software and use it, possibly for work purposes, having the chance to use an integrated AI tool without having to rely on third-party service by "bouncing" their project between the two is certainly an important feature.

Finally, Stable Diffusion occupies a middle ground. It serves as an accessible introduction to AI for a broad user base and offers a decent level of control, although it could benefit from improvements in feedback clarity, error handling and ease of reversing actions. Also, unlike most of the other tools, Stable Diffusion only generates one image at a time, making the overall creative process more time con-

	STABLE DIFFUSION	ADOBE FIREFLY (PHOTOSHOP)	DALL.E 3	MIDJOURNEY
Strive for consistency	3	2	5	4
Seek universal usability	3	2	5	1
Offer informative feedback	2	1	2	5
Design dialogs to yield closure	4	2	2	4
Prevent errors	1	1	3	1
Permit easy reversal of actions	1	4	5	4
Keep users in control	4	1	2	5
Reduce short-term memory load	3	4	5	1
SCORE	21	17	29	25

Table 1. Heuristics evaluation of the selected AI models.

suming. While it may be surpassed by other tools in terms of features, its unique position as an open-access option sets it apart in the landscape of generative AI tools.

4 CONCLUSIONS

We highlighted how the field of generative AI is evolving very rapidly with diametrically opposed approaches from various service providers. At the same time, some collateral issues that the emergence of these new tools is bringing about, related to the ecological and ethical sphere, were brought to attention. The minimum common denominator between the rapid uncontrolled, multi-modal evolution and associated issues is the centrality of the role of designers in being able to guide or not guide this flow by allocating proper attention to the necessary concerns.

However, the paper pointed out that very little has been done so far in this direction.

Indeed, regardless of the various approaches that have been shown, none of them present elements aimed at counteracting the issues highlighted. In all the case studies analysed, for example, there is no indication of the energy consumption required to generate an image. Only Stable Diffusion explicitly discloses the time required, while Midjourney – to which one subscribes by paying for hours of GPU work – reports how time consuming it will be to generate the image in those terms. Furthermore, no service indicates the sources from which the images are generated, nor does it warn the user in any way about possible misbehaviour. In general, almost all services simply block certain words but do not actively act to fight any attitudes that are on the borderline of what is permissible, i.e., that cannot be banned but could still lead to non-moral uses of the graphic content produced.

As a society we are moving toward ecological awareness and an increased concern about the advancement of technological capabilities. Implementing such features would certainly be an important step for a generative AI service. Careful UX design in this context can be instrumental in providing a coherent response to the issues expressed in the introduction. If legislation struggles to keep up with the very rapid evolution of AI, and the inordinate use of AI could contribute to the ever-increasing energy footprint of technology in our lives, the role of interface designers might contribute to being able to counter these deleterious phenomena. To achieve positive results, it is not necessary to disrupt the functioning of these tools, but proper information about the risks involved in using generative AI tools or their ecological impact in order to make users aware of these topics may be sufficient. One could also consider implementing some gentle nudging strategies to direct users toward good behaviours, taking inspiration from more established digital tools such as Google Maps, which introduced the eco-friendly route in 2021 as a default choice, calculating it through parameters like road incline and traffic congestion (Green Nudges n.d.). AI tools could ask for confirmation before performing tasks they consider unethical, or they

could communicate the energy usage needed to generate the request before starting the process. This may allow users to define their requests taking into account energy demand as well, together with usual parameters such as size, format and visual style. Midjourney, which bases part of its success on social community dynamics, could somehow highlight the most virtuous users in terms of energy conservation, thus enticing people to use the service consciously. These are just some of the possibilities in this regard; what is certain is that the road to further growth of generative AI is paved and that in the future we cannot ignore such matters.

The main issue in generative AI can be summed up as a lack of transparency and interpretability: the generative process is not easily interpretable by humans, its sources are rarely identifiable, its environmental impact is unknown and its outputs are quite unpredictable. This condition leads to a reduction in trust, which instead is a theme directly connected to the main dimensions of UX (i.e. utility and usability) (Lew and Schumacher 2020). Further exploration on ways to enhance interaction with generative AI could study strategies such as employing visualisation tools, leveraging attention models or producing explanations in tandem with the outputs (Bandi et al. 2023).

Designers are thus presented with a strategic chance to influence the development of generative AI by adopting an approach that is multidisciplinary, impactful and conscientious. Their role involves deeply engaging with the intricacies of these sophisticated AI systems, comprehending their mechanisms and evaluating the wider systemic effects of their work.

ACKNOWLEDGEMENTS

The work was carried out under the supervision of M. Menichinelli, who helped define the concept and revise the contribution. A. Vacanti and F. Burlando contributed to the definition of the methodology, writing and investigation. A. I. Paz contributed to the editing and heuristic investigation.

BIOGRAPHIES

Annapaola Vacanti, PhD
Università Iuav di Venezia;
Elisava, Barcelona School of Design and Engineering (UVic-UCC)

Designer and research fellow at Università Iuav di Venezia, and visiting researcher at Elisava, Annapaola Vacanti works on human-technology interaction, exploring the intersection between human factors, technological development and the latter's social and ecological impact. She received her PhD in Design from the Università di Genova in 2022. Parallel to her academic career, since 2018 she has been the art director and organiser of TEDxGenova, an autonomous event operating under official TED license for the local dissemination of valuable ideas.

Francesco Burlando, PhD
Università di Genova;
Elisava, Barcelona School of Design and Engineering (UVic-UCC)

Designer and research fellow at the Università di Genova, visiting researcher at Elisava, and visiting professor at Beijing University of Chemical Technology. His research focuses on fostering the design of innovative technological systems and, especially, the relationship between them and the users for whom they are intended. He received his doctorate from the Università di Genova in 2022 with a thesis entitled "More Than Humanoid. Practices and tools for the design of human* robots".

Alejandro Ivan Paz Ortiz, PhD
Elisava, Barcelona School of Design and Engineering (UVic-UCC)

Iván Paz is an Associate Professor at Elisava, Barcelona School of Design and Engineering (UVic-UCC). He likes to investigate how the intersection between sciences, arts and humanities offers new creative possibilities. In particular, he studies how we interact with machine learning algorithms and the use of code as an interface for real-time creation. His work uses custom-developed software and electronics. He has collaborated with universities, designed interactive systems for festivals and presented works and performances around America and Europe.

Massimo Menichinelli, PhD
Elisava, Barcelona School of Design and Engineering (UVic-UCC)

Massimo Menichinelli is an Associate Professor at Elisava, Barcelona School of Design and Engineering (UVic-UCC), who holds a Doctor of Arts in New Media (Aalto University) and a Master of Science in Industrial Design (Politecnico di Milano). Since 2005 he has worked and researched on the role of design in developing ecosystems and infrastructures that support collaborative, distributed and open processes. He researches how to visualize and co-design design processes, how to measure the impact of design activities, how to develop and support social and digital innovation initiatives, how to develop and research platforms and how to promote the collaboration and creation of ecosystems between university and industry.

ENDNOTES

1. <https://www.crayon.com/>
2. <https://www.freepik.com/ai/image-generator>
3. <https://stablediffusionweb.com/StableDiffusionXL>

REFERENCES

- Bandi, A., P. V. S. R. Adapa, and Y. E. V. P. K. Kuchi. 2023. "The power of generative AI: A review of requirements, models, input-output formats, evaluation metrics, and challenges". *Future Internet* 15 (8): 260, 1-60. <https://doi.org/10.3390/fi15080260>
- Caramiaux, Baptiste, and Sarah Fdili Alaoui. 2022. "Explorers of Unknown Planets: Practices and Politics of Artificial Intelligence in Visual Arts." In *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, issue CSCW2, article 477, 1-24. New York: ACM. <https://doi.org/10.1145/3555578>
- Cattabriga, Andrea, and Vladan Joler. 2023. "Decentering Design With AI". *Diid — Disegno Industriale Industrial Design* 80 (12): 10-19. <https://doi.org/10.30682/diid8023a>
- Crawford, Kate. 2021. *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press.
- Crawford, Kate, and Trevor Paglen. 2021. "Excavating AI: The Politics of Images in Machine Learning Training Sets." *AI & Society* 36, no. 4: 1105-1116. <https://doi.org/10.1007/s00146-021-01162-8>
- Dove, Graham, Kim Halskov, Jodi Forlizzi, and John Zimmerman. 2017. "UX Design Innovation: Challenges for Working with Machine Learning as a Design Material". In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 278-288. New York: ACM. <https://doi.org/10.1145/3025453.3025739>
- Green Nudges. n.d. "Promoting greener driving: Google Maps' fuel-efficient routes." *Green Nudges*. Accessed 03/04/2024. <https://www.green-nudges.com/google-maps/>
- Hintemann, Ralph, and Simon Hinterholzer. 2022. *Data centers 2021: Data center boom in Germany continues - Cloud Computing Drives the Growth of the Data Center Industry and Its Energy Consumption*. Berlin: Borderstep Institute. <https://doi.org/10.13140/RG.2.2.31826.43207>
- Holmquist, Lars Erik. 2017. "Intelligence on Tap: Artificial Intelligence as a New Design Material." *Interactions* 24 (4): 28-33. <https://doi.org/10.1145/3085571>
- Lew, Gavin, and Schumacher, Robert M. 2020. *AI and UX: Why artificial intelligence needs user experience*. New York: Apress.
- Luccioni, Alexandra Sasha, Yacine Jernite, and Emma Strubell. 2024. "Power Hungry Processing: Watts Driving the Cost of AI Deployment?". In *FAccT '24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 85-99. New York: ACM. <https://doi.org/10.1145/3630106.3658542>
- Manovich, Lev and Emanuele Arielli. 2024. "Artificial Aesthetics: Generative AI, art and visual media." *Manovich.net*. Accessed 03/04/2024. <http://manovich.net/index.php/projects/artificial-aesthetics>
- Mugunthan, Tarun. 2023. "Overcoming the Articulation Barrier in Generative AI Using Hybrid Interfaces." Nielsen Norman Group. Accessed 02/01/2024. <https://www.nngroup.com/articles/ai-articulation-barrier>
- Nielsen, Jakob. 1994. "Heuristic Evaluation." In *Usability Inspection Methods*, edited by Jakob Nielsen and Robert L. Mack. New York: Wiley.
- Nielsen, Jakob. 2023a. "UX Needs a Sense of Urgency About AI." *UX Tigers*. Accessed 02/01/2024. <https://www.uxtigers.com/post/ux-urgency-ai>
- Nielsen, Jakob. 2023b. "Classic Usability Important for AI." *UX Tigers*. Accessed 03/04/2024. <https://www.uxtigers.com/post/classic-usability-ai>
- Paoletti, Ingrid. 2021. *Siate materialisti!*. Torino: Einaudi.

- Pasquinelli, Matteo, and Vladan Joler. 2021. "The Nooscope Manifested: AI as Instrument of Knowledge Extractivism". *AI & Society* 36: 1263-1280. <https://doi.org/10.1007/s00146-020-01097-6>
- Shneiderman, Ben, Catherine Plaisant, Maxine Cohen, Steven Jacobs, and Niklas Elmqvist. 2016. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. 6th ed. Boston: Pearson.
- Shumailov, Iliia, Zichao Shumaylov, Yiren Zhao, Yarin Gal, Nicolas Papernot, and Ross Anderson. 2023. "Model Dementia: Generated Data Makes Models Forget." *arXiv e-prints*, arXiv:2305.17493. <https://doi.org/10.48550/arXiv.2305.17493>
- Thiel, David. 2023. "Identifying and Eliminating CSAM in Generative ML Training Data and Models". *Stanford Digital Repository*. <https://doi.org/10.25740/kh752sm9123>
- Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi. 2017. "Transparent, Explainable, and Accountable AI for Robotics." *Science Robotics* 2: eaan6080. <https://doi.org/10.1126/scirobotics.aaan6080>
- Zamfirescu-Pereira, J.D., Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. "Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts." In *CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, article 4437, 1-21. New York: ACM. <https://doi.org/10.1145/3544548.3581388>

A. Vacanti, F. Burlando, A. I. Paz Ortiz
i M. Menichinelli

*Reptes i responsabilitats del disseny
d'UX en models d'IA de text
a imatge: Discussió a través d'una
avaluació heurística comparativa*

Traducció al Català

PARAULES CLAU

Política d'IA, IA Generativa, Text a Imatge, Experiència
d'Usuari, Interfície Gràfica d'Usuari, Avaluació Heurística.

RESUM

El ràpid avenç i l'adopció de la IA generativa en àmbits creatius i professionals marquen una evolució significativa en la interacció tecnològica i plantegen debats fonamentals sobre la seva integració i usabilitat i sobre el seu impacte social més ampli. Aquest article aprofundeix en els múltiples reptes i habilitats que la IA generativa presenta, centrant-se en concret en les eines de conversió de text a imatge i la seva integració en les pràctiques laborals quotidianes d'artistes, dissenyadors, arquitectes i investigadors. A mesura que aquests creadors incorporen a la seva feina la imprevisibilitat i ambigüitat inherents a la IA, es fa imperatiu un examen crític de les seves implicacions ètiques, legislatives i mediambientals. Destaquem que els aspectes relacionats amb l'experiència d'usuari (UX) i la interfície d'usuari (UI) d'aquestes tecnologies revolucionàries continuen estant majoritàriament poc explorats. Jakob Nielsen, pioner de la UX, ha advertit fa poc que els professionals d'aquest camp s'han d'adaptar als ràpids avenços de la IA o es quedaran obsolets per a una indústria que cada vegada és més determinant en l'àmbit creatiu contemporani. Nielsen estableix un paral·lelisme entre l'era actual de la IA i l'auge de les *puntcom* en els anys noranta i fa palesa la

lentitud amb què els professionals de l'experiència d'usuari es van adaptar a internet, que va provocar una pèrdua d'oportunitats i la manca d'un disseny centrat en l'usuari. Com a conseqüència, la usabilitat de les eines d'IA generativa és una preocupació primordial, sobretot perquè els usuaris no experts cada cop utilitzen més aquests sistemes. En aquest article s'analitzen els problemes d'usabilitat, concretament la barrera d'articulació en què els usuaris tenen dificultats per comunicar eficaçment les instruccions a la IA, i la consegüent necessitat d'interfícies més intuïtives i accessibles.

Per mitjà d'una anàlisi comparativa i una avaluació heurística de quatre models capdavanters (Midjourney, Dall-E, Stable Diffusion i Adobe Firefly), que presenta un cas pràctic únic a efectes comparatius gràcies als seus elements divergents d'interfície d'usuari i experiència d'usuari, volem oferir una visió de l'estat actual de les eines d'IA i proposar vies per fer-ne un ús responsable i eficaç.

1 INTRODUCCIÓ

L'any 2023 la IA generativa ha estat el terme de moda i alhora, el principal objecte de crítica, ja que la intel·ligència artificial, uns setanta anys després que s'encunyés el terme, ha començat a introduir-se a les pràctiques laborals quotidianes dels creadors, en el sentit més ampli del terme. La IA generativa consisteix a crear objectes multimèdia amb xarxes neuronals artificials que inclouen text, veu, música, models 3D, conjunts de dades i codi. Termes com ara *imatge generativa*, *imatge d'IA* o *IA visual* fan referència específicament a objectes visuals i abracen imatges estàtiques que imiten diverses formes mediàtiques, des de fotografies fins a dibuixos, i imatges dinàmiques que s'assemblen a animacions i vídeos (Manovich 2024). Artistes, dissenyadors, arquitectes i investigadors estan entrelaçant la IA en la seva feina, utilitzant-ne deliberadament la imprevisibilitat i l'ambigüitat inherents com a forma d'expressió (Caramiaux i Fdili Alaoui 2022). En alguns contextos, la IA s'ha definit com un nou material de disseny que ofereix un gran potencial creatiu (Dove et al. 2017), i també com un mitjà adequat per a l'experimentació (Holmquist 2017).

De totes maneres, hem de reconèixer que dissenyar amb IA vol dir treballar amb sistemes complexos que determinen resultats que són inherentment menys predictibles que els associats a materials i eines tradicionals i passius. Tal com indiquen Cattabriga i Joler (2023), l'autonomia de la IA, caracteritzada per la seva capacitat de fer tasques de formes que poden no seguir estrictament les instruccions facilitades, requereix una comprensió i un compromís més amplis per part dels dissenyadors, més enllà de considerar-la simplement una eina o aplicació. Els dissenyadors han de combinar la seva pràctica amb una comprensió global de les ramificacions polítiques, ètiques i socials dels sistemes d'IA per garantir-ne un ús responsable i descobrir-ne noves vies de creixement professional.

La contínua evolució i el perfeccionament de les eines d'IA generativa, concretament les de conversió de text a imatge, com ara Midjourney, DALL-E i Stable Diffusion, està impulsant la ràpida aparició de noves qüestions en la intersecció entre la interacció humana i els sistemes tecnològics. Aquestes qüestions giren principalment al voltant de tres grans àrees: els reptes legislatius i ètics de la integració de la IA, l'impacte mediambiental del desplegament d'aquestes tecnologies i les consideracions d'usabilitat que garanteixen que aquestes eines siguin accessibles i eficaces per als usuaris previstos.

1.1. El dilema legislatiu i social
A mesura que augmenta la producció de continguts generats per IA, es fa cada cop més evident el paper social de la informació visual digital. Aquests continguts estan arrelant a les pràctiques, habilitats, normes i institucions que regeixen la creació i difusió de la informació, afegint capes de complexitat a la forma en què la informació es contextualitza, interpreta i utilitza en diverses activitats i constructes socials. La interacció dels sistemes d'IA amb la societat comporta diverses i variades implicacions, com l'extracció de coneixements i la explotació del treball creatiu (Pasquinelli i Joler 2020).

El fonament dels sistemes d'IA generativa de text a imatge és una àmplia col·lecció de dades en brut en forma d'imatges, essencials per aprendre i generar continguts nous. A través del procés d'etiquetatge s'assigna a aquestes imatges, mancades de significat inherent, un aspecte central del rendiment i els resultats de la IA. La categorització i l'etiquetatge de

les imatges els concedeix un significat cultural i social, i incrusta supòsits i estereotips socials a aquests conjunts de dades massives (Crawford i Paglen 2021). Els biaixos, estereotips o imprecisions inherents al procés d'etiquetatge es poden amplificar i perpetuar en diferents aplicacions, i això té importants conseqüències en el món real i introdueix conceptes erronis en la representació visual. La mida i l'opacitat d'aquests conjunts de dades plantegen importants reptes als organismes reguladors i als investigadors dedicats a garantir un ús i responsable de la IA. Recentment, per exemple, s'ha descobert que LAION-5B, un enorme conjunt de dades d'entrenament que conté 5000 milions d'imatges estretes d'internet, inclou enllaços a imatges d'abús infantil i això podria permetre que models populars d'IA, com ara Stable Diffusion, creesin continguts perjudicials (Thiel 2023). Mentrestant, ja ens enfrontem a l'anomenat *problema del bucle sintètic*, que fa que els sistemes d'IA s'entrenin amb continguts generats per la IA (Shumailov et al. 2023). Pel que sembla, la barreja de dades sintètiques i dades creades per l'ésser humà fa que els models fallin, que produeixin resultats imprecisos i que no siguin capaços de detectar el que es genera i el que no.

La proliferació descontrolada d'aplicacions web d'IA de lliure accés complica encara més la tasca de regulació, ja que les lleis de drets d'autor s'enfronten a enormes desafiaments. Els conjunts de dades d'entrenament fan un ús extensiu d'obres artístiques de diverses disciplines i actualment això es considera generalment un ús legítim, tot i que és evident que aquest concepte s'està estenent més enllà del seu suposat abast original, consistent a permetre que els creadors es basin en continguts ja existents. A més, cal tenir en compte que, segons els marcs jurídics actuals, les entitats no humanes no poden ser titulars de drets d'autor, un principi que ara s'estén a les creacions generatives d'IA. Com a resultat, ningú reclama legalment l'autoria de la infinitat d'obres generades pels sistemes d'IA, independentment de les indicacions o instruccions humanes proporcionades. S'estan generant milers de milions d'obres sense una autoria clara i això posa en dubte el mateix concepte de propietat artística (Crawford 2021).

1.2. El dilema mediambiental
Tal com passa amb totes les eines de l'àmbit digital, la IA sovint es percep com una entitat intangible. Aquesta consideració per part del públic dona lloc al seu ús desmesurat i desinformat. Tanmateix, la IA requereix un ecosistema per funcionar, i aquest ecosistema està format per infraestructures físiques com ara el dispositiu de l'usuari, el servidor i la xarxa, que comporten una certa despesa energètica (Crawford 2018; Paoletti 2021). Segons les dades més recents, els últims anys el consum elèctric dels centres de dades de tot el món ha augmentat entre un 20 % i un 40 % cada any, que comporta un 1-1,3 % del consum mundial total d'electricitat i un 1 % de les emissions de gasos d'efecte hivernacle provocades pel sector energètic des del 2022 (Hintemann i Hinterholzer 2022).

És un fet àmpliament reconegut que el procés d'entrenament de la IA a través de gegantines bases de dades consumeix molta energia. Tanmateix, és menys coneguda la quantitat d'energia que es consumeix en el procés de generació de continguts a través de la inferència. Tot i que el processament d'una sola instància durant la inferència exigeix considerablement menys computació en comparació amb l'entrenament del mateix model, la inferència es produeix molt més sovint (potencialment fins a milers de milions de vegades al dia per als models que sustenten els serveis d'ús generalitzat). El model menys eficient energèticament per generar imatges consumeix la quantitat d'energia equivalent a 950 càrregues de telèfons intel·ligents (11,49 kWh), que equival aproximadament a una càrrega per cada imatge produïda. Tanmateix, el consum d'energia varia significativament entre els diferents models de generació d'imatges, depenen en gran mesura de la mida de les imatges que produeixen (Luccioni et al. 2023). En una seu social que, amb raó, fa cada vegada més atenció a les qüestions mediambientals i al consum conscient, el creixement exponencial de l'ús excessiu de solucions digitals, entre les quals durant l'últim any han aparegut eines d'IA generativa, s'erigeix en una paradoxal contradicència.

1.3.El dilema de la usabilitat
Cada sistema d'IA té un cert grau d'autonomia i capacitat de presa de decisions, una característica integral del seu disseny (Wachter et al. 2017). Aquesta autonomia és essencial, ja que permet que la IA faci tasques complexes de forma independent. Tanmateix, la utilitat dels sistemes d'IA generativa depèn de la capacitat de l'usuari per comprendre i utilitzar eficaçment aquestes tecnologies. La manca de comprensió en el maneig d'aquests sistemes autònoms mina paradoxalment la seva pretesa eficàcia i augment creatiu. Per tant, a mesura que la IA generativa segueixi evolucionant, també

caldrà posar l'accent en les qüestions d'interacció per mitjà d'un disseny que n'augmenti la transparència i la interpretabilitat. La popularitat que han pogut assolir les eines d'IA generativa ha fet que en poc temps un gran nombre d'usuaris, sovint no experts, comencin a utilitzar-les i això fa encara més urgent la necessitat de facilitar-ne l'ús. Això contrasta amb la complexitat de les possibilitats que ofereix la IA generativa, i hi ha el risc de caure en la paradoxal situació de disposar d'una tecnologia molt potent a l'abast de tothom, però no poder utilitzar-la plenament. No fem referència a la qualitat o la noblesa del seu ús, sinó a la barrera que probablement es crearà entre les expectatives de l'usuari i el resultat real. Seria com ser al davant del geni d'una làmpada que no parla la mateixa llengua que tu.

Jakob Nielsen (2023a), pioner de la UX, adverteix que els professionals del sector s'han d'adaptar als vertiginosos avenços de la IA o es quedaran obsolets per a un sector que cada vegada és més determinant en l'àmbit creatiu contemporani. Nielsen estableix un paral·lelisme entre l'era actual de la IA i l'auge de les *puntcom* els anys noranta i fa palesa la lentitud amb què els professionals de l'experiència d'usuari es van adaptar a internet, que va provocar una pèrdua d'oportunitats i la manca d'un disseny centrat en l'usuari. Va predir que en el futur faltarien experts en UX amb experiència en usabilitat de la IA, i va destacar les implicacions professionals d'aquest camp. La IA generativa comporta l'arribada del tercer paradigma d'interfície d'usuari en la història de la informàtica, que va passar de la interacció basada en ordres (primer paradigma d'interfície d'usuari) a l'especificació de resultats basada en la intenció (segon paradigma d'interfície d'usuari) (Nielsen, 2023a). Aquest canvi de paradigma revoluciona la forma en què els usuaris interactuen amb la tecnologia, passant de dir a l'ordinador com fer les tasques a especificar el resultat desitjat (Zamfirescu-Pereira et al. 2023). A diferència dels paradigmes anteriors, els usuaris dels sistemes basats en la intenció comuniquen els seus objectius i el sistema determina els passos necessaris per assolir aquests resultats. Tot i que és extremadament eficient en termes de temps, l'abstracció d'ordres específiques a intencions més àmplies pot enfosquir el procés i dificultar que els usuaris corregeixin o perfeccionin els resultats. En segon lloc, sorgeix un repte notable en forma de barrera d'articulació (Mugunthan 2023). Per generar imatges d'alta qualitat, els usuaris han d'escriure instruccions complexes i articulades, i això requereix la capacitat de comunicar-se amb precisió per escrit. Els usuaris no només han de descriure el tema de la imatge sinó també articular conceptes abstractes relacionats amb l'estètica visual. Aquest requisit representa una tasca sofisticada, especialment difícil per a les persones que no estan familiaritzades amb estils artístics específics o que no tenen el vocabulari tècnic necessari per expressar la seva visió amb eficàcia.

2 AVANTGUARDA EN UI/UX DE LES EINES D'IA DE TEXT A IMATGE

Tot i que al mercat apareixen noves funcions i models a un ritme increïblement ràpid, actualment es pot accedir als models d'IA de text a imatge a través de tres enfocaments principals que descrivim breument a l'apartat següent, en què proporcionem un mapa de les principals funcions disponibles en el moment d'escriure aquest article i oferim exemples reals per a cadascun d'aquests enfocaments.

2.1. Aplicació i/o pàgina web específica
Els generadors de text a imatge que s'executen en pàgines web específiques solen presentar una interfície extremadament simplificada que permet que els usuaris escriguin la seva petició i, sovint, afegeixin algunes especificacions a través d'una interfície gràfica d'usuari (GUI) que varia molt d'una eina a una altra. A l'escenari més bàsic, a l'usuari només se li presenta un quadre d'entrada on pot escriure la seva petició. De vegades, el quadre d'entrada ofereix suggeriments d'escriptura i exemples per ajudar els usuaris inexperts. Tanmateix, les interfícies més recents solen incloure elements de GUI, com ara quadres de petició negativa, botons, commutadors, caselles de verificació, menús desplegables i controls lliscants. Aquests elements se solen utilitzar per millorar el control dels usuaris sense imposar-los la necessitat d'escriure peticions llargues i redundants (la redundància s'utilitza a les peticions per forçar el model a generar una característica desitjada de la imatge).

De moment, hi ha una gran varietat i manca d'estandardització a les GUI de les eines d'IA. El generador gratuït Craiyon¹ només ofereix un

A. VACANTI ET AL.

nombre limitat de botons per triar l'estil artístic de la imatge i una casella d'entrada per a les peticions negatives. No obstant, ajuda l'usuari buscant inspiracions rellevants a la seva base de dades d'imatges generades mitjançant la recerca de paraules clau en la petició (vegeu la Fig. 1, on vam escriure “un alvocat amb una cara somrient” i van aparèixer alvocats antropomorfitzats abans de prémer el botó de dibuixar).

En canvi, Freepik AI Image Editor² vol reduir la longitud i complexitat de la petició amb un menú lateral detallat amb diverses opcions que l'usuari pot triar per generar la imatge amb relació a la proporció, l'estil el color, la il·luminació i l'enquadrament (Fig. 2). Les opcions seleccionades apareixen com a banderes a l'enquadrament d'entrada just abans de la petició.

Stable Diffusion³, probablement l'eina de conversió de text a imatge en codi obert més famosa, sembla que està adreçada a usuaris professionals ja que la seva interfície ofereix un nivell de control més elevat que, al seu torn, no fa suggeriments visuals ni escrits. Utilitza diversos elements de GUI per controlar diferents aspectes de la generació: per exemple, quadres d'entrada de peticions i peticions negatives, menú desplegable d'estils, controls lliscants d'ajustaments avançats com llavor, escala d'orientació i alçada/amplada, i un botó per triar velocitat sobre qualitat o viceversa (Fig. 3).

2.2. Integració al software
La integració d'eines d'IA generativa de text a imatge al *software* professional vol augmentar el procés creatiu introduint un nou mètode dinàmic per a la creació de continguts. En un ecosistema de *software* típic, aquestes eines se solen integrar com un mòdul específic o una funció accessible a través de la barra d'eines o els menús principals i això en garanteix el fàcil descobriment sense interrompre el flux de treball de l'usuari. La interfície consisteix en un camp d'entrada de text, sovint acompanyat d'indicacions o exemples útils, i possiblement funcions com ara autocompletar o suggeriments sintàctics per ajudar a elaborar peticions eficaces. Com que en aquest cas la generació d'imatges forma part d'un procés creatiu més ampli, el flux de la UX se centra a permetre que els usuaris perfeccionin o incorporin les imatges al projecte per mitjà d'eines d'edició manual. Per tant, l'eina generativa s'ha d'integrar perfectament al conjunt més ampli d'eines del *software*. A més, a causa de la naturalesa nova de la IA generativa, hi sol haver funcions d'ajuda, tutorials o consells integrats, en forma d'ajuda en línia, vídeos tutorials o galeries d'exemples.

Adobe ha desenvolupat el seu propi model generatiu, Firefly, que és accessible des d'una interfície web i des dels seus programes més utilitzats, com ara Photoshop i Illustrator. A Photoshop (Fig. 4), la GUI està integrada a l'entorn nadiu, probablement amb l'objectiu de reduir la corba d'aprenentatge dels usuaris existents. L'eina es limita a utilitzar un senzill quadre d'entrada de peticions. El model sempre genera tres imatges, que es mostren de forma destacada a l'àrea principal del llenç, i l'usuari pot anar canviant entre les diverses opcions o generar més versions. Les imatges generades es poden manipular fàcilment en una capa específica, mantenint la funcionalitat principal del *software*.

Queda clar que l'accés a l'ús d'aquest tipus d'IA està reservat principalment a aquelles persones que ja utilitzen el *software*. De fet, Photoshop ja és complex per si mateix i és poc probable que un usuari comenci a utilitzar-lo només per aprofitar aquesta nova funció.

2.3. Integració en altres plataformes
La integració d'eines generatives de text a imatge en plataformes representa un enfocament matísat, centrat a millorar la interacció de l'usuari sense complicar gaire l'estructura de serveis existent. Aquestes integracions s'han de dissenyar perquè semblin extensions naturals de les plataformes originals i permetin que els usuaris interactuïn amb funcionalitats complexes d'IA a través d'interfícies i interaccions familiars. Malgrat tot, aquesta elecció pot resultar un arma de doble tall. La decisió de basar-se en plataformes preexistents implica inevitablement basar-se també en la UX i la UI d'aquestes últimes. Els dissenyadors no tenen control absolut sobre l'experiència d'usuari i, per tant, l'èxit de la IA generativa dependrà en gran mesura de factors sobre els quals tenen un control insignificant.

En la seva última versió, DALL-E 3 s'ha integrat a ChatGPT, la reeixida IA de text a text. En aquest cas, l'elecció resulta especialment encertada, ja que la plataforma nadiua en què s'integra l'eina ja està dissenyada per a aquest ús específic, amb l'única variant que fins aleshores només podia generar un resultat textual. DALL-E 3 es basa en la UX i la UI de ChatGPT (Fig. 5), que proporcionen una interfície minimalista en què la interacció amb el *software* es fa essencialment a través de peticions conversacionals senzilles en un entorn similar al d'un xat. Els usuaris poden obtenir imatges

TEMES DE DISSENY #40

ESTUDI DE CAS

166

167

Francesco Burlando, PhD
Universit  di Genova;
Elisava, Facultat de Disseny i Enginyeria de
Barcelona (UVic-UCC)

Francesco Burlando  s dissenyador i becari d’investigaci  a la Universit  di Genova, investigador visitant a Elisava i professor visitant de la Beijing University of Chemical Technology. La seva investigaci  se centra a fomentar el disseny de sistemes tecnol gics innovadors i, especialment, la relaci  entre aquests i els usuaris als quals estan destinats. L’any 2022 es va doctorar a la Universit  di Genova amb una tesi titulada “M s que humanoides. Pr ctiques i eines per al disseny de robots humans”.

Alejandro Iv n Paz Ortiz
Elisava, Facultat de Disseny i Enginyeria de
Barcelona (UVic-UCC)

Iv n Paz  s professor associat a Elisava, Facultat de Disseny i Enginyeria de Barcelona (UVic-UCC). Li agrada investigar com la intersecci  entre ci ncies, arts i humanitats ofereix noves possibilitats creatives. En concret, estudia com interactuem amb els algorismes d’aprenentatge autom tic i l’ s del codi com a interf cie per a la creaci  en temps real. El seu treball utilitza *software* i electr nica desenvolupats a mida. Ha col·laborat amb universitats, ha dissenyat sistemes interactius per a festivals i ha presentat obres i actuacions per Am rica i Europa.

Massimo Menichinelli
Elisava, Facultat de Disseny i Enginyeria de
Barcelona (UVic-UCC)

Massimo Menichinelli  s professor associat a Elisava, Facultat de Disseny i Enginyeria de Barcelona (UVic-UCC), doctor en Nous Mitjans (Aalto University) i m ster en Disseny Industrial (Politecnico di Milano). Des del 2005 investiga i treballa en el paper del disseny en el desenvolupament d’ecosistemes i infraestructures que donen suport a processos col·laboratius, distribuïts i oberts. Investiga com visualitzar i codissenyar processos de disseny, com mesurar l’impacte de les activitats de disseny, com desenvolupar i donar suport a iniciatives d’innovaci  social i digital, com desenvolupar i investigar plataformes i com promoure la col·laboraci  i la creaci  d’ecosistemes entre la universitat i la ind stria.

FIGURES I TAULES
NOTES FINALS
REFER�NCIES

- https://www.crayon.com/
- https://www.freepik.com/ai/image-generator
- https://stablediffusionweb.com/StableDiffusionXL

Veure llistat complet de refer ncies a la p gina 165.



PALABRAS CLAVE
RESUMEN

El r pido avance y la adopci n de la IA generativa en  mbitos creativos y profesionales marcan una evoluci n significativa en la interacci n tecnol gica, planteando debates fundamentales sobre su integraci n, usabilidad e impacto social m s amplio. Este art culo profundiza en los m ltiples retos y oportunidades que la IA generativa presenta, centr ndose en particular en las herramientas de conversi n de texto en imagen y su integraci n en las pr cticas laborales cotidianas de artistas, dise adores, arquitectos e investigadores. A medida que estos creadores incorporan a su trabajo la imprevisibilidad y ambigüedad inherentes a la IA, se hace imperativo un examen cr tico de las implicaciones  ticas, legislativas y medioambientales. Destacamos c mo los aspectos relacionados con la experiencia de usuario (UX) y la interfaz de usuario (UI) de estas tecnolog as revolucionarias siguen estando poco explorados en su mayor a. Jakob Nielsen, pionero de la UX, ha advertido recientemente de que los profesionales de este campo deben adaptarse a los r pidos avances de la IA o se quedar n obsoletos para una industria que cada vez es m s determinante en el  mbito creativo contempor neo. Nielsen establece un paralelismo entre la era actual de la IA y el auge de las *puntocom* en los a os 90, poniendo de relieve la lentitud con la que los profesionales de la experiencia de usuario se adaptaron a internet, lo que provoc  la p rdida de oportunidades y la falta de un dise o centrado en el usuario. En consecuencia, la usabilidad de las herramientas de IA generativa es una preocupaci n primordial, sobre todo porque los usuarios no expertos utilizan cada vez m s estos sistemas. En este art culo se analizan los problemas de usabilidad, concretamente la barrera de articulaci n en la que los usuarios tienen dificultades para comunicar eficazmente las instrucciones a la IA, y la consiguiente necesidad de interfaces m s intuitivas y accesibles.

Mediante un an lisis comparativo y una evaluaci n heur stica de cuatro modelos punteros (Midjourney, Dall-E, Stable Diffusion y Adobe Firefly), que presenta un caso pr ctico  nico a efectos comparativos debido a sus elementos divergentes de interfaz de usuario y experiencia de usuario, pretendemos ofrecer una visi n del estado actual de las herramientas de IA y proponer v as para su uso responsable y eficaz.

1

1
INTRODUCCI N

La IA generativa ha sido tanto el t rmino de moda como el principal objeto de cr tica en 2023, ya que la inteligencia artificial, unos setenta a os despu s de que se acu ara el t rmino, ha empezado a introducirse en las pr cticas laborales cotidianas de los creadores, en el sentido m s amplio del t rmino. La IA generativa consiste en crear objetos multimedia con redes neuronales artificiales que incluyen texto, voz, m sica, modelos 3D, conjuntos de datos y c digo. T rminos como *imagen generativa*, *imagen de IA* o *IA visual* se refieren espec ficamente a objetos visuales y abarcan im genes est ticas que imitan diversas formas medi ticas, desde fotograf as hasta dibujos, e im genes din micas que se asemejan a animaciones y v deos (Manovich 2024). Artistas, dise adores, arquitectos e investigadores est n entretejiendo la IA en su trabajo, utilizando deliberadamente su imprevisibilidad y ambigüedad inherentes como forma de expresi n (Caramiaux y Fdili Alaoui 2022). En algunos contextos, la IA se ha definido como un material de dise o novedoso que ofrece un gran potencial creativo (Dove et al. 2017), as  como un medio adecuado para la experimentaci n (Holmquist 2017).

Sin embargo, debemos reconocer que dise ar con IA significa trabajar con sistemas complejos que determinan resultados que son inherentemente menos predecibles que los asociados a materiales y herramientas tradicionales y pasivos. Como se alan Cattabriga y Joler (2023), la autonom a de la IA, caracterizada por su capacidad para realizar tareas de formas que pueden no seguir estrictamente las instrucciones facilitadas, requiere una compresi n y un compromiso m s amplios por parte de los dise adores, m s all  de considerarla una simple herramienta o aplicaci n. Los dise adores deben combinar su pr ctica con una compresi n global de las ramificaciones pol ticas,  ticas y sociales de los sistemas de IA para garantizar un uso responsable y descubrir nuevas v as de crecimiento profesional.

La continua evoluci n y perfeccionamiento de las herramientas de IA generativa, en particular las de conversi n de texto en imagen, como Midjourney, DALL-E y Stable Diffusion, est  impulsando la r pida aparici n de nuevas cuestiones en la intersecci n entre la interacci n humana y los sistemas tecnol gicos. Estas cuestiones giran principalmente en torno a tres grandes  reas: los retos legislativos y  ticos de la integraci n de la IA, el impacto medioambiental del despliegue de estas tecnolog as y las consideraciones de usabilidad que garantizan que estas herramientas sean accesibles y eficaces para los usuarios previstos.

1.1. El dilema legislativo y social
A medida que aumenta la producci n de contenidos generados por IA, se hace cada vez m s evidente el papel social de la informaci n visual digital. Estos contenidos se est n arraigando en las pr cticas, habilidades, normas e instituciones que rigen la creaci n y difusi n de la informaci n, a adiendo capas de complejidad a la forma en que la informaci n se contextualiza, interpreta y utiliza en diversas actividades y constructos sociales. La interacci n de los sistemas de IA con la sociedad conlleva varias y diversas implicaciones, como la extracci n de conocimientos y la explotaci n del trabajo creativo (Pasquinelli y Joler 2020).

El fundamento de los sistemas de IA generativa de texto a imagen es una amplia colecci n de datos en bruto en forma de im genes, esenciales para aprender y generar contenidos novedosos. A estas im genes, desprovistas de significado inherente, se les asigna un significado a trav s del proceso de etiquetado, un aspecto central del rendimiento y los resultados de la IA. La categorizaci n y el etiquetado de las im genes les confiere un significado cultural y social, e incrusta supuestos y estereotipos sociales en estos conjuntos de datos masivos (Crawford y Paglen 2021). Los sesgos, estereotipos o impresiones inherentes al proceso de etiquetado pueden amplificarse y perpetuarse en diversas aplicaciones, lo que tiene importantes consecuencias en el mundo real e introduce conceptos err neos en la representaci n visual. El tama o y la opacidad de estos conjuntos de datos plantean importantes retos a los organismos reguladores y a los investigadores dedicados a garantizar un uso justo y responsable de la IA. Recientemente, por ejemplo, se ha descubierto que LAION-5B, un enorme conjunto de datos de entrenamiento que contiene 5000 millones de im genes extra das de internet, contiene enlaces a im genes de abuso infantil, lo que podr a permitir que modelos populares de IA como Stable Diffusion crearan contenidos perjudiciales (Thiel 2023). Mientras tanto, nos enfrentamos ya al llamado *problema del bucle sint tico*, que hace que los sistemas de IA se

entrenen con contenidos generados por la IA (Shumailov et al. 2023). Al parecer, la mezcla de datos sint ticos y datos creados por el ser humano hace que los modelos fallen, que produzcan resultados imprecisos y que no sean capaces de detectar lo que se genera y lo que no.

La proliferaci n descontrolada de aplicaciones web de IA de libre acceso complica a n m s la tarea de regulaci n, ya que las leyes de derechos de autor se enfrentan a enormes desaf os. Los conjuntos de datos de entrenamiento hacen un uso extensivo de obras art sticas de diversas disciplinas, lo que en la actualidad se considera generalmente un uso leg timo, aunque es evidente que este concepto se est  extendiendo m s all  de su supuesto alcance original, consistente en permitir a los creadores basarse en contenidos ya existentes. Adem s, hay que tener en cuenta que, seg n los marcos jur dicos actuales, las entidades no humanas no pueden ser titulares de derechos de autor, un principio que ahora se extiende a las creaciones generativas de IA. Como resultado, nadie reclama legalmente la autor a del sinfin de obras generadas por los sistemas de IA, independientemente de las indicaciones o instrucciones humanas proporcionadas. Se est n generando miles de millones de obras sin una autor a clara, lo que pone en tela de juicio el concepto mismo de propiedad art stica (Crawford 2021).

1.2. El dilema medioambiental
Como ocurre con todas las herramientas pertenecientes al  mbito digital, la IA se percibe a menudo como una entidad intangible. Esta side-raci n por parte del p blico da lugar a su uso desmesurado y desinformado. Sin embargo, la IA requiere un ecosistema para funcionar, y ese ecosistema se compone de infraestructuras f sicas como el dispositivo del usuario, el servidor y la red, que conllevan un cierto gasto energ tico (Crawford 2018; Paoletti 2021). Seg n los datos m s recientes, en los  ltimos a os el consumo el ctrico de los centros de datos de todo el mundo ha aumentado entre un 20 % y un 40 % cada a o, lo que supone un 1-1,3 % del consumo mundial total de electricidad y un 1 % de las emisiones de gases de efecto invernadero provocadas por el sector energ tico desde 2022 (Hintemann y Hinterholzer 2022).

Es un hecho ampliamente reconocido que el proceso de entrenamiento de la IA a trav s de gigantescas bases de datos consume mucha energ a. Sin embargo, se sabe menos sobre la cantidad de energ a que se consume en el proceso de generaci n de contenidos a trav s de la inferencia. Aunque el procesamiento de una sola instancia durante la inferencia exige considerablemente menos computaci n en comparaci n con el entrenamiento del mismo modelo, la inferencia se produce con mucha m s frecuencia (potencialmente hasta miles de millones de veces al d a para los modelos que sustentan los servicios de uso generalizado). El modelo menos eficiente energ ticamente para generar im genes consume la cantidad de energ a equivalente a 950 cargas de tel fonos inteligentes (11,49 kWh), lo que equivale aproximadamente a una carga por cada imagen producida. Sin embargo, el consumo de energ a var a significativamente entre los distintos modelos de generaci n de im genes, dependiendo en gran medida del tama o de las im genes que producen (Luccioni et al. 2023). En una sociedad que, con raz n, presta cada vez m s atenci n a las cuestiones medioambientales y al consumo consciente, el crecimiento exponencial del uso excesivo de soluciones digitales, entre las que durante el  ltimo a o han aparecido herramientas de IA generativa, se erige en una parad jica contratendencia.

1.3. El dilema de la usabilidad
Cada sistema de IA posee un cierto grado de autonom a y capacidad de toma de decisiones, una caracter stica integral de su dise o (Wachter et al. 2017). Esta autonom a es esencial, ya que permite a la IA realizar tareas complejas de forma independiente. Sin embargo, la utilidad de los sistemas de IA generativa depende de la capacidad del usuario para comprender y utilizar eficazmente estas tecnolog as. La falta de compresi n en el manejo de estos sistemas aut nomos socava parad jicamente su pretendida eficacia y aumento creativo. Por tanto, a medida que la IA generativa siga evolucionando, tambi n habr  que hacer hincapi  en las cuestiones de interacci n mediante un dise o que aumente la transparencia y la interpretabilidad. La popularidad que han podido alcanzar las herramientas de IA generativa ha hecho que en poco tiempo un gran n mero de usuarios, a menudo no expertos, empiecen a utilizarlas, lo que hace a n m s urgente la necesidad de facilitar su uso. Esto contrasta con la complejidad de las posibilidades que ofrece la IA generativa, y existe el riesgo de caer en la parad jica situaci n de disponer de una tecnolog a muy potente al alcance de todos, pero sin poder utilizarla plenamente. No nos referimos a la calidad o nobleza de su uso, sino a la barrera que probablemente se crear  entre las expectativas del

usuario y el resultado real. Sería como estar ante el genio de una lámpara que no habla el mismo idioma que tú.

Jakob Nielsen (2023a), pionero de la UX, advierte de que los profesionales del sector deben adaptarse a los vertiginosos avances de la IA o se quedarán obsoletos para un sector que cada vez es más determinante en el ámbito creativo contemporáneo. Nielsen establece un paralelismo entre la era actual de la IA y el auge de las *puntocom* en los años 90, poniendo de relieve la lentitud con la que los profesionales de la experiencia de usuario se adaptaron a internet, lo que provocó la pérdida de oportunidades y la falta de un diseño centrado en el usuario. Predijo que en el futuro escasearían los expertos en UX con experiencia en usabilidad de la IA, y subrayó las implicaciones profesionales de este campo. La IA generativa supone la llegada del tercer paradigma de interfaz de usuario en la historia de la informática, que pasó de la interacción basada en comandos (primer paradigma de interfaz de usuario) a la especificación de resultados basada en la intención (segundo paradigma de interfaz de usuario) (Nielsen, 2023a). Este cambio de paradigma revoluciona la forma en que los usuarios interactúan con la tecnología, pasando de decirle al ordenador cómo realizar las tareas a especificar el resultado deseado (Zamfirescu-Pereira et al. 2023). A diferencia de los paradigmas anteriores, los usuarios de los sistemas basados en la intención comunican sus objetivos y el sistema determina los pasos necesarios para alcanzar esos resultados. Aunque es extremadamente eficiente en términos de tiempo, la abstracción de comandos específicos a intenciones más amplias puede oscurecer el proceso, dificultando que los usuarios corrijan o perfeccionen los resultados. En segundo lugar, surge un reto notable en forma de barrera de articulación (Mugunthan 2023). Para generar imágenes de alta calidad, los usuarios tienen que escribir instrucciones complejas y articuladas, lo que requiere la capacidad de comunicarse con precisión por escrito. Los usuarios no solo tienen que describir el tema de la imagen, sino también articular conceptos abstractos relacionados con la estética visual. Este requisito representa una tarea sofisticada, especialmente difícil para las personas que no están familiarizadas con estilos artísticos específicos o que carecen del vocabulario técnico necesario para expresar su visión con eficacia.

2 VANGUARDIA EN UI/UX DE LAS HERRAMIENTAS DE IA DE TEXTO A IMAGEN

Aunque en el mercado aparecen nuevas funciones y modelos a un ritmo increíblemente rápido, en la actualidad se puede acceder a los modelos de IA de texto a imagen a través de tres enfoques principales que describimos brevemente en el siguiente apartado, en el que proporcionamos un mapa de las principales funciones disponibles en el momento de escribir este artículo y ofrecemos ejemplos reales para cada uno de estos enfoques.

2.1. Aplicación y/o página web específica
Los generadores de texto a imagen que se ejecutan en páginas web específicas suelen presentar una interfaz extremadamente simplificada que permite a los usuarios escribir su petición y, a menudo, añadir algunas especificaciones a través de una interfaz gráfica de usuario (GUI) que varía mucho de una herramienta a otra. En el escenario más básico, al usuario solo se le presenta un cuadro de entrada donde puede escribir su petición. A veces, el cuadro de entrada ofrece sugerencias de escritura y ejemplos para ayudar a los usuarios inexpertos. Sin embargo, las interfaces más recientes suelen incluir elementos de GUI, como cuadros de petición negativa, botones, conmutadores, casillas de verificación, menús desplegados y controles deslizantes. Estos elementos se suelen utilizar para mejorar el control de los usuarios sin imponerles la necesidad de escribir peticiones largas y redundantes (la redundancia se utiliza en las peticiones para forzar al modelo a generar una característica deseada de la imagen).

Por el momento, hay una gran variedad y falta de estandarización en las GUI de las herramientas de IA. El generador gratuito Craiyon¹ solo ofrece un número limitado de botones para elegir el estilo artístico de la imagen y una casilla de entrada para las peticiones negativas. Sin embargo, ayuda al usuario buscando inspiraciones relevantes en su base de datos de imágenes generadas mediante la búsqueda de palabras clave en la petición (véase la Fig. 1, donde escribimos “un aguacate con una cara sonriente” y aparecieron varios aguacates antropomorfizados antes de pulsar el botón de dibujar).

TEMES DE DISSENY #40

En cambio, Freepik AI Image Editor² pretende reducir la longitud y complejidad de la petición mediante un menú lateral detallado con varias opciones entre las que el usuario puede elegir para generar la imagen con relación a la proporción, el estilo, el color, la iluminación y el encuadre de la misma (Fig. 2). Las opciones seleccionadas aparecen como banderas en el cuadro de entrada justo antes de la petición.

Stable Diffusion³, probablemente la herramienta de conversión de texto en imagen de código abierto más famosa, parece estar dirigida a usuarios profesionales por el mayor nivel de control que ofrece su interfaz, que a su vez no realiza sugerencias visuales ni escritas. Utiliza varios elementos de GUI para controlar diferentes aspectos de la generación: por ejemplo, cuadros de entrada de peticiones y peticiones negativas, menú desplegable de estilos, controles deslizantes de ajustes avanzados como semilla, escala de orientación y altura/anchura, y un botón para elegir velocidad sobre calidad o viceversa (Fig. 3).

2.2. Integración en el software
La integración de herramientas de IA generativa de texto a imagen en el *software* profesional pretende aumentar el proceso creativo, introduciendo un nuevo método dinámico para la creación de contenidos. En un ecosistema de *software* típico, estas herramientas suelen integrarse como un módulo específico o una función accesible a través de la barra de herramientas o los menús principales, lo que garantiza su fácil descubrimiento sin interrumpir el flujo de trabajo del usuario. La interfaz consiste en un campo de entrada de texto, a menudo acompañado de indicaciones o ejemplos útiles, y posiblemente funciones como autocompletar o sugerencias sintácticas para ayudar a elaborar peticiones eficaces. Dado que en este caso la generación de imágenes forma parte de un proceso creativo más amplio, el flujo de la UX se centra en permitir a los usuarios perfeccionar o incorporar las imágenes al proyecto mediante herramientas de edición manual. Por tanto, la herramienta generativa debe integrarse perfectamente en el conjunto más amplio de herramientas del *software*. Además, debido a la naturaleza novedosa de la IA generativa, suelen estar presentes funciones de ayuda, tutoriales o consejos integrados, en forma de ayuda en línea, vídeos tutoriales o galerías de ejemplos.

Adobe ha desarrollado su propio modelo generativo, Firefly, accesible desde una interfaz web y desde sus programas más utilizados, como Photoshop e Illustrator. En Photoshop (Fig. 4), la GUI está integrada en el entorno nativo, probablemente con el objetivo de reducir la curva de aprendizaje de los usuarios existentes. La herramienta se limita a utilizar un simple cuadro de entrada de peticiones. El modelo genera siempre tres imágenes, que se muestran de forma destacada en el área principal del lienzo, y el usuario puede ir cambiando entre las diversas opciones o generar más versiones. Las imágenes generadas pueden manipularse fácilmente en una capa específica, manteniendo la funcionalidad principal del *software*.

Está claro que el acceso al uso de este tipo de IA está reservado principalmente a quienes ya utilizan el *software*. De hecho, Photoshop ya es complejo de por sí y es poco probable que un usuario empiece a utilizarlo simplemente para aprovechar esta nueva función.

2.3. Integración en otras plataformas
La integración de herramientas generativas de texto a imagen en plataformas representa un enfoque matizado, centrado en mejorar la interacción del usuario sin complicar en exceso la estructura de servicios existente. Estas integraciones deben diseñarse para que parezcan extensiones naturales de las plataformas originales, permitiendo a los usuarios interactuar con funcionalidades complejas de IA a través de interfaces e interacciones familiares. No obstante esta elección puede resultar un arma de doble filo. La decisión de basarse en plataformas preexistentes implica inevitablemente basarse también en la UX y la UI de estas últimas. Los diseñadores no tienen control absoluto sobre la experiencia de usuario y, por tanto, el éxito de la IA generativa dependerá en gran medida de factores sobre los que tienen un control insignificante.

En su última versión, DALL-E 3 se ha integrado en ChatGPT, la exitosa IA de texto a texto. En este caso, la elección resulta especialmente acertada, ya que la plataforma nativa en la que se integra la herramienta ya está diseñada para ese uso específico, con la única variante de que hasta entonces solo podía generar un resultado textual. DALL-E 3 se basa en la UX y la UI de ChatGPT (Fig. 5), que proporcionan una interfaz minimalista donde la interacción con el *software* se realiza esencialmente a través de peticiones conversacionales sencillas en un entorno similar al de un chat. Los usuarios pueden obtener imágenes generadas como si simplemente

ESTUDIO DE CASO

172

estuvieran conversando con alguien. Gracias a esta elección, los usuarios que ya utilizaban ChatGPT pueden aprovechar un nuevo potencial, y para los que utilizaban DALL-E en la interfaz específica es muy fácil migrar a la nueva plataforma.

Como ya se ha indicado, la interfaz minimalista permite a los usuarios escribir todas las especificaciones para la generación de imágenes en lenguaje conversacional a través de un cuadro de entrada. También es posible adjuntar una imagen a la petición mediante un botón con forma de clip. Una vez generada la imagen, es posible leer la petición final que DALL-E ha utilizado haciendo clic en el resultado, que es una extensión del concepto expresado por el usuario. Por ejemplo, hemos dado instrucciones a la herramienta para “dibujar un aguacate con una cara sonriente” y ChatGPT ha ampliado la petición de esta forma: “Un aguacate alegre con una gran cara sonriente. El aguacate debe ser verde y estar maduro, y su hueso marrón, visible. Tiene los ojos grandes y redondos, y una gran sonrisa de felicidad. El aguacate está cortado por la mitad, erguido, y su piel tiene una superficie texturizada y rugosa, típica de un aguacate Hass”. Para seguir editando, basta con continuar la conversación.

El enfoque de Midjourney, que se integra en la aplicación de mensajería social Discord a través de canales específicos (Fig. 6), presenta una tendencia totalmente distinta. Este genera una interacción comunitaria que convierte el uso de la IA en un proceso social complejo. De hecho, no es raro ver a miembros utilizar la producción de otro usuario como punto de partida para intentar generar algo nuevo, modificándola a través de diferentes peticiones. Además de estas dinámicas, Midjourney permite a los usuarios una gran libertad para operar dentro de la generación de imágenes.

De hecho, ofrece seis comandos diferentes para generar nuevas imágenes, trece parámetros para dirigir la producción hacia ciertas características requeridas y seis herramientas adicionales para modificar aún más la imagen generada. Sin embargo, a diferencia de otras herramientas de IA generativa, la mayoría de estas funciones no se muestran al usuario de forma explícita. La mayoría de los comandos se activan añadiendo algún texto específico al final de la petición, como se puede ver en la parte superior de la Fig. 6. Además, estos comandos suelen seguir un lenguaje de codificación específico. Esta característica convierte a Midjourney en una herramienta adecuada para usuarios experimentados que disponen del tiempo y la habilidad necesarios para informarse sobre su funcionamiento. No obstante, incluso sin saber utilizar todo el potencial de la IA, es posible producir imágenes simplemente introduciendo la petición gracias a diversas posibilidades adicionales que se ofrecen por medio de botones (Fig. 7) y que permiten, por ejemplo, variar la imagen generada, o solo una parte de ella, o hacer una nueva imagen con una vista más amplia. Cabe suponer que Midjourney se encuentra entre los sistemas de IA más difíciles de usar, ya que solo los usuarios más experimentados son capaces de dominar todas las funciones y pueden controlar cómo reaccionará la IA al uso combinado de varios comandos a la vez.

3 EVALUACIÓN HEURÍSTICA COMPARATIVA

3.1. Metodología
Para profundizar en el conocimiento de la usabilidad de las herramientas de IA generativa, se ha realizado un análisis comparativo de cuatro modelos punteros: Stable Diffusion 2.0, Adobe Firefly (integrado en el *software* Photoshop 2024), Dall-E 3 y Midjourney V6. Hemos utilizado el famoso método de inspección de la evaluación heurística, introducido por Jakob Nielsen (1994), para ayudar a los equipos de diseño de UX a identificar los problemas de usabilidad de cada interfaz. La heurística seleccionada incluye las ocho reglas de oro de la interfaz definidas por Shneiderman et al. (2016). Estos principios se han adaptado y aplicado a los aspectos específicos de las herramientas de IA para facilitar nuestro análisis:

- Esfuerzo por la constancia:* La interfaz para introducir peticiones, ajustar parámetros y ver resultados es constante en toda la herramienta. La terminología, la presentación y los elementos de diseño son uniformes.
- Búsqueda de una usabilidad universal:* La herramienta se adapta a una amplia gama de usuarios, desde artistas visuales hasta principiantes, tiene en cuenta sus distintos conocimientos técnicos y necesidades creativas, incluyendo variables o modos ajustables, y garantiza la accesibilidad a personas con discapacidad.

A. VACANTI ET AL.

3. Respuesta informativa: La interfaz ofrece una respuesta inmediata y clara cuando los usuarios introducen peticiones o ajustan parámetros. Cuando la IA está procesando una imagen, muestra indicadores de progreso. Al finalizar, los resultados se muestran con claridad y cualquier error o limitación en el proceso de generación se comunica de forma eficaz.

4. Diseño de diálogos que lleven a una conclusión: El proceso de creación de imágenes se organiza en pasos claros y lógicos. Después de cada acción, el usuario tiene la sensación de haber completado la tarea, tanto si se trata de finalizar una petición como de generar una imagen. Queda claro cuándo se ha finalizado una tarea o cuándo el usuario debe pasar al siguiente paso.

5. Prevención de errores: Las interfaces de introducción de peticiones y ajuste de parámetros están diseñadas para minimizar los errores del usuario. Para ello pueden validar las peticiones, sugerir correcciones u orientar a los usuarios para que eviten problemas conocidos que puedan dar lugar a malos resultados. Se dan orientaciones claras sobre los tipos de peticiones que funcionan bien y las limitaciones del modelo de IA.

6. Posibilidad de revertir acciones fácilmente: Los usuarios pueden deshacer fácilmente las acciones o revisar las peticiones sin tener que empezar desde cero. Si un usuario no está satisfecho con una imagen generada, puede modificar la petición o los parámetros y volver a generar la imagen de forma rápida y eficaz.

7. Los usuarios tienen siempre el control: Los usuarios sienten que guían el proceso creativo y que la IA sirve de herramienta para hacer realidad su visión. Los usuarios controlan todos los parámetros posibles, incluidos estilo, contenido y formato, manteniendo el carácter intuitivo y la sencillez de la interfaz.

8. Reducción de la carga de memoria a corto plazo: Los usuarios no necesitan recordar continuamente información en distintas partes de la herramienta. La información esencial es visible o fácilmente recuperable, especialmente cuando se trata de peticiones complejas o de múltiples parámetros. Cada herramienta de IA se evalúa empleando un baremo del 1 (muy mala) al 5 (muy buena) para cada heurística.

En una evaluación heurística, los expertos revisan un sistema en función de una serie de criterios para detectar problemas de usabilidad. En el contexto de nuestro artículo, los cuatro autores se centraron en las cuatro herramientas de IA generativa seleccionadas valiéndose de su experiencia. El estudio realizado en diciembre de 2023 consistió en navegar por las herramientas y utilizarlas libremente para generar y seguir editando diversas imágenes y probar distintas funciones. Después de otorgar calificaciones individuales, por medio de un debate activo, compartimos nuestras impresiones y experiencias individuales con estas herramientas para llegar a un consenso sobre un marco de evaluación común que garantizase que nuestras conclusiones fuesen exhaustivas y consensuadas entre todos los autores. En el siguiente apartado, presentamos una síntesis de nuestra evaluación sin mencionar la calidad de los resultados visuales, ya que no son el objeto de esta investigación.

3.2. Resultados y discusión
[Véase tabla 1]
A primera vista, los resultados muestran que DALL-E es el modelo más usable, mientras que Firefly resulta ser el más complejo y menos abordable. El primero obtiene muy buenos resultados gracias a la sencillez de la interfaz tipo chat y a la posibilidad de seguir editando las imágenes pidiendo modificar el modelo en lenguaje natural. El enfoque minimalista de ChatGPT, que oculta la complejidad de la IA tras conversaciones sencillas, aunque hace que el proceso sea fácil para el usuario, disminuye la libertad de actuación del mismo. Si bien es cierto que esa facilidad de uso resulta rentable para generar más tráfico en la plataforma, limitar la libertad del usuario y dejar así más libertad a la IA generativa puede acarrear diversos problemas en las imágenes, como la reparación excesiva de ciertos sesgos y la dificultad para obtener peticiones muy precisas. Además, a veces la herramienta genera una sola imagen y otras veces más, lo que provoca confusión, ya que los usuarios no son conscientes del motivo.

Las opciones de diseño de DALL-E son radicalmente opuestas a las de Midjourney, que presenta claramente la mayor curva de aprendizaje y requiere un conocimiento experto de sus comandos y tipos de peticiones. No obstante, Midjourney es, con diferencia, el mejor modelo en términos de control del usuario. De hecho, es posible dar instrucciones muy precisas al

173

This special issue of *Temes de Disseny* explores the ethical, theoretical and practical roles of designers in the 21st century, focusing on design justice, community engagement, digitalization, AI, healthcare and social justice. In today's world, beset by multifaceted crises, these changes must trigger a reflection on the strategic direction of design practice to keep its social and economic importance in society relevant and productive.

Increasingly, designers are focusing on social and ethical considerations, ensuring their processes and outcomes align with sustainability, well-being, privacy and broader societal impact. This shift has sparked a growing interest in examining the complex role of designers through various lenses. This issue of *Temes de Disseny* delves into these topics through original research papers, case studies and a provocation.

Temes de Disseny #40

El Rol dels Dissenyadors a la Societat:
Perspectives Ètiques, Teòriques i Pràctiques.

Aquest número especial de *Temes de Disseny* analitza les funcions ètica, teòrica i pràctica dels dissenyadors el segle XXI, amb un interès especial en la justícia del disseny, la participació de la comunitat, la digitalització, la IA, l'atenció sanitària i la justícia social. Al món actual, sacsejat per crisis multidimensionals, aquests canvis han de donar lloc a una reflexió sobre la orientació estratègica de la pràctica del disseny per mantenir la seva importància social i econòmica en la societat d'una manera rellevant i productiva.

Els dissenyadors cada cop se centren més en consideracions ètiques i socials i s'asseguren que els seus processos i resultats estiguin en harmonia amb la sostenibilitat, el benestar i la privacitat i aconseguixin un ampli impacte social. Aquest gir ha despertat un creixent interès en la revisió de la complexa funció dels dissenyadors des de diferents perspectives. El número de *Temes de Disseny* que teniu a les mans aprofundeix en aquests temes per mitjà d'articles d'investigació originals, estudis de cas i una provocació.

Temes de Disseny #40

El Rol de los Diseñadores en la Sociedad:
Perspectivas Éticas, Teóricas y Prácticas.

Este número especial de *Temes de Disseny* analiza las funciones ética, teórica y práctica de los diseñadores en el siglo XXI, con especial interés en la justicia del diseño, la participación de la comunidad, la digitalización, la IA, la atención sanitaria y la justicia social. En el mundo actual, sacudido por crisis multidimensionales, estos cambios deben dar lugar a una reflexión sobre la orientación estratégica de la práctica del diseño para mantener su importancia social y económica en la sociedad de un modo relevante y productivo.

Los diseñadores se centran cada vez más en consideraciones éticas y sociales y se aseguran de que sus procesos y resultados estén en armonía con la sostenibilidad, el bienestar y la privacidad y consigan un amplio impacto social. Este giro ha despertado un creciente interés en la revisión de la compleja función de los diseñadores desde diferentes perspectivas. El presente número de *Temes de Disseny* profundiza en estos temas a través de artículos de investigación originales, estudios de caso y una provocación.