

MECHANISMS, PATTERNS, PROCESSES: ASSESSING VISUAL GENERATIVE AI AND HUMAN CREATIVITIES AT ONCE

Carlo Turri

Accademia di Belle Arti di Verona (Italy)

Abstract

In early 2025, a legal protection granted by the US Copyright Office to an image composed entirely of AI generated material marked a clear distinction between the human and artificial creativities, positioning the latter as tool to be guided by someone's intent and imagination. Generative visual AI is just a tool then, easy enough to use, yet challenging to master. The seed alone, one of the core parameters common to all image generative systems, is a 32-bit random value allowing users to visualize the same textual prompt into 4 billion possible different ways. Ignoring the calculation time, the simple act of gazing to each single generated image would require approximately 250 years. Combined with limitless stylistic variations and combinations, it is clear how 'inhuman' generative AI is, by utterly surpassing both the physical and cognitive capabilities of even the most skilled and expert group of individuals. Nevertheless, such sophisticated tools have been embedded in the daily practices of professionals, students, and educators. Whereas it seems that this technology will further evolve to match specific disciplinary needs as well as to improve user-machine interactions, we argue that much of the efficiency and the proficiency in using such instruments is to be found in people's hands and minds. Building on Guilford's concept of divergent and convergent thinking, and Goleman's four-stage model of creativity we have investigated such behaviours in a practical case scenario on multiple users, by highlighting significant correlations with the inputs and the responses to and from the software, as well as their interplay with the creators' agency. Based on both quantitative and qualitative data, the subsequent observations of emerging patterns may offer valuable insights in understanding the unfolding connections when the sentient being and the intelligent tool when are operating together, thus contributing to the definition to methodological guidelines in support of the creative work of artists and designers.

Keywords: *Visual Generative AI, education experience, open-source AI tool, cognitive psychology.*

1. Theoretical framework

To properly investigate the nuanced interactions between human agency and artificial intelligence within the 'Suerte' experiment, it is necessary to outline the established psychological theories of creativity at its base.

Historically, the study of human creativity has been dominated by two distinct but complementary perspectives: the macro-level observation of the creative timeline, and the micro-level analysis of cognitive generation. By examining the stage-based model formalized by Wallas (1926) and later popularized by Goleman et al. (1992), combined with the cognitive approach of divergent and convergent thinking proposed by Guilford (1950, 1967), we can observe AI-assisted workflows through a lens that brings human-related factors back into play. Furthermore, the work of Todd Lubart (2001) adds a valuable bridge between these paradigms, offering a dynamic approach highly applicable to contemporary technology-aided artistic practices.

The first attempt to map the temporal evolution of a creative idea is attributed to Graham Wallas (1926). This holistic approach was later expanded and adapted to multiple contexts by Daniel Goleman et al. (1992), who unfolded creativity into four sequential phases: *preparation* (gathering information and consciously framing the problem), *incubation* (a period of unconscious processing where the mind wanders away from the problem), *illumination* (the sudden emergence of an insight or a strong inspiration), and *verification* (the final conscious elaboration and validation of the emerged ideas).

In contrast to describing the act of creating as a linear and time-based process, the psychologist J.P. Guilford (1950, 1967) shifted the focus toward the specific cognitive patterns involved in problem-solving. In this perspective, creativity is grounded in a fluctuating balance between two distinct

and complementary types of thought: divergent and convergent thinking. The first mode implies the ability to generate a wide array of different ideas for an open-ended problem; conversely, the latter involves the analytical process of evaluating different options to identify the most proper way to solve it.

Originally, the model did not provide a specific order nor iterations between diverging and converging thinking. However, when these principles were linked to the key elements characterizing a given context (such as size, duration, and disciplinary field), the Guilford model proved to be extremely helpful and effective in deciphering creative processes.

This is the case with the Double Diamond Model, devised by the British Design Council (2005), which chains two consecutive divergent/convergent phases: the first couple is necessary to properly address the problem (discover and define, respectively), while the second pair (develop and deliver) leads toward the best solution. Most importantly, such an approach not only provides a rational and believable understanding of most design journeys, but it also provides a robust methodology and evaluation model widely applied in design learning environments (van Boeijen et al., 2020).

For decades, the temporal stage framework proposed by Wallas and the cognitive action model proposed by Guilford were treated as somewhat parallel discourses. However, researchers like Todd Lubart (2001) offered a way to bridge this gap, arguing that a true understanding of human creativity requires a synthesis of both perspectives. In Lubart's perspective, creativity is not a strictly linear progression through key-stages, nor is it a simple toggle between divergent and convergent acting. Instead, he framed the creative process as a highly dynamic and cyclical flow, in which divergent and convergent thinking can be considered sub phases that occur within every single macro-stage defined by the Wallas model.

This integrative approach seemed to be particularly fitting for setting the ground to the 'Suerte' experiment. In fact, when designers and artists interact with visual generative AI tools, they are no longer navigating a human-paced and predefined path. They are instead engaging in rapid, high-frequency interaction cycles with the machine, in which artist's prompts and the AI's outputs as intertwined phases of divergent generation and convergent selection, occurring at every step of the creative task.

2. The experiment

Arranged in a workshop formula and inclusive of technical set-up and training, this experience was shared with post-graduate students at the Verona Arts Academy. The experiment's name, 'Suerte' (the Spanish word for luck and fortune), was chosen to address both the functional randomness at the base of all deep generative neural networks and the chosen theme of Tarot design.

2.1. Technological environment

The technological context in which this investigation takes place should be properly addressed by scrutinizing its multiple different levels.

While it is not possible to address each of them with the necessary depth here, it is important to point out that most of the technology employed in this experiment is linked to an open-source cosmos comprising Stable Diffusion models (Rombach et al., 2022) and their myriad software integrations, laptop-sized hardware, free-to-use user interfaces, and a broad, unarticulated corpus of online knowledge.

It is also relevant to underline that, as in this case, relying on open-source digital ecosystems comes with a wide range of issues, ranging from annoying installation procedures to steep learning curves, and from fruitless troubleshooting on Reddit chats to inspiring and ground-breaking innovations made available on Hugging Face and other repositories.

Most importantly, open-source technology has the merit of being inherently more accessible when we attempt to understand what is 'running under the hood': the specific algorithms and computing principles might remain obscure to most artists and designers, but such complex yet fragmented knowledge remains stored and accessible in academic journals, conference proceedings, and through a vast array of projects and platforms spread across the web.

The 'Suerte' experiment was primarily based on Invoke AI Community Edition, which runs Stable Diffusion-derived models in combination with several auxiliary components such as ControlNet (Zhang et al., 2023) and IP-Adapters (Ye et al., 2023), up-scaler modules, as well as other editing features like drawing, masking, and filtering. Occasionally, other tools such as Adobe Photoshop and Midjourney were employed as well.

It is worth noting that other software alternatives were taken into consideration. However, after several comparisons, Invoke AI stood out for being relatively easy to set up and manage. Most importantly, it demonstrated the capability to combine ease of use, versatility, good levels of accuracy, and overall effectiveness in the management of image generation and manipulation processes. We hypothesized that this capacity for high interaction to and from the AI system, combined with the faculty of recalling and branching any specific step in the production, would likely provide a wider variety of factors to be observed.

Despite the general flaws derived from the intricacy of working with open-source technology, Invoke AI Community Edition has proven to be successfully handled and appreciated both by novices and by those already experienced in image editing.

2.2. Creative task

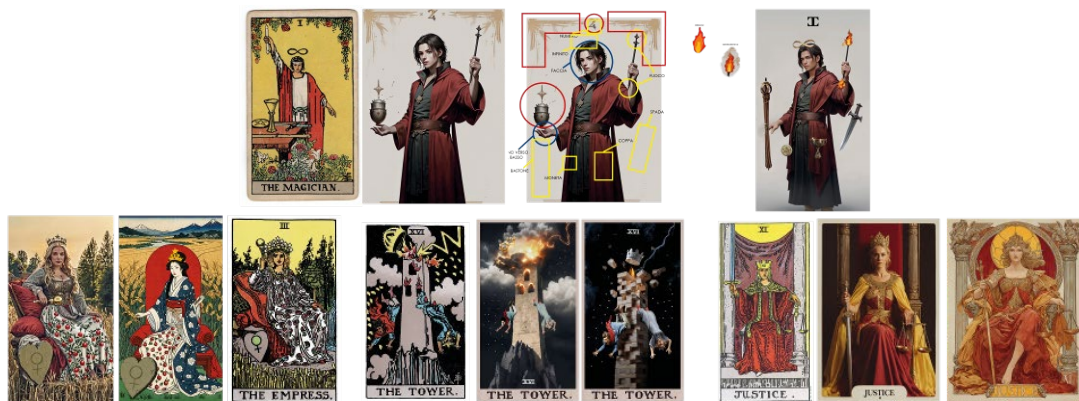
Students have been challenged on multiple levels: setting aside the pedagogical ones, the ‘Suerte’ design brief was set up to combine both constraining factors and high creative freedom.

The redesign of Tarot cards seemed to be a good opportunity to combine and stress such contrasting elements, which were necessary to observe the intense interplay between artists and the AI tools and workflows. In all Tarot decks, and with particular emphasis on the Major Arcana, each card pictures a complex system of signs (verbal, visual, and numeric) to which specific meanings are linked. In the most renowned and popular Tarots, like the Rider-Waite (1909) deck used for this experiment, the bonds between the image and the surrounding symbology can be deep and complex to the point that almost every visual element, ranging from subtle details to overall image composition, becomes relevant for proper and effective fortune-telling.

After randomly picking a card from the deck, each student had the task of recreating all the visual features and other key information (words, symbols, as well as decorative elements) with maximum adherence to the original source. The redesign had to be done at least twice per card: all the assigned manipulations had to preserve the original arrangement as much as possible, but one set had to specifically aim for a photorealistic look, while the others, being much more open to experimentation, could recall any other stylistic approaches. On average, each student committed to the task for more than 8 hours.

Both processes were expected to be quite challenging to fulfil, since the preservation of the card’s visual richness could not be achieved exclusively via textual prompting; it would have required working alongside in-painting and out-painting techniques. The difficulty in obtaining multiple and specific visual elements required an intense exchange between human participants and the AI dimension, thus creating a sufficiently reliable information base to be later scrutinized.

Figure 1. Examples from tarot redesign: process and outcomes.



2.3. Data collection

The design experience was closely overseen, student by student. Despite being highly interesting to report, the gathered impressions of the ongoing activity had to be explored in a more consistent form.

One method was the request, for each project, to produce a mostly visual representation of the key phase’s students went through, such as preliminary exploration, iteration over certain images, as well as moments of stillness and inspiration. A specific template or guidelines for compiling such a document was not provided, as we believe that the uncharted territory of human and AI interaction should also be told via the spontaneous voices of those who are directly experiencing such a vast shift in image creation practices.

In parallel to this qualitative approach, the students’ experience was measured via a questionnaire designed to plot multiple relevant aspects, such as a technological inquiry, a reflection on the creative process, and specific feedback on the proposed exercise. The survey also helped in shedding some light on the intricate panorama of the available AI tools, services, and user bases.

Most of the questions, however, were designed to ponder the overall relationship between human agency and technological behaviour, before and throughout the experiment. Particularly, participants were asked to assess the divergent and convergent phases they encountered in the process, as well as to provide feedback on the four stages of the creative process defined by Goleman-Wallas, thus helping to identify the most prominent factors in AI-based workflows.

3. Analysis

Students were asked to reflect on the visual generative AI creative processes they had previously experienced, as well as the one established in the ‘Suerte’ activity. The responses underline a clear awareness of creative paths where individual agency and machine randomness coexist. Even when specifically reflecting on the advanced technological framework (Invoke AI) introduced for the experiment, although in lesser proportions, such a compound response is still very evident.

Similar traits also emerged when scrutinizing the correlations between two distinct and opposite parameters, where nearly half of the participants who felt significant distress in handling the project also recalled an elevated capability in controlling and steering the process toward the desired directions.

This apparent contradiction is also confirmed when extending the comparison to the negative and positive spins (the feelings of instability and surprise, respectively) emerging within an experience of this kind: overall, the data suggests that the inner unpredictability of visual generative AI workflows seems to equally provide both creative benefits and obstacles.

We then investigated the presence of Guilford’s (1950, 1967) convergent and divergent thinking modes, discovering that the two phenomena share the field with nearly mathematical precision, both in this experience as well as during previous ones. Overall, working with visual generative AI clearly implies the recognition, and therefore the acceptance, of both explorative and focusing dynamics.

The analysis of Goleman-Wallas creativity patterns (Goleman et al., 1992; Wallas, 1926) also provided quite clear answers. Overall, the framework was confirmed: participants recalled having been involved at least ten times in each of the four steps outlined by the model. But here the distribution does not appear to be uniformly divided, with fewer “preparation” and “verification” moments in favour of a larger presence of the “illumination” and “incubation” stages.

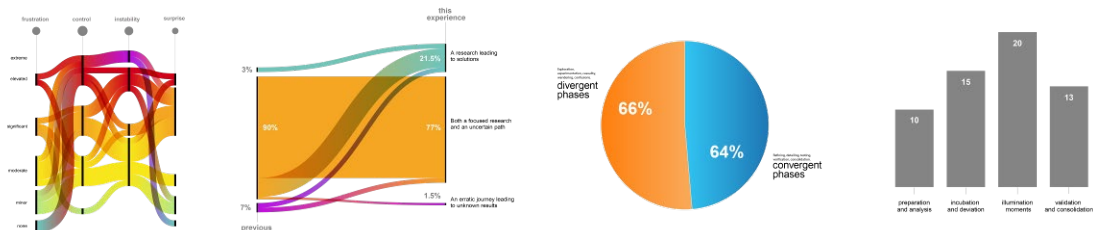
Such unevenness can be attributed to several factors, most of which cannot be entirely separated from the specificities given by the experiment (theme, tasks, technologies, time). Nevertheless, it is possible to argue that, similarly to other art and design workflows, AI-driven ones surely begin with clear goals/visions that culminate in their finalization/polishing. However, here the larger portion of the process appears to be less influenced by straightforward and rational approaches, and more dependent on welcoming and/or refusing what is produced by AI systems.

It is interesting to note that whereas the incubation phase in human creativity occurs mainly on an unconscious level, such interiorization cannot be directly applied to AI-generative-related practices in the exact same macro-level terms formulated by Goleman et al. (1992). Instead, aligning with Lubart’s (2001) integrative model, a similar roaming-free activity of the imaginative mind can be traced not in the user’s reasoning itself, but rather in their approach to using the tool, allowing the generation mechanism to operate toward more explorative and less constraining behaviours.

While it might be a stretch to say that the latent space acts as a sort of collective visual unconscious from which to draw striking images, we can observe that, in comparison with other technology-aided scenarios, these new tools are providing artists with a much wider degree of freedom to deviate from original ideas to discover new ones.

Lastly, participants were asked to reflect on this experience both in technical and methodological learning terms. The average responses were positive (scoring about 7 out of 10 each); considering the educational environment in which the experiment took place, such results suggest that even in a field entirely dominated by sheer technological advancements, exploring human-machine relationships to stimulate self-awareness and critical thinking is considered by novices to be equally relevant to acquiring new technical prowess.

Figure 2. From left to right: sentiment distribution in AI visual generative AI practices; overall impressions AI workflows before and after the experiment; divergent and convergent thinking experienced during the experiment; recorded creative stages during the ‘Suerte’ workshop.



4. Conclusions

The limited number of participants does not suffice to reach solid conclusions about such a branched and mutable scenario, but it still gives multiple opportunities to reflect on the interactions between the human mind and generative tools.

The Goleman-Wallas creativity model (Goleman et al., 1992; Wallas, 1926) surely becomes relevant in framing generative AI-aided processes as practices that remain fundamentally based on human agency and informed decision-making. However, differently from even the most technologically advanced scenarios (e.g., visual effects, CG animation), there is no gradual progression between its four stages. Such disconnection is probably related to the extreme speed and versatility offered by AI generative systems, which, in combination with a high degree of unpredictability, tend to favor straightforward, cyclical approaches: overall, opting for an extensive horizontal generation of visual solutions to pick from seems to be more effective than the in-depth pursuing of specific ones.

On the other hand, the alternation of diverging and converging thinking (Guilford, 1950), whether applied for initial general exploration or focused on detailing, seems to be omnipresent in all the processes we observed. It is tempting to argue that the converging phase stands primarily on the human side, whereas the AI is responsible for diverging and creating variety; but at a closer look, it is the human agency that deliberately steers the process in both directions, either by leveraging the high level of machine randomness or restricting it (via controllers, low-denoising values, and image manipulation).

Similarly, to the previous models, the Double Diamond's paradigm (British Design Council, 2005) deeply resonated in the experiment. However, to further validate its effectiveness in this field, we need to consider how drastically AI technologies can compress into minutes what in the past would have required many working hours to accomplish. Interestingly, such extreme time-shortening does not actually speed up the overall process itself; on the contrary, it seems to multiply the cycles between problems to be identified and counter-solutions to be found. In other words, in the 'Suerte' experiment, we clearly witnessed long chains of micro-diamond cycles which intensify the overall picture-making process, possibly leading toward better visual outcomes.

Having realized that such natural dynamics have a distinct role also in generative AI-related practices, further questions arise: when do these phenomena precisely take place? Is there any recurring macro-correlation useful to spot broader phases in the process? And if so, can they be properly modelled and shared as methodological blueprints?

AI technologies will keep changing in vaguely foreseeable ways, and so will the disciplines impacted by them. In the face of a complex equation full of variables, the capability to intertwine problem-solving and exploratory thinking (human creativity) will remain, in the perspective of truly enriching art and design fields, the key constant to look at.

References

- British Design Council. (2005). *The Double Diamond: A universally accepted depiction of the design process*. Design Council. <https://www.designcouncil.org.uk>
- Goleman, D., Kaufman, P., & Ray, M. (1992). *The creative spirit*. New York Dutton.
- Guilford, J. P. (1950). Creativity. *American Psychologist*, 5(9), 444–454. <https://doi.org/10.1037/h0063487>
- Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
- Lubart, T. I. (2001). Models of the creative process: Past, present and future. *Creativity Research Journal*, 13(3-4), 295-308. https://doi.org/10.1207/S15326934CRJ1334_07
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10684–10695). <https://doi.org/10.1109/CVPR52688.2022.01042>
- van Boeijen, A., Daalhuizen, J., Zijlstra, J., & van der Schoor, R. (2020). *Delft Design Guide: Perspectives, Models, Approaches, Methods* (Revised ed.). Amsterdam: BIS Publishers.
- Wallas, G. (1926). *The art of thought*. Harcourt, Brace and Company.
- Ye, H., Zhang, J., Liu, S., Han, X., & Yang, W. (2023). *IP-Adapter: Text compatible image prompt adapter for text-to-image diffusion models*. arXiv. <https://doi.org/10.48550/arXiv.2308.06721>
- Zhang, L., Rao, A., & Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3836–3847). <https://doi.org/10.1109/ICCV51070.2023.00355>