

Creative Text-to-Image Generation: Suggestions for a Benchmark

Irene Russo

Istituto di Linguistica Computazionale “Antonio Zampolli” (ILC-CNR), Pisa

irene.russo@ilc.cnr.it

Abstract

Language models for text-to-image generation can output good quality images when referential aspects of pictures are evaluated. The generation of creative images is not under scrutiny at the moment, but it poses interesting challenges: should we expect more creative images using more creative prompts? What is the relationship between prompts and images in the global process of human evaluation?

In this paper, we want to highlight several criteria that should be taken into account for building a creative text-to-image generation benchmark, collecting insights from multiple disciplines (e.g., linguistics, cognitive psychology, philosophy, psychology of art).

1 Introduction

Creativity is generally defined as the ability to produce new work that departs from existing practices and is appropriate, e.g., “normally fitted or adapted to the resolution of problems or difficulties existing within defined constraints” (Carter, 2004). As a peculiar human feature, creativity has been investigated by multiple disciplines (psychology, aesthetics, linguistics) to find recurring patterns and regularities in the motivated and intentional breaking or bending of rules that every creative act implies.

Creativity varies in time and space as every culturally determined concept. The evaluation of creativity during the Renaissance differed from today’s practices because cultural expectations have changed. In Eastern cultures, the focus is on the creative act *per se* instead of on the final result (Lubart, 1990).

The analysis of creativity is also dependent on the media: linguistic and visual creativity are both the result of complex psychological processes in the creator’s mind but give rise to radically different perceptual experiences for the receiver. Indeed,

linguistic and visual outcomes require different interpretation processes.

To make the understanding of creativity more complicated, the artificial generation of creative instances such as texts, music, images – under the name of computational creativity – poses further challenges to creativity’s definition. Sometimes the artist cooperates with the automatic system (as in generative art), while in other cases the result is independent of human agency, such as in text-to-image generation systems.

Today, computational models for text-to-image generation based on unsupervised deep learning methods can output realistic images, translating human written textual descriptions of variable length into images. Text-to-image (T2I, henceforth) generation model aims to generate photorealistic images semantically consistent with the text descriptions.

Starting with the generation of images from single labels or keywords, these models can handle more complex linguistic descriptions. Recent works mainly try to understand how much these images are referentially coherent and complete (Saharia et al., 2022). The evaluation of the referential aptness is more straightforward than the evaluation of creativity, which depends more on subjective variability in judgments and needs well-posed questions to be adequately isolated from other co-occurring variables that influence the aesthetic experience.

The generation of creative images is not under scrutiny at the moment, but it raises interesting questions: should we expect more creative images when using more creative prompts? What is the relationship between captions and images in the holistic process of evaluation? Can we have creativity without agency and authorial responsibility?

In this paper, we want to highlight several criteria that should be taken into account for building a text-to-image generation benchmark that addresses

those questions, collecting and discussing insights from multiple disciplines (e.g., linguistics, cognitive psychology, philosophy, psychology of art).

The remainder of this article is structured as follows: in Section 2, we describe relevant works from generative art and psychology of art. Section 3 focuses on the definition of creativity in language. We discuss how it is realized at different levels, together with examples that can be included in the benchmark. In Section 4, visual creativity is presented under the lens of the findings from psychology of art. Section 5 introduces the T2I available systems usable at the moment for generating images using textual prompts. Finally, in Section 6, we delineate the key features of human evaluation of creativity for automatically generated images that would help to answer important research questions before concluding in Section 7.

2 Related Work

The definition of a benchmark for creative T2I generation is a practical effort deeply influenced by theoretical questions previously addressed by other disciplines. In this section, we briefly report their relevant findings.

2.1 Generative Art

Generative art is a way to create art that requires a system “set into motion with some degree of autonomy contributing to or resulting in a completed work of art” (Galanter, 2003). It uses agents and is based on unpredictability, a key feature of creativity (Boden and Edmonds, 2009).

Galanter (2003) theorizes the system as self-contained enough to operate autonomously, so the artist’s role is to limit this autonomy. For example, an artist could intervene by acting on parameters, filtering the final outputs, or interactively modifying the system through feedback. However, when these systems are black boxes with opaque internal operations, such as deep learning models, it is difficult for the artist to act on them (Dorin et al., 2012). The limitations of deep learning models have been investigated: autonomous deep learning systems created for emulating arts (for example, through style transfer) are not able to reproduce the creative process, and output images with bias (Srinivasan and Uchino, 2021). Another important limitation is that generative art introduces randomness as part of its creative process, while in deep learning methodologies, randomness is a constitutive feature of the

design process. However, human intervention can intentionally change the degree of randomness in the training phase or the generation process.

Under these premises, we believe that T2I generated images could be perceivable with features attributable to generative art products, e.g., valuable as more or less creative, novel, or pleasant.

Another issue raised by generative art is the role of agency. Creativity is an agential disposition that produces new and valuable things thanks to the know-how of a human agent. Mechanical search and trial and error procedures are not creative. This view is endorsed by Paul and Stokes (2018) who argue that judgments about the creativity of an object implicitly refer to the generative processes involving agency. But computer-based generative art defies or at least causes a rethinking of the notion of agency (Wheeler, 2018).

The output of T2I generation systems is potentially art if the possibility of human interaction is made transparent in this scenario. An appropriate benchmark would go in this direction, investigating the interactions between the types of linguistic prompts and the perceived effects of generated images.

2.2 Computational Aesthetics

Computational aesthetics is a field of study interested in the convergence and generability of aesthetic judgments (Hönig, 2005; Bo et al., 2018). It focuses on the automatic assessment of beauty in human creative products, starting with the datasets of human judgments used to train specialized algorithms. The goal is to develop automatic systems that replicate the evaluation performances of human experts.

It is relevant for evaluating T2I generation outputs because researchers found convergences in the subjects’ aesthetic experiences and discover dimensions that constitute regularities manageable with algorithms. Even if the evaluation of creativity is independent of the assessment of aesthetic properties, sometimes the distinction is blurred since aesthetics is an aspect through which creativity is manifested and can be evaluated. A computational approach can illuminate the interplay between an image’s perceived creativity and aesthetic value.

We aim to investigate if T2I systems can be globally compared in terms of the aesthetic appreciation of their outputs and how the creativity of the prompt affects the human evaluation of this aspect

for the automatically generated images.

2.3 Neuroaesthetics

Reflections about the value and the evaluation of cultural objects such as paintings and pictures have been for centuries the object of study of aesthetics (Carroll, 1999). Human-produced images such as pictures and paintings can be evaluated as creative when presented in a context that clarifies their nature as cultural objects. Throughout history, the predominant criteria defining beauty and pleasure evolved, also influenced by new scientific discoveries in other fields, such as psychology of art. Nowadays, aesthetic theories are deeply influenced by experimental results from psychology, and the emerging field of neuroaesthetics presents promising results about universal regularities in the perception of aesthetic features (Nadal and Chatterjee, 2019).

According to neuroaesthetics, the aesthetic experience is composed of bottom-up perceptual habits (e.g., the tendency to identify objects when viewing artworks) and top-down control mechanisms that involve high-level cognition processes that attribute meanings to images (Cupchik et al., 2009). In general terms, perceptual fluency is enhanced by the amount of information, symmetry, and figure-ground contrast that increases the subjective pleasantness of an image. On the other hand, the complexity of an image is composed of the number of elements, differences in elements, and patterns in their arrangement. From experimental evidence, we know that the relationship between image complexity and pleasantness ratings forms an inverted-U shape graph: people increasingly like art as it goes from very simple to more complex until a peak when pleasantness ratings begin to fall again (Berlyne et al., 1968).

Without denying the influence of social and historical contexts on the perception of such features as beauty, novelty, and creativity, general principles about the perception of complexities in the composition of abstract paintings made clear that too much complexity is negatively correlated with aesthetic appreciation.

These general principles are declined in different ways when the object is a picture, a figurative, or an abstract painting. Evaluating a picture requires comparing the concrete reality that it reproduces and the artist's interpretation of that reality. With paintings, it is sometimes irrelevant the reality por-

trayed since the artist could deliberately distort it. In those cases, the title of the work acts as a framework for the interpretation. The relationship between the artistic work and the title is essential for evaluating creativity because it sets the boundaries of the evaluation process (see Section 2.4).

At the moment, it is not possible to control high-level properties of the generated images, such as the complexity or the symmetry of the composition, but we are confident it will be possible to a certain degree in the near future. Playing with these parameters will add an interesting dimension to T2I generation models' benchmark.

2.4 Psychology of Art

A series of experiments in the psychology of art investigated title/artwork relationship in the viewer experience (Russell, 2003; Franklin et al., 1993). They agreed on the fact that the title provided by the artist supports the interpretation process, making the image partly dependent on the verbal context. The title is a guide to painting's meaning, affecting attention and interpretation, increasing coherence, and enhancing aesthetic experience.

The viewers use the title to determine the artist's intentions, and different kinds of titles guide the viewer in different ways. Descriptive titles summarise the painting or picture in a short and neutral declarative sentence (e.g., *Woman planting flowers*). In contrast, elaborative titles use abstract words or metaphors not anchored to the image (e.g., *The Satin Tuning Fork*), forcing a metaphorical interpretation. This distinction is easy to understand for representational artworks. In the case of abstract paintings that do not contain a recognizable object, the distinction is between labeling titles void of meaning (e.g., *Studio n.5*) and titles that guide the processing of visual content (e.g., *From Pale Hands to Weary Skies*).

Differences between the two types of paintings (representational vs. abstract) also reflect the interpretation process guided by titles. For representational artworks, elaborative titles increase aesthetic experience more than descriptive titles but not the understanding of them (Millis, 2001). Depending on the time allocated for processing the abstract painting, elaborative titles increased the understanding when the time slot was 60 sec, while in the 1-sec scenario, descriptive titles helped more. The 60-sec exposure did not affect the aesthetic experience (Leder et al., 2006).

To reproduce these results on automatically generated images, we plan to include descriptive and elaborative prompts in the benchmark for evaluating T2I generation systems.

3 Creativity in Language

Linguistic creativity is a multi-dimensional construct, a distinctive trait of human beings not necessarily limited to literary texts but also retrievable in daily conversations (Carter, 2004). Thanks to corpus linguistics, creativity in language uses is an investigable topic: corpora represent the average level against which to measure novelty. Since we know the regularities of language, creative linguistic usages seem something that can be measured and organized along a cline. They gradually outdistance themselves from the norms.

However, very creative usages could not be attested even if the corpus is reputed representative for the language investigated. Moreover, the same corpus could not contain with significance frequency widely used idiomatic expressions that would be wrongly recognized as creative. For this reason, corpus linguistic methodologies should be used with care in the study of linguistic creativity.

The creative exploitation of linguistic means is retrievable at various levels that differ by granularity: if creativity at the morphological level concerns single words or pairs of words, when the focus is on the metaphors, syntagmatic units, or whole sentences are the objects of the analysis. In principle, multiple instances of creative language can be found in a single example: a creative word resulting from blending can be inserted in a metaphorical pattern that the reader recognizes as a flash fiction story (e.g., a very short story, limited in length, see 3.3).

In the following sub-paragraphs, short descriptions of creativity at various levels are reported, with examples and theoretical stances from works on these topics. As an illustrative purpose, we include just metaphors and analogies because they are among the most frequent figures of speeches exploited creatively to illustrate how they can be realized at different linguistic levels.

For each level, we distinguish when possible between denotational and connotational creative exploitations. The first kind aims to introduce new words or collocations that identify a new referent in the world. We produce connotational creative exploitations when we want to communicate about

an abstract property, a state of mind, a moral disposition, etc. While with denotational creativity the generated image should keep some of its referential properties (e.g., objects mentioned in the prompt should be identifiable in the image) this is not necessary for connotational creativity that, on the contrary, should produce more abstract images.

3.1 Morphological Creativity

Morphological creativity has been widely investigated to understand when and how morphological rules are exploited (Dal and Namer, 2018; van Marle, 1985). The new coinages can be playful and involve irregular means of word formation (e.g., blends or the import of affixes from other languages). New words are created by the speaker/writer on the fly to cover some communicative needs and are understandable thanks to linguistic and extra-linguistic contexts.

Under the umbrella of morphological creativity – broadly designating the coinage of new words – we posit two distinct morphological processes: the creation of new words with productive morphological rules and the creation of new words with irregular means of word formation, such as blends or the import of affixes from other languages. Neologisms can be created exploiting these paths; the study of *hapax legomena* in corpora can produce a list of attested examples, some of them included in dictionaries at later stages. Cook and Stevenson (2010) created a dataset of recently coined blendings (words such as *staycation*, *Japanimation*) for English to perform experiments for the automatic identification of source words.

A T2I benchmark could contain a couple of them, but genuinely creative examples are the ones not included for sure in training sets, i.e., non-words or pseudowords used in psycholinguistic experiments as distractors or fillers (words such as *rooned*, *lilf*, *aurene*). As a consequence, we refer to psycholinguistic and computational experiments as a data source.

The ARC Nonword database (Rastle et al., 2002) contains more than 350,000 nonwords and pseudohomophones that are orthographically or phonotactically legal, organized on the basis of several psycholinguistic dimensions such as bigram frequencies and phonological neighbours. Several experiments showed that readers attribute semantic meanings to non-words in proper linguistic contexts (Humphries et al., 2007).

3.2 Syntactic Creativity

In the generativist literature on the topic, syntactic creativity is a deviation that needs to be explained. Multiple studies focus on syntactic creativity in child language, as a sign of imperfect acquisition of syntactic rules (Lieven et al., 2003). Examples that deviate – i.e. are not explainable by – generative models are residually investigated as marked or non-standard use, salient from a sociolinguistic point of view.

However, the intentional exploitation of syntactic rules is a common feature also in literary writing. In this case, the violation of rules is functional to some pragmatic effects on the reader/hearer that still need proper experimental investigations (Lecercle, 1990).

As examples of syntactic creativity, Hampe and Schönefeld (2007) propose verbs used with an argument structure much more typically associated with that of other verb classes, as in example 1:

1. *He supported them through the entrance door* (vs. *push through the door*).

The evaluation of creative examples of syntactic usages requires, in several cases, a referential interpretation. With these examples we can not test how much T2I systems can be creative but we can understand if they are good at interpreting and unpacking information from non-standard examples. As such, whether they should be included or not in the benchmark is questionable.

3.2.1 Collocational Creativity

The existence of collocational creativity is a debated issue, especially because if it is associated with rarity in a corpus, there are no decisive methodologies to find creative collocations or rate created ones (Dillon, 2006). Apart from being striking because rare, a creative collocation needs to be apt and tailored for a unique communicative moment.

In order to get conceptually unusual collocations, a good benchmark for the evaluation of creativity in T2I systems could take into account cognitive psychology experiments that use as stimuli adj+noun pairs that are more complex to process (Murphy, 1990). Also adv+verb and verb+complement creative collocations are interesting, but no datasets are available for them. Creative collocations can be metaphorical in nature (e.g., *a diamond-encrusted book*). Metaphors draw analogies between domains

not normally linked, treating something as something else by means of similarities, and are more often realized by a sentence (*This book is a gem*). T2I generation systems can produce images that are coherent with prompts containing metaphors when they refer to concrete aspects of the object. In this case, there is denotational creative exploitation. When the function of the metaphor is the evaluation of some abstract properties, the output of the T2I generation systems will signal the lack of understanding, as in Figure 1 and Figure 2:

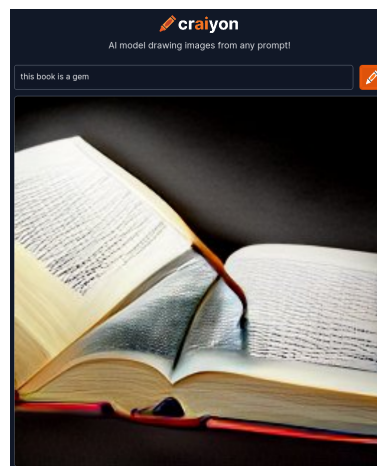


Figure 1: *This book is a gem* according to DALL-E mini

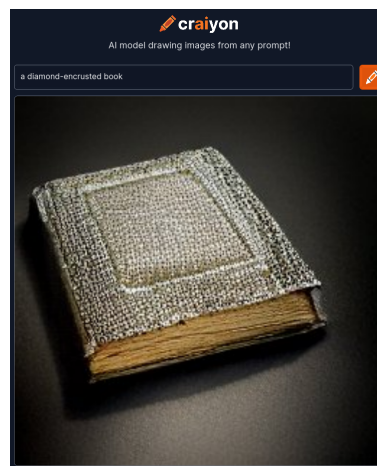


Figure 2: *A diamond-encrusted book* according to DALL-E mini

3.3 Textual Creativity

Textual creativity is the most complex type of linguistic creativity because it encompasses all the previous levels and is more subject to social and cultural expectations. It concerns the creation of fiction and alternative worlds. The literary language uses involve a complex patterning at the linguistic

level, often spanning several pages. At the same time, the interplay with the creativity of the plot – how it violates readers’ expectations – is an additional source of perceived creativity. The interplay between phenomena at different linguistic and structural levels and how they contribute to the perception of creativity in texts is an underinvestigated topic.

Since the length of the prompt in T2I systems is limited, one possibility is to use as an example of textual creativity flash fiction (between 250-750 words) (Masih, 2009) or ‘Twitter fiction’ (140 characters max) (Raguseo, 2010), concise stories that, because of their shortness, are not based on a complex plot.

4 Text-to-Image Generation Systems

In this section, we briefly introduce T2I systems, linking models based on them if available.

GLIDE (Guided Language to Image Diffusion for Generation and Editing) (Nichol et al., 2022) is a system based on a diffusion model and two guidance strategies, e.g., CLIP and classifier-free guidance. The diffusion model is trained with 3.5 billion parameters on the same dataset as DALL-E (Ramesh et al., 2022), using a text encoder to condition natural language descriptions. When compared with other systems, it is preferred by human evaluators for photorealism and caption similarity. GLIDE is able to produce artistic renderings of novel concepts (see Figure 3).



“a painting of a fox in the style of starry night”

Figure 3: Image generated by GLIDE

However, when the text prompt defies world knowledge (e.g., “a mouse hunting a lion”), GLIDE fails (see Figure 4). The system is not able to handle complex textual prompts; for this reason, the model has editing capabilities that allow users to

improve model samples until they match the complex prompt.



“a mouse hunting a lion”

Figure 4: Image generated by GLIDE

The authors released a smaller diffusion model on a filtered dataset based on CLIP without editing capabilities¹.

In January 2021, OpenAI introduced DALL-E, a neural network (12-billion parameter version of GPT-3) that creates images from text, using as training set text-image pairs (Ramesh et al., 2021). One year later, DALL-E 2 (Ramesh et al., 2022) was released. It generates more realistic and accurate images with greater resolution using diffusion. Unfortunately, none of those systems is open source, but a smaller model of DALL-E is available online through an online interface².

Imagen is a text-to-image diffusion model that does not use only image-text data for training but is based on large transformer generic language models (Saharia et al., 2022) that produce realistic images with a good image-text alignment. It consists of a text encoder that maps text to a sequence of embeddings and a set of conditional diffusion models that map the embeddings to images of increasing resolutions. A Pytorch implementation is freely available³.

For the evaluation, the authors introduce Draw-Bench⁴, a benchmark of 200 text prompts designed to probe different semantic properties of the models, such as compositionality, spatial relations, rare words, and more creative prompts that, according

¹<https://github.com/openai/glide-text2im>

²<https://www.craiyon.com>

³<https://github.com/lucidrains/imagen-pytorch>

⁴<https://docs.google.com/spreadsheets/d/1y7nAbmR4FREi6npBlu-Bo3GF\dwDOPYJc617rBOxIRHY/edit#gid=0>

to the authors, “push the limits of models’ ability to generate highly implausible scenes well beyond the scope of the training data.” However, these prompts do not explicitly contain instances of creative language.

Parti (Pathways Autoregressive Text-to-Image model) (Yu et al., 2022) is an autoregressive text-to-image generation model that outputs photorealistic image generation coherent with world knowledge. It is complementary to Imagen because it explores different families of generative models (autoregressive vs. diffusion). For Parti, text-to-image generation is a sequence-to-sequence modeling problem analogous to machine translation: it outputs sequences of image tokens instead of text tokens in another language. Parti uses a powerful image tokenizer, ViT-VQGAN, to encode images as sequences of discrete tokens, and takes advantage of its ability to reconstruct such image token sequences as high-quality, visually diverse images. A Pytorch implementation is freely available ⁵.

As part of this project, the authors released PartiPrompts (P2) ⁶, a rich set of over 1600 prompts in English that constitute a holistic benchmark. P2 can be used to measure model capabilities across various categories and challenging aspects. It contains 52 examples such as *A high resolution photo of a chicken working out in a gym* labeled as *Imagination* to test if T2I systems are able to reproduce a not realistic state of affairs.

5 How to Structure the Evaluation Process

The images generated by T2I systems, thanks to prompts included in the benchmark – both paintings and pictures, when it is possible to specify the type of output – should be evaluated by human subjects to answer research questions about the relationship between artificially generated outputs and perceived creativity.

The evaluation of artificially generated images presents both similarities and differences with respect to the evaluation of images – pictures, drawings, paintings – created by humans.

In the following paragraph, valuable criteria for collecting judgments are listed and discussed, reporting, if necessary, how they would impact the selection of prompts included in the benchmark.

⁵<https://github.com/lucidrains/parti-pytorch>

⁶<https://github.com/google-research/parti/blob/main/PartiPrompts.tsv>

5.1 Comparative collection of graded judgments on creativity

One of the aims of the evaluation process is to collect converging judgments on creativity. In this case, the search for a good inter-annotator agreement apparently contrasts with the subjective nature of aesthetic judgments that also include the perception of creativity. Nowadays, The gold standard in measuring creativity is the Consensual Assessment Technique (CAT, henceforth) (Amabile, 1982) which concerns the assessment of the creative performance on a real task such as writing a poem or creating a collage. CAT is a product-based subjective assessment technique built on a consensual definition of creativity as the quality of products or responses judged to be creative by appropriate observers.

This technique is based on the availability of experts that know the domain and act as judges that reach good inter-rater reliability, ranging from 0.70 to 0.9 (Long and Wang (2022)). There is mixed evidence about the convergence of non-expert raters: their agreement on the evaluation of visual products in art tasks is higher (and more correlated with experts’ judgments) than on the evaluation of written output (Kaufman et al., 2008).

CAT was originally designed to compare parallel creative products created in response to the same prompt. Inspired by the design of experiments based on CAT, the evaluation of creativity in T2I generation systems should split the outputs into pairs or small groups of items, rated comparatively by the same annotator with Likert-style evaluation (1-5 or 1-7).

5.2 Collection of creativity judgments that take into account the verbal prompts and their properties

Creativity concerns the meaningful breaking of rules. The recipient of the creative act is in charge of attributing meaning to the creative object, which often requires an interpretation process with a cognitive cost.

Images can not be evaluated out of context. For this reason, it is essential to structure the evaluation phase on linguistic description-image pairs. Therefore, two types of questions are proposed, one addressing the denotational level and another the connotational level of the image:

- How much does the following image represent the content of the associated linguistic

description?

- How much does the linguistic description inspire the following image?

These questions aim to force a graded evaluation that compare the results across linguistic prompts with different level of creativity and across different models. One of the working hypotheses testable concerns the idea that more creative connotational linguistic descriptions should generate more creative images.

It is important to include in the prompt connotational and denotational examples, creative and non-creative examples and, among the creative prompts, include examples located along a cline.

5.3 Collection of aesthetic judgments that correlate with creativity judgments

From experimental studies in psychology of art, we know that the perception of aesthetic properties partially influences the perception of creativity. Following [Niu and Sternberg \(2001\)](#), each generated image can be evaluated by asking for comparative judgments on different dimensions of an artistic product: creativity (the degree to which the image is creative), likeability (the degree to which the judge likes it), appropriateness (the degree to which the image is coherent with the textual prompt). There is a correlation between creativity and likeability for drawings and collages produced by humans. We expect that when abstract representations are generated, the influence of likeability on creative judgments would be more substantial.

5.4 Collection of judgments from annotators with different or no expertises

The optimal evaluation process should involve different types of annotators. Instead of involving just people with expertise in art, as proponents of CAT suggest, we plan to ask for judgments from people that are more or less aware of artificial generation in order to understand if technical knowledge influences the evaluation. Also, results from a Turing test scenario, where people do not know which image is artificially generated, could shed light on the limitations of creative T2I generation.

6 Conclusions and Future Work

Creativity is contextually and historically framed and depends on the medium. Nowadays, an unprecedented occasion for investigating the topic is

represented by T2I generation models that combine linguistic inputs with visual outputs. However, while the evaluation of the referential quality and coherence of the automatically generated images has been investigated, there are no papers extensively discussing the role and the evaluation of creativity in T2I generation systems.

Is the output of T2I generation systems perceived as creative by humans? Is creativity a property that could be computationally mimicked and empirically increased in models that generate artificial instances? Should we expect more creative images using more creative prompts? What is the relationship between prompts and images in the process of human evaluation?

These are several of the questions that could be addressed with a proper benchmark. In this paper, we critically revised multidisciplinary works that could help to design a good benchmark for the evaluation of creativity in T2I generation systems and highlight several criteria that could shape the evaluation process.

References

- T. M. Amabile. 1982. Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, 43:997–1013.
- Daniel E. Berlyne, John C. Ogilvie, and L C Parham. 1968. The dimensionality of visual complexity, interestingness, and pleasingness. *Canadian journal of psychology*, 22 5:376–87.
- Yihang Bo, Jinhui Yu, and Kang Zhang. 2018. Computational aesthetics and applications. *Visual Computing for Industry, Biomedicine and Art*, 1.
- Margaret A. Boden and Ernest A. Edmonds. 2009. What is generative art? *Digital Creativity*, 20:21 – 46.
- Noël Carroll. 1999. *Philosophy of Art : A Contemporary Introduction*. Routledge, New York.
- Ronald Carter. 2004. *Language and creativity : the art of common talk*. Routledge, New York.
- Paul Cook and Suzanne Stevenson. 2010. Automatically identifying the source words of lexical blends in english. *Computational Linguistics*, 36:129–149.
- Gerald C. Cupchik, Oshin Vartanian, Adrian P. Crawley, and David J. Mikulis. 2009. Viewing artworks: Contributions of cognitive control and perceptual facilitation to aesthetic experience. *Brain and Cognition*, 70:84–91.

- Georgette Dal and Fiammetta Namer. 2018. Playful nonce-formations in french: Creativity and productivity. In Sabine Arndt-Lappe, Angelika Braun, Claudine Moulin, and Esme Winter-Froemel, editors, *Expanding the Lexicon: At the crossroads of innovation, productivity, and ludicity*. De Gruyter.
- George L Dillon. 2006. Corpus, creativity, cliché: Where statistics meet aesthetics. *Journal of Literary Semantics*, 35(2):97–103.
- Alan Dorin, Jonathan McCabe, Jon McCormack, Gordon Monro, and Mitchell Whitelaw. 2012. A framework for understanding generative art. *Digital Creativity*, 23:239 – 259.
- Margery B. Franklin, Robert C. Becklen, and Charlotte Lackner Doyle. 1993. The influence of titles on how paintings are seen. *Leonardo*, 26:103 – 108.
- Philip Galanter. 2003. What is generative art? complexity theory as a context for art theory. In *In GA2003 – 6th Generative Art Conference*.
- Beate Hampe and Doris Schönefeld. 2007. Syntactic leaps or lexical variation? – more on “creative syntax”. In Stefan Th. Gries Anatol Stefanowitsch, editor, *Corpora in Cognitive Linguistics: Corpus-Based Approaches to Syntax and Lexis*, pages 127–157. Mouton de Gruyter.
- Florian Hönig. 2005. Defining computational aesthetic. In B. Gooch W. Purgathofer L. Neumann, M. Sbert, editor, *Computational Aesthetics in Graphics, Visualization and Imaging*.
- Colin J. Humphries, Jeffrey R. Binder, David A. Medler, and Einat Liebenthal. 2007. Time course of semantic processes during sentence comprehension: An fmri study. *NeuroImage*, 36:924–932.
- J. Kaufman, John Baer, Jason C. Cole, and Janel D. Sexton. 2008. A comparison of expert and nonexpert raters using the consensual assessment technique. *Creativity Research Journal*, 20:171 – 178.
- Jean-Jacques Lecercle. 1990. *The Violence of Language*. Routledge, New York.
- Helmut Leder, C. Carbon, and Ai-Leen Ripsas. 2006. Entitling art: Influence of title information on understanding and appreciation of paintings. *Acta psychologica*, 121 2:176–98.
- Elena Lieven, Heike Behrens, J. Speares, and Michael Tomasello. 2003. Early syntactic creativity: a usage-based approach. *Journal of child language*, 30 2:333–70.
- Haiying Long and Jue Wang. 2022. Dissecting reliability and validity evidence of subjective creativity assessment: A literature review. *Educational Psychology Review*.
- Todd Lubart. 1990. Creativity and cross-cultural variation. *International Journal of Psychology*, 25:39–59.
- Jaap van Marle. 1985. *On the paradigmatic dimension of morphological creativity*. Foris, Dordrecht.
- T. L. Masih. 2009. *Field guide to writing flash fiction: Tips from editors, teachers, and writers in the field*. Rose Metal Press, Brookline, MA.
- Keith K. Millis. 2001. Making meaning brings pleasure: the influence of titles on aesthetic experiences. *Emotion*, 1 3:320–9.
- Gregory L. Murphy. 1990. Noun phrase interpretation and conceptual combination. *Journal of Memory and Language*, 29:259–288.
- Marcos Nadal and Anjan Chatterjee. 2019. Neuroaesthetics and art’s diversity and universality. *Wiley interdisciplinary reviews. Cognitive science*, 10 3:e1487.
- Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. 2022. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. In *ICML*.
- Weihua Niu and Robert J. Sternberg. 2001. Cultural influences on artistic creativity and its evaluation. *International Journal of Psychology*, 36:225–241.
- Elliot Samuel Paul and Dustin Stokes. 2018. Attributing creativity. In Berys Gaut and Matthew Kieran, editors, *Creativity and Philosophy*. Routledge.
- Carla Raguseo. 2010. Twitter fiction: Social networking and microfiction in 140 characters. *TESL-EJ*, 16.
- Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical text-conditional image generation with clip latents. *ArXiv*, abs/2204.06125.
- Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-shot text-to-image generation. *ArXiv*, abs/2102.12092.
- Kathleen Rastle, Jonathan Harrington, and Max Coltheart. 2002. 358,534 nonwords: The arc nonword database. *Quarterly Journal of Experimental Psychology*, 55:1339 – 1362.
- Philip A. Russell. 2003. Effort after meaning and the hedonic value of paintings. *British journal of psychology*, 94 Pt 1:99–110.
- Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L. Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, Seyedeh Sara Mahdavi, Raphael Gontijo Lopes, Tim Salimans, Jonathan Ho, David Fleet, and Mohammad Norouzi. 2022. Photorealistic text-to-image diffusion models with deep language understanding. *ArXiv*, abs/2205.11487.

Ramya Srinivasan and Kanji Uchino. 2021. Biases in generative art: A causal look from the lens of art history. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*.

Michael Wheeler. 2018. Talking about more than heads. the embodied, embedded and extended creative mind. In Berys Gaut and Matthew Kieran, editors, *Creativity and Philosophy*. Routledge.

Jiahui Yu, Yuanzhong Xu, Jing Yu Koh, Thang Luong, Gunjan Baid, Zirui Wang, Vijay Vasudevan, Alexander Ku, Yinfei Yang, Burcu Karagol Ayan, Benton C. Hutchinson, Wei Han, Zarana Parekh, Xin Li, Han Zhang, Jason Baldridge, and Yonghui Wu. 2022. Scaling autoregressive models for content-rich text-to-image generation. *ArXiv*, abs/2206.10789.