

Erwin Feyersinger; Lukas Kohmann; Michael Pelzer

Fuzzy Ingenuity: Creative Potentials and Mechanics of Fuzziness in Processes of Image Creation with AI-Based Text-to-Image Generators

2023

<https://doi.org/10.25969/mediarep/22317>

Veröffentlichungsversion / published version

Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Feyersinger, Erwin; Kohmann, Lukas; Pelzer, Michael: Fuzzy Ingenuity: Creative Potentials and Mechanics of Fuzziness in Processes of Image Creation with AI-Based Text-to-Image Generators. In: *IMAGE. Zeitschrift für interdisziplinäre Bildwissenschaft*. Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, Jg. 19 (2023), Nr. 1, S. 135–149. DOI: <https://doi.org/10.25969/mediarep/22317>.

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under a Deposit License (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual, and limited right for using this document. This document is solely intended for your personal, non-commercial use. All copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute, or otherwise use the document in public.

By using this particular document, you accept the conditions of use stated above.

Erwin Feyersinger, Lukas Kohmann, and Michael Pelzer

Fuzzy Ingenuity: Creative Potentials and Mechanics of Fuzziness in Processes of Image Creation with AI-Based Text-to-Image Generators

Abstract: This explorative paper focuses on fuzziness of meaning and visual representation in connection with text prompts, image results, and the mapping between them by discussing the question: How does the fuzziness inherent in artificial intelligence-based text-to-image generators such as DALL·E 2, Midjourney, or Stable Diffusion influence creative processes of image production – and how can we grasp its mechanics from a theoretical perspective? In addressing these questions, we explore three connected interdisciplinary approaches: (1) Text-to-image generators give new relevance to Hegel’s notion of language as ‘the imagination which creates signs’. They reinforce how language itself inevitably acts as a meaning-transforming system and extend the formative dimension of language with a technology-driven facet. (2) From the perspective of speech act theory, we discuss this explorative interaction with an algorithm as performative utterances. (3) In further examining the pragmatic dimension of this interaction, we discuss the creative potential arising from the visual feedback loops it includes. Following this thought, we show that the fuzzy variety of images which DALL·E 2 presents in response to one and the same text prompt contributes to a highly accelerated form of externalized visual thinking.

Introduction

The newest generation of text-to-image generators not only challenges our traditional notions of design processes and conceptual flows in creating visual art, but also poses disruptive questions in regard to the theoretical intersection between language and visibility as well as to the nature of artistic intentionality: Like a wizard trying to find the right words for an unknown magic spell, prompt engineers permutate their wordings to generate specific results. The process

behind this extends the formative dimension of language (i.e., the way in which we use language to not only describe, but also make sense of the world and construct meaning) with a technology-driven facet. As a result, it highlights how language itself inevitably acts as a meaning-transforming system. Focusing on the mechanics of fuzziness and sharpening in relation to text prompts, image results, and the mapping between them, this paper presents a collection of related comments and ideas that explore the fuzziness inherent in AI-based text-to-image generators such as DALL·E 2, Midjourney, or Stable Diffusion.

When we discuss ‘fuzziness’ as a technical term in the context of this paper, we draw on a long theoretical tradition addressing the indirect and sometimes diffused relation between (conceptual) ideas and their perceivable realization in concrete objects that essentially traces back to Plato’s remarks on the theory of forms (cf., for instance, PATTERSON 1985). At the same time, we reference interdisciplinary research in the field of “Fuzzy Logic” as outlined by Lotfi A. Zadeh (1965)¹ and appropriate its core concepts to AI-based image generation and art in a wider sense. In doing so, we extend upon stimulating thoughts put forth by Hanns-Werner Heister (2021) in regard to the application of Fuzzy Logic from the perspective of the science of music. We argue that the concept of fuzziness – or more precisely: artful interactions between complementary mechanics of fuzziness and sharpening (cf. HEISTER 2021: x) – are an essential axis of analysis that can help us in better understanding some of the characteristics and creative potentials inherent in processes of image creation using tools such as DALL·E 2, Midjourney, or Stable Diffusion.

The aspects of fuzziness thus discerned can relate to

1. the mapping between captions and visuals in the training process of image generators (Input),
2. the hidden algorithmic structures mapping input to output (Machine Learning),
3. the mapping between text prompts and visuals in the generation process (Output).

In particular, we want to examine how mechanics of fuzziness pertaining to the latter area of ‘output’ influence creative processes of image production and how we can grasp them with various theoretical approaches. In pursuing these goals, we want to explore three connected interdisciplinary perspectives: Lukas Kohmann starts by analyzing the interaction between humans and image generators against the background of Wittgenstein’s and Hegel’s *theories of language* and discusses to what extent the process of turning text prompts into images carries *imaginative qualities*. Connecting questions of artistic imagination to the

1 As an introduction and overview to the concept and accompanying research in a wider sense, cf., e.g., NGUYEN/WALKER 2005.

perspective of speech act theory, Erwin Feyersinger further examines the explorative interactions inherent in prompt engineering as *performative utterances* – and discusses whether they can be regarded as conversational processes. In further outlining the pragmatic dimension of these interactions, Michael Pelzer eventually investigates the creative potential arising from the *visual feedback loops* they include – and explores how AI-based visualization tools transform existing concepts of artistic ingenuity.

The Process of Image Generation through the Lens of Hegel's Concept of "the Imagination which Creates Signs"

In the interaction with text-to-image generators, users enter text prompts describing the idea of an image that is to be generated by the system. Thus, we provide the generator with letters forming words, i.e., signs containing meaning, from which the generator is meant to retrieve said meaning. On an abstract or perhaps only superficial level, the ordinary dialogue between two human interlocutors can be described in a similar way. This sort of interaction has always been problematic and is aptly formulated by Ludwig Wittgenstein as follows: "But if you say: 'How am I to know what he means, when I see nothing but the signs he gives?' then I say: 'How is he to know what he means, when *he* has nothing but the signs either?'" (WITTGENSTEIN 1998: No. 504, original emphasis).¹²

Image generators such as Midjourney confront us with 'fantastic' image creations within a brief computation time, based only on a given text prompt. Apparently, a non-arbitrary process of practical sign comprehension takes place on the generator's side. The impression of a human-like understanding of signs arises, namely the recognition of relations to the objects referred to by the prompt as well as the relations between the signs themselves, that is, the illusion of a metaphysical reference as "the imagination which creates signs" (HEGEL 2007: §457).¹³ Through the use of signs, reality can be interpreted and understood. For philosopher Georg W.F. Hegel, language merely fulfills a denotation function hereby. The world of language thus forms a second, higher existence, in which the sensations, views, and ideas of the mind are contained (cf. HEGEL 2007: §459). The meaning of signs, however, is not definitive but has an intermediary function toward an immediate understanding of signs.

2 Original: "Wenn man aber sagt: 'Wie soll ich wissen, was er meint, ich sehe ja nur seine Zeichen', so sage ich: 'Wie soll er wissen, was er meint, er hat ja auch nur seine Zeichen'" (WITTGENSTEIN 1998: No. 504, original emphasis).

3 Original: "Zeichenmachende Phantasie" (HEGEL 1970b: §459).

Now that it has been forgotten what names properly are, viz. *externalities which of themselves have no sense*, and only get signification as *signs*, and now that, instead of names proper, people ask for terms expressing a sort of definition, which is frequently changed capriciously and fortuitously, the denomination (HEGEL 2007: §459, original emphases).^[4]

Although text-to-image generators seem to be able to interpret text prompts adequately (most of the time) and to generate corresponding images, this process cannot be understood as “the imagination which creates signs” (HEGEL 2007: §457) or as *imagination* in general, as Hegel theorized it. Any character is a sign that is intrinsically meaningless to a computer and only gains meaning by an allocation given to it by a user who subsequently interprets it in a particular way. Computers generally process any sign. In a repeated playful interaction with a text-to-image generator, it can be observed that the perceived natural linguistic quality of the interaction can increasingly be deconstructed. Text-to-image generators ultimately utilize a vocabulary that, contrary to actual linguistic conventions, does not address the “*ideational realm*” (HEGEL 2007: §459, original emphasis).^[5] Thus, no actual reference to the material world is established but merely read into it *a posteriori* by a human recipient.

It may be concluded that, viewed through the lens of Hegel’s theory, such technology does not yet have a proper understanding of reality. In logical consequence, it should be added that the use of the term ‘understanding’ is within a category that imposes a false demand on the image-generator. While it can be used to generate images based on textual descriptions, it is not capable of ‘understanding’ reality in all its depth and complexity. According to Hegel’s theory of intelligence, the mindlessness of sign-processing reason lies in the “*indifference of content to form*” since the mind is regarded as a “‘lot’ of *forces*” (HEGEL 2007: §445, original emphases).^[6] Form and substance are inseparable. Content is the enveloping of form and form is nothing other than the enveloping of content (cf. HEGEL 1970a: §133). The formalism in terms of semantics can only be conceived from the perspective of the programmer who determines the training dataset. Therefore, what is implemented is always derivative (cf. BLANKE 2007: 293). Sign processing replicates the dichotomy of form and content without being able to reflect or change this. Tobias Blanke concludes in this regard:

4 Original: “Namen als solche sind, nämlich für sich *sinnlose Äußerlichkeiten*, die erst als *Zeichen* eine Bedeutung haben, seit man statt eigentlicher Namen den Ausdruck einer Art von Definition fordert und dieselbe sogar häufig auch wieder nach Willkür und Zufall formiert, ändert sich die Benennung” (HEGEL 1970b: §459, original emphases).

5 Original: “*Reiche des Vorstellens*” (HEGEL 1970b: §459, original emphasis).

6 Original: “Die *Kraft* ist zwar die *Unendlichkeit* der Form, des Inneren und Äußerer, aber ihre wesentliche *Endlichkeit* enthält die *Gleichgültigkeit des Inhalts* gegen die Form. Hierin liegt das Vernunftlose, was durch diese Reflexionsform und die Betrachtung des Geistes als einer Menge von *Kräften* in denselben sowie auch in die Natur gebracht wird” (HEGEL 1970b: §445, original emphases).

First, there is a lack of understanding of the collectivity of intelligence. According to Hegel, intelligence is not tested by solving a series of combinatorial tasks on a piece of paper, but by exposing oneself to the knowledge of the public. Second, a rationality built on the formal substitutability of signs is overwhelmed in dealing with the inconsistent relation between thinking and observation. Machines have no imagination (BLANKE 2007: 292, our translation).¹⁷

A text-to-image generator relies on the data that was available at the time of training and is limited by that same data. This adds to the already existing linguistic fuzziness. Not only are the words ambiguous in their meaning, but both the user and the image generator address completely different systems with the signs used. ‘Meaning’ is open to a wider range of interpretations, allowing for a multitude of ‘correct’ image outputs. Novel or unknown inputs can therefore lead to random and, to a human observer, completely unrelated outputs. The generator has difficulties in dealing with such inputs in a meaningful way. It generates images based on trained statistical patterns and concatenations of words and images. According to Hegel, it may be argued that such a system lacks the ability to add new or unexpected aspects of reality, that is, truly imaginative creative aspects. Everything that is generated is always merely derivative and lacks a true reference to the world. Being capable of cognition not only means having knowledge, but also intuiting, conceiving, remembering, imaging, and so on (cf. HEGEL 2007: §445).

However, we should ask ourselves whether this claim is in line with our modern understanding of art, especially one that includes the experience of the recipient. According to Hegel, tools such as DALL-E 2 would not be capable of grasping the text prompt linguistically in the way the human author has written it. Not only because of the – let us borrow the term from Philip J. Tichenor, George A. Donohue, and Clarice N. Olien (1970: 16off.) – “knowledge gap”, but because the system processes the characters entirely differently than a human would. On this account, as we could conclude with Blanke, DALL-E 2 is not capable of imagination in terms of a sign-making ability – even if we shift our attention away from natural language observation and instead try to conceptualize the generator’s output as a symbol. Hegel says that intelligence is a form of imagination that expresses itself as a symbolizing, allegorizing, or poetic imagination, but whose creations still lack material existence (cf. SIMON 1996: 261). Intelligence thus means also being able to refer to objects that are not themselves part of the physical world, but only constitute meaning for themselves through reference

7 Original: “Es fehlt erstens am Bewusstsein der Kollektivität von Intelligenz. Nach Hegel wird die Intelligenz nicht getestet, indem man auf einem Stück Papier eine Reihe von kombinatorischen Aufgaben löst, sondern indem man sich dem Wissen der Allgemeinheit aussetzt. Zweitens ist eine Vernunft, die auf der formalen Substituierbarkeit von Zeichen aufgebaut ist, überfordert, geht es um den inkonsistenten Zusammenhang von Denken und Anschauung. Maschinen haben keine Phantasie” (BLANKE 2007: 292).

to denomination categories for different physical things. However, the question that inevitably arises is how the generator's stochastic image-making relates to the *artistic* imagination: According to Hegel, the artistic "imagination [Phantasie]" is to be differentiated from the "purely passive imagination [Einbildungskraft]" because "imagination [Phantasie]" itself is something that actively creates (HEGEL 1975: 281-288).^[8] Nevertheless, this artistic imagination is not entirely independent of the ability to comprehend the world and shape our understanding of it – and thus remains connected to both language and the sign-making imagination.

Prompt Engineering as a Monologic Series of Speech Acts

If we understand text-to-image generators as incapable of genuine artistic imagination, how can we then theorize the strong positive and negative reactions their widespread public introduction in 2022 has created? How can we understand the fascination for what is perceived as new aesthetic qualities and a new utility of automated image generation? How can we contextualize the outcry by artists who experience these new tools as a threat to their livelihood and their skills? Here, we propose to shift the perspective from semantics and Hegel's language philosophy to the perspective of pragmatics. At least in the early experimental phase of text-to-image generators of 2022 and 2023, we can understand the interaction with the machine as a monologic succession of speech acts that, in a constant feedback loop, are refined based on the output by the machine and how much it conforms with the user's expectations. For interfaces directly based on text inputs such as DALL·E 2, but to some degree also for parametrized apps such as Lensa that replace text inputs with predefined input options, speech act theory is a promising approach for conceptualizing text-to-image generators because it allows us to consider the *performativity* of the interaction as well as the pragmatic aspects of its similarity and dissimilarity to natural language use.

Apart from the explicit command "/imagine" used with the text-to-image generator Midjourney, in most cases the input is an indirect speech act that usually consists of a sequence of phrases describing content, style, medium, and other aspects of the intended images. In an interplay of fuzziness and sharpening, the user, after assessing the results, tweaks either the verbal statement, generates more variations (if the text-to-image generator offers this option), or modifies parts or aspects of the image by inpainting or outpainting. Users thus learn over time how to phrase the input to achieve results closer to their intentions. Returning to Wittgenstein's quote above, the inner workings of the algorithms

8 Original: "Die Phantasie ist schaffend" (HEGEL 1970C: 263).

and how they ‘understand’ the text input remains a black box phenomenon to the individual user which nonetheless ‘works’ – and often leads to highly convincing results. From a practical point of view, it might even seem irrelevant how the model arrives at an image. However, as the results are often unexpected and still seem to exhibit a fuzzy but close enough ‘understanding’ of not only natural language commands but even the user’s intention, it does add to the fascinating qualities of the engagement with a text-to-image generator. This interaction can then be perceived as a *bidirectional* or *dialogic* form of communication, as evidenced by how users describe their experience, for example, artist Bokar N’Diaye in a YouTube explainer video: “You realize that you can refine the way you talk to the machine. It becomes a kind of a dialog” (as quoted in vox 2022: n.pag.). Both the fuzziness of the results in relation to the input and the process of sharpening further inputs is not only productive in the creation process but also highly mesmerizing.

From the perspective of speech act theory, we can describe this explorative intentional interaction with an algorithm as “performative utterances” (AUSTIN 1962: 6). Prompt engineers permute their wordings to make the machine generate a specific result, a process that can border on a mysterious experience, as the following statement by artist Mario Klingemann in the same explainer video demonstrates: “What I love about prompting: for me [...] it has something like magic where you have to know the right words for the spell” (as quoted in vox 2022: n.pag.). This fascination surrounding a performative exploration of the interface is also reflected in how users share unusual discoveries. For example, as Giannis Daras and Alexandros G. Dimakis (2022) point out in a preprint article, some seemingly made-up words almost consistently result in images of the same entities, such as “*Apoploe vesrreaitais*” repeatedly generating images of birds. However, this could just be caused by a proximity of these expressions to existing words, as, for example, two bird species are named after Mount Apo. Similarly, misspelled words often still lead to appropriate results – a further aspect of fuzziness. Current text-to-image generators are comparable to common conversational user interfaces (CUIs) such as Siri and Alexa that allow request-response interactions. However, unlike these CUIs and especially current AI-based chatbots such as ChatGPT, which appear to be able to respond to a variety of speech acts, text-to-image generators are designed to perform only one specific task again and again, i.e., to generate an image. Even if users perceive and describe the iterative interaction with the machine as a dialog, it only consists of a series of monologic illocutionary acts, which can be classified as *directives*, i.e., “attempts (of varying degrees, and hence more precisely, they are determinates of the determinable which includes attempting) by the speaker to get the hearer to do something” (SEARLE 1975: 11).

To the current AI-based image generators, these directives remain single unconnected commands and they (unlike ChatGPT) do not take earlier requests into consideration when a new input is entered. To the user, in contrast, the process may appear continuous, which can cause frustration as intended results can only be achieved by trial and error. Examining various speech acts in the interaction with CUIs, Minha Lee (2020) emphasizes how the use of natural language may lead to wrong expectations. Misunderstanding the communication with a text-to-image generator as an anthropomorphized dialogue can likewise be frustrating, especially if the users are not experienced with writing effective prompts. Richard W. Janney (1999) also cautions that perceiving an interaction with a computer as an I-You relationship is problematic – especially from the intentionalist perspective of speech act theory. Because of a computer’s lack of intentions, its speech acts cannot have illocutionary force. However, it is questionable whether this also means that a computer “cannot recognise or process the intentions of a human user” (JANNEY 1999: 73) and that the user’s inputs equally have no illocutionary force. Precisely because the users can experience the interplay of fuzziness and sharpening as a dialog and because the way the generator reacts to their illocutionary intent is often highly productive, speech act theory is, despite these caveats, a fitting approach to understanding creative potentials of text-to-image generators.

How AI-Based Visualization Tools Impact Artistic Ingenuity and Visual Thinking

Turning to the current debate of intentionality and, indeed, questions of authorship and artistic embodiment regarding AI-generated visuals among designers and the artistic community in general, the far-reaching implications of the observations we discussed above become evident. Indeed, the transformations that AI-based image generators such as DALL-E 2, Midjourney, or Stable Diffusion bring about for the work of illustrators and visual artists are already in full swing. Concerns that core aspects of traditional creation processes in these fields will quickly be superseded by new AI technologies are palpable: Indeed, parts of the artistic community have adopted a defensive (and even openly dismissive) stance, with hashtags such as “#noaiart” and “#artbyhumans” trending on Instagram and Twitter (cf., e.g., BRANDON 2022) and the “No to AI generated Images” slogan (and visual label) being used in widespread (social) medial protest (cf., e.g., ELIAÇIK 2023).

How closely core points behind these protests are related to some of the questions and concepts we have touched upon above becomes obvious once we take a closer look at some of the arguments put forth in the pertaining discourse. For

instance, in late December 2022, the online database and art book publisher *3dtotal* joined the discussion by tweeting a statement that echoed many concerns voiced by the wider art community. Part of these concerns were legal copyright questions (arising in connection with the way in which existing visual art has been used to train AI-based image generators) and fears of a rise of “AI prompt artists that can tackle the workload of teams of artists” (3DTOTAL 2022: n.pag.). However, the authors tellingly also brought up implications of a possible “reduction in creative careers and a lack of true innovation in media” as well as a potential loss of the expressive function of art as a powerful tool to “capture some of the personality of the artist” that “should not be automated by a computer”.⁹¹

In many ways, the wider debate on the relation between design and technology that constitutes the background of these concerns is not new, but it has recently reached a highly accelerated quality – and it goes beyond the level of practical implications. Discussing an “increasing distance between technologists and designers”, which he observed as early as 1985, Richard Buchanan famously criticized “a general attitude that technology is only an applied science, rather than a part of design art” (BUCHANAN 1985: 4). This observation, extending our view to the wider societal context and the relation between design and the philosophy of science, rings even louder in the face of today’s AI-based image generators. While parts of the art community, as exemplified above, regard the technology behind these new, AI-based tools as existential competition and actively distinguish it from an understanding of creativity that is deeply rooted in the human perspective (and more traditional tools of its expression), Buchanan (1985: 4f.) notably called for an integration of technology and design – and highlighted the role that rhetorical theory could play in facilitating it. In that sense, we should not stop at discussing the transformations brought about by AI-based technology in the field of design from a purely descriptive point of view, but rather (along with considering important concerns and valid critique) also ask for potential benefits and possible ways of productive integration in existing mechanics of artistic creation.

Diving deeper into the pragmatic dimension of the interaction between image generators such as DALL-E 2 and human users, we might indeed argue that there is an added creative potential arising from the visual feedback loop it includes. If artistic activity indeed is, as Rudolf Arnheim suggested, “a form of reasoning in which perceiving and thinking are indivisibly intertwined” (ARNHEIM 1969: v), it is crucial to explore how AI-based visualization tools transform existing concepts of artistic strategy and ingenuity: At a bare minimum, text-to-image

91 See above. In response to the post, many creatives uttered their support for its arguments, but it also evoked contrary reactions such as “let AI be free to learn, let creatives use it as a tool” (@MADEBYRASA, December 21, 2022, quoted after 3DTOTAL 2022) or “history has shown what happens to people who stand in the way of progress” (@Charleywarliet, December 21, 2022, quoted after 3DTOTAL 2022).

generators seem to have the potential to speed up the process of prototyping visual drafts, thereby accelerating visual feedback loops between perceiving and thinking which are, according to Arnheim, crucial for creative, productive activity in general. In addition, the variety of four different images which DALL·E 2 presents in almost immediate response to one and the same text prompt (cf. fig. 1) can introduce an element of conceptual and compositional fuzziness that might even provide specific variations or combinations of elements that the designer has not previously thought of or imagined. This is particularly important since the act of artfully illustrating and representing (existing) concepts and visual ideas is just one part of the skills required by designers. A core aspect of their work takes place on the conceptual level, too. It consists in finding translations, metaphors, recontextualizations, and new compositions (cf. FAUCONNIER/TURNER 2002) that help us see a topic in different and engaging ways. Notably, Arnheim (1969: 116-134) also highlighted the importance of images in concept formation – including a notion of “experiments with drawings” (ARNHEIM 1969: 120ff.).

Following these thoughts, the fuzzy variations of image outcomes which DALL·E 2 produces in response to a text prompt not only contribute to an accelerated form of externalized visual thinking, they also introduce an element of fuzzy serendipity that invites experimentation and has the potential to add a creative surplus to the visual idea the designer strives to form: By producing a spectrum of possible ‘visual answers’ to a text input given by the user, the image generator might function, to an extent, as an artificial ‘sparring partner’ to brainstorm, prototype, and refine visual ideas as well as conceptual and stylistic approaches to a given topic or idea.



Figure 1: Cluster of images created with DALL·E 2 in February 2023 using the text prompt “human creativity”

Consider the example provided above (fig 1): It shows a cluster of visualizations created by DALL·E 2 in response to the text prompt “human creativity”. We are presented with a selection of four variations that are quite different in conceptual content and style, resulting in a diverse array of visual representations based upon the same prompt. In addition, there is a diachronic element to

this variation, as running the same text prompt again can yield utterly different results. In essence, the relation between the text prompt and the visuals created is not precise, it is fuzzy – and while complex, elaborately designed text prompts can strategically guide and narrow down the extent of this fuzziness, a certain degree of vagueness and imprecision will always remain.

Adapting some of Hanns-Werner Heister’s (2021) thoughts on using core concepts of Fuzzy Logic to elucidate the artwork process can be a first step towards a deeper understanding of the creative potential inherent in this intrinsic fuzziness: Heister describes artistic processes in terms of a “multi-dimensional, multi-layered, involved (encapsulated) and folded [...] [dialectic of] fuzziness and sharpening” (HEISTER 2021: x) that applies principles of similarity, filtering, crystallization, blurring, and variation (cf. HEISTER 2021: 17-20) “for intentionally compositional-artistic utilization of fuzziness in its different facets of art” (HEISTER 2021: x). In essence, this complex dialectic is also at play when we use AI-based tools such as DALL·E 2 to create visuals, iteratively permutating our prompts to guide and (re)sharpen the spectrum of fuzzy visual results that is being generated.

While Heister’s theory is explorative and mostly developed in connection to the field of music, it convincingly manages to relate the concept of fuzziness to general mechanics of creativity and innovation, arguing that (in relation to artistic processes) “fuzziness is necessary in an at least double sense: it is inevitable, and it is necessary for changes, developments, variations of the given” (HEISTER 2021: 1). These thoughts lead us back to the observations we made about processes of imagination at the very outset of this paper – and they also closely correspond to Giambattista Vico’s (1979 [1711-1712]) remarks on the roots of ingenuity in general. According to Vico, understanding and aptly assessing any situation requires a ‘flexible’ use of reason – a kind of ‘fuzzy logic’ – that he discusses within the framework of his theory of *ingenium*. This *ingenium* is characterized as a capacity of thinking that perceives *the similar in the different*: it is the ability to discover similarities in seemingly foreign concepts and differences in what appears to be similar (cf. VICO 1979: 135). In short, we might say: It is the capacity to come up with (and think in) metaphorical connections and distinctions.

According to Vico, *ingenium* in this sense constitutes the actual cognitive faculty of the human being. He argues that the analytical-deductive method of Descartes is only able to dissect what has already been found, while finding something new is the task of the *ingenium* (cf. VICO 1947 [1709]: 46-47; see also FUCHS 2020: 76). Using an inventive and combinatorial topology, we can put problems and facts in a new and unexpected light, uncover hidden connections, and open up new perspectives on issues. As a matter of fact, Vico himself pointed out that ‘ingenious’ thinking often makes use of metaphors and analogies, which he considered not only as aesthetic and artful forms of representation but

also as an inventive way of generating new ideas (cf. FUCHS 2020). As it creates changing variations of possible structures of visual meaning rather than one-dimensional, ‘precise’ translations, the element of fuzziness inherent in the image creation process with text-to-image generators thus also has the potential to catalyze the invention of new metaphors – and, in extension, processes of creative ingenuity in general.

Following this train of thought, even deeper implications regarding theories of knowledge and cognition might be considered. In contrast to sheer rationalism, Vico tellingly advocated the intertwining quality of logical and sensual aspects inherent in metaphorical thinking: Metaphors and analogies can make evident what is otherwise only abstract – an idea that can be tracked back to Aristotle’s *Poetics* (1995 [335 a.d.]: 4-8). This epistemological concept of metaphorical thinking is closely related to Arnheim’s thoughts we referred to further above – and generally theorizes that ‘fuzzy’ (visual) processes of creation open up an expanded space of understanding and interpretation compared to a more sober, abstract access. The focus here is not to achieve utmost precision, but to create meaningful images that carry orientating power. To *define* something *as something* allows less leeway than to judge something as *similar to something* (cf. BRANDSTÄTTER 2008: 23). In this sense, the dispersion of possible outcomes presented by images generators such as DALL·E 2, Midjourney, or Stable Diffusion carries a unique creative potential: It creates a number of similar, but different renditions of a conceptual input given via a text prompt – and thus provides potential impulses for new connections between seemingly different things.

Conclusion

We have discussed various aspects in which the current generation of text-to-image generators transform existing visual creation processes and artistic ingenuity, and how the ‘fuzzy’ variety of images that tools like DALL·E 2, Midjourney, or Stable Diffusion present in response to one and the same text prompt contributes to a highly accelerated form of externalized visual thinking. This form of visual thinking is aided by the bimodal, seemingly conversational nature of the text-to-image interface, appearing as a sequence of natural language speech acts, which can be both a mesmerizing experience and a source of frustrated expectations. Hegel’s rather narrow concept of what it means to have ‘imagination’ and to use art as a means of relating to the world makes the creative space in which text-to-image generators operate seem extremely small. However, examining the variability and possibilities of image creation through other viewpoints illustrates how large this supposedly small space actually is. Not being able to understand exactly what is meant, i.e., the vagueness of the linguistic interaction, is crucial

for opening up this space. While new text-to-image generators challenge our traditional notions of design processes and pose various disruptive questions in both theory and practice, it is crucial to also examine the creative potential inherent in the technology behind them. The explorative thoughts collected in this paper present a first rough approach towards examining the distinctive mechanics of ‘fuzzy ingenuity’ in that context – and can hopefully lead to further and deeper discussions of the topic.

Bibliography

- 3DTOTAL (@3dtotal): 3dtotal has Four Fundamental Goals ... Tweet on *Twitter*. December 21, 2022. <https://twitter.com/3dtotal/status/1605597714187575297> [accessed February 16, 2023]
- ARISTOTLE: *Poetics*. Edited and translated by Stephen Halliwell. Harvard [Harvard University Press] 1995 [335 a.d.]
- ARNHEIM, RUDOLF: *Visual Thinking*. Berkeley [University of California Press] 1969
- AUSTIN, JOHN L.: *How to Do Things with Words*. Oxford [Oxford University Press] 1962
- BLANKE, TOBIAS: Hegels “Artificial Intelligence”. In: ANDREAS ARNDT; et al. (eds): *Hegel-Jahrbuch*, vol. 2007, no. 1. Berlin [Akademie Verlag] 2007, pp. 292-297
- BRANDON, ELISSAVETA M.: Fueled by the AI Frenzy, #artbuyhumans is the New #nofilter. In: *Fast Company*. December 21, 2022. <https://www.fastcompany.com/90826292/AI-frenzy-art-by-humans-is-the-new-no-filter> [accessed February 16, 2023]
- BRANDSTÄTTER, URSULA: *Grundfragen der Ästhetik*. Köln [Böhlau] 2008
- BUCHANAN, RICHARD: Declaration by Design: Rhetoric, Argument, and Demonstration in Design Practice. In: *Design Issues*, 2(1), 1985, pp. 4-22
- DARAS, GIANNIS; ALEXANDROS G. DIMAKIS: Discovering the Hidden Vocabulary of DALLE-2. *arXiv:2206.00169*. June 1, 2022. <https://arxiv.org/abs/2206.00169> [accessed February 16, 2023]
- ELIAÇIK, ERAY: Does ArtStation Become PromptStation? In: *DataConomy*. January 5, 2023. <https://dataconomy.com/2022/12/no-to-AI-generated-images-artstation> [accessed February 16, 2023]
- FUCHS, BRIGITTA: Vico über rhetorische und szientifische Evidenz. In: OLAF KRAMER; CARMEN LIPPARDT; MICHAEL PELZER (eds): *Rhetorik und Ästhetik der Evidenz*. Berlin [de Gruyter] 2020, pp. 67-82
- FAUCONNIER, GILLES; MARK TURNER: *The Way we Think: Conceptual Blending and the Mind's Hidden Complexities*. New York [Basic Books] 2002
- HEGEL, GEORG WILHELM FRIEDRICH: *Aesthetics: Lectures on Fine Art*. Vol 1. Translated by Thomas M. Knox. Oxford [Clarendon Press] 1975 [1835-1838]

- HEGEL, GEORG WILHELM FRIEDRICH: *Enzyklopädie der philosophischen Wissenschaften im Grundrisse. Erster Teil: Die Wissenschaft der Logik. Mit den mündlichen Zusätzen.* 13th edition. Frankfurt /M. [Suhrkamp] 1970a [1817]
- HEGEL, GEORG WILHELM FRIEDRICH: *Enzyklopädie der philosophischen Wissenschaften im Grundrisse. Dritter Teil. Die Philosophie des Geistes. Mit den mündlichen Zusätzen.* 11th edition. Frankfurt /M. [Suhrkamp] 1970b [1930]
- HEGEL, GEORG WILHELM FRIEDRICH: *Hegel's Philosophy of Mind.* Translated by William Wall. Oxford [Clarendon Press] 2007 [1930]
- HEGEL, GEORG WILHELM FRIEDRICH: *Vorlesungen über die Ästhetik I.* 16th edition. Frankfurt/M. [Suhrkamp] 1970c [1835-1838]
- HEISTER, HANNS-WERNER: *Music and Fuzzy Logic: The Dialectics of Idea and Realizations in the Artwork Process.* Berlin [Springer] 2021
- JANNEY, RICHARD: Computers and Psychosis. In: JONATHON P. MARSH; BARBARA GORAYSKA; JACOB L. MEY (eds): *Humane Interfaces: Questions of Method and Practice in Cognitive Technology.* Amsterdam [Elsevier] 1999, pp. 71-79
- LEE, MINHA: Speech Acts Redux: Beyond Request-Response Interactions. In: *Proceedings of the 2nd Conference on Conversational User Interfaces.* Bilbao [ACM] 2020. <https://doi.org/10.1145/3405755.3406124> [accessed February 16, 2023]
- NGUYEN, HUNG T.; WALKER, ELBERT A.: *A First Course in Fuzzy Logic.* 3rd edition. Boca Raton [Chapman and Hall/CRC] 2006
- PATTERSON, RICHARD: *Image and Reality in Plato's Metaphysics.* Indianapolis [Hackett] 1985
- SEARLE, JOHN R.: A Taxonomy of illocutionary Acts. In: GUNDERSON, KEITH (ed): *Language, Mind, and Knowledge.* Minneapolis [University of Minnesota Press] 1975, pp. 344-369
- SIMON, JOSEF: Zeichenmachende Phantasie: Zum systematischen Zusammenhang von Zeichen und Denken bei Hegel. In: *Zeitschrift für philosophische Forschung*, 1/2, 1996, pp. 254-270
- TICHENOR, PHILIP. J.; GEORGE. A. DONOHUE; CLARICE. N. OLIEN: Mass Media Flow and Differential Growth in Knowledge. In: *Public Opinion Quarterly*, 34(2), 1970, pp. 159-170
- VICO, GIAMBATTISTA: *De nostri temporis studiorum ratione. Vom Wesen und Weg der geistigen Bildung.* Translated by Walter F. Otto. Bad Godesberg [Küpper] 1947 [1709]
- VICO, GIAMBATTISTA: *Liber Metaphysicus: Risposte.* Translated by Stephan Otto and Helmut Viechtbauer. Munich [Wilhelm Fink] 1979 [1711-1712]
- VOX [@vox]: The Text-to-Image Revolution, Explained. Video on *YouTube*. January 1, 2022. <https://www.youtube.com/watch?v=SVcsDDABekM> [accessed February 16, 2023]

WITTGENSTEIN, LUDWIG: *Philosophical Investigations / Philosophische Untersuchungen*.
2nd edition. Translated by Gertrude E.M. Anscombe. Cambridge [Blackwell]
1998 [1953]

ZADEH, LOTFI A.: Fuzzy Sets. In: *Information and Control*, 8(3), June 1965, pp. 338-353