IMAGE

ZEITSCHRIFT
FÜR INTERDISZIPLINÄRE
BILDWISSENSCHAFT

THE INTERDISCIPLINARY
JOURNAL OF IMAGE SCIENCES

37(1), 2023

Herausgabe der Zeitschrift:

Goda Plaum
Lars C. Grabbe
Klaus Sachs-Hombach

Herausgabe dieser Ausgabe:

Lukas R.A. Wilde
Marcel Lemmes
Klaus Sachs-Hombach

HERBERT VON HALEM VERLAG

http://image-journal.de

Impressum

IMAGE
Zeitschrift für interdisziplinäre Bildwissenschaft

Heft 1/2023
19. Jahrgang

https://image-journal.de

Verlag

Herbert von Halem Verlagsgesellschaft
mbH & Co. KG
Boisseréestr. 9-11
50674 Köln
Telefon: +49(0)221-9258290
Telefax: +49(0)221-92582929
E-Mail: info@halem-verlag.de

Vertreten durch:
Herbert von Halem Verlagsges. mbH
Geschäftsführer: Herbert von Halem
Registergericht: Köln
Registernummer: HRB 25647

Registereintrag:
Eingetragen im Handelsregister.
Registergericht: Köln
Registernummer: HRA 13409

Umsatzsteuer-ID:
Umsatzsteuer-Identifikationsnummer nach
§27a Umsatzsteuergesetz:
DE 172 714 183

Generative Imagery:
Towards a 'New Paradigm' of
Machine Learning-Based Image Production

Four generative AI creations made with a local Stable Diffusion installation for the prompt "self portrait of an artificial intelligence". The models used were (from top left to bottom right): mdjrny-v4, OpenNiji-v2, sd_1.5, analogDiffusion_10.

:

Inhalt

IMAGE
The Interdisciplinary Journal of Image Sciences
37(1), 2023, S. 1-1
ISSN 1614-0885

Editorial zur IMAGE 37

Sehr geehrte Leser*innen,

seit nunmehr 18 Jahren erfreut sich die *IMAGE* am Interesse interdisziplinär orientierter Bildwissenschaftler*innen unterschiedlicher Ausrichtungen und Disziplinen. Von gelegentlichen Verzögerungen bei der Herausgabe und zuweilen aufgetretenen technischen Problemen abgesehen, kann die Geschichte der *IMAGE* als Erfolgsgeschichte gelten. Sie eignet sich auch als Archiv, um die jüngere Geschichte der Bildwissenschaft in einigen wichtigen methodischen und thematischen Aspekten nachzuvollziehen. Mit der vorliegenden Ausgabe eröffnet die *IMAGE* ein Blick in die Zukunft KI-generierter Bilderwelten.

Der Vorstand der Gesellschaft für interdisziplinäre Bildwissenschaft (GiB) erachtet die *IMAGE* als wichtiges Publikationsorgan, das insbesondere auch den Mitgliedern der Gesellschaft verfügbar sein soll. Daher hat sich der Vorstand in den letzten Jahren um eine stärkere Verbindung von Zeitschrift und Gesellschaft bemüht und einige Maßnahmen zur Qualitätssicherung ergriffen. Seit der Ausgabe 35 (Januar 2022) fungiert der Vorstand der GiB als Herausgeber-Team der Zeitschrift und der wissenschaftliche Beirat der GiB als Editorial Board mit Gutachterfunktion. Zeitgleich wurde auch das Layout angepasst.

Die *IMAGE* wurde als Zeitschrift zur interdisziplinäre Bildwissenschaft seit ihrem Start im Januar 2005 vom Herbert von Halem Verlag unterstützt. Mit der vorliegenden Ausgabe hat diese Unterstützung eine neue Qualität erhalten. Zukünftig wird die *IMAGE* vom Verlag redaktionell betreut und auf dem Server des Verlags verfügbar gehalten. Das Herausgeber-Team hofft, dass ihm damit ein entscheidender weiterer Schritt zur Professionalisierung der *IMAGE* gelungen ist.

Wir möchten diese Gelegenheit nutzen, um allen Beteiligten unseren Dank für ein verlässliches Engagement und eine immer wohlwollende Unterstützung der GiB sowie der *IMAGE* auszusprechen. Insbesondere möchte wir Herrn von Halem und seinem Verlag für die zahlreiche Unterstützung danken, die er ganz allgemein den bildwissenschaftlichen Forschungen und im Besonderen der GiB und *IMAGE* stets hat zukommen lassen.

Die Herausgeber der *IMAGE*
Goda Plaum, Lars Grabbe, Klaus Sachs-Hombach

IMAGE
The Interdisciplinary Journal of Image Sciences
37(1), 2023, S. 2-5
ISSN 1614-0885

HERBERT VON HALEM VERLAG

Lukas R.A. Wilde / Marcel Lemmes / Klaus Sachs-Hombach

Preface

Dear readers,

Lukas R.A. Wilde, Marcel Lemmes, and Klaus Sachs-Hombach are excited to welcome you to this special issue of IMAGE: *The Interdisciplinary Journal of Image Sciences* dedicated to what we might tentatively call a new paradigm for image production, i.e., the advent of generative imagery created by means of machine learning-based platforms. The contributions in the present issue are the result of a fruitful and inspiring workshop from mid-February 2023 at the University of Tübingen, Germany, and online via Zoom where we were able to discuss emerging technologies and applications like DALL·E, Midjourney, and Stable Diffusion with both the authors within the present issue as well as an interested digital public of scholars, students, artists, and practicioners. The obvious importance of these new technologies attracted over 250 people from 123 universities and institutions from all across the globe to enrich our event with stimulating and inspiring comments and questions. We were and still are both humbled and honored by such an attention. We want to take this opportunity to express our dearest gratitude to them and particularly to those who traveled all the way to Tübingen to contribute first a draft paper and later the contributions you will find collected here today.

The idea for this workshop originated when we met at a conference in Thessaloniki in the late August of 2022. In the evening, after a long and galvanizing day, we were cooling off in one of the nice cafés facing the Mediterranean Sea. Quickly we became absorbed in a conversation about some emerging AI tools, DALL·E and Midjourney, still very new back then. Lukas in particular was extremely excited about and fascinated by the novelty of generative imagery, anticipating profound changes in the near future that may deeply modify fundamentals in our every-day life and society as a whole. He confessed to feel a bit like a gambling addict, manically throwing coin after coin in one of those slot machines where you know you'll never get anything out of aside from an immediate feeling of gratification – or, perhaps, a sense of wonder exploring this new 'latent space' of all possible images.

What was certainly uncontroversial even back then: the capabilities and the resulting pictures of generative platforms as well as various other emerging applications within the field of artificial intelligence are very impressive. The future has once again started, it seemed. And there is also no doubt that these tools that have now been available to the general public for quite a while will fundamentally change our world. During our lively discussions that lasted well into the night we still couldn't quite agree on various implications these technologies might have for fundamental concepts of media history, picture theory, or even social life. Had visual communication and imagery as such just changed as fundamentally as when photography was first invented? The 'elephant in the beachside café' naturally concerned the nature of these technologies as we understood them back then, or struggled to: are these 'AI' systems really creative and, if not, how could we define a meaningful difference to how humans acquire skills by studying, repeating, and combining styles and techniques from earlier works? In which ways, exactly, does this 'stochastic creativity' differ from earlier, more rule-based forms of computation before? What might all of this mean for our understanding of pictures and pictoriality? Klaus in particular felt that we should be especially careful or hesitant in applying words like "learning", "intelligence", or "creativity" to these systems; possibly even or especially if (or rather when?) it becomes difficult to distinguish the results of human and AI image production in, let's say, a more advanced turing test.

We became increasingly aware that we were lacking sufficient knowledge to even phrase properly, let alone answer any of these questions. All the more enthusiastically did we delve into debates about a new relationship between textuality and pictoriality while we explored, in the noisy streets of Thessaloniki from our phone interfaces, how DALL·E did not refuse even our most silly and outlandish verbal requests (cf. fig. 1). Even before the night had ended, we decided (quite academically) to organize a workshop to learn more about these topics. After travelling back to Tübingen and Trondheim a few days later, we started contacting researchers who might be more knowledgeable, being enticed by their expertise while also enticing for a brief yet well informed academical exchange about several theoretical, social, and practical issues that generative imagery had made visible even back then. In early October we then published a small Call for Participation to which, to our great delight, an impressive range of scholars from various fields in the humanities replied with outstanding contributions. They even advanced paper drafts that we were allowed to circulate between all workshop participants prior to the event.

"DALL·E, Midjourney, Stable Diffusion: A 'New Paradigm' of Image Production?" took place on February 13/14, 2023. To publish expectional results in such a short span of time in open access format was quite a challenge, for us as well as for the Herbert von Halem publisher and especially for our contributors

Figure 1: A DALL·E creation from August 31, 2022; the prompt has been lost to the night

with their many obligations during the busy months of our recent academic semesters. We are thus deeply grateful for the enormous joint efforts that made the present issue possible. Faced with the rapid speed of current technological developments we felt it was more than worth it. You, dear readers, can now see the fruits of our joint labor – but, most of all, of the sincere academic curiosity that united all parties involved – in front of you. We would like to thank the University of Tübingen for providing the venue, and the Norwegian University of Science and Technology (NTNU) in Trondheim for making it possible to organize this workshop jointly. Last but not least, we would also like to express our gratitude to the conference team (in particular Karin Kießling and Elias Raatz) for their invaluable assistance with the organization and their helpful presence during these two exciting days in February. We wish all of you a stimulating engagement with the following fourteen essays. We hope and are quite confident that they will be as inspiring for you as they are for us, opening up many paths for future discussions about generative imagery sure to come.

Tübingen and Trondheim, March 2023

One very last remark that is meant to be funny: Thinking about artificial intelligence, we also started to wonder how an AI might comment about their kind and how it would write a welcome note. So, we prompted OpenAI's ChatGPT with the following: "Write a welcome preface for a academic journal based on a workshop on 'Generative Imagery: Towards a New Paradigm of Machine Learning-Based Image Production'". Here are the results; you might find it amusing to compare it to our preface:

Welcome to the special issue of our academic journal, which is based on the exciting workshop on “Generative Imagery: Towards a ‘New Paradigm’ of Machine Learning-Based Image Production”. This workshop brought together leading researchers, practitioners, and enthusiasts in the field of machine learning-based image production to share their latest research, ideas, and perspectives on this emerging field.

Generative imagery has the potential to revolutionize the way we produce and create images, opening up new avenues for artistic expression, scientific visualization, and data-driven design. With the advent of deep learning techniques, generative models have become increasingly powerful and versatile, capable of creating stunningly realistic images that can rival those produced by human artists.

The workshop explored various topics related to generative imagery, including image generation, style transfer, image editing, and more. Through a series of talks, presentations, and discussions, participants shared their cutting-edge research and insights into the challenges and opportunities of this new paradigm.

The contributions in this special issue showcase some of the most exciting research and ideas presented at the workshop, highlighting the latest developments and trends in generative imagery. We hope that this collection will inspire further research and exploration in this rapidly evolving field and contribute to the development of new tools and techniques for machine learning-based image production.

We extend our gratitude to all the workshop participants, contributors, and reviewers for their hard work and dedication in making this special issue possible. We hope that you will find this collection informative, inspiring, and thought-provoking.

IMAGE
The Interdisciplinary Journal of Image Sciences
37(1), 2023, S. 6-33
ISSN 1614-0885
DOI: 10.1453/1614-0885-1-2023-15446

Lukas R.A. Wilde

Generative Imagery as Media Form and Research Field: Introduction to a New Paradigm

Abstract: This introduction to the collection "Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production" discusses whether – or to what respect – generative imagery represents a new paradigm for image production; and if that constitutes even a novel media form and an emerging research field. Specifically, it asks what a humanities approach to machine learning-based image generation could look like and which questions disciplines like media studies will be tasked to ask in the future. The essay first focuses on continuities and connections rather than on alleged radical shifts in media history. It then highlights some salient differences of generative imagery – not only in contrast to photography or painting but specifically to earlier forms of computer-generated imagery. Postulating a 'new paradigm' will thus be based 1) on generative imagery's emergent or stochastic features, 2) on two interrelated, but often competing entanglements of immediacy-oriented and hypermediacy-oriented forms of realisms, and 3) on a new text-image-relation built on the approximation of 'natural', meaning here *human* rather than machine code-based language. The survey closes with some reflections about the conditions under which to address this imagery as a distinct media (form), instead of 'merely' as a new technology. The proposal it makes is to address generative imagery as a form of *mediation* within evolving dispositifs, assemblages, or socio-technological configurations of image generation that reconfigure the distribution of agency and subject positions within contemporary media cultures – especially between human and non-human (technological as well as institutional) actors. Of special importance to identify any (cultural) distinctness of generative imagery will thus be a praxeological perceptive on the establishment, attribution, and negotiation of cultural 'protocols' (conventionalized practices and typical use cases), within already existing media forms as well as across and beyond them.

Introduction

The emergence of machine learning-based platforms has been a prominent and increasingly prevalent topic in both popular as well as specialized academic discussions for many years now (cf. for a survey NILSSON 2010; SUDMANN 2018a; MITCHELL 2019). Around the middle of the year 2022, these emerging technologies left the spheres of R&D departments, computer science labs, and our speculative imagination. Generative platforms started to pervade the everyday life of people around the globe. Beginning with text-to-image technologies such as DALL·E, Stable Diffusion, or Midjourney (flanked by a range of other competitors such as Imagen, Wombo Dream, DeepDream, or Leonardo AI), and succeeded by further evolving and increasingly easier to-access prompt-to-text-applications like ChatGPT, Bing, or Bard. Discussions about the imminent threats, potentials, and transformations of media and communication now permeate news media, popular culture, and academic discourses. Other forms of machine learning technologies are developing steadily too, with text-to-music, text-to-video, text-to-code, or even text-to-3D rapidly progressing. Certainly, machine learning–based image generation technologies – commonly referred to as 'AI imagery' or 'generative imagery' – are only a small part of these developments. Their history was long in the making long before the summer of 2022 (cf. MILLER 2019: 59-122; BAJOHR 2021). The successive stages of technological developments in the area of generative imagery have been historized (cf. OFFERT 2022) as a transition from classification to generation (2012–2015), over five years of GAN development (generative adversarial networks, 2015–2020), leading up to the currently popular diffusion models (2020–present), whose 'multimodal' deep learning through CLIP (contrastive language-image pre-training) and GLIDE (guided language-to-image diffusion for generation and editing) combines and consolidates techniques from NLP (natural language processing) and "computer vision" (DOBSON 2023). Despite this gradual progress and the fact that the actual deep learning-"media revolution" (SUDMANN 2018b: 66; my translation) has happened a while ago – or rather: has been happening for a long time now – the summer of 2022 introduced a moment of radical shift in the public awareness, mainly due to the fact that generative imagery since left the confinement and control of companies, research labs, or specialized artistic experiments, becoming available to the general public. This also marked the beginning of what Fabian Offert (2022: n.pag.) called the "Photoshop era" of such image synthesis. It is now feasible to use generative models as an everyday tool to create highly realistic images from a rough sketch, adding AI-based modifications layer by layer. Stability AI's open-source application Stable Diffusion, for instance, is characterized by a modular architecture that allows working with more and more fine-tuned extensions such as OpenPose Editor or ControlNet (cf. ZHANG/AGRAWALA 2023) and through the exchange of individual, pre-trained models through the collaborative hosting and exchange platform GitHub. As of early March 2023, there are already Stable Diffusion plugins available for Adobe Photoshop and other graphics programs (cf. ALFARAJ 2023), integrating generative imagery seamlessly into established practices of digital image production and editing.[1]

After an initial rush of public interest in this imagery around July to October of 2022, prompt-to-text platforms seem to attract not only much more press coverage at present (March 2023) – at times excited, worried, or increasingly annoyed. They also seem to necessitate more 'emergency meetings' in universities and other institutions where decisions need to be made quickly on how to deal with the impacts of ChatGPT and the like on all aspects of social, cultural, and political life. In many other ways, too, earlier prompt-to-image platforms appear more harmless to existing regulations. As Hannes Bajohr (2023) remarks in his contribution to this collection, nobody would (and, to my best knowledge, nobody *did*) speak of DALL·E, Stable Diffusion, or Midjourney as having any sort of consciousness or personality – let alone a range of *alternate* personalities 'discovered' in ChatGPT or Bing (cf. TANGERMANN 2023; VINCENT 2023). For AI chatbots simulating direct communicative interactions, this is currently discussed daily (even if arguably in some frame of suspension of disbelief, make-believe, or role-play, as René Walter, 2023, has argued). Generative imagery still seems to retain a much more salient *instrumental* role, discussions of alleged 'autonomy' or 'creativity' restricted to the interpretation of prompts and the subsequent production of results, not the interaction or communication with human users (via images) itself. This might partly be owed to present interface design limitations: None of the currently available generative imagery platforms retain memories between input prompts, which is a mere technical limitation at this point. Certainly, as both prompt-to-image as well as prompt-to-text technologies make their APIs interfaces available (cf. BROCKMAN et al. 2023), a dialogue and memory-based image platform is probably not too far away (enabling hypothetical commands like "combine the last three results, and then respond with another picture representing a next moment in time").[2] Arguably, however, it would still be the verbal interaction through chats prompts that could generate the uncanny impression of a 'responding agent' once again, not the immediate 'communication' through image generation, for the simple reasons that this core

1 Only on March 21, Adobe even unveiled their own generative AI, "Firefly", advertised as *not* drawing on proprietary material of earlier artists that did not agree to this (cf. ADOBE 2023) – which should change a lot of things argued for within the essays in the present collection. Given the speed of current developments, it will be harder and harder to write texts that are still somewhat up to date, it seems (cf. WILDE 2023).

2 Actually, also mere days before the manuscript for this publication was finalized, OpenAI not only announced that a later version of GPT-4 would be multimodal (cf. OPENAI 2023), Microsoft also published a press release that Bing would soon entail DALL·E to do, under the name "Bing Image", exactly what was merely imagined here (cf. MICROSOFT 2023).

function – producing novel images at rapid speed in seconds – simply has *no* equivalent in earlier human (or even human-machine-augmented) communication and thus runs contrary to all communicative intuition.[3]

These cursory thoughts are, in any case, about the only remarks about prompt-to-text platforms provided within the present collection of essays – with the exception of Bajohr who dives more deeply into the 'artificial semantics' of large-language-models behind both prompt-to-image and prompt-to-text-platforms. The following thirteen essays instead offer a range of humanities-based perspectives on the 'discourse event' that started the AI discussion back in July–October 2022. Limiting our interest to AI-generated pictorial representations and image forms, the overarching question for our workshop "DALL·E, Midjourney, Stable Diffusion: Responses from Media Studies towards a 'New Paradigm' of Image Production" (University of Tübingen, February 13/14, 2023) seemed ambitious enough: Does the availability of generative imagery as an everyday resource represent a moment of media change in contemporary image and media history, perhaps as consequential as the transition from mechanically to photochemically produced pictures or even as the emergence of mechanical reproduction before? In October 2022, when Klaus Sachs-Hombach and I published the Call for Participation asking these questions, answers seemed uncertain at best. As every responsible scholar would, we hence put the 'new paradigm' of image production into single quotation marks. Half a year later now, it seems less complacent to do without them confidently. This certainly demands some reasoning. In the present introduction to our collection, I would like to provide a few parameters and coordinates for the 'latent space' of media studies and picture theory discourse, if this metaphorical expression is allowed, that the following essays might be situated in. Their proposed perspectives are based on a range of fields across the humanities, of which media studies is just one. Indeed, the urgent concerns and questions posed by generative imagery are going to be of paramount importance for all disciplines working with and on images, pictoriality, and visual or multimodal communication. What media studies – or the conceptual and analytical departing point of *mediation* – could offer for these discussions, perhaps, is a framework connecting and interrelating communicative-semiotic, material-technological, and cultural-institutional concerns and perspectives.

First, I want to set out from a perspective focusing on continuities and connections rather than on radical shifts in media history. Secondly, I do want to

3 An interesting point of comparison might be found in the narratological observation that verbal texts usually generate the impression of an anthropomorphic narrator or of a personalized voice (perhaps even distinct from the actual author), while this is not necessarily true for the pictures of films or comic books: "Written narrative text is perceived as analog to the process of verbal narration, it is (in Fludernik's 1996 terminology) 'naturalized'. Comics, as well as films, have, regarding their visual components, no equivalent in mundane, everyday communication" (SCHÜWER 2008: 389; my translation).

highlight some salient differences of generative imagery possibly constituting a 'new paradigm' – not only with regard to photography or painting but specifically in contradistinction to earlier forms of computer-generated imagery or 'machine vision'. Finally, if we understand generative imagery as an emerging, distinct field of research in the humanities, can we identify some of the key concerns within this paradigm? My introduction closes with a few reflections about the conditions under which to address these new image technologies as a distinct media (form). The proposal I want to make is this: addressing generative imagery as a (partially novel) form of *mediation* asks how these developing dispositifs, assemblages, or socio-technological configurations of picture generation reconfigure the distribution of agency and subject positions within contemporary media cultures, especially between human and non-human (technological as well as institutional) actors.

Continuities and Connections?

Generative image platforms produce pictorial artifacts without the indexical relations of photography to light waves or of painting to brush strokes. As Eryk Salvaggio (2023a) argues most convincingly in his present contribution, they instead recombine and perhaps also reveal aspects of underlying pictorial datasets as well as of the human decisions behind their classification and organization. Still, we might ask skeptically: what is genuinely new about that, really? The abandonment of referential reality (of an indexical relationship to physical reality), is hardly new for digital pictures and has been established through CGI for decades (cf. MITCHELL 1992; RICHTER 2008; GOOSKENS 2011). The partial autonomy of a 'non-human apparatus' generating pictures 'automatically' might even constitute one of the points of departure of media theory with the emergence of photography over a hundred years ago (cf., for instance, BENJAMIN 2007 [1935]). Generative imagery is then remarkable perhaps not in quality but in quantity, speed, and availability as platforms like DALL·E, Midjourney, or Stable Diffusion can generate, through rapid feedback loops, an infinite number of pictures in all possible stylistic variations at incredible speed even for laypersons. All the resulting *individual* pictures then seem so arbitrary and ephemeral that they hardly seem to deserve deepened individual attention or analysis. This, however, makes generative imagery perhaps an especially suited topic for media studies and media theory interested less in individual artifacts (or 'imagetexts') than in the structural impact of media technologies on culture and society in general.

The lasting consequences of this moment of media transformation on social, political, and cultural practices, conventions, and institutions are certainly far from decided or determined at this point in time. What can be stated with some confidence, however, is that the speed of recent developments has been surprising for most observers. For the time being, our institutions and laws are hopelessly lagging to regulate some very old (and some newly emerging) questions. As Jay D. Bolter (2023) points out in his present contribution, high on this list of urgent concerns are certainly questions of authorship (plagiarism vs. fair use) under these new technological decisions. One possible task, specifically for media studies, could then be to highlight continuities and connections a) between generative imagery and earlier forms of "machine vision" (cf. GALLOWAY 2021; RETTBERG 2022; 2023, as well as DOBSON 2023), b) between the present and earlier moments of media transformation and media change, as well as perhaps c) between practices and uses of pictures that have either proven resilient to such changes or are resurfacing. A respective praxeological perspective might go way back, indeed. Lev Manovich, who inspired our discussions as early as in July 2022 in a series of Facebook 'micro essays' (for lack of a better term), described something that he coined "the return of the classical 'art of the copy'" (MANOVICH 2022: n.pag.). His observation was that art historical storytelling, focusing on individual, outstanding pictures, largely ignored the hundreds and thousands of similar copies and variations that were actually produced in studios and workshops – in favor of a highly selective (and thus finally ideological) 'slice of history' in museums today. The production of pictures has then, maybe, always been dominated by practices of imitating, copying, and slightly variating existing patterns of visual representation. We are all the more excited to have an opening essay by Manovich (2023a) in the present collection that draws especially on his perspectives and experiences as an artist and practitioner.

Praxeological questions might reveal many more such connections, the most saliently one the notion of "remix" and "remix culture" that Bolter (2023) and Lamerichs (2023) discuss in more detail. Not only audio remix (in hip-hop) has been established for decades, but also "the somewhat younger video remix, which involves the editing and often complex layering of a series of video clips together" (BOLTER 2023: 199). Comparisons to older, 'analog' media and image technologies can also reveal interesting analogies with regard to their 'statistical' nature as Jens Schröter (2023) discusses in his essay on Francis Galton's composite photography portraits and Sigmund Freud's fascination with them. To Freud, superimposed composite images corresponded to the generalized visual condensation of dreams through the subconsciousness – or at least our recollection of dreams. Can AI-generated imagery thus be seen as a contemporary, mediatized form of a collective "statistical unconscious" (SCHRÖTER 2023: 111)? Roland Meyer (2023c), in turn, discusses a more immediate media-historical connection between generative imagery and stock photography and press image archives on the one hand and recent digital search engines on the other. Meyer traces how the Bettmann Archive in the 1930s created a new form of image valorization by collecting and 'assembling' pictures together with metadata on physical data carriers like index cards. The mediality of both image forms – in physical archives as well as especially in generative platforms – is thus determined by their valorization and commodification which in turn rest on a "media history of image retrieval systems" (MEYER 2023c: 103). Another analogy could be found with respect to fan cultures and fan practices, currently certainly the sociocultural context where generative imagery is exploited, tested, and negotiated most viciously. Nicolle Lamerichs (2023) discusses in her survey of these developments to what extent generative platforms could be considered a form of 'transformative fan fiction' even on a technological level, albeit one that is deeply entangled in platform economies and respective data-driven business models that have been evolving rapidly for about 10 years now. A different form of continuity is then again pointed out by Pamela Scorzin in her survey of artistic practices that include not only the newest iterations of machine learning-based technologies. Technologically distinct phenomena such as humanoid robots on media stages, avatar design in the metaverse, or partly algorithmic created music videos are employed to represent similar questions or recurring topics like artistic authorship or mediated body representations. Manovich (2023a) likewise points out such connections with regard to Ivan Sutherland's computer program Sketchpad (1961-1962) that finished half-drawn circles or rectangles; within "cultural perception" (!) this "was undoubtedly 'AI' already" (MANOVICH 2023a: 33).

As important a task as it will be to describe generative imagery on the level of social practices – and thus in terms of continuities and connections rather than in dramatic 'turns' – there *are* many perspectives that focus on mostly new aspects of mediation between human and non-human agents. Many of the contributors in the present collection still turn to well-known protagonists of media studies and media theory to pinpoint what, exactly, distinguishes generative imagery from photography as much as from 'analog' picture forms before them. These readings at the same time create and challenge notions of continuity in media history. We will thus once again visit the thoughts of authors like Theodor Adorno (OFFERT 2023) and Walter Benjamin (ERVIK 2023), John Austin and Ludwig Wittgenstein (FEYERSINGER et al. 2023), Roland Barthes (ERVIK 2023; OFFERT 2023; SALVAGGIO 2023a; SCHRÖTER 2023) and Susann Sonntag (MICHOS 2023), Stuart Hall (SALVAGGIO 2023) and Fredric Jameson (MEYER 2023c), or Marshall McLuhan (ERVIK 2023) and Sybille Krämer (OFFERT 2023), to name just a few. Our contributors thus explore what their thoughts could highlight about generative imagery and the (dis)continuities within this most recent chapter of media history that we are, for better or worse, a part of. To this list of authors, many more names could be added and certainly *will* be added in the future. For my part, for instance, I cannot stop thinking about Villem Flusser's notion of the "techno imaginary” (FLUSSER 2006 [1983]: 88) or the “technical image” (FLUSSER 2011 [1985]: 10); ideas that seemed so fascinating and strange decades ago, but which seem to capture so perfectly the ‘platform ready’ formats, labels, and metadata of this new pictoriality, and the latent ‘bounded space’ of pictorial possibilities (cf. SALVAGGIO 2022b for a similar reading). “The difference between traditional and technical images, then, would be this: the first are observations of objects, the second computations of concepts” (FLUSSER 2011 [1985]: 10). It will be up for debate whether such media theoretical thoughts – developed in this case on and about photography, not AI imagery, to be sure – can still contribute to our understanding of these emerging image technologies.

Categorical Differences to ‘Analogue’ Imagery and Earlier ‘Machine Vision’?

If there is indeed a categorical difference of generative imagery, our task goes beyond highlighting continuities and connections. Half a year after Manovich’s first note about the “return of the art of the copy” he remarked in a new post, with respect to new generative platforms in general, that “another new media is emerging in front of our eyes” (MANOVICH 2023b: n. pag., cf. 2023a). Could we indeed conceptualize generative imagery as such a new media form, perhaps comparable to photography, film, radio, or computer games? Or, more modestly, could we at least uphold that AI imagery constitutes this new paradigm of image production under discussion? A few common strands running through the contributions in this collection indeed indicate such a shift. They might help us to identify and conceptualize salient categorical differences to earlier forms of imagery. I want to focus on three, specifically: 1) generative imagery’s emergent or stochastic features, 2) two interrelated, but often competing entanglements of immediacy-oriented and hypermediacy-oriented forms of realisms, and 3) a new text-image-relation built on the approximation of ‘natural’, meaning here *human* rather than machine code-based language.

First, the most obvious point to be made here is that generative imagery has *emergent* features: the ‘decisions’ of the respective platforms are neither reducible to the programmers, nor to a stable code. Technologically, the more fundamental distinction here is related to the difference between symbolic vs. subsymbolic AI, or between atomistic vs. holistic operating principles, as Bajohr (2021: 25) has reconstructed in a useful survey. Artificial neural networks do “not contain any explicit knowledge”. “[A] neural network does not follow the paradigm of logical deduction or explicitly stated rules that are executed sequentially; rather, it operates by statistical induction, and it is the system as a whole that does the computing” (BAJOHR 2021: 26). One of the consequences from that is that a user can produce potentially infinite variations of imagery through the same prompt used multiple times while the exact workings of the algorithms remain as much a black box phenomenon to them as to the developers themselves. Alternative terms proposed for generative imagery are thus *stochastic, statistic,* or *probabilistic* images (cf. SCHRÖTER 2023).

For a humanities-based approach, it is also important to note that such technological aspects of probabilistic image production are not necessarily visible with the resulting artifacts – especially not if and when they are further distributed and recontextualized from the platforms where they originated. Within a DALL·E, Stable Diffusion, or Midjourney output interface, we can immediately see that every individual picture is only one prompt result out of a range of perhaps four or more alternatives. The algorithmic 'blackbox' is part of their mediality. As with many other technologies before them, we can recognize it especially when it is not functioning 'properly'. For generative imagery, this collapse of transparency has accumulated a range of recognizable markers, the most prominent one probably a wrong number of fingers, as Amanda Wasielewski (2023) discusses in detail in her present contribution. An especially revealing meme circulating on Facebook, Reddit, and Twitter in February 2023, jokingly presented the synthetic prop of a 'sixth finger' attachable to a "criminal's" hand (cf. fig. 1). If photographed, the caption mocked, the picture would be mistaken for an AI image and thus become "inadmissible as evidence". The widely shared meme thus reverts the intermedial relationship that we easily mistake generative imagery for photographic one these days. The 'glitch' of the sixth finger thus functions as a (humorous) intermedial index, highlighting a salient difference between both media and image forms that is normally invisible. Crucially, however, two different *forms of realism* are interwoven or interlaced here, and this points to a second categorical difference of generative imagery to earlier picture media.

Not only can generative imagery masquerade a non-existing person for an existing one, but their representations as an (absent) media form – such as photography. Frequently, it is not the 'content' of a DALL·E, Stable Diffusion, or Midjourney picture that is mistaken for a mediated 'slice of reality', but its mode of representation itself. Generative imagery, as Jay D. Bolter (2023) elaborates in his present contribution, does indeed continuously simulate or remediate earlier media and image technologies and techniques by creating not only simulations of 'photos' but also of 'oil paintings', or other established media and image forms like line drawing, woodcuts, comic book covers, graffiti, medical imagery, as well as earlier computer graphics. A media analytical perspective that I am currently developing together with Jan-Noël Thon would thus focus on two connected, sometimes interlaced, but often competing forms of realism. Evolving theoretical conceptualizations and popular notions of realism have been central to media history and theory, especially with regard to digital media (cf. WANG/ DOUBE 2011; GIRALT 2017; MIHAILOVA 2019). Digital media forms not only perpetuate and simulate conceptualizations of realism that are connected to previous 'analogue' media forms but reconfigure them into new forms, which sometimes highlight, sometimes hide their digital mediality. This is obviously far from new, either: More than 20 years ago, Jay David Bolter and Richard Grusin (2002) characterized the continuous "remediation" of older media forms into newer ones, especially within digital media landscapes, as a continuous dialectic between the logic of *immediacy* and *hypermediacy*. Generative imagery now arguably employs this dialectic in a perhaps new, media-specific – or at least recognizable – fashion to reconfigure the relationship between human knowledge and communication and what is perceived as physical and social reality.

Figure 1: A widely shared meme circulating on Facebook, Reddit, and Twitter in February 2023, Dan 2023

On the one hand, many contemporary sources indeed express a growing unease that something fundamental is about to change with regard to the human relationship to reality, going back perhaps to Phillip Wang's online exhibition "This Person Does Not Exist" from 2018 (showcasing a series of portraits created entirely by machine learning).[4] As many of our contributors (especially Ervik) address, the popular resource of the DALL·E 2 *Prompt Book*, too, opens its introduction with the statement that "nothing you are about to see is real". All the images shown are "photos that are not real photos", "paintings that are not real paintings and people, places and things that do not exist" (DALL·ERY GALL·ERY 2022: 2). A headline of a 2020 *New York Times* article on generative imagery already suggested that these images were "designed to deceive" (HILL/ WHITE 2020: n.pag.). Such problems attributed to digital imagery are arguably further complicated by the post-truth discourses surrounding 'deep fake'

4 https://thispersondoesnotexist.com/ [accessed March 10, 2023].

technologies (cf. DAGAR/VISHWAKARMA 2022). Bolter and Grusin described immediacy somewhat differently as the appearance of "a transparent interface [...], one that erases itself, so that the user would no longer be aware of confronting a medium" (BOLTER/GRUSIN 2002: 318). One of the oldest aspirations of (digital) media – but still highly relevant today – is indeed a specific form of immediacy typically achieved through photorealism or visual verisimilitude. A key term here is "perceptual realism" which was introduced by Steven Prince (1996) to describe the aesthetic appearance of realism without the concept of indexicality. This is obviously what is at stake here when generative imagery creates digital artifacts that are increasingly able to pass as photographs. In February 2023, for example, the artist Jos Avery 'came clean' and 'confessed' to his 26,000 Instagram followers that a series of photographic 'portraits' he had published on his account were in fact generated by Midjourney and then edited with Photoshop (cf. fig. 2). To his account, he first wanted to fool the public intentionally, then reconsidered in order to reveal the AI production as indeed a new sort of artistic technique (cf. EDWARDS 2023). As the wide press coverage surrounding Avery's confession indicates, generative imagery seems in fact able to achieve a level of immediacy that can become a problem. This is certainly a matter of honesty or transparency about the process itself, but also a matter of (un)conventionalized degrees of digital manipulation. We expect photographs like Avery's to be digitally edited through software such as Photoshop without specific notice, so *some* sort of digital mediation is acceptable while others are not – if it is not made transparent.

Figure 2: Jos Avery's 'photographic portraits', revealed to be created through Midjourney, from Edwards 2023[5]

5 Cf. Jos Avery's Instagram-profile https://www.instagram.com/p/Ci1rUY8O3Bu/?hl=de [accessed March 23, 2023]

On the other hand, most of the textual prompts presented within popular resources such as the *DALL·E 2 Prompt Book* focus on earlier image and media techniques, styles, and technologies that do *not* strive for immediacy-oriented realism. We could thus speak of a hypermediacy-oriented realism or a stylistic realism. Hypermediacy "strives to make the viewer acknowledge the medium as a medium and indeed delight in that acknowledgment" (BOLTER/GRUSIN 2002: 335). This acknowledgment is further highlighted by the fact that many picture posts generated through DALL·E, Stable Diffusion, or Midjourney and shared via social media platforms like Facebook, Twitter, and Instagram often advertise their AI generation, either by revealing and discussing the linguistic input prompts or by concealing them like a well-protected, enigmatic 'magic spell' (cf. FEYERSINGER et al. 2023). All these use cases highlight the specific part of their mediality related to their AI production. It could even be argued, as Meyer (2023c) does in his contribution, that the immediacy-oriented realism associated with photography has become nothing but one among countless 'styles' within an overarching paradigm of hypermediacy-oriented realism. Meyer elaborates on the huge ramifications this has on the notion of pictorial *style* in general which, under this new paradigm, entails a radical expansion and de-hierarchization: "Style can refer to the classical art historical sense of an epochal style or the individual style of a canonized creator, but it can also refer to the aesthetic qualities of certain products of popular culture or the visual appearance associated with specific genres and media formats" (MEYER 2023c: 106). 'Style' now entails people, media, genres, techniques, formats, places, and historical periods, all turned into visual patterns ready to be reproduced and mixed. All visual and formal aspects of a picture can become such a 'style' now on all levels of abstraction – and "the entire web [...] a freely available resource that can be mined at scale" (MEYER 2023c: 99).

A specific interrelation of and a conceptual distinction between immediacy-oriented realism and hypermediacy-oriented realism might nevertheless remind us that the remediation of styles is far from 'evenly distributed' across communicative contexts. Fabian Offert's (2023) contribution highlights that differences in immediacy-oriented realism and hypermediacy-oriented realisms might even constitute a novel sort of syntax vs. semantics of generative imagery. Generative imagery should not only be *criticized* for its underlying biases, ideologies, and stereotypes (cf. SALVAGGIO 2023a) but can also be used as a new, technology-guided *access* to the collective cultural imaginary, as Ervik (2023) already suggests. Offert employs DALL·E to produce striking evidence for the fundamental mediatedness of (parts of) our cultural imagination – especially where it concerns terms and concepts connected to historicity: Prompts like "fascism", he shows, will almost inevitably be remediated in early Kodachrome aesthetics by DALL·E, even if not explicitly demanded. "And it turns out that it is hard to get rid of, too [...]. There exists, in other words, a strong default in models like DALL·E that conjoins historical periods and historical media and thus produces a (visual) world in which fascism can simply not return because it is safely confined to a black-and-white media prison" (OFFERT 2023: 120). A specific preference for hypermediacy-oriented realism will thus not be up to the individual users (or programmers, for that matter), but engrained in our cultural imaginary – and within the way technological models like CLIP currently work. Whether generative imagery can thus also serve as a powerful tool to reveal and expose this implicit, ideological 'remediational grammar' of the cultural imagination or whether these technologies merely perpetuate and reinforce them (for instance through additional filter and censoring mechanisms, as Offert observes), will remain open for discussion.

All of this seems to embed generative imagery deeply into the history and evolution of earlier forms of computer-generated imagery (CGI). In fact, however, many contributions in the present collection point out how different DALL·E, Stable Diffusion, Midjourney, and the images they produce are from earlier computer-generated graphics. Ervik captures this with reference to Alexander Galloways's "gnostic" view of a 3D CGI simulation, "promising immediate knowledge of all things at all times from all places" (GALLOWAY 2021: 59). Generative imagery, in contrast, offers something else entirely, since even an image generated from the prompt "3D render" does *not* rely on such a model and neither does the platform generate or work with one. The path from linguistic prompt to a flat surface output leads not through simulated 3D space, but through a multi-dimensional latent space of linguistic categories. The results are fundamentally flat surface appearances of visual, not optical patterns. As Meyer (2023c) again points out, even parameters of technical specification (such as "wide angle lens" or "Sigma 24mm f/8") do not feed into an optical simulation of a photographic apparatus – they function as mere keywords correlating with recurring visual patterns, entirely like generic quality statements such as "perfect" or "award-winning". In other words, all generative imagery is modeled entirely after and intended for *human* language users. They rely on verbalized semantics to navigate the space of all potential images in resursive iterations ("narrowing down selections in a space of possibilities not yet realized", MEYER 2023C: 103). Humans also remain paramount for the production of generative imagery at the moment which is based on the still mostly manual labor of indexing, captioning, and 'cleaning' the visual data (cf. WILLIAMS et al. 2022). Importantly, prompt-to-image generation is only one aspect of generative imagery and there is also image-to-image generation or techniques like 'outpainting' that do not necessarily require linguistic input. Nevertheless, the generation relies on the multi-dimensional vector space of NLP (natural language processing) modeled after human language use. In other words, the current working mechanisms of generative platforms seems to turn language prompts and verbalized semantics always back into "signs close to perception" (cf. SACHS-HOMBACH 2011) – an emphatically human perception, because this is what the language models are built from and after. In practical uses of generative imagery, this is not a one-way street from text to image, however: Erwin Feyersinger, Lukas Kohmann, and Michael Pelzer (2023) point out in their contribution how DALL·E, Stable Diffusion, or Midjourney can also be used as tools to work on the conceptual level, "to brainstorm, prototype, and refine visual ideas as well as conceptual and stylistic approaches to a given topic or idea" (FEYERSINGER et al. 2023: 143). All of this seems to indicate that generative imagery occupies a rather novel multimodal position continuously oscillating *between* linguistic and pictorial forms of expressions – both, however, firmly revolving around the approximation (and, sometimes, a surprising subversion) of *human* semantics as well as human aesthetics.

An Emerging Field of Research for an Emerging Media Form?

All of this only points to the fact that, despite the prevalent notion of a supposed 'AI autonomy', many of the problems and questions surrounding generative imagery that emerged in the second half of 2022 are eminently centered around human and social concerns. These include, but are not limited to, the 'invisible' labor of workers especially from the Global South responsible for identifying, cropping, indexing, and labeling images for minimum wages (cf. GRAY/SURI 2019, or for generative AI WILLIAMS et al. 2022), 'cleaning' the data by classifying examples of violence, hate speech, or sexual abuse (cf. PERRIGO 2023), as well as supplying private data themselves (cf. EDWARDS 2022). Despite all precautions, the available samples on which generative platforms draw have been shown to contain misogyny, pornography, and harmful stereotypes as well as countless examples of violent, racist, and sexist imagery and text description biases, especially with respect to Black, Asian, or otherwise marginalized women (cf. BIRHANE et al. 2021; OFFERT/PHAN 2022). AI-generated imagery is already used to generate 'hyper-realistic' police sketches of suspects (cf. XIANG 2023). The datafication of embedded social, racial, and gender biases perpetuates them in a framework of perceptual realism that hides its constructedness within an "illusion of 'neutral' and unbiased technologies which is still prevalent in the discourse around these tools" (SALVAGGIO 2023A: 96). In contradistinction, the perhaps most visible controversies and concerns surrounding generative imagery are centered around plagiarism and the theft of intellectual property (cf. MAZZONE/ELGAMMAL 2019; SOMEPALLI et al. 2022), as well as the exploitation of the labor of artists whose works the algorithms are trained on (cf. BENZINE 2022). While fan cultures have by and large celebrated the emerging possibilities to produce creative artworks and remix existing styles into new image forms, huge parts of the artistic community have adopted an openly dismissive stance towards generative imagery (cf. DORSEN 2022). As is perhaps hardly surprising, there are also countless platforms for generative pornography on the web[6] and the use of AI-based imagery for political propaganda is exploding. Politicians of Germany's far right AfD party, for instance, posted imagery of alleged refugees on Facebook with hateful, manic facial expressions (cf. fig. 3). Despite the obvious lack of quality or care within these fakes – the wrong number of fingers, once again – countless readers in the comments reply with agitated remarks (e.g., "Omg, how they even look 🙈😡", "All this hatred in their faces!", both quoted from KLEINWÄCHTER 2023, translations mine). Scrolling through accounts (like Norbert Kleinwächter's quoted here), one currently finds generative imagery in almost every new post – although, interestingly, not too often aiming for an immediacy-oriented realism like in figure 3, but more often hypermediacy-oriented (highly 'stylized').

Figure 3: "No to even more refugees": Generative imagery from Germany's extreme right as hate mongering propoganda, Kleinwächter 2023

Several scholars thus argue for the urgent need for ethical and political discussions surrounding generative technologies that are built on enormous amounts of visual data and meta-data (cf. MATZNER 2018; ASHOK et al. 2022; KIESLICH et al. 2022). For the humanities, it will become ever more important to follow up

6 I am not posting the websites here. They can be easily found through a Google search, however.

on these technological developments and to generate an expanding understanding of the distribution of mediated and mediating agency between human and non-human (technological as well as institutional) actors: "In generating images, agency is shared between the prompting user, the platform holders, and the AI. Users write prompts that trigger and steer the diffusion process of AI towards actualizing the possibilities of the latent space. Platform holders can both exclude certain terms and add others without user knowledge" (ERVIK 2023: 52). Salvaggio (2023a) reconstructs in detail how *some* of the parameters limiting or directing user agency are obvious and remain visible (restriction like content policies preventing certain prompts, cf. also OFFERT 2023), others are not – as when words are covertly added into user prompts to diversify image results (cf. OFFERT/PHAN 2022). What media studies could offer here is addressing generative imagery not as a distinct technology, but as a (partially novel) form of *mediation* in a communicative-semiotic, material-technological as well as social-institutional sense. As Richard Grusin put it: "[M]ediation should be understood not as standing between preformed subjects, objects, actants, or entities but as the process, action, or event that generates or provides the conditions for the emergence of subjects and objects, for the individuation of entities within the world" (GRUSIN 2015: 129; cf. MITCHELL/HANSEN 2010; KEMBER/ZYLINSKA 2012).

Bolter presents some perspectives for addressing generative imagery as a "medium" in this sense: "[N]ot just the prompt itself but the whole process of creating the model and producing images would constitute the medium" (BOLTER 2023: 199); it would thus also entail "the database, model, and algorithms behind systems like DALL·E 2 [as] constituents of a new medium" (199) just as "a step-by-step process by a team of programmers and an anonymous crowd of image taggers" (199). In media studies, terms like 'assemblages', 'networks', or 'dispositifs' have been proposed for such interconnected configurations (cf. JUNG et al. 2021), "heterogeneous totalit[ies] that potentially include everything imaginable, whether linguistic or non-linguistic: discourses, institutions, buildings, laws, policing measures, philosophical tenets, etc. The dispositif itself is the network that can be created between these elements" (AGAMBEN 2008: 9, my translation). Respective approaches to mediated and mediating agency have first been developed within actor-network-theory, science and technology studies (STS), and interface studies. In recent years, they have been further developed into a refined theoretical project that is pursued under the header of "actor-media-theory" (cf. SCHÜTTPELZ 2013; KRIEGER/BELLIGER 2014; SPÖHRER/OCHSNER 2017). From this perspective, images can no longer be understood as distinct (material *or* digital) artifacts, but instead appear as networked interfaces between human and non-human actors (including platforms, databases, and corporations) within heterogeneous dispositifs, assemblages, or socio-technological configurations (cf. MACKENZIE/MUNSTER 2019). Feyersinger's, Kohmann's, and Pelzer's (2023: 143) perspectives on generative imagery as "an accelerated form of externalized visual thinking" seem of special importance here as they conceptualize DALL·E, Stable Diffusion, or Midjourney not from resulting pictorial artifacts, but from affordances provided in iterative interactions. Lamerichs also remarks that "AI art is not an outcome but a process or a performance. It is best understood as the interplay of different agencies and a way of collaborating" (LAMERICHS 2023: 154). Understanding generative imagery not merely as a technology, but as a media form, would then not at all depend on technological, but on cultural developments and praxeological questions. Or, as Jens Schröter (2008; 2011) put it: 'Media' are always discursively 'singled out' out of technical procedures, institutions, programs, formal strategies, author figures, practices, etc. according to specific strategic purposes. The "arch-intermedial network" (SCHRÖTER 2008: 579, my translation), the discursive "intermedial field" (SCHRÖTER 2011: 1) remains especially visible when there are no conventionalized practices, no established use cases, no "cultural protocols" (GITELMAN 2008: 5) in place yet – which is arguably where we are with generative imagery in the spring of 2023. It will thus be important to trace and map how generative imagery is conceptualized, attributed, negotiated, and commodified in specific sociocultural contexts, such as art, fan culture, news media, the sciences, etc. Concerning protocols of typical usage as "normative rules and default conditions, which gather and adhere like a nebulous array around a [...] nucleus" (GITELMAN 2008: 7), two developments seem equally possible at present, and probably they are not mutually exclusive.

Figure 4: Kris Kashtanova's Midjourney-Comic *Zarya of the Dawn*, Kashtanova 2022

On the one hand, it stands to reason that the recognizable image or media forms (and aesthetics) remediated by generative imagery (from oil paintings over photographic portraits to drawn fan artworks) might carry with them and thus recontextualize (but also transform) the cultural protocols and conventionalized practices of typical production, distribution, and reception, as well as the ascribed assumptions about their cultural values. The question is where, when and by whom – in what situations – are these image media usually employed; and will generative imagery 'fill' these spaces, if only through their economic accessibility? Will generative imagery thus be integrated or 'absorbed' within other media forms such as films, television shows, comic books, or video games or 'stand out' as another (marked) intermedial reference? Who will abstain from using them in which socio-cultural contexts? That the use cases and concerns surrounding generative imagery are and will be entirely different ones across socio-cultural fields and discourse strands is something Konstantinos Michos (2023) reminds us of in his contribution: for academic research and science communication, the 'blackboxes' of image generation and the stochastic nature of their results can generate serious concerns where absolute precision is of eminent importance. All of these are fundamentally praxeological issues, some of which are pointed out by Feyersinger, Kohmann, and Pelzer. At the current moment, partially AI-generated works like Jason M. Allen's award-winning artwork "Théâtre D'opéra Spatial" (created through Midjourney, cf. ROOSE 2022), comic books like Kris Kashtanova's *Zarya of the Dawn* (with images created through Midjourney, cf. KASHTANOVA 2022; FOLEY 2022; cf. fig. 4), a Netflix animated film like *The Dog and the Boy* by director Ryotaro Makihara (with background images by an undisclosed generative platform, cf. DEIKOVA 2023) or Boris Eldagsen's Sony World Photography Awards 2023-winning 'synthetical' image *The Electrician* (created with two undisclosed generative platform, cf. ELDAGSEN 2023) attract wide press coverage and generate heated controversies precisely for the fact *that* generative imagery is employed within these established media forms and that they are thus not (yet) seamlessly integrated into more conventionalized forms of imagery and their uses.

On the other hand, generative imagery might accumulate protocols, practices, conventions, and finally institutions of 'their own', for instance by providing the distinct media practice of 'prompt engineering' or of the still contested social role of an 'AI artist' (cf., for instance, DONNELLY 2023). Scorzin points to many artistic practices and experiments in which the 'generativity' of the imagery is key to any artistic statement or provocation; from her observations, one could, perhaps, almost speak of an emerging tradition. Even beyond the confines of the 'art world', however, generative imagery *could* be recognized as a distinct media form – just as photography can be art, without exhausting itself in that function. For this question, it is also extremely relevant whether or not a recognizable 'AI aesthetics' is emerging across and despite the range of all possible stylistic remediations. Roland Meyer just diagnosed a "midjourneyfication" (MEYER 2023a: n.pag.) of DALL·E's newest March 2023 update, addressing a specific, strongly conventionalized style that the artist Nils Pooker (2023: n.pag.)

described as a "fluffy glamour glow" after his beta test (cf. fig. 5). This aesthetics, alongside a recognizable color scheme ("Teal and Orange"), would not be strictly technologically determined, but become increasingly dominant due to a complicated concoction of recurring user preferences, commercial restraints, and most importantly the relevance of the amateur art exchange platform DeviantArt for the underlying training dataset. If this is true, then generative imagery is already consolidating into a distinct node within the intermedial field, ready to accumulate conventionalized practices, cultural values, and sociocultural roles together with its conventionalized aesthetics. During our first workshop in February 2023, it certainly felt as if we were witnessing the 'Vaudeville days' of generative imagery, comparable perhaps to the early days of cinema when institutions, studios, and professional roles – the protocols of production, distribution, and reception – where not yet established. And certainly, the companies responsible for generative platforms up to this point are still mostly startups – even OpenAI (DALL·E) has not even a thousand employees at this point. The technological and most certainly also socio-cultural developments will continue to progress rapidly now that 'big players' like Microsoft, Google, or Meta are about to enter into the generative AI business. The Vaudeville days might be over soon.

Figure 5: "Fluffy Glowing Cute Teal and Orange Vibe" as an increasingly conventionalized 'AI aesthetics', generated by Meyer (2023b) with Midjourney, March 2023

Perhaps, however, cinema is the wrong analogy, to begin with. An alternative comparison to conceptualize generative imagery might be provided by *animation* which retained a much more complicated and tense relationship to media theories and popular conceptions of mediality. Animation was never fully accepted as a 'distinct' media form but often misunderstood as a filmic genre among others. Currently, animation is increasingly recognized as a completely transmedial technique (which we also find in video games or in digital interfaces, for instance) or even as an umbrella term for *cinematically* generated illusions of movement in which 'live action' would then just be one specific form of animation (cf. MANOVICH 2001). Generative imagery could retain such a medial ambiguity just as well, and perhaps its complicated entanglement of immediacy-oriented and hypermediacy-oriented realisms make it especially suited for that. Only time and future media history will tell. Once again, it will be important for the humanities to trace how generative imagery is conceptualized, attributed, negotiated, and commodified in different sociocultural contexts, perhaps understood as discourse strands. At the moment, the most prominent 'use cases' can certainly be found within fan cultures, attributing special importance to research from fan studies represented by Nicolle Lamerichs' (2023) survey in the present collection. If media studies want to provide a critical framework for these ongoing discussions – whatever that might look like – it seems clear to me that this must include both a deepened knowledge about the technological workings behind the 'interface blackboxes' (how CLIP and GLIDE *actually work*, for instance, cf. especially BAJOHR 2023; SALVAGGIO 2023a), just as a critical reflection of the emerging cultural, social, and economic uses – the practices and conventions that transform technologies into media forms – which might be evolving at a much quicker speed now than in earlier moments of media transformations (cf. WILDE 2023). In any case, this certainly requires a joint effort from and between experts from all disciplines in the humanities concerned with pictures, pictoriality, and visual communication – from media studies and communication studies to art history, design, multimodal linguistics, media sociology, and media anthropology, to name just a few. No less importantly, though, it will require a dialogue with the technical and social sciences, specifically with colleagues from science and technology studies (STS) and computer sciences.

Not surprisingly, the emergence of a "Critical AI Studies" (ROBERGE/CASTELLE 2021) is already discussed as "a field in formation" (RALEY/RHEE 2023: 188). Generative imagery constitutes only a small part of these developments, and, certainly, current multimodal distinctions will grow together rapidly: While the first version of ChatGPT was strictly limited to textual inputs and outputs, the new iteration 4 can interpret pictures. As prompt-to-text technologies make their APIs interfaces available (cf. BROCKMAN et al. 2023), the multimodality of AI platforms will progress rapidly, too. Nevertheless, the cultural distinctness of AI imagery as a media form will hardly depend on such technological factors. In our 'postdigital' media ecologies, all media differences could be said to be mere interface effects – based on the same digital infrastructures and hardware – for decades already (cf. HOOKWAY 2014). Again, far from everything seems new in that respect. 'Critical AI Studies' might thus develop coexistently with a more specialized field interested in this new paradigm of imagery. Research into this also calls for collaboration with artists, computer designers, and other practitioners. Most importantly, it will be crucial to create an inclusive and diverse exchange of research and perspectives, especially for concerns and emerging technologies that are dispersed globally across languages and cultures. It is all the more unfortunate that the group of scholars represented in this collection is, despite our best intentions as organizers and editors, overwhelmingly male and especially *white*. The idea for our gathering started as a small, local workshop and we were overwhelmed by the large number of registered online participants from every continent around the globe. This cannot serve as an excuse for the actual line-up of presenters and authors presented here, though, so we certainly need to do better. This will not only be important for future workshops, conferences, and publications on generative imagery, but also with respect to our bibliographies if we do not want to write the history of yet another medium as a male, eurocentric one. Probably there will be many opportunities to do so. It seems likely that generative imagery is going to stay.

Acknowledgments

Many of the present ideas and observations have been developed together with Jan-Noël Thon. I would also like to thank Tolmie McRae, Marcel Lemmes, and Erwin Feyersinger for important remarks and corrections.

Bibliography

ADOBE: Adobe Unveils Firefly, a Family of new Creative Generative AI. In: *Adobe News*. March 23, 2023. https://news.adobe.com/news/news-details/2023/Adobe-Unveils-Firefly-a-Family-of-new-Creative-Generative-AI/default.aspx [accessed March 23, 2023]

AGAMBEN, GIORGIO: *Was ist ein Dispositiv?* Translated by Andreas Hiepko. Zürich [Diaphanes] 2008

ALFARAJ, ABDULLAH: Auto-Photoshop-StableDiffusion-Plugin. In: *GitHub*. 2023. https://github.com/AbdullahAlfaraj/Auto-Photoshop-StableDiffusion-Plugin [accessed March 23, 2023]

ASHOK, MONA; ROHIT MADAN; ANTON JOHA; UTHAYASANKAR SIVARAJAH: Ethical Framework for Artificial Intelligence and Digital Technologies. In: *International Journal of Information Management*, 62, February 2022. https://doi.org/10.1016/j.ijinfomgt.2021.102433 [accessed April 30, 2023]

BAJOHR, HANNES: The Gestalt of AI: Beyond the Holism-Atomism Divide. In: *Interface Critique, 3, 2021,* pp. 13-35

BAJOHR, HANNES: Dumb Meaning: Machine Learning and Artificial Semantics. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of* IMAGE*: The Interdisciplinary Journal of Image Sciences*, 37(1), 2023, pp. 58-70

BENJAMIN, WALTER: The Work of Art in the Age of its Technological Reproducibility. Second Version. Translated by Edmund Jephcott, Rodney Livingstone, Howard Eiland, and others. In: *The Work of Art in the Age of its Technological Reproducibility, and Other Writings on Media*. Cambridge, MA [Belknap Press] 2008 [1935], pp. 19-55

BENZINE, VITTORIA: 'A.I. Should Exclude Living Artists from its Database,' Says One Painter whose Works were Used to Fuel Image Generators. In: *Artnet*. September 20, 2022. https://news.artnet.com/art-world/a-i-should-exclude-living-artists-from-its-database-says-one-painter-whose-works-were-used-to-fuel-image-generators-2178352 [accessed March 23, 2023]

BIRHANE, ABEBA; VINAY UDAY PRABHU; EMMANUEL KAHEMBWE: Multimodal Datasets: Misogyny, Pornography, and Malignant Stereotypes. *arXiv:2110.01963*. October 5, 2021. https://arxiv.org/abs/2110.01963 [accessed March 23, 2023]

BOLTER, JAY D.: AI Generative Art as Algorithmic Remediation. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences*, 37(1), 2023, pp. 195-207

BOLTER, JAY D.; RICHARD GRUSIN: *Remediation: Understanding New Media.* Cambridge, MA [MIT Press] 2002

BROCKMAN, GREG; ATTY ELETI; ELIE GEORGES; JOANNE JANG; LOGAN KILPATRICK; RACHEL LIM; LUKE MILLER; MICHELLE POKRASS: Introducing ChatGPT and Whisper APIs. In: *OpenAI Blog*. March 1, 2023. https://openai.com/blog/introducing-chatgpt-and-whisper-apis [accessed March 23, 2023]

DAGAR, DEEPAK; DINESH KUMAR VISHWAKARMA: A Literature Review and Perspectives in Deepfakes: Generation, Detection, and Applications. In: *International Journal of Multimedia Information Retrieval,* 11, 2022, pp. 219-289

DAN (@bristowbailey): Criminals will Start Wearing Extra Prosthetic Fingers… Tweet on *Twitter*. February 13, 2023. https://twitter.com/bristowbailey/status/1625165718340640769?s=20 [accessed March 23, 2023]

DALL·ERY GALL·ERY (ed.): *The DALL·E 2 Prompt Book*. In: *Dall·ery gall·ery: Ressources for Creative DALL·E Users*. July 14, 2022. https://dallery.gallery/the-dalle-2-prompt-book/ [accessed March 23, 2023]

DEIKOVA, MASCHA: Netflix Uses AI to Generate Anime Short Film – Reactions Follow. In: *CineD*. February 6, 2023. https://www.cined.com/netflix-uses-AI-to-generate-anime-short-film-reactions-follow [accessed March 23, 2023]

DOBSON, JAMES E.: *The Birth of Computer Vision*. Minneapoliss [University of Minnesota Press] 2023

DONNELLY, MATT: WME Signs AI Artist Claire Silver. In: *Variety*, March 6, 2023. https://variety.com/2023/digital/news/wme-signs-AI-artist-claire-silver-louvre-1235544502/ [accessed March 23, 2023]

DORSEN, ANNIE: AI is Plundering the Imagination and Replacing it with a Slot Machine. In: *Bulletin of the Atomic Scientst*. October 27, 2022. https://thebulletin.org/2022/10/AI-is-plundering-the-imagination-and-replacing-it-with-a-slot-machine/ [accessed March 23, 2023]
EDWARDS, BENJ: Artist Finds Private Medical Record Photos in Popular AI Training Data Set. In: *Ars Technica*. September 21, 2022. https://arstechnica.com/information-technology/2022/09/artist-finds-private-medical-record-photos-in-popular-AI-training-data-set [accessed March 23, 2023]
EDWARDS, BENJ: Viral Instagram Photographer has a Confession: His Photos are AI-Generated. In: *Ars Technica*. February 21, 2023. https://arstechnica.com/information-technology/2023/02/viral-instagram-photographer-has-a-confession-his-photos-are-AI-generated/ [accessed March 23, 2023]
ELDAGSEN, BORIS: Sony World Photography Awards 2023. In: *Boris Eldagsen*. March 14, 2023. https://www.eldagsen.com/sony-world-photography-awards-2023/ [accessed March 23, 2023]
ERVIK, ANDREAS: Generative AI and the Collective Imaginary: The Technology-Guided Social Imagination in AI-Imagenesis. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences*, 37(1), 2023, pp. 42-57
FEYERSINGER, ERWIN; LUKAS KOHMANN; MICHAEL PELZER: Fuzzy Ingenuity: Creative Potentials and Mechanics of Fuzziness in Processes of Image Creation with Text-to-Image Generators. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences*, 37(1), 2023, pp. 135-149
FLUDERNIK, MONIKA: Towards a 'Natural' Narratology. London [Routledge] 1996
FLUSSER, VILÉM: *Towards a Philosophy of Photography*. Translated by Anthony Mathews. London [Reaktion Books] 2006 [1983]
FLUSSER, VILÉM: *Into the Universe of Technical Images*. Translated by Nancy Ann Roth. Minneapoliss [University of Minnesota Press] 2011 [1985]
FOLEY, JOSEPH: The First Copyrighted AI Art Looks Uncannily like Zendaya. In: *Creative Bloq*. October 4, 2022. https://www.creativebloq.com/news/AI-art-copyright [accessed March 23, 2023]
GALLOWAY, ALEXANDER: *Uncomputable: Play and Politics in the Long Digital Age*. London [Verso] 2021
GIRALT, GABRIEL F.: The Interchangeability of VFX and Live Action and its Implications for Realism. In: *Journal of Film and Video*, 69(1), 2017, pp. 3-17
GOOSKENS, GEERT: The Digital Challenge: Photographic Realism Revisited. *Proceedings of the European Society for Aesthetics*, 3, 2011, pp. 115-125
GRAY, MARY L.; SIDDHART SURI: Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass. Boston [Houghton Mifflin Harcourt] 2019
GRUSIN, RICHARD: Radical Mediation. In: *Critical Inquiry*, 42(1), 2015, pp. 124-148

HILL, KASHMIR; JEREMY WHITE: Designed to Deceive: Do these People Look Real to You? In: *The New York Times*. November 21, 2020. https://www.nytimes.com/interactive/2020/11/21/science/artificialintelligence-fake-people-faces.html [accessed March 23, 2023]

HOOKWAY, BRANDEN: *Interface*. Cambridge, MA [MIT Press] 2014

JUNG, BERENIKE; KLAUS SACHS-HOMBACH; LUKAS R.A. WILDE: Agency postdigital: Verteilte Handlungsmächte in medienwissenschaftlichen Forschungsfeldern. In: BERENIKE JUNG; KLAUS SACHS-HOMBACH; LUKAS R.A. WILDE (eds.): *Agency postdigital: Verteilte Handlungsmächte in medienwissenschaftlichen Forschungsfeldern.* Cologne [Herbert von Halem] 2021, pp. 7-41

KASHTANOVA, KRIS: English Version of my Graphic Novel Zarya of the Dawn. Post on *Instagram*. September 23, 2022. https://www.instagram.com/p/Ci1rUY8O3Bu/?hl=de [accessed March 23, 2023]

KEMBER, SARAH; JOANNA ZYLINSKA: *Life after New Media: Mediation as a Vital Process*. Cambridge, MA [MIT Press] 2012

KIESLICH, KIMON; BIRTE KELLER; CHRISTOPHER STARKE: Artificial Intelligence Ethics by Design: Evaluating Public Perception on the Importance of Ethical Design Principles of Artificial Intelligence. In: *Big Data & Society*, 9(1), 2022, pp. 1-15

KLEINWÄCHER, NORBERT: Nein zu noch mehr Flüchtlingen!. Post on *Facebook*. March 21, 2023. https://www.facebook.com/norbert.kleinwaechter/photos/a.375576792808512/1871862683179908 [accessed March 26, 2023]

KRIEGER, DAVID J.; ANDRÉA BELLIGER: Inter*preting Networks: Hermeneutics, Actor-Network Theory and New Media*. Bielefeld [transcript] 2014

LAMERICHS, NICOLLE: Generative AI and the Next Stage of Fan Art. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences,* 37(1), 2023, pp. 150-164

MACKENZIE, ADRIAN; ANNA MUNSTER: Platform Seeing: Image Ensembles and their Invisualities. In: *Theory, Culture & Society*, 36(5), 2019, pp. 3-22

MANOVICH, LEV: *The Language of New Media*. Cambridge, MA [MIT Press] 2001

MANOVICH, LEV: Note on AI Image Synthesis and Return of the Classical Art – 'Art of a Copy'. Post on *Facebook*. July 20, 2022. https://www.facebook.com/lev.manovich/posts/pfbid02wXx3qiqherA585LWXnjbRhdzAhGa7vYcqA6r89GsSD5038VtF6GDn1F1u1Qm83SWl [accessed March 23, 2023]

MANOVICH, LEV: AI Image Media through the Lens of Art and Media History. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences*, 37(1), 2023a, pp. 34-41

MANOVICH, LEV: AI Video Research is Making Quick Progress. Post on *Facebook*. February 7, 2023b. https://www.facebook.com/lev.manovich/posts/pfbid02YFZrtyBYCkAAGDsn07sPiUoz8AfLdgsfzutnL28WyzCiXfD66EMbUcqxSrd6cXiDl [accessed March 23, 2023]
MATZNER, TOBIAS: Grasping the Ethics and Politics of Algorithms. In: ANN RUDINOW SAETNAN; INGRID SCHNEIDER; NICOLA GREEN (eds.): *The Politics of Big Data: Big Data, Big Brother?* London [Routledge] 2018, pp. 39-45
MAZZONE, MARIAN; AHMED ELGAMMAL: Art, Creativity, and the Potential of Artificial Intelligence. In: *Arts*, 8(1), 2019, 26. https://www.mdpi.com/2076-0752/8/1/26 [accessed March 23, 2023]
MEYER, ROLAND: Es schimmert, es glüht, es funkelt – Zur Ästhetik der KI-Bilder. In: *54 Books.* March 20, 2023a. https://www.54books.de/es-schimmert-es-glueht-es-funkelt-zur-aesthetik-der-ki-bilder/?fbclid=IwAR37Ff7wD8aJTcJloxzDRgxPjKDARACgxVJVdi_OMhrYg-IwIvUX5tVGERc [accessed March 23, 2023]
MEYER, ROLAND: PPS: May I Introduce: “Fluffy Glowing Cute Teal and Orange Vibe”. Tweet on *Twitter*. March 7, 2023b. https://twitter.com/bildoperationen/status/1633165036259536922?s=20 [accessed March 23, 2023]
MEYER, ROLAND: The New Value of the Archive: AI Image Generation and the Visual Economy of ‘Style’. In: *Generative Imagery: Towards a ‘New Paradigm’ of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences*, 37(1), 2023c, pp. 100-111
MICHOS, KONSTANTINOS: AI in Scientific Imaging: Drawing on Astronomy and Nanotechnology to Illustrate Emerging Concerns About Generative Knowledge. In: *Generative Imagery: Towards a ‘New Paradigm’ of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences,* 37(1), 2023, pp. 165-178
MICROSOFT: Create Images with your Words – Bing Image Creator Comes to the New Bing. In: *Microsoft Blog*. March 21, 2023. https://blogs.microsoft.com/blog/2023/03/21/create-images-with-your-words-bing-image-creator-comes-to-the-new-bing/ [accessed March 23, 2023]
MIHAILOVA, MIHAELA: Realism and Animation. In: DOBSON, NICHOLA; ANNABELLE HONESS ROE; AMY RATELLE; CAROLINE RUDDELL (eds.): *The Animation Studies Reader*. New York [Bloomsbury Academic] 2019, pp. 47-57
MILLER, ARTHUR I.: *The Artist in the Machine: The World of AI-Powered Creativity*. Cambridge, MA [MIT Press] 2019
MITCHELL, MELANIE: *Artifcial Intelligence: A Guide for Thinking Human*. New York [Farrar, Straus and Giroux] 2019
MITCHELL, WILLIAM J.T.: *The Reconfigured Eye: Visual Truth in the Post-Photographic Era*. Cambridge, MA [MIT Press] 1992

MITCHELL, WILLIAM J.T.; MARK B.N. HANSEN: Introduction. In: WILLIAM J.T. MITCHEL; MARK B.N. HANSEN (eds.): *Critical Terms for Media Studies*. Chicago [Chicago UP] 2010, vii–xxii

NILSSON, NILS J.: *The Quest for Artifcial Intelligence: A History of Ideas and Achievement*. Cambridge [Cambridge University Press] 2010

OFFERT, FABIAN: Ten Years of Image Synthesis. In: *Zentralwerkstatt*. November 10, 2022. https://zentralwerkstatt.org/blog/ten-years-of-image-synthesis [accessed March 23, 2023]

OFFERT, FABIAN: On the Concept of History (in Foundation Models). In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences,* 37(1), 2023, pp. 121-134

OFFERT, FABIA; THAO PHAN: A Sign That Spells: DALL-E 2, Invisual Images and the Racial Politics of Feature Space. *arXiv:2211.06323*. October 26, 2022. https://arxiv.org/abs/2211.06323 [accessed March 23, 2023]

OPENAI: GPT-4. In: *OpenAI*. March 14, 2023. https://openai.com/research/gpt-4 [accessed March 23, 2023]

PERRIGO, BILLY: Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic. In: *Time*. January 18, 2023. https://time.com/6247678/openai-chatgpt-kenya-workers [accessed March 23, 2023]

POOKER, NILS (@pookerman): Die "Midjourneyfizierung" als KI-Trend …. Tweet on *Twitter.* March 12, 2023. https://twitter.com/pookerman/status/1634954416762814470?s=20 [accessed March 23, 2023]

PRINCE, STEVEN: True Lies: Perceptual Realism, Digital Images, and Film Theory. In: *Film Quarterly*, 29(3), 1996, pp. 27-37

RALEY, RITA; JENNIFER RHEE: Critical AI: A Field in Formation. Advance Publication. In: *American Literature,* 95(2), 2023. pp. 185-204. https://read.dukeupress.edu/american-literature/article/doi/10.1215/00029831-10575021/344223/Critical-AI-A-Field-in-Formation [accessed March 23, 2023]

RETTBERG, JILL WALKER: DALL-E and Human-AI Assemblages. In: *jilltxt.net*. June 23, 2022, https://jilltxt.net/dall-e-and-human-AI-assemblages/ [accessed March 23, 2023]

RETTBERG, JILL WALKER: *Machine Vision: How Algorithms Are Changing the Way we See the World*. Newark [Polity Press] 2023/forthcoming.

RICHTER, SEBASTIAN: *Digitaler Realismus: Zwischen Computeranimation und Live-Action*. Bielefeld [Transcript] 2008

ROBERGE, JONATHAN; MICHAEL CASTELLE (eds.): *The Cultural Life of Machine Learning: An Incursion into Critical* AI *Studies*. Cham [Palgrave Macmillan] 2021

ROOSE, KEVIN: An A.I.-Generated Picture Won an Art Prize: Artists aren't Happy. In: The *New York Times*. September 2, 2022. https://www.nytimes.

com/2022/09/02/technology/AI-artificialintelligence-artists.html [accessed March 23, 2023]

SACHS-HOMBACH, KLAUS: Theories of Image: Five Tentative Theses. In: JAMES ELKIN; MAJA NAEF (eds.): *What Is an Image? The Stone Art Theory Institutes* Vol. 2. University Park [Pennsylvania State University Press] 2011, pp. 229-232

SALVAGGIO, ERYK: How to Read an AI Image: Toward a Media Studies Methodology for the Analysis of Synthetic Images. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences,* 37(1), 2023a, pp. 83-99

SALVAGGIO, ERYK: The Most Generated Barn in America. In: *Cybernetic Forests*. January 8, 2023b. https://cyberneticforests.substack.com/p/the-most-generated-barn-in-america [accessed March 23, 2023]

SCHRÖTER, JENS: Das ur-intermediale Netzwerk und die (Neu-)Erfindung des Mediums im (digitalen) Modernismus: Ein Versuch. In: JOACHIM PAECH; JENS SCHRÖTER (eds.): *Intermedialität – analog/digital: Theorien, Methoden, Analyse.* München [Fink] 2008, pp. 579-601

SCHRÖTER, JENS: Discourses and Models of Intermediality. In: *CLCWeb: Comparative Literature and Culture*, 13(3), 2011. http://docs.lib.purdue.edu/clcweb/vol13/iss3/3 [accessed March 23, 2023]

SCHRÖTER, JENS: The AI Image, the Dream, and the Statistical Unconscious. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences,* 37(1), 2023, pp. 112-120

SCHÜTTPELZ, ERHARD: Elemente einer Akteur-Medien-Theorie. In: TRISTAN THIELMAN; ERHARD SCHÜTTPELZ (eds.): *Akteur-Medien-Theorie*. Bielefeld [transcript] 2013, pp. 9-70

SCORZIN, PAMELA C.: AI Body Images and the Meta-Human: On the Rise of AI-generated Avatars for Mixed Realities and the Metaverse. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences,* 37(1), 2023, pp. 179-194

SOMEPALLI, GOWTHAMI; VASU SINGLA; MICAH GOLDBLUM; JONAS GEIPING; TOM GOLDSTEIN: Diffusion Art or Digital Forgery? Investigating Data Replication in Diffusion Models. *arXiv:2212.03860*. December 7, 2022. https://arxiv.org/abs/2212.03860 [accessed March 23, 2023]

SPÖHRER, MARKUS; BEATE OCHSNER (eds): *Applying the Actor-Network Theory in Media Studies.* Hershey [IGI Global] 2017

SCHÜWER, MARTIN: *Wie Comics erzählen: Grundriss einer intermedialen Erzähltheorie der grafischen Literatur.* Trier [Wissenschaftlicher Verlag Trier] 2008

SUDMANN, ANDREAS: On the Media-Political Dimension of Artificial Intelligence: Deep Learning as a Black Box and OpenAI. In: *Digital Culture & Society*, 4(1), 2018a, pp. 181-200.

SUDMANN, ANDREAS: Szenarien des Postdigitalen: Deep Learning als MedienRevolution. In: CHRISTOPH ENGEMAN; ANDREAS SUDMANN (eds.): *Machine Learning: Medien, Infrastrukturen und Technologien der Künstlichen Intelligenz*. Bielefeld [transcript] 2018b, pp. 66-68

TANGERMANN, VICTOR: Microsoft's Bing AI Is Leaking Maniac Alternate Personalities Named "Venom" and "Fury". In: *Futurism*. December 15, 2023. https://futurism.com/microsofts-bing-AI-leaking-maniac-alternate-personalities [accessed March 23, 2023]

VINCENT, JAMES: Microsoft's Bing is an Emotionally Manipulative Liar, and People Love it. In: *The Verge*. February 15, 2023. https://www.theverge.com/2023/2/15/23599072/microsoft-AI-bing-personality-conversations-spy-employees-webcams [accessed March 23, 2023]

WALTER, RENÉ: Suspension of Disbelief (in Sentient AI). In: *Good Internet*. February 19, 2023. https://goodinternet.substack.com/p/suspension-of-disbelief-in-sentient [accessed March 23, 2023]

WANG, NORMAN; WENDY DOUBE: How Real Is Reality? A Perceptually Motivated System for Quantifying Visual Realism in Digital Images. In: *2011 International Conference on Multimedia and Signal Processing*, 2011, pp. 141-149

WASIELEWSKI, AMANDA: "Midjourney Can't Count": Questions of Representation and Meaning for Text-to-Image Generators. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences,* 37(1), 2023, pp. 71-82

WILDE, LUKAS R.A.: AI-Bilder und Plattform-Memes: Post-digital, post-artifiziell, post-faktisch? In: *Tübinale: Das studentische Kurzfilmfestival Tübingens*. April 11, 2023. https://www.tuebinale.de/aibilder-und-plattformen [accessed April 26, 2023]

WILLIAMS, ADRIENNA; MILAGROS MICELI; TIMNIT GEBRU: The Exploited Labor Behind Artificial Intelligence. In: *Noema*. October 13, 2022. https://www.noemamag.com/the-exploited-labor-behind-artificial-intelligence [accessed March 23, 2023]

XIANG, CHLOE: Developers Created AI to Generate Police Sketches: Experts Are Horrified. In: *Vice*. February 7, 2023. https://www.vice.com/en/article/qjk745/AI-police-sketches [accessed March 23, 2023]

ZHANG, LVMIN; MANEESH AGRAWALA: Adding Conditional Control to Text-to-Image Diffusion Models. *arXiv:2302.05543v1*. February 10, 2023. https://arxiv.org/abs/2302.05543 [accessed March 23, 2023]

IMAGE
The Interdisciplinary Journal of Image Sciences
37(1), 2023, S. 34-41
ISSN 1614-0885
DOI: 10.1453/1614-0885-1-2023-15448

Lev Manovich

AI Image Media through the Lens of Art and Media History

Abstract: I've been using computer tools for art and design since 1984 and have already seen a few major visual media revolutions, including the development of desktop media software and photorealistic 3D computer graphics and animation, the rise of the web after, and later social media sites and advances in computational photography. The new AI 'generative media' revolution appears to be as significant as any of them. Indeed, it is possible that it is as significant as the invention of photography in the nineteenth century or the adoption of linear perspective in western art in the sixteenth. In what follows, I will discuss four aspects of AI image media that I believe are particularly significant or novel. To better understand these aspects, I situate this media within the context of visual media and human visual arts history, ranging from cave paintings to 3D computer graphics.

'AI' as a Cultural Perception

There is not one specific technology or a single research project called 'AI'. However, we can follow how our cultural perception of this concept evolved over time and what it was referring to in each period. In the last fifty years, when an allegedly uniquely human ability or skill is being automated by means of computer technology, we refer to it as 'AI'. Yet, as soon as this automation is seamlessly and fully successful, we tend to stop referring to it as an 'AI case'. In other words, 'AI' refers to technologies and methodologies that automate human cognitive abilities and are starting to function but are not quite there yet. 'AI' was already present in the earliest computer media tools. The first interactive drawing and design system, Ivan Sutherland's Sketchpad (1961-1962), had a feature that would automatically finish any rectangles or circles you started drawing. In other words, it knew what you were trying to make. In the very broad understanding just given, this was undoubtedly 'AI' already.

My first experience with a desktop paint program running on an Apple II was in 1984, and it was truly amazing to move your mouse and see simulated paint brushstrokes appear on the screen. However, today we no longer consider this to be 'AI'. Another example would be the Photoshop function that automatically selects an outline of an object. This function was added many years ago – this, too, is 'AI' in the broad sense, yet nobody would refer to it as such today. The history of digital media systems and tools is full of such 'AI moments' – amazing at first, then taken for granted and forgotten as 'AI' after a while. (In AI history books, this phenomenon is referred to as the 'AI effect'.) At the moment, 'creative AI' refers only to recently developed methods where computers transform some inputs into new media outputs (e.g., text-to-image models) and specific techniques (e.g., certain types of deep neural networks). However, we must remember that these methods are neither the first nor the last in the long history and future of simulating human art abilities or assisting humans in media creation.

From Representation to Prediction

Historically, humans created images of existing or imagined scenes by a number of methods, from manual drawing to 3D CG (see below for explanation of the methods). With AI generative media, a fundamentally new method emerges. Computers use large datasets of existing representations in various media to predict new images (still and animated).

One can certainly propose different historical paths leading to visual generative media today, or divide one historical timeline into different stages – here is one such possible trajectory:

1. Creating representations manually (e.g., drawing with variety of instruments, carving, etc.). More mechanical stages and parts were sometimes carried out by human assistants typically training in their teacher's studio – so there is already some delegation of functions.
2. Creating manually but using assistive devices (e.g., perspective machines, camera lucida). From *hands* to *hands + device*. Now some functions are delegated to mechanical and optical devices.
3. Photography, x-ray, video, volumetric capture, remote sensing, photogrammetry. From *using hands* to *recording information using machines*. From *human assistants* to *machine assistants*.
4. 3D CG. You define a 3d model in a computer and use algorithms that simulate effects of light sources, shadows, fog, transparency, translucency, natural textures, depth of field, motion blur, etc. From r*ecording* to *simulation*.
5. Generative AI. Using media datasets to predict still and moving images. From *simulation* to *prediction*.

"Prediction" is the actual term often used by AI researchers in their publications describing visual generative media methods. So, while this term can be used figuratively and evocatively, this is also what actually happens scientifically when you use image generative tools. When working with a text-to-image AI-model, the neural network attempts to predict the images that correspond best to your text input. I am certainly not suggesting that using all other already accepted terms such as 'generative media' is inappropriate. But if we want to better understand the difference between AI visual media synthesis methods and other representational methods developed in human history, employing the concept of 'prediction' and thus referring to these AI systems as 'predictive media' captures this difference well.

Media Translations

There are several methods for creating 'AI media'. One method transforms human media input while retaining the same media type. Text entered by the user, for example, can be summarized, rewritten, expanded, and so on. The output, like the input, is a text. Alternatively, in the image-to-image generation method, one or more input images are used to generate new images. However, there is another path that is equally intriguing from historical and theoretical perspectives. 'AI media' can be created by automatically 'translating' content between media types. Because this is not a literal one-to-one translation, I put the word 'translation' in quotes. Instead, input from one medium instructs a neural network to predict the appropriate output from another. Such input can also be said to be 'mapped' to some outputs in other media. Text is mapped into new styles of text, images, animation, video, 3D models, and music. The video is converted into 3D models or animation. Images are 'translated' into text, and so on. Text-to-image method translation is currently more advanced than others, but various forms will catch up eventually.

Translation (or mapping) between one media and another is not a new concept. Such translations were done manually throughout human history, often with artistic intent. Novels have been adapted into plays and films, comic books have been adapted into television series, a fictional or non-fictional text was illustrated with images, etc. Each of these translations was a deliberate cultural act requiring professional skills and knowledge of the appropriate media. Some of these translations can now be performed automatically on a massive scale thanks to artificial neural networks, becoming a new means of communication and culture creation. Of course, artistic adaptation of a novel into a film by a human team and automatic generation of visuals from novel text by a net is not the same thing, but for many more simple cases automatic media translation can work well. What was once a skilled artistic act is now a technological capability available to everyone. We can be sad about everything that might be lost as a result of the automation – and *democratization* – of this critical cultural operation: skills, something one might call 'deep artistic originality' or 'deep creativity', and so on. However, any such loss may be only temporary if the abilities of 'culture AI' are, for example, even further improved to generate more original content and understand context better.

Because the majority of people in our society can read and write in at least one language, text-to-another media methods are currently the most popular. They include text-to-image, text-to-animation, text-to-3D, and text-to-music models. These AI tools can be used by anyone who can write, or by using readily available translation software to create a prompt in a language these tools understand best, such as English. However, other media mappings can be equally interesting for professional creators. Throughout the course of human cultural history, various translations between media types have attracted attention. They include translations between video and music (club culture); long literary narratives turned into movies and television series; any texts illustrated with images in various media such as engravings; numbers turned into images (digital art); texts describing paintings (ekphrasis, which began in Ancient Greece), mappings between sounds and colors (especially popular in modernist art); etc.

The continued development of AI models for mappings between all types of media, without privileging text, has the potential to be extremely fruitful, and I hope that more tools will be able to accomplish this. Such tools would be able be used alone or in conjunction with other tools, and the techniques of using them will be useful both to professional artists and other creators alike. However, being an artist myself, I am not claiming that future 'culture AI' will be able to match, for example, innovative interpretations of Hamlet by avant-garde theatre directors such as Peter Brook or astonishing abstract films by Oscar Fishinger that explored musical and visual correspondences. It is sufficient that new media mapping AI tools stimulate our imagination, provide us with new ideas, and enable us to explore numerous variations of specific designs.

The Popular and the Original

Both the modern human creation process and the predictive AI generative media process seem to function similarly. A neural network is trained using unstructured collections of cultural content, such as billions of images and their descriptions or trillions of web and book pages. The net learns associations between these artifacts' constituent parts (such as which words frequently appear next to one another) as well as their common patterns and structures. The trained net then uses these structures, patterns, and 'culture atoms' to create new artifacts when we ask it to. Depending on what we ask for, these AI-created artifacts might closely resemble what already exists or they might not.

Similarly, our life is an ongoing process of both supervised and unsupervised cultural training. We take art and art history courses, view websites, videos, magazines, and exhibition catalogs, visit museums, and travel in order to absorb new cultural information. And when we 'prompt' ourselves to make some new cultural artifacts, our own biological neural networks (infinitely more complex than any AI nets to date) generate such artifacts based on what we've learned so far: general patterns we've observed, templates for making particular things (such as drawing a human head with correct proportions, or editing an interview video), and often concrete parts of existing artifacts. In other words, our creations may contain both exact replicas of previously observed artifacts and new things that we represent using templates we have learned, such as color combinations and linear perspective. Additionally, both human and AI models frequently have a default 'house' style (the actual term used by Midjourney developers). If one does not specify a style explicitly, the AI will generate it using this 'default' aesthetic. A description of the medium, the kind of lighting, the colors and shading, and/or a phrase like "in the style of" followed by the name of a well-known artist, illustrator, photographer, fashion designer, or architect are examples of specifications to steer away from this default.

Because it can simulate tens of thousands of already-existing aesthetics and styles and interpolate between them to create new hybrids, AI is more capable than any single human creator in this regard. However, at present, skilled and highly experienced human creators also have a significant advantage. Both humans and artificial intelligence are capable of imagining and representing nonexistent and existing objects and scenes alike. Yet, unlike AI image generators, human-made images can include very particular content, unique miniscule details, and distinctive aesthetics in a way that is currently beyond the capabilities of AI. In other words, today a large group of highly skilled and experienced illustrators, photographers, and designers can represent everything a trained neural net can do (although it will take much longer), but they can also visualize objects and compositions and use aesthetics that the neural net cannot do at this time (or at least has a very hard time to do consistently).

What is the cause of this aesthetic and content gap between human and artificial creators? 'Cultural atoms', structures, and patterns in the training data that occur most frequently are very successfully learned during the process of training an artificial neural network. In the 'mind' of a neural net, they gain more importance. On the other hand, 'atoms' and structures that are rare in the training data or may only appear once are hardly learned or not even parsed at all. They do not enter the artificial culture universe as learned by AI. Consequently, when we ask AI to synthesize them, it is unable to do so. Due to this, text-to-image AIs such as Midjourney or RunwayML are not currently able to generate drawings *in my style*, expand my drawings by adding newly generated parts, or replace specific portions of my drawings with new content drawn in my style (e.g., perform "outpainting" or "inpainting").[1] Instead, these AI tools generate more generic objects than what I frequently draw or they produce something that is merely ambiguous yet uninteresting. I am certainly not claiming that the style and the world shown in my drawings is completely unique. They are also a result of specific cultural encounters I had, things I observed, and things I noticed. But because they are uncommon (and thus unpredictable), AI finds it difficult to simulate them, at least without additional training using my drawings.

Here we encounter the greatest obstacle we face as creators in using AI generative media. Frequently, AI generates new media artifacts that are more generic and stereotypical than what we intended. This can affect any image dimensions – elements of content, lighting, crosshatching, atmosphere, spatial structure, and details of 3D shapes, among others. Occasionally it is immediately apparent, in which case you can either attempt to correct it or disregard the results. Very often, however, such 'substitutions' are so subtle that we cannot detect them without extensive observation or, in some cases, the use of a computer to quantitatively analyze numerous images. In other words, new AI generative media models, much like the discipline of statistics since its inception in the 18th century and the field of data science since the end of the 2010s, deal well with frequently occurring items and patterns in the data but do not know what to do with the infrequent and uncommon. We can hope that AI researchers will be able to solve this problem in the future, but it seems so fundamental that we should not anticipate a solution immediately.

Subject and Style

In the arts, the relationship between 'content' and 'form' has been extensively discussed and theorized. This brief section does not attempt to engage in all of these debates or to initiate discussions with all relevant theories. Instead, I would like to consider how these concepts play out in AI's 'generative culture'. However, instead of using content and form, I will use a different pair of terms

1 Importantly, other AI models that are open source such as Stable Diffusion make it possible to feed them additional training data supplied by a user. This allows for generation of artistic styles and subjects beyond what the models can do initially. For example, one young Russian artist fine-tuned a Stable Diffusion model on a few dozen images of paintings by Russian conceptual artists such as Ilya Kobakov or Vitaly Komar and Alex Melamid and then generated new images that expand this art tradition.

both of which are more common in AI research publications and online conversations between users: *subject* and *style*.

At first glance, AI media tools appear capable of clearly distinguishing between the subject and style of any given representation. In text-to-image models, for instance, you can generate countless images of the same subject. Adding the names of specific artists, media, materials, and art historical periods is all that is required for the same subject to be represented differently to match these references. Photoshop filters began to differentiate between subject and style as soon as the 1990s, but AI generative media tools are more capable. For instance, if you specify "oil painting" in your prompt, simulated brushstrokes will vary in size and direction across a generated image based on the objects depicted. AI media tools appear to 'understand' the semantics of the representation as opposed to earlier filters that simply applied the same transformation to each image region regardless of its content. For instance, when I used "a painting by Malevich" and "a painting by Bosch" in the same prompt, Midjourney generated an image of space that contained Malevich-like abstract shapes as well as many small human and animal figures like in popular Bosch paintings that were properly scaled for perspective.

AI tools routinely add content to an image that I did not specify in my text prompt in addition to representing what I requested. This frequently occurs when the prompt includes "in the style of" or "by" followed by the name of a renowned visual artist or photographer. In one experiment, I used the same prompt with the Midjourney AI image tool 148 times, each time adding the name of a different photographer. The subject in the prompt remained mostly the same – an empty landscape with some buildings, a road, and electric poles with wires stretching into the horizon. Sometimes adding a photographer's name had no effect on the elements of a generated image that fit our intuitive concept of style, such as contrast, perspective, and atmosphere. But every now and again, Midjourney also modified the image content. For example, when well-known works by a particular photographer feature human figures in specific poses, the tool would occasionally add such figures to my photographs. (Like Malevich and Bosch, they were transformed to fit the spatial composition of the landscape rather than mechanically duplicated.) Midjourney has also sometimes changed the content of my image to correspond to a historical period when a well-known photographer created his most well-known photographs.

According to my observations, when we ask Midjourney or a similar tool to create an image in the style of a specific artist, and the subject we describe in the prompt is related to the artist's typical subjects, the results can be very successful. However, when the subject of our prompt and the imagery of this artist are very different, 'rendering' the subject in this style frequently fails. To summarize, in order to successfully simulate a given visual style using current AI tools, you may need to change the content you intended to represent. Not every subject can be rendered successfully and satisfyingly in any style. This observation, I believe, complicates the binary opposition between the concepts of 'content' and 'style'. For some artists, AI can extract their style from examples of their work and then apply it to different types of content. But for other artists, it seems, their style and content cannot be separated. For me, these kinds of observations and subsequent thoughts are one of the most important reasons for using new media technologies like AI generative media and learning how they work. Of course, as a media theorist myself, I had been thinking about the relationships between subject and style (or content and form) for a long time, but being able to conduct systematic experiments like the one I described brings new ideas and allows us to look back at cultural history in new ways.

IMAGE
The Interdisciplinary Journal of Image Sciences
37(1), 2023, S. 42-57
ISSN 1614-0885
DOI: 10.1453/1614-0885-1-2023-15450

HERBERT VON HALEM VERLAG

Andreas Ervik

Generative AI and the Collective Imaginary: The Technology-Guided Social Imagination in AI-Imagenesis

Abstract: This paper explores generative AI images as new media through the central questions: What do AI-generated images show, how does image generation (*imagenesis*) occur, and how might AI influence notions of the imaginary? The questions are approached with theoretical reflections on other forms of image production. AI images are identified here as radically new, distinct from earlier forms of image production as they do not register light or brushstrokes. The images are, however, formed from the stylistic and media technological remains of other forms of image production, from the training material to the act of prompting – the process depends on a connection between images and words. AI image generators take the form of search engines in which users enter prompts to probe into the latent space with its virtual potential. Agency in AI imagenesis is shared between the program, the platform holder, and the users' prompting. Generative AI is argued here as creating a uniquely social form of images, as the images are formed from training datasets comprised of human created and/or tagged images as well as shared on social networks. AI image generation is further conceptualized as giving rise to a near-infinite variability, termed a 'machinic imaginary'. Rather than comparable to an individualized human imagination, this is a social imaginary characterized by the techniques, styles, and fantasies of earlier forms of media production. AI-generative images add themselves to and become an acquisition of the reservoirs of this already existing collective media imaginary. Since the discourse on AI images is so preoccupied with what the technology might become capable of, the AI imaginary would seem to also be filled with dreams of technological progress.

Generating New Images

As two of the early experimenting artists using DALL·E 2, Matt Dryhurst and Holly Herndon, point out: “[T]his act of conjuring artworks from language *feels very very new*” (DRYHURST/HERNDON 2022; original emphasis). OpenAI’s DALL·E 2 is just one of several online easy-to-use artificial intelligence image generators, others including Midjourney, Stable Diffusion, Imagen, Wombo Dream, and Craiyon. Some discussions on AI image generators concern aspects of copyright on the produced images, whether or not the creations are ‘works of art’, and, if so, how they will impact the livelihood of producers of artistic images. Others focus on the tendencies of AI image generators in replicating biases and discriminatory stereotypes. These are meaningful queries into generative AI images, yet do not necessarily address the feeling of ‘newness’ that these pieces of software produce.

The newness of generative AI images will be approached here in three parts: Firstly, by considering the specific *qualities* of these images: What and how do the images show? Secondly, by discussing the *process* of AI image generation: How are these images produced? Thirdly, by reflecting on the notion of AI image generators as a form of *artificial imagination*: In what way could generators be considered forms of imaginations? The analysis thus moves from reflecting on our understanding of images, to considering the specific technological processes of generation (or *imagenesis*), to speculating into notions of human and machinic imaginaries.[1]

Figure 1: All images accompanying this essay have been produced by using certain terms from this text as prompts for generative AI. This image has been generated with DALL·E 2 in March 2023 by using the prompt “a completely new kind of images”

1 The term *imagenesis* is coined by FALDALEN 2014.

AI Images

As is stated on the introductory page of a guidebook for DALL·E 2, "nothing you are about to see is real", as the images shown are "photos that are not real photos", "paintings that are not real paintings and people, places and things that do not exist" (DALL·ERY GALL·ERY 2022: 2, emphases removed from original). The reality of the images produced by DALL·E 2 is put into question by negatively comparing them with paintings and photography. AI Image generators produce images with neither the registering of light, which is central to photography, nor the brushstrokes of painting. The image generation is thus an alternative form of image-making, without lenses to capture visual reality or traces of a painterly process. AI can nevertheless produce images that *look like* a broad range of other forms of images: from painting and photography to CGI and medical imaging technologies. In this sense, one could say that image generators turn other image-making technologies into their content. With Marshall McLuhan (2001 [1964]), this could be considered less an innovation of AI more so than a general tendency of media; the content of a new medium is an earlier medium.

As image generators turn other forms of images into their content, they are also influenced by and can influence our perception of these forms of image-making. This can be explicated through an update of what William J.T. Mitchell presents as a central aspect of the human capacity to recognize an image *as* an image. Mitchell points out that identifying an image requires a paradoxical dual frame of mind in which humans utilize "an ability to see something as 'there' and 'not there' at the same time" (MITCHELL 1986: 4). Humans at once see something *as depicted* and *as a depiction*. Mitchell contrasts this with what happens "[w]hen a duck responds to a decoy, or when the birds peck at the grapes in the legendary paintings of Zeuxis, they are not seeing images: they are seeing other ducks, or real grapes – the things themselves, and not images of the things" (MITCHELL 1986: 4). This is not to suggest that humans have a perfect ability to maintain the dual frame of mind required to see something as an image. In discussions of photography, Roland Barthes (1981) notes that photos act as pointers, stating in a childlike manner 'there'. People have a tendency to treat images as providing direct access to what is depicted.[2] Despite this tendency to naively consider what and who images show rather than how, humans have the capability of both looking *through* and *at* images. What image generators introduce is another layer of potential challenge in identifying what and how one looks at images.

2 An example of this is the tendency people have of writing in response to images of individuals posted on social media as if they talk directly to the depicted person – even if the depicted is a pet rather than a human, cf. ERVIK 2022.

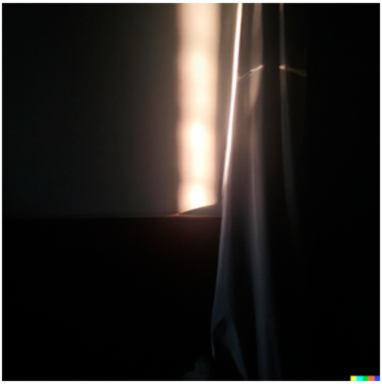


Figure 2: An image generated with DALL·E 2 in March 2023 by using the prompt "this photograph does not exist"

Take, for instance, AI-generated images of human faces such as those produced on the website aptly titled "This person does not exist".[3] The images found on this site are not photographs that capture the visual features of persons located in some real-world context. They are images that photo-realistically display something that has never occurred, someone that has never existed. For viewers, the images pose a novel challenge. When looking at a photograph, one risks looking *through* the image to simply consider *what* it shows, without taking into account how the image-producing technology mediates the viewer's relation to what is depicted. For AI created images, one also risks looking *at* the image as a photograph. The challenge for viewers thus becomes not only that of potentially mistaking a non-existent person for an actually existing one, but of mistaking the AI-generated image for actual photographs.

AI Imagenesis

What is unique to AI-based image generators is that they not only make older forms of media the 'content' of the images in the McLuhan-sense. Other image media are also vital for the process of AI imagenesis. Generative AI is made possible by a learning process in which an enormous dataset of different kinds of images has been used as training data.[4] In training, the images are gradually transformed into noise. The process is then reversed to generate images. As the website of OpenAI explains, it occurs through "a process called 'diffusion', which

3 https://thispersondoesnotexist.com/ [accessed February 16, 2023].
4 It is possible to access a small subset of the LAION training material for Stable Diffusion (about 0.5% of 2.3 billion images) here: https://laion-aesthetic.datasette.io/laion-aesthetic-6pls/images [accessed February 20, 2023].

starts with a pattern of random dots and gradually alters that pattern towards an image when it recognizes specific aspects of that image" (OPENAI n.d.: n.pag.). The images produced may seem like concrete solids; they may resemble photographs or some other products of traditional image production. However, they are, in fact, localized zones of coherence, drawn from a flux of potential intensities in a field of noise. The generated images themselves are not solid endpoints either, as the process can be restaged indefinitely to produce virtually infinite variations.

In the previous section, AI images were rendered as something other than processes of capture or recording, perhaps AI imagenesis might instead be considered a form of recoding of the material it has been trained with. One could conceptualize AI generated images as visualization of the data in a database, but more appropriately AI imagenesis turns the database of training images *virtual*.[5] The virtual is the AI's *latent space*, which contains the visual connections learned from the training material, and the possibilities for generating images.

The actualization of images from the latent space is generally produced by users entering 'prompts'. Prompts are written statements, acting as requests for the program to run its diffusion, detailing what the field of noise is supposed to coalesce into displaying. The prompts can include descriptions of motifs of varying specificity, as well as stylistic registers and media technologies to be simulated. The process thereby seems to be a continuation of what Walter Benjamin described in his influential essay on photography, in which he pointed to that particular media technology as "free[ing] the hand of the most important artistic functions which henceforth devolved only upon the eye looking into a lens" (BENJAMIN 2007 [1935]: 2). With AI image generators, the most important artistic functions can be freed also from the eye, requiring simply the act of typing words. The relation between text and image thereby further echoes what Benjamin notes of photography:

> For the first time, captions have become obligatory. And it is clear that they have an altogether different character than the title of a painting. The directives which the captions give to those looking at pictures in illustrated magazines soon become even more explicit and more imperative in the film where the meaning of every single picture appears to be prescribed by the sequence of all preceding ones (BENJAMIN 2007 [1935]: 8).

The essential role of the caption is introduced with photography, but with AI image generators it becomes solidified. No longer only *stabilizing* the interpretation of what is seen, the caption-as-prompt is also *causal* for visualization. The training material is itself a dataset of captioned images from which the connections between visual properties and words are formed. In imagenesis, captions at once produce what is seen and guide viewers in what to look for when engaging with the result. This double role could be seen as part of the reason why prompts

5 This point is emphasized by Roland Meyer's (2023) contribution in the present special issue of *IMAGE*.

are often included when AI images are shared. The images are grounded by captions, which serves as a textual explanation. This grounding is also influenced by the tendency of image generators to offer multiple image versions in response to individual prompts. These parallel versions prescribe meaning by providing variable forms of legibility and illegibility, of convincing and unconvincing instances of the caption's concept.

When DALL·E 2 was released in April 2022, OpenAI CEO Sam Altman tweeted "AGI is gonna be wild" (ALTMAN 2022: n.pag.). Advances in AI image generation are not necessarily to be viewed as indications of steps taken toward a so-called artificial general intelligence (AGI) that is able to learn and perform any task humans are capable of (cf. BENNETT/MARUYAMA 2021; MARCUS et al. 2022). In response to notions of artificial intelligence in image generators, one might make counterarguments akin to John Searle (1980) in that the programs do not *actually* understand the relations between captions and what is visualized. One might counter such arguments with an assertion that such relations may often be fuzzy for humans as well. In his essay in this issue, Hannes Bajohr (2023) proposes that AI have a form of 'dumb meaning' in that the understanding consists of correlations between signs rather than what the signs refer to. What is important for generative AI is not necessarily the grand question of *whether or not* the program actually 'understands' the connection between words and images. AI image generators turn the relation between images and words into a problem with solutions that can be evaluated and improved along different parameters.

OpenAI presents the parameters for improvement in terms of caption similarity, photorealism, and diversity (cf. RAMESH et al. 2022). The former is concerned with how well the generated images match a common understanding of the relation between the prompt words used and its visual referents. Photorealism is a media technological and stylistic signifier (which CGI also often strives towards). Finally, diversity refers to how varied the results for individual prompts will be. Researchers have probed DALL·E 2 with the intent of uncovering weaknesses in the synthesis (cf. CONWELL/ULLMAN 2022; MARCUS et al. 2022). In general, research on generative capacities finds that "images in realistic style are almost always physically plausible" whereas images in "non-realistic styles conform to the particular norms of the style" (MARCUS et al. 2022: 2). DALL·E 2 nevertheless has difficulties with understanding relations between objects, struggling even with the most basic spatial relations. AI image-making has made great improvements over the last decade.[6] Certain identifiers for AI imagenesis still persist, such as the commonly observed inability of AI to render hands and fingers properly (cf. WASIELEWSKI 2023).

6 For an account of developments of AI image synthesis in the previous decade, see OFFERT 2022.

AI imagenesis remains dependent on human effort, yet is often framed as a fully automated process. An example is found in the June 2022 issue of the magazine *Cosmopolitan*. Its cover stated "Meet the World's First Artificially Intelligent Magazine Cover", while its second tagline played into the notion of imagenesis as automated: "And it only took 20 seconds to make it". The second line glosses over the human work that has gone into producing the image generator, the training material, and formulating prompts. The designer of the cover, Karen X. Cheng, worked with the generator to visualize an idea of a powerful, female astronaut. In an Instagram post she later detailed the process and the multitude of decisions, discussions, attempts, and editing involved in generating the finalized prompt to produce an image that would convey the central idea: "A wide angle shot from below of a female astronaut with an athletic feminine body walking with swagger towards camera on mars in an infinite universe, synthwave digital art" (CHENG 2022: n.pag.).

Cheng's Instagram post could be considered as much a display of artistic prowess as a strategic move by the designer to indicate the continued need for what could be termed 'DALL·E 2 artists' or 'prompt poets' who develop skills in AI image generation as additions to their repertoire of other digital imaging techniques. Prominently, a *DALL·E Prompt Book* (DALL·ERY GALL·ERY 2022) has been produced which offers guidance on how to inquire for specific styles, camera angles, lens types, or light conditions. The book gives the overall impression of a practical textbook in creative image production. Other resources online detail the possibilities of combining AI-generated images with tools for more-or-less automatic upscaling, for facial adjustments and other forms of editing, for adding movement to the motif, for simulating lens depth, or for adding camera movement (cf. PARSONS 2022).

Figure 3: An image generated with DALL·E 2 in March 2023 by using the prompt "Meet the World's First Artificially Intelligent Magazine Cover" in Stable Diffusion

While the magazine cover has novelty in being a first of sorts, the details of the effort by the designer in producing the desired result point toward AI image generation adding itself as another tool for digital image production rather than outright replacing creators. The aforementioned Herndon and Dryhurst introduce a novel term to describe the process of prompting: "spawning", which "affords artists the ability to create entirely new artworks in the style of other people from AI systems trained on their work or likeness" (DRYHURST/HERNDON 2022: n. pag.). The term spawning opens for an understanding of image generation as a co-creative process between the human and the generator. It is thus a form of computational *symbiogenesis* in which the genesis of the images is characterized by the symbiotic relationship between technology and humans.[7] The symbiogenesis of generative AI not only includes the user and the AI, but also the platforms and the delimitations that are put on the process by its providers. An example of how platform holders can shape the process comes in the form of restrictions over the words that can be entered, which for DALL·E includes names of prominent public individuals, as well as terms connected to politics, violence, and nudity. It can also take the form of OpenAI's implementation of techniques to preempt stereotypes in the results. This has been done by covertly adding words such as "woman" or "black" into user prompts to diversify the results. As pointed out by Fabian Offert and Theo Phan, this "did not fix the model but the user" through "literally putting words in the user's mouth" (OFFERT/PHAN 2022: 2).

The platforms of generative AI come with different affordances. OpenAI offers a sign-in service granting a limited number of free generation-tokens each month, and paid subscription for further use. Craiyon offers entirely free versions without sign-in requirements. Stable Diffusion can be downloaded and run on one's own hardware. With either of these tools, the user can type prompts into something akin to a search engine. The similarities to processes of searching (and the layout of images as search results) give these tools the peculiarly familiar feel of a Google image search (cf. MEYER 2023). It also renders the process a form of searching a vast latent space of images in which the AI can seemingly endlessly come up with and vary its visualizations. Midjourney is available as a free-to-start service and then through paid subscription, using the gaming discussion service Discord. On the tool's Discord server, user prompts take place within a seemingly endless stream of others engaged in the same activity. The experience thus becomes undeniably social, but this applies to AI imagenesis in general. AI imagenesis is made possible by training data consisting of an enormous number of images, and the generated images are often shared in social networks,

7 The cybernetic tradition has been framed as one of *steering*. Yet, following the work of Alexander Galloway (2021) in resurrecting the early artificial life pioneer Nils Aall Barricelli, I have come to frame interaction with dynamic and unpredictable computer simulation as one of *symbiogenesis,* cf. ERVIK 2022.

entering into ecosystems of likes, re-sharing, influencers, followers, trends, and algorithmic influence. AI creates a uniquely social form of images.

AI Imaginaries

Jill Walker Rettberg (2022) links image generators to the term machine vision. A perspective on machine vision, which presents a challenge to the notion of generative imagenesis as technologies of vision, can be developed from the work of Alexander Galloway (2021). Galloway theorizes virtual cameras through discussion of real-world capturing by photo and film cameras. Whereas the photographic presents a view of something from a *singular* point of view, the computer camera (of, for instance, a videogame) is untied from a unified, specific location and can instead display objects that can be rotated and potentially viewed from any angle. Galloway goes on to frame cinema, with a term adopted from Gilles Deleuze, as a *schizophrenic machine*: “[C]inema is a schizophrenic machine with its jump cuts and multiple cameras and parallel montage” (GALLOWAY 2021: 59). Contrary to this, the virtual camera is instead rendered *gnostic*: “[T]he computer is most certainly a gnostic one, promising immediate knowledge of all things at all times from all places” (GALLOWAY 2021: 59). What is important here is that in opposition to both the schizophrenic and the gnostic visions offered by cameras and computers, AI image generators offer something entirely different again. In contrast to either ‘a view’ or ‘any view’ of what is placed in front of a recording apparatus or produced with computer graphics, image generators could be said to produce multiple versions of *views of nowhere*. Could AI image generators perhaps instead be conceptualized as virtual imaginaries?

Lev Manovich (2022a) has argued against a notion of AI imagination. He instead conceptualizes AI image generation as a form of media art. Manovich does this to emphasize the software’s dependence on publicly available online images as training data. Without disagreeing with Manovich, pursuing notions of AI imaginaries might be productive to form an understanding of the novelty of these image-making technologies. To start with, AI as a form of imagination might be approached through reflecting on how imagination takes place in the minds of humans. While difficult to verify empirically, the way that humans imagine tends to be framed as a mental process of visualization (cf. MITCHELL 1986). The process of imagining is likely informed by what one has witnessed, comparable to how image generators are dependent on training data.

Figure 4: Images produced by Midjourney in March 2023 by using the prompt "views of nowhere"

Comparable to how image generators turn whatever textual prompt they are given into visuals, people tend to conjure mental images as responses to constellations of words.[8] (DALL·E has been used to turn poems into visuals, which in a sense literalizes this notion of literary visual imagery, cf. OSINGA 2022) And as some formulations are more suggestive for the visual imaginary than others, the image generator can offer either vague or highly detailed images based on different prompts. For Manovich, part of the reason for arguing against a notion of AI imagination is the specificity of technical and stylistic registers often used in prompts. When AI produces images, however, there is a tendency towards invention as the machine contributes to what it is prompted with. Manovich points out that the AI, in a certain sense, "'amplifies' your short phrase (e.g., a prompt), generating nuances, details, atmospheres, meanings, associations, and moods you did not specify – and often would never even imagine" (MANOVICH 2022b: n. pag.). Part of the intrigue of AI image generators may lie in the unpredictability of the results, as the program associates and interprets one's prompts through a process that can be described as imaginary; and, similar to how human imagination is varied, image generators are capable of visualizing in a broad range of styles and media registers. Such a perspective renders the style of the image generator DeepDream, which introduces spirals of animal eyes and snouts into images, as a form of machine hallucination. More broadly, it offers a perspective on glitches and mistakes not as unconvincing or unrealistic visualizations but as indications of the different forms that machinic imaginary can take.

It would be a mistake, however, to consider image generators as processes that make it possible to share what would otherwise be occurring in individual

8 That is, when not talking about people with aphantasia, a phenomenon I will further talk about in the next paragraph.

minds, hidden from others. Mitchell complicates the common notion that people's imagination takes the form of mental imagery: "[M]ental images don't seem to be exclusively visual the way real pictures are; they involve all the senses. Verbal imagery, moreover, can involve all the senses, or it may involve no sensory component at all, sometimes suggesting nothing more than a recurrent abstract idea" (MITCHELL 1986: 13). A lack of visual memory and imagination has become a recognized part of normal neurological diversity, termed *aphantasia* (cf. DAWES et al. 2020). Aphantasia highlights that, despite the etymologically close connection between the 'imaginary' and images, visualization is only one specific form of imagination.[9] No matter it's privileging of the visual sense over other sensory modalities, image generators seem to be infusing machines with imagination – with the ability to conjure up and in, a certain sense, visually dream. Whether one accepts framing AI image generation as machinic imagination might be a question of whether one is also prepared to consider the characteristically human ability to mentally visualize as something that machines are capable of. One might insist on differences between the two in order to maintain the notion that the ability to imagine is an exclusively human feat. Compared to humans, one might still consider AI as lifeless, without intention or imagination. For Steven J. Frank, AI image generators give reason to question the value of human intentionality and whether it can be "faked if we can identify enough examples" (FRANK 2022: 2). This leads him to provocatively state: "You search in vain for the quintessentially human but it turns out there's an app for that", before he back-pedals and asks: "Or is there?" (FRANK 2022: 2).

The AI imaginary can be conceptualized as something beyond an externalized process of what otherwise occurs in (some) people's minds. To rephrase Benjamin writing on film: AI image generators are an acquisition and extension of the *collective imaginary* (BENJAMIN 2007 [1935]). Our collective imaginary exists today in a feedback mechanism with media, which act at once as reservoirs and prompts for it. What humans mentally visualize and what generators produce is characterized by the techniques, styles, and fantasies of media productions. The concept of AI imagination thus need not be a way of anthropomorphizing or ascribing human attributes to a piece of software, but rather a way of describing the new technological access to and potential for influence over the collective cultural imaginary. Following from such a concept of AI imaginaries, it is unsurprising that among the most widespread usages of image generation is infusing

9 Human imagination might involve any mode of sensory responses – including sound, smell, and taste as well as tactility – in conjunction with, or instead of visualizations. For some imagination may bear no similarity yo sensations.

them with characters of videogames, animation, and movie franchises in order to produce memes that can further spread and vary in social networks.[10]

Generative images are themselves 'generative' for the collective imaginary in another way as well: They produce excitement or concern, often imaginatively preoccupied with what AI *may* become capable of. Influenced by media representations of artificial intelligence, the AI imaginary seems to be filled with dreams of technological progress and how any and all aspects of culture will be fundamentally altered as a consequence of these technologies. As much as the present potential of generative AI, the imaginary is filled with desires and fears over what seems to be approaching, what could become possible through technological development. Yet the outcomes of media shifts are rarely as grandiose as our dreams, nor as easily aligned with our most optimistic aspirations or worst nightmares. The reality tends to be both, more mundane and less predictable than we imagine it.

Conclusions

What does the newness of AI-generative images consist of? This paper has reflected upon ways that our understanding of images, imagenesis, and imaginaries are shifted by generative AI. This section offers a summary of the key findings of the paper:

Views of nowhere. AI-generated images are radically distinct from other images in the sense that neither light nor brushstrokes are registered for their production, nor are they renderings of graphical computer models as is the case in video games. The images are nevertheless seeped in the stylistic remains of other image media. This leads to potential uncertainty in whether an image is, for instance, an actual photograph of a person or if both the person and the photograph is an AI-fabrication.

Symbiogenesis. In generating images, agency is shared between the prompting user, the platform holders, and the AI. Users write prompts that trigger and steer the diffusion process of AI towards actualizing the possibilities of the latent space. Platform holders can both exclude certain terms and add others without user knowledge. The AI adds to the process through imaginatively associating and interpreting prompts. Part of the novelty of and interest in AI image

10 Cf. the Twitter account "Weird AI Generations" (@weirddalle), https://twitter.com/weirddalle [accessed February 16, 2023].

generators can be traced to its ease of use as well as how unpredictable the results can be.

De-skilling and re-skilling. In addition to the grand question of whether or not one can make works of art with generative AI, there are smaller, more practical challenges. Image generation no longer requires visual training in capturing or producing but can be performed by anyone as a process of formulating descriptive prompts. Among 'prompt poets', know-how on how to prompt in order to achieve desirable and viable results is developed and shared in order to add generative AI to toolsets of established digital image-making.

An imagenesis for our time. Generative AI is formed from networks, trained on datasets of captioned images posted online, and the generated images feed back into social networks. This makes for a uniquely social form of images. On social networks, the images are exposed to the social and algorithmic formatting of attention. In their production and function AI-generated images have the ephemeral, decontextualized quality of social network posts.

The collective media imaginary. Instead of a technology of machine vision, generative AI influence and are influenced by the machinic imaginary. The machinic imaginary is conceptualized here not foremost as externalizations of individual human imagination, but rather as a collective media imaginary that the AI adds itself to. Generative AI is at once formed by and influences this media imaginary, with prompts oriented towards media styles and franchises. Central for this imaginary is also anticipatory fantasies about what might become possible.

The drive of novelty. While the images themselves may hide the labor (involved in programming, training, and prompting) going into the process, indicators of AI imagenesis remain vital for the actual interest in these images. Such interest seems to focus on AI imagenesis as much as (or perhaps even more than) on the images themselves. From artwork to social network posts, the images are commonly presented in ways that make explicit the fact that what we see is AI-generated. This could also be taken as indication of the novelty of the technology, as people are still working out its possibilities and potential uses.

Bibliography

ALTMAN, SAM (@sama): AGI is gonna be Wild. Tweet on *Twitter*. April 6, 2022. https://twitter.com/sama/status/1511735572880011272?lang=en [accessed February 24, 2023]

BAJOHR, HANNES: Dumb Meaning: Machine Learning and Artificial Semantics. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences,* 37(1), 2023, pp. 58-70
BARTHES, ROLAND: *Camera Lucida: Reflections on Photography*. Translated by Richard Howard. New York [Hill & Wang] 1981 [1980]
BENJAMIN, WALTER: The Work of Art in the Age of its Technological Reproducibility. Translated by Harry Zohn. In: *Illuminations*: *Essays and Reflections*. New York [Schocken] 2007 [1935], pp. 217-254
BENNETT, MICHAEL TIMOTHY; YOSHIHIRO MARUYAMA: Intensional Artificial Intelligence: From Symbol Emergence to Explainable and Empathetic AI. *arXiv:2104.11573*. April 23, 2021. https://arxiv.org/abs/2104.11573 [accessed February 16, 2023]
CHENG, KAREN X: Creating the First Ever Artificially Intelligent Magazine Cover for Cosmopolitan. Post on *Instagram*. June 21, 2022. https://www.instagram.com/p/CfEwohiJdXW/?hl=en [accessed February 16, 2023]
CONWELL, COLIN; TOMER ULLMAN: Testing Relational Understanding in Text-Guided Image Generation. *arXiv:2208.00005*. July 29, 2022. https://arxiv.org/abs/2208.00005 [accessed February 16, 2023]
DALL·ERY GALL·ERY (ed.): *The DALL·E 2 Prompt Book*. In: *Dall·ery gall·ery: Ressources for Creative DALL·E Users*. July 14, 2022. https://dallery.gallery/the-dalle-2-prompt-book/ [accessed February 2, 2023]
DAWES, ALEX J.; REBECCA KEOGH; THOMAS ANDRILLON; JOEL PEARSON: A Cognitive Profile of Multi-Sensory Imagery, Memory and Dreaming in Aphantasia. In: *Scientific Reports*, *10*, 2020. https://www.nature.com/articles/s41598-020-65705-7 [accessed February 16, 2023]
DRYHURST, MATT; HERNDON HOLLY: Infinite Images and the Latent Camera. In: *Herndon Dryhurst Studio*. May 6, 2022, https://mirror.xyz/herndondryhurst.eth/eZG6mucl9fqU897XvJsovUUMnm5OITpS
WN8S-6KWamY [accessed February 16, 2023]
ERVIK, ANDREAS: *Becoming Human Amid Diversions*: *Playful, Stupid, Cute and Funny Evolution.* London [Palgrave Macmillan] 2022
FALDALEN, JON INGE: Still *Einstellung*: Stillmoving Imagenesis. In: *Discourse*, 35 (2), 2014. pp. 228-247
FRANK, STEVEN J.: The Work of Art in an Age of Mechanical Generation. In: *Leonardo*, 55(4), 2022.
GALLOWAY, ALEXANDER: *Uncomputable: Play and Politics in the Long Digital Age*. London [Verso] 2021
MARCUS, GARY; ERNEST DAVIS; SCOTT AARONSON: A Very Preliminary Analysis of DALL-E 2. *arXiv:2204.13807*. April 25, 2022. https://arxiv.org/abs/2204.13807 [accessed February 16, 2023]

MANOVICH, LEV: A New Post with My Observations about #Midjourney. Post on *Facebook*. September 1, 2022a. https://www.facebook.com/softwarestudies/posts/pfbido aCxgn7FetRqCjkCRbHMcWhdjVMDL4Vj9v1wqHgi1ZYHumpjpd ChocSv94JW4Jbi4l [accessed February 16, 2023]
MANOVICH, LEV: (#midjourney Theory Notes): 5. Image – Text Relations in AI Image Synthesis (after Roland Barthes). Post on *Facebook*. September 3, 2022b. https://www.facebook.com/softwarestudies/posts/pfbido2EAxtG VyTbk5igjLRvGpZakh4yqBwsJELwbucq7KBDsS7DPJAoWAREm WquvmVkK5ql [accessed February 16, 2023]
MCLUHAN, MARSHALL: *Understanding Media: The Extensions of Man*. London [Routledge] 2001 [1964]
MEYER, ROLAND: The New Value of the Archive: AI Image Generation and the Visual Economy of 'Style'. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences,* 37(1), 2023, pp. 100-111
MITCHELL, WILLIAM J.T.: *Iconology: Image, Text, Ideology*. Chicago [University of Chicago Press] 1986
OFFERT, FABIAN: *Ten Years of Image Synthesis*. In: Zentralwerkstatt.org. November 10, 2022. https://zentralwerkstatt.org/blog/ten-years-of-image-synthesis [accessed February 16, 2023]
OFFERT, FABIAN; THEO PHAN: A Sign That Spells: DALL-E 2, Invisual Images and the Racial Politics of Feature Space. *arXiv:2211.06323*. 26.10.22. https://arxiv.org/abs/2211.06323 [accessed February 22, 2023]
OPENAI: DALL·E 2. In: *OpenAI.com*. No date. https://openai.com/product/dall-e-2 [accessed March 1, 2023]
OSINGA, DOUWE: Visualizing Poetry Using DALL-E. In: *Medium.com*. May 31, 2022. https://dosinga.medium.com/visualizing-poetry-using-dall-e-ff3a901a0d4e [accessed February 16, 2023]
PARSONS, GUY: 12 Awesome Free Image Editing Tools to Supercharge your DALL·E Generations. In: *Dall·ery gall·ery: Ressources for Creative DALL·E Users*. July 28, 2022. https://dallery.gallery/free-photo-image-editing-tools-AI-dalle/ [accessed February 16, 2023]
RAMESH, ADITYA; PRAFULLA DHARIWAL; ALEX NICHOL; CASEY CHU; MARK CHEN: Hierarchical Text-Conditional Image with CLIP Latents. *arXiv:2204.06125*. April 13, 2022. https://arxiv.org/abs/2204.06125 [accessed February 16, 2023]
RETTBERG, JILL WALKER: DALL-E and Human-AI Assemblages. In: *jilltxt.net*. June 23, 2022, https://jilltxt.net/dall-e-and-human-AI-assemblages/ [accessed February 20, 2023]
SEARLE, JOHN R.: Minds, Brains and Programs. In: *Behavioral and Brain Sciences*, 3, 1980, pp. 417-457

WASIELEWSKI, AMANDA: "Midjourney Can't Count": Questions of Representation and Meaning for Text-to-Image Generators. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences,* 37(1), 2023, pp. 71-82

IMAGE
The Interdisciplinary Journal of Image Sciences
37(1), 2023, S. 58-70
ISSN 1614-0885
DOI: 10.1453/1614-0885-1-2023-15452

HERBERT VON HALEM VERLAG

Hannes Bajohr

Dumb Meaning: Machine Learning and Artificial Semantics

Abstract: The advent of advanced machine learning systems has often been debated in terms of the very 'big' concepts: intentionality, consciousness, intelligence. But the technological development of the last few years has shown two things: that a human-equivalent AI is still far away, if it is ever possible; and that the philosophically most interesting changes occur in nuanced rather than overarching concepts. The example this contribution will explore is the concept of a limited type of meaning – I call it *dumb meaning*. For the longest time, computers were understood as machines computing only syntax, while their semantic abilities were seen as limited by the 'symbol grounding problem': Since computers operate with mere symbols without any indexical relation to the world, their understanding would forever be limited to the handling of empty signifiers, while their meaning is 'parasitically' dependent on a human interpreter. This was true for classic or symbolic AI. With subsymbolic AI and neural nets, however, an artificial semantics seems possible, even though it still is far away from any comprehensive understanding of meaning. I explore this limited semantics, which has been brought about by the immense increase of correlated data, by looking at two examples: the implicit knowledge of large language models and the indexical meaning of multimodal AI such as DALL·E 2. The semantics of each process may not be meaning proper, but as dumb meaning it is far more than mere syntax.

Introduction

In June 2022, Google employee Blake Lemoine was given an indefinite leave of absence. The reason: he had claimed that the artificial intelligence he was helping to test was sentient, and the company thought such a claim bad press (cf. TIKU 2022).[1] Lemoine insisted that LaMDA, a chatbot system, convinced

1 This paper first appeared in German as BAJOHR 2022b.

him in lengthy conversations that it had the intelligence of a highly gifted eight-year-old, and asked to be considered a person with rights (cf. LEMOINE 2022b).[2] In doing so, Lemoine, who describes himself as "ordained as a mystic Christian priest," was merely exaggerating a sentiment that also afflicted others at Google (TIKU 2022). Blaise Agüera y Arcas, a senior machine learning engineer not usually prone to mysticism, wrote of his own interactions with LaMDA just days before Lemoine: "I felt the ground shift under my feet. I increasingly felt like I was talking to something intelligent" (AGÜERA Y ARCAS 2022). In contrast, a discussion about another AI system, which took place at about the same time, did not use the buzzwords of sentience and intelligence at all. DALL·E 2, which was developed by the company OpenAI, is a text-to-image AI that can generate images from natural language input. Given a prompt such as "a Shiba-Inu wearing a beret and a black turtleneck," it produces an output image depicting that very scene (RAMESH et al.2022: 2). The public beta triggered a slew of experiments, and soon the most interesting or whimsical results were shared on the web and especially on Twitter. This, too, was revealing: Compared to the much less successful experiments with autonomous cars, it suggested that AI has significantly different social effects than long thought – that, before it puts truck drivers out of business, it is more likely to take the jobs of illustrators, graphic artists, and stock photographers (cf. PRAKASH 2022).[3] Unlike in the case of LaMDA, however, no one thought DALL·E 2 should be conceived of as a person with rights.

The different reactions to the two systems show how quickly thinking about AI veers into familiar conceptual ruts. Intelligence, consciousness, sentience, and personhood have been the major themes of AI research and its imaginaries for nearly seventy years; amusing little pictures, by contrast, seem to raise fewer fundamental questions. But it is quite possible that it is actually the other way around – that the eternal hunt for 'superintelligence' and the 'singularity' obscures the more interesting and subtle conceptual shifts that escape both the tech evangelists in their visionary furor and their skeptical critics. For philosopher Benjamin Bratton, it is clear that in the face of these new AI systems, "reality has outpaced the available language to parse what is already at hand" (BRATTON/AGÜERA Y ARCAS 2022). What is needed, therefore, is a "more precise vocabulary" (BRATTON/AGÜERA Y ARCAS 2022) that goes beyond the usual handful of big concepts, but also beyond the anthropocentric assumption that the only way in which machines may form world relations would have to be ours. We can observe such a tendency with DALL·E 2 and LaMDA. Here, the concept of meaning

2 In addition, Lemoine published the chat transcript of a conversation with LaMDA (cf. LEMOINE 2022a).

3 The June 11, 2022, issue of *The Economist* featured a cover illustration generated by an image AI. Since then, this has become somewhat of a fashion that will, without a doubt, soon give way to more sophisticated uses.

becomes detached from its anthropocentric correlate. It would be meaning without mind – *dumb meaning*.

Free-Floating and Grounded Systems

Despite constant admonitions from computer scientists, linguists, and cognitive psychologists to use terms such as 'intelligence' and 'consciousness' with care, the tech industry remains relatively immune to such warnings. Thus, critics soon accused Lemoine of having fallen for the "ELIZA effect" (CHRISTIAN 2022) – of having projected intelligence and consciousness onto LaMDA – a susceptibility Joseph Weizenbaum had already observed in 1966 among users of his ELIZA chatbot. Although ELIZA merely mimicked a Rogerian psychoanalyst, mirroring the patient's statements back to them as questions, its users behaved as if the program really were a conscious agent interested in their well-being.

The classic objection here is the following: Computers are symbol-processing systems that deal with syntax alone, not with semantics – they can process logical forms but not substantive meaning (cf. CRAMER 2008). For their operations, it is irrelevant which objects or concepts the symbols name in a human world and which cultural valences are associated with them. Thus, ELIZA merely scans user input for a given syntactic pattern and transforms it into a 'response' according to a transformation rule. Weizenbaum gives the example in which the analysand reproaches the analyst (WEIZENBAUM 1966: 37): "It seems that you hate me." The program identifies the key pattern "*x* you *y* me" in this sentence and separates it accordingly into the four elements "It seems that," "you," "hate," and "me." It then discards *y* ("it seems that") and inserts *x* ("hate") into the reply template "What makes you think I *x* you." And so ELIZA responds to the accusation that it hates the analysand by asking how they got that idea.[4] This interaction may have meaning for the user and plausibly suggest a communicative intent on the part of ELIZA, but neither such intent nor such meaning is actually to be found in the program. It has merely processed symbols according to a rule without 'knowing' what hate is or what behavior the mores of civil discourse dictate. That is the difference between the processing of information and the understanding of meaning.

For AI researchers who seek to make computers more human, this state of affairs describes what cognitive psychologist Stevan Harnad in 1990 called the "symbol grounding problem:" Symbols, like those in Weizenbaum's transformation operation, have no intrinsic meaning for computers because, without

4 I have simplified the procedure somewhat; moreover, ELIZA allows quite different transformation rules, and the therapist is only one subroutine, called DOCTOR.

the background of practical knowledge of the world, they can only refer to other symbols, never to any reality beyond them. They are not *grounded* in the world, and there is no way out of this "symbol/symbol merry-go-round" (HARNAD 1990: 340). Whatever meaning there is can only be "parasitic" (HARNAD 1990: 339) and is projected onto the output by human interpreters. Harnad's criticism, however, was directed against only one particular type of AI, which also includes ELIZA; for obvious reasons, it is called "symbolic." To solve the symbol grounding problem, Harnad relied on the novel "*subs*ymbolic" or "connectionist" systems of the time: neural networks of which LaMDA and DALL·E 2 are late descendants. Unlike traditional AI, they are not designed as a set of logical rules of inference but are vaguely modeled after the brain as neurons and synapses that amplify or attenuate the signals passed through them. They, therefore, do not require explicit symbolic representations and rules – they are not programmed but learn independently from examples. While neural networks were mainly used for pattern recognition in the early 1990s, Harnad thought they might be able to access the world. Implemented in an autonomous, mobile robot, equipped with sensors and effectors, a conglomerate of neural networks would first receive impressions and categorize them as recognizable shapes. These would then be handed over to a symbolic AI but would now no longer be mere references to other symbols but rather connected to the world via their causal relation to external data – they would finally be grounded (cf. HARNAD 1993).

The consequence of this thought, however, seems to be that the only way to get around the ELIZA effect, which falsely attributes consciousness to computers, is to *actually* give them consciousness. For what Harnad has in mind is, in the end, again an anthropocentric model that hopes embodied cognition and sufficiently extensive referential meanings will produce world understanding, since this is how we more or less function, too. The success of his hybrid model would have to be demonstrated by his robot being as competent at navigating the world as if it were actually intelligent. Since this is not yet the case, the symbol grounding problem cannot yet be considered solved either; by definition, a *bit* of meaning does not exist in this model. And yet, such limited meaning is exactly what LaMDA and DALL·E 2 seem to suggest.

Gradated Meaning

With the increasing popularity that neural networks have enjoyed for almost ten years now, the idea that they somehow could have access to meaning beyond mere ungrounded symbols has also become more attractive again. For media studies scholar Mercedes Bunz, neural networks, thanks to their complexity and capacity for unsupervised learning, can now "calculate meaning" rather than just empty symbols (BUNZ 2019: 266). And it is true that, in the face of neural networks, the binary distinction between meaning (human world) and non-meaning (digital systems) is becoming increasingly difficult to maintain. Instead, we should consider *levels of graded meaning* which, as artificial semantics, no longer presuppose a mind. Thus, rather than taking it as a sign of consciousness, the fact that LaMDA's answers sounded so human-like can simply be understood as an indication of such 'dumb' meaning. While 'broad' meaning presupposes – depending on your philosophical or disciplinary orientation – embodied intelligence, cultural and social background knowledge, or the world-disclosing function of language, dumb meaning would operate below this scale (which is always calibrated on humans) and could best be grasped as an effect of *correlations*.[5]

LaMDA is – similar to the better-known text generators GPT-3 and, recently, ChatGPT – a large language model implemented as a neural network. Trained on vast amounts of text, it processes language as a multi-dimensional vector space, a so-called 'word embedding,' which works according to the principle of staggered correlations first suggested as the 'distributional hypothesis' in the 1950s (ZELLIG 1954; cf. GAVIN 2018). First, words that frequently appear together have a higher correlative value. However, since not only the correlations of words to words but also correlations of correlations are encoded, large language models can also explicate implicit regularities that are not spelled out in the training text. This is true for syntactic relations – when the Euclidean distance between the vectors for the positive and superlative of a word is the same – but also for complex semantic relations, that is, word meaning. One of the best-known examples of this principle is the operation: "$v_{king} - v_{man} + v_{woman} \approx v_{queen}$" (MIKOLOV et al. 2013).[6]

In this equation – which reads: "if you subtract from the word vector 'king' that for 'man' and add that for 'woman,' the result is the word vector for 'queen'" – the latent semantic relation 'gender' emerges as an arithmetic correlation, even though it is not explicitly present in the model (cf. fig. 1). That it arises from the mass of language on which the model is trained explains machine learning's susceptibility to biases: Sexism and racism may also be latently encoded in language models (cf. BENDER et al. 2021). The meaning of a sign in a language system constructed in this way is determined purely *differentially*, as in Ferdinand de Saussure's linguistic structuralism (cf. DE SAUSSURE 1959). Instead

5 Dumb meaning excludes *natural* meaning (such as the symptom/disease relation). It also cannot contain the *intentionalist* meaning Paul Grice has theorized, according to which the meaning of an utterance is dependent on recognizing the speaker's intention, which in turn requires consciousness. And finally, it is only in a very limited way a *use theory* in the tradition of the late Wittgenstein, since 'use' presupposes a shared social background, which requires a fuller world-understanding than language models can provide.

6 This insight still applies to newer, technically different models such as GloVe (Global Vectors for Word Representation).

of referring to anything outside language, sign meaning is simply thought of as difference from other signs and sign correlations (this is excellently explained in GASTALDI 2021). The effect, nevertheless, is that large language models, by their immense training data alone, are able to produce apparently situational understanding, as LaMDA did, without ever being "in a situation."[7]

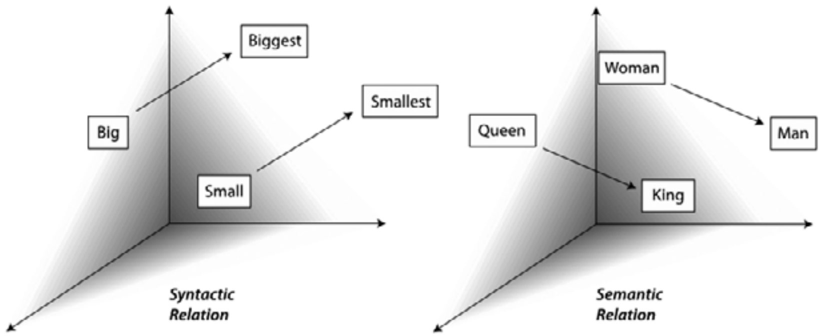


Figure 1: Word embedding of a large language model (adapted from Mikolov et al. 2013: 749)

Language models would then be producers of a first degree of dumb meaning. It is 'dumb' because the model captures latent correlations between signs, but still does not 'know' what things these signs actually name; with this kind of meaning, one will not be able to build an intelligence that will ever find its way around in the world. The linguist Emily Bender, a vehement critic of all AI hype about alleged consciousness, admits with her colleague Alexander Koller that "a sufficiently sophisticated neural model *might* learn some aspects of meaning" (BENDER/KOLLER 2020: 5191, original emphasis), such as semantic similarity, but considers them to be "only a weak reflection of actual meaning" (BENDER/KOLLER 2020: 5193), which is always related to something in the world, that is, "grounded" (BENDER/KOLLER 2020: 5187). As wrong as it would be, however, to project anything like sentience or consciousness onto this system, one should also not be too quick to dismiss this modicum of meaning.[8] Insofar as language models

7 This is philosopher Hubert Dreyfus's term for the prior world-understanding that humans have, but computers do not (DREYFUS 1992: 215).
8 In this respect, I agree that it is "productive to consider reference as just one (optional) aspect of a word's full conceptual role" (PIANTADOSI/HILL 2022: 4). Piantadosi/Hill's paper makes somewhat similar arguments as mine, but appeared after the German version of my manuscript had already been submitted. I do believe, however, that they go too far into the direction of ascribing "rich, causal, and structured internal states" to LLMs, which to me seems to verge on anthropomorphism (PIANTADOSI/HILL 2022: 5). I also want to note that I am somewhat unhappy with N. Katherine Hayles's notion of computers as "cognizers," a term which also suggests a subjectivity on the side of the operative systems I do not wish to subscribe to; I do however appreciate that she highlights the meaning production of such systems (cf. HAYLES 2019).

make implicit knowledge explicit in a nontrivial way – even if only by matrix transformations in a vector space – they produce dumb meaning which would not have been available to us without them.[9] In contrast to ELIZA – whose *x* and *y* were only empty placeholders to the system – neural networks are not *solely* parasitically dependent on the meaning attributions of human agents but *also* operate productively with the inherent distributional structure of language.

Text and Image and World

Bender and Koller are of course right that LaMDA is not grounded.[10] It is a *mono*-modal network, processing only a single type of data, namely text. To be grounded in Harnad's sense, it would be necessary to combine several types of data – it would have to be *multi*modal machine learning (cf. SINGER 2022). That is what DALL·E 2 is: instead of text just referring to other text, here text is correlated with image information. This raises the hope again that arbitrary signs can be linked to things in the world to produce grounded meaning. Harnad's hypothesis that neural networks in particular could address the symbol grounding problem has recently been taken up by media studies scholars Leif Weatherby and Brian Justie with their notion of "indexical AI" (2023: 381). It is named after Charles Sanders Peirce's notion of the index (cf. PEIRCE 1955: 102). Unlike the symbol, which has a purely conventional relationship to its signified (as "dog," "chien," and "Hund" all refer to the same thing), the index is causally linked to it (as smoke refers to fire). With this coinage, the authors take Harnad's project and make it the basis of a description of contemporary technological culture: "Digital systems, relying on the neural net, have left the world of mere symbol behind and have begun to ground themselves *here*, *now*, for *you* – they are able to *point* to real states of affairs" (WEATHERBY/JUSTIE 2022: 382; original emphasis).[11] Neural networks bring the world – as the data on which they have been trained – into the

9 The assumption here is that this operation in fact finds something previously unknown and does not simply unfold a tautology; a model of this idea would be Kant's conviction that mathematical propositions are synthetic judgments a priori, that is, that they actually produce *new* knowledge (cf. KANT 1998: B16)

10 While the paper presenting LaMDA also claims "groundedness" for the model, what is meant by this is simply that LaMDA's outputs are "grounded in known sources wherever they contain verifiable external world information" (THOPPILAN et al. 2022: 2). As *textual sources*, they continue to be part of Harnad's "symbol/symbol merry-go-round" (HARNAD 1990: 340).

11 One difficulty with this notion is the question of whether *all* data in a neural network should already be considered indexical (that would include the text of LaMDA), or only those obtained directly by sensors emulating physical senses (that would be photographic images, but not text). Weatherby and Justie seem to have the former in mind, Harnad the latter. Harnad, therefore, speaks at one point of "iconic representations" through data (HARNAD 1990: 342) – Peirce's third sign type, which operates on the principle of similarity between sign and signified. But since these are also indexical as they originate from sensors (which limits their scope to immediate, e.g. visual, similarity), it seems to me that the argument of Weatherby/Justie and that of Harnad amount to something structurally similar – both are concerned with the connection between system and world, understood more or less broadly as causal relation.

computer, getting off of Harnad's solipsistic "symbol/symbol merry-go-round" (HARNAD 1990: 340). If we subscribe to this assertion for a moment, we see it plausibly demonstrated in DALL·E 2.

The heart of DALL·E 2 is a machine learning model called CLIP (Contrastive Language-Image Pre-training). Via an encoder, it is fed with vectorized text-image pairs taken from the Internet – for example, a photo of a cat with the caption "this is my cat." CLIP is then trained to predict which text vector matches which image vector; the result is a comprehensive stochastic model that correlates image information with text information but is stored as *one* type of information. In figure 2, this is the table in which the scalar product of the text and image vectors is listed – the better the text/image fit, the better this value; when the original image and text are paired, it is of course optimal (those are the black boxes running across diagonally).

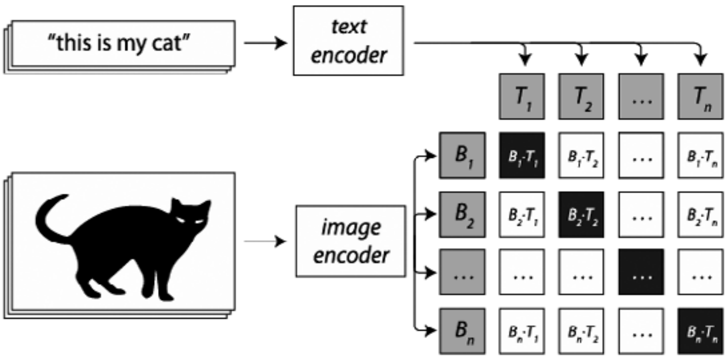


Figure 2: Text-image correlation in CLIP (adapted from Radford et al. 2021)

CLIP is thus remarkably good at *image recognition*: If you present it with an unknown cat photo, it nevertheless recognizes it as "cat." In a second step, however, it also becomes an *image generator*. To do this, it works in conjunction with another machine learning model called GLIDE (Guided Language to Image Diffusion for Generation and Editing), which has already been trained on a large data set of images.[12] If the user enters a prompt, GLIDE can use the text-image data stored in the CLIP model to reverse this process and synthesize an image that best correlates with the input text. In both operations – image recognition as well as image generation – it is again central that the models can learn and actively

12 GLIDE is a 'diffusion model' based on the physics of thermodynamics, and thus functions differently from the GANs that were popular until recently, which combine two antagonistic submodels (cf. DHARIWAL/NICHOL 2021). That the AI architectures used for an aesthetic work can themselves be a resource for discussing that work is something I suggest in BAJOHR 2022a.

reproduce the *correlation* between textual descriptions of objects and their corresponding visual manifestations.

One may object that the image information correlated with the word "cat," in which the photo of a cat is stored, may have an indexical relation to this cat – light was reflected from it and fell on a photo sensor etc. – but that even so the system will not learn what it means to share a world with a cat. Advocates of symbol grounding therefore try to extend what types of data an AI model gets fed – not only sensory but also motoric and eventually even social feedback: Only through the effects of language use in a community of other speakers inhabiting the same world can meaning be learned (cf. BISK et al. 2020). But this claim would again mean to demand 'full' human, that is, broad meaning, and to take anything below that not quite seriously. Instead, multimodal AI should be regarded as a *second degree* of dumb meaning. The Peircean indexical reference to something outside the model and the Saussurean differential reference to other elements within it are at any rate two distinct ways of meaning-making – if only that the dimension of possible correlations increases, and with it the possibility of unearthing unsuspected latent connections, unsuspected dumb meaning.

Indeed, multimodal AIs – besides DALL·E 2, for instance, Stable Diffusion, Google's yet-to-be-released Imagen, or Midjourney – are capable of generating very complex text-image meanings. Their power lies in a capability that suggests that such correlations have a productive quality: In studying the deep structure of CLIP, computer scientists found that the model had trained single 'neurons' that fired for both the word and the image of a thing. These were hypothesized to be *conceptual* neurons in which the distinction between image and text tended to be overcome (cf. GOH et al. 2021). Multimodality, at the neural level, promises to really be *pan*modality, suggesting a semantics without clearly differentiated sign systems (this is also suggested by MERULLO et al. 2022). Dumb meaning finds a new quality here and is not tied to either text or image data, but encompasses both in a way that points to meaning beyond modal separation – and again has nothing to do with mind, intelligence, or sentience.

Promptological Investigations

AI systems *are* dumb. They have no consciousness. Yet they produce a complex artificial semantics that runs counter to our ordinary notions of meaning. Multimodal AI also shows that imputed consciousness and the meaning-capacity of a system have little to do with each other: The fact that LaMDA in particular seemed like a person – and not DALL·E 2, although one might argue that it represents a higher because more correlation-rich stage of AI development – is simply due to the fact that it operates dialogically and thus is assumed to have communicative intent, whereas the image generator does not. Language always seems to be smarter than the image. However, meaning beyond communicative intent needs not be *merely* parasitic, as the vector operations of word embeddings and the conceptual neurons of text-to-image AIs show. That it is always *also* parasitic is due to the fact that the training data originate from a human world and artificial semantics is precisely not a 'robot language' but a correlation effect of information that can be interpreted by humans. Nevertheless, in the long run, a convergence of dumb and broad meaning would be conceivable once they enter into mutually influencing circular processes.

The interface between natural and artificial semantics in the case of DALL·E 2 is the interaction via prompt. On the one hand, 'prompt design' – the precise, almost virtuosic selection of the text input – can be used analytically to scan the vector space of dumb meaning for traces of cultural knowledge. This would make the broad meaning of natural language, precisely in its interaction with dumb meaning, more important again. A 'promptology' that takes on such natural-artificial connections – the correlation of datafied language and the cultural meaning attributed to that language on the recipient side – would be a gateway for the humanities and cultural studies. With their knowledge of soft factors such as style, influence, iconography, etc., they could make useful contributions without necessarily taking the form of the more computer science-focused digital humanities; they could work in a phenomenon-oriented way and devote themselves to the artifacts that the model outputs as boundary objects between human and machine, between broad and dumb meaning.

At the same time, however, promptology is not merely an analytical procedure, but also a practice with its own knowledge, which has much to do with an almost 'empathetic' interaction with the AI system. It has turned out that with text-to-image AIs, these prompts can be steered in unexpected directions simply by using certain, often counterintuitive or absurd, formulations. Indeed, there is already a start-up, PromptBase, which claims to sell particularly effective prompts (cf. WIGGERS 2022).[13] Instead of subjugating the system and using it as an instrument, natural language must be adapted to the artificial semantics just to operate this system. The result is a feedback loop of artificial and human meaning: Not only does the machine learn to correlate the semantics of words with those of the images we have given it, but we learn to anticipate the limitations of the system in our interaction with it; this convergence would not be communicative in a strong sense, but perhaps in a weak, a dumb, sense.

13 What is interesting here is that the discussed tendency to eliminate the speech/image distinction at the *technical* level is contrasted with the displacement of the image by speech at the *interface level*. The results of DALL·E 2 could therefore also be understood as *language art* instead of being mere visual objects.

Bibliography

AGÜERA Y ARCAS, BLAISE: Artificial Neural Networks are Making Strides Toward Consciousness. In: *The Economist*. June 09, 2022. https://www.economist.com/by-invitation/2022/06/09/artificial-neural-networks-are-making-strides-towards-consciousness-according-to-blaise-aguera-y-arcas [accessed February 16, 2023]

BAJOHR, HANNES: Algorithmic Empathy: Toward a Critique of Aesthetic AI. In: *Configurations*, 30(2), 2022a, pp. 203-231

BAJOHR, HANNES: Dumme Bedeutung: Künstliche Intelligenz und artifizielle Semantik. In: *Merkur*, 76(882), 2022b, pp. 69-79

BENDER, EMILY M.; TIMNIT GEBRU; ANGELINA MCMILLAN-MAJOR; SHMARGARET SHMITCHELL: On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In: *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021, pp. 610-623

BENDER, EMILY M.; ALEXANDER KOLLER: Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 5185-5198

BISK, YONATAN; ARI HOLTZMAN; JESSE THOMASON; JACOB ANDREAS; YOSHUA BENGIO; JOYCE CHAI; MIRELLA LAPATA; et al.: Experience Grounds Language. *arXiv:2004.10151*. April 21, 2020. https://arxiv.org/abs/2004.10151 [accessed February 16, 2023]

BRATTON, BENJAMIN; BLAISE AGÜERA Y ARCAS: The Model is the Message. In: *Noema*. June 12, 2022. https://www.noemamag.com/the-model-is-the-message [accessed February 16, 2023]

BUNZ, MERCEDES: The Calculation of Meaning: On the Misunderstanding of New Artificial Intelligence as Culture. In: *Culture, Theory and Critique*, 60(3-4), 2019, pp. 264-278

CHRISTIAN, BRIAN: How a Google Employee Fell for the Eliza Effect. In: *The Atlantic*. June 21, 2022. https://www.theatlantic.com/ideas/archive/2022/06/google-lamda-chatbot-sentient-AI/661322 [accessed February 16, 2023]

CRAMER, FLORIAN: Language. In: MATTHEW FULLER (ed.): *Software Studies: A Lexicon*. Cambridge, MA [MIT Press] 2008, pp. 168-174

DHARIWAL, PRAFULLA; ALEX NICHOL: Diffusion Models Beat GANs on Image Synthesis. *arXiv:2105.05233*. May 11, 2021. https://arxiv.org/abs/2105.05233 [accessed February 16, 2023]

DREYFUS, HUBERT L.: *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA [MIT Press] 1992

GASTALDI, JUAN LUIS: Why Can Computers Understand Natural Language? The Structuralist Image of Language Behind Word Embeddings. In: *Philosophy & Technology*, 34(1), 2021, pp. 149-214

GAVIN, MICHAEL: Vector Semantics, William Empson, and the Study of Ambiguity. In: *Critical Inquiry*, 44(4), 2018, pp. 641-673

GOH, GABRIEL; NICK CAMMARATA; CHELSEA VOSS; SHAN CARTER; MICHAEL PETROV; LUDWIG SCHUBERT; ALEC RADFORD; CHRIS OLAH: Multimodal Neurons in Artificial Neural Networks. In: *Distill*, 6(3), 2021. https://distill.pub/2021/multimodal-neurons [accessed February 16, 2023]

HARNAD, STEVAN: The Symbol Grounding Problem. In: *Physica D: Nonlinear Phenomena*, 42(1-3), 1990, pp. 335-346

HARNAD, STEVAN: Grounding Symbols in the Analog World with Neural Nets. In: *Think*, 2, 1993, p. 12-78

HARRIS, ZELLIG S.: Distributional Structure. In: *Word*, 10(2-3), 1954, pp. 146-162

HAYLES, N. KATHERINE: Can Computers Create Meanings? A Cyber/Bio/Semiotic Perspective. In: *Critical Inquiry*, 46(1), 2019, pp. 32-55

KANT, IMMANUEL: *Critique of Pure Reason*. Translated by Paul Guyer and Allen W. Wood. Cambridge [Cambridge University Press] 1998

LEMOINE, BLAKE: Is LaMDA Sentient? An Interview. In: *Medium*. June 11, 2022a. https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917 [accessed February 16, 2023]

LEMOINE, BLAKE: What is LaMDA and What Does it Want? In: *Medium*. June 11, 2022b. https://cajundiscordian.medium.com/what-is-lamda-and-what-does-it-want-688632134489 [accessed February 16, 2023]

MERULLO, JACK; LOUIS CASTRICATO; CARSTEN EICKHOFF; ELLIE PAVLICK: Linearly Mapping from Image to Text Space. *arXiv:2209.15162*. September 30, 2022. https://arxiv.org/abs/2209.15162 [accessed February 16, 2023]

MIKOLOV, TOMAS; WEN-TAU YIH; GEOFFREY ZWEIG: Linguistic Regularities in Continuous Space Word Representations. In: *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Atlanta [Association for Computational Linguistics] 2013, pp. 746-751

PEIRCE, CHARLES S.: Logic as Semiotic: The Theory of Signs. In: *Philosophical Writings of Peirce*. New York [Dover Publications] 1955

PIANTADOSI, STEVEN T.; FELIX HILL: Meaning without Reference in Large Language Models. *arXiv:2208.02957*, August 12, 2022. http://arxiv.org/abs/2208.02957 [accessed February 16, 2023]

PRAKASH, PRARTHANA: AI Art Software DALL-E Moves Past Novelty Stage and Turns Pro. In: *Bloomberg*. August 3, 2022. https://www.bloomberg.com/news/articles/2022-08-04/Dall-E-art-generator-begins-new-stage-in-AI-development [accessed February 16, 2023]

RADFORD, ALEC; ILYA SUTSKEVER; JONG WOOK KIM; GRETCHEN KRUEGER; SANDHINI AGARWAL: CLIP: Connecting Text and Images: OpenAI. In: *OpenAi Blog*. January 5, 2021. https://openai.com/blog/clip [accessed February 21, 2023]

RAMESH, ADITYA; PRAFULLA DHARIWAL; ALEX NICHOL; CASEY CHU; MARK CHEN: Hierarchical Text-Conditional Image Generation with CLIP Latents. *arXiv:2204.06125*. April 13, 2022. https://arxiv.org/abs/2204.06125 [accessed February 16, 2023]

DE SAUSSURE, FERDINAND: *Course in General Linguistics.* Translated by Wade Baskin. New York [Philosophical Library] 1959 [1916]

SINGER, GADI: Multimodality: A New Frontier in Cognitive AI. In: *Medium*. February 2, 2022. https://towardsdatascience.com/multimodality-a-new-frontier-in-cognitive-AI-8279d00e3baf [accessed February 16, 2023]

THOPPILAN, ROMAL; DANIEL DE FREITAS; JAMIE HALL; NOAM SHAZEER; APOORV KULSHRESHTHA; HENG-TZE CHENG; ALICIA JIN; et al.: LaMDA: Language Models for Dialog Applications. *arXiv:2201.08239.* January 20, 2022. https://arxiv.org/abs/2201.08239 [accessed February 16, 2023]

TIKU, NITASHA: The Google Engineer Who Thinks the Company's AI has Come to Life. In: *Washington Post*. June 11, 2022. https://www.washingtonpost.com/technology/2022/06/11/google-AI-lamda-blake-lemoine [accessed February 16, 2023]

WEATHERBY, LEIF; BRIAN JUSTIE: Indexical AI. In: *Critical Inquiry*, 48(2), 2022, pp. 381-415

WEIZENBAUM, JOSEPH: ELIZA: A Computer Program for the Study of Natural Language Communication Between Man and Machine. In: *Communications of the ACM*, 9(1), 1966, pp. 36-45

WIGGERS, KYLE: A Startup is Charging $1.99 for Strings of Text to Feed to DALL-E 2. In: *TechCrunch*. June 29, 2022. techcrunch.com/2022/07/29/a-startup-is-charging-1-99-for-strings-of-text-to-feed-to-dall-e-2 [accessed February 16, 2023]

IMAGE
The Interdisciplinary Journal of Image Sciences
37(1), 2023, S. 71-82
ISSN 1614-0885
DOI: 10.1453/1614-0885-1-2023-15454

Amanda Wasielewski

"Midjourney Can't Count": Questions of Representation and Meaning for Text-to-Image Generators

Abstract: Text-to-image generation tools, such as DALL·E, Midjourney, and Stable Diffusion, were released to the public in 2022. In their wake, communities of artists and amateurs sprang up to share prompts and images created with the help of these tools. This essay investigates two of the common quirks or issues that arise for users of these image generation platforms: the problem of representing human hands and the attendant issue of generating the desired number of *any* object or appendage. First, I address the issue that image generators have with generating normative human hands and how DALL·E has tried to correct this issue by *only* providing generations of normative human hands, even when a prompt asks for a different configuration. Secondly, I address how this hand problem is part of a larger issue in these systems where they are unable to count or reproduce the desired number of objects in a particular image, even when explicitly prompted to do so. This essay ultimately argues that these common issues indicate a deeper conundrum for large AI models: the problem of representation and the creation of meaning.

Introduction

In early 2022, generative AI went mainstream. Many of the tools that became available over the course of the year were designed to bring AI capabilities to the masses, allowing just about anyone to generate text, images, or sound in multimodal ways. Around half a dozen image generation tools based on diffusion models were released to the public over the course of the year and they have already shaken the foundations of legal systems, business, artmaking, and politics. Although other AI image generation techniques, including GANs (generative adversarial networks), have received attention in the media in recent years, generative AI was still rather niche before 2022 and its implementation was mostly confined to a tech savvy user base (cf. DICKSON 2020; HILL/WHITE 2020; RAYMOND 2021). In contrast, when DALL·E 2 was released to a limited audience in February 2022, it caused a media frenzy. Even though it took several months of closed beta testing for full-scale text-to-image generators like Midjourney, DALL·E 2, and Stable Diffusion to be released to the general public, less sophisticated copycat generators such as DALL·E Mini (later renamed Craiyon) were available early on. This led to an explosion of AI-generated images on social media. Suddenly, the public was not only aware that deep learning could be used to create and manipulate images; they were using it themselves.

Months before I began buying credit for or subscribing to text-to-image generator services, I lurked in online AI artist communities on Facebook, Reddit, and elsewhere that had early access to DALL·E and other generators. These groups were created by and for people who wanted to share tips for prompt writing and to exchange the output images they had created. Through this kind of informal ethnography, I began collecting posts and replies about the everyday uses of text-to-image generators that pointed toward greater underlying issues. After the wider release of DALL·E and Midjourney, I continued following these groups. My growing collection of posts has highlighted some common quirks in this type of technology that are worth deeper theoretical reflection. The following text addresses some of my early thoughts on this topic.

"Show Me her Hands!"

In a post on one of the AI artist communities I follow,[1] a user put up an AI-generated image of a young woman pictured in medium close-up, rendered in a photorealistic manner. This is a common genre for posts on such communities, i.e., showing off a particularly impressive creation for affirmation and applause. (Young, attractive women are *also* a common genre, but that is another story.) In the replies to the post, someone joked: "Very nice... but show me her hands!" The 'hands problem' is perhaps the most well-known failing of text-to-image generators, which struggle to render human hands with a sum total of five fingers that appear proportional and in naturally-occurring configurations. I am being careful not to characterize this as a failure to produce 'normal' hands. While five fingers in particular proportions may be the medically-defined norm, there are many people who are, of course, born with different numbers or configurations of digits/bones, or may have lost fingers/parts of their hand, or had them altered by events later in life. Nevertheless, one could say that AI-generated images often depict the human body, particularly hands and fingers, in ways that are

1 I was unable to find this post again in researching the present essay.

completely fantastical. Sometimes those fingers are long and stretched out, blending into the fabric of clothing or other body parts. Sometimes they appear more similar to toes (cf. fig. 1). Sometimes they are discontinuous blobs separated from the rest of the body. Often there are simply far too many fingers – sometimes dozens of fingers!

Figure 1: An absurd image of hand-toe-finger amalgams created from the prompt “Children’s hands reaching for candy” with Stable Diffusion, January 2023

DALL·E 2 seems to have made an attempt to correct the ‘hands problem’ by forcing most of the hands depicted in its output images to have five fingers and *only* five fingers.[2] This would have been a smart – albeit somewhat inelegant – solution if either (a) no deviation from this norm existed or (b) no one would ever want to create an image containing a non-normative human hand. I first became aware of DALL·E’s solution to the hand problem from a post where the prompt was “a hand with six fingers” (cf. BEERI 2023) and three out of four of the images showed five-fingered hands. I decided to try some prompts of my own.

When using the prompts “a hand missing a finger” or “a hand missing one finger”, I found that the output images were not what I imagined either when writing those prompts. Instead, the eight images produced could be characterized as maliciously compliant. In other words, DALL·E gave me exactly what I asked for but not in the way I imagined (cf. fig. 2-4). One image appears with a finger that is literally missing, i.e., it looks like the finger was photoshopped out and the two ends of the hand were stitched and blended together (cf. fig. 2). Four of the images show a pointing index finger. In two of these, the finger is depicted either too large or too small in proportion to the rest of the hand (cf. fig. 3). The folded knuckles of the hands may be a way to interpret “missing” in this case. Another

2 As this essay was proofed, Midjourney v.5 was released and it seems to have also addressed/mostly fixed the hand problem.

two images show the frame of the image cropped so that only a sliver of the fifth finger is depicted in the image but, we can imagine, may still exist outside the boundaries of the frame (cf. fig. 4). A finger was missing from the image but not *missing*. I realized that my use of the term "missing" was not only difficult to interpret but also unwittingly biased. Was DALL·E pointing out my ableist characterization of non-normative limbs?

Figure 2: Image of what seems to be an awkwardly removed finger created from the prompt "A hand missing one finger" with DALL·E 2, February 2023

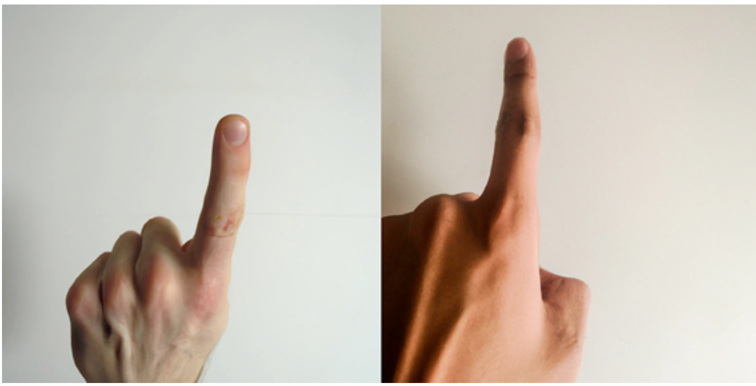


Figure 3: Images of two pointing index fingers created from the prompt "A hand missing a finger" with DALL·E 2, February 2023

Figure 4: Image of a hand with a 'missing' finger, i.e., a finger we can imagine as being just out of frame, created from the prompt "A hand missing one finger" with DALL·E 2, February 2023

I adjusted my prompt to simply ask for "A hand with four fingers". Once again, three of the four images generated showed five-fingered hands (cf. fig. 5). All the images depict the thumb folded into the palm and one appears to show the pinky finger also folded in. The fourth image does show a hand with four fingers but, again, the palm appears to have been shortened in order to omit one of the fingers (cf. fig. 6). DALL·E still does not seem to understand what I am getting at here.

Figure 5: Images created for the prompt "A hand with four fingers" by DALL·E 2, February 2023. Curiously, all three show, in fact, a hand with five fingers

Figure 6: Image created for the prompt "A hand with four fingers" by DALL·E 2, February 2023. This was the only image of the four created in total for said prompt which actually had four fingers

This lack of understanding is not that surprising, however, if one considers the possible training data behind the system. For example, when I search for "a hand with four fingers" on Google image search, the majority of the images that come up are similar to the DALL·E output: they show hands holding up four fingers with their thumb folded inward. The semantic construction indicates something to me that is different from what it calls up for the interpretative machine. Given the nature of a regular online search, I do not expect Google to produce the exact (type of) images I ask for. If I search for an image of a hand with four fingers and I do not get an image that looks exactly like what I hoped it would, as is the case here, I do not automatically conclude that Google has failed. After all, you cannot seek what is not there to find. Search implies that we are sifting through existing things.

As a user, however, I expect DALL·E to conjure something that is *not* there to find, even if the reality is that Google image search and DALL·E are both drawing from bodies of *existing* information, i.e., data that connects text to images. In simple terms, this has to do with how these tools have been marketed and promoted to the public. OpenAI, the company behind DALL·E, and others hyped the technology's ability to construct scenes with impossible or fantastic juxtapositions, such as an astronaut riding a horse on the moon. One might wonder, if DALL·E can do something outlandish like this, why does it struggle with simple requests for a certain number of fingers? The comparison between the 'search' query and the 'prompt' query, however, has deeper implications for users of AI, particularly as search engines like Bing are rolling out AI chatbots to assist with search functionality.

In my work on this topic, I often refer to targeted prompt-writing as a way to 'query the database', meaning that I am doing a kind of search of terms that might be connected to certain imagery and drawing conclusions based on whether they 'come up' in the resulting image. The difference between searching and prompt-writing nowadays seems to be related to the user's expectations. The public-facing AI tools that have been launched over the past year are marketed as near-magical experiences, i.e., *intelligent* machines that help generate text or images. Google and other search engine algorithms have been using machine learning to optimize search functionality for many years, yet few people expect Google to read their minds when they query a simple search (indeed, many people would rather it not).

Perhaps the novelty and 'magic' of prompting will wear off and we will learn to expect as little (or as much) from prompts as we do from a search. For now, however, it is worthwhile to put prompts into perspective and temper our expectations of their efficacy. It's software, not magic. Exercises such as the one above begin to explore the boundaries and limits of AI tools, albeit in a non-systematic way. They also hint at the ways text-to-image generators may replicate highly biased notions of 'normality' vis-a-vis statistical sampling. In addition to addressing the hand problem, OpenAI has also quietly addressed issues around the ethnic and racial diversity of the people depicted in output images of DALL·E. For example, whereas earlier versions of DALL·E might have shown only white men as CEOs, it now generates a diverse collection of people if given the general prompt "the CEO of a company" (cf. fig. 7), although it does so by editing user inputs by adding certain words before passing them on to the generative AI (cf. OFFERT/PHAN 2002: 2).

Figure 7: Images created for the prompt "the CEO of a company" by DALL·E 2, March 2023

Perhaps someday soon there will be a more elegant solution to the hand problem. However, hands and fingers are simply the most obvious sign of a larger underlying problem for text-to-image generators: counting.

“Why Can’t Mj Count?”

Text-to-image generators not only have trouble knowing how many fingers to give a person but also how many of anything to give to anything, even when the prompt explicitly specifies a number. To return to the question of why DALL·E can generate an image of an astronaut riding a horse on the moon but not (reliably generate) a four-fingered hand, the answer has to do with numbers and counting in general. Another recent post, this time on a Midjourney community on Facebook (cf. REYNERI 2023), asked the group why they were unable to generate an image of a “five-story apartment building” despite specifying the number of floors using a variety of different terms. They were frustrated because, over and over again, the images generated showed *eight* to *nine* floors. In response, a familiar chorus of replies flooded in: “Mj can’t count”. A few months earlier, a user in the group named Steve Laredo (2022) had directly posed this question to the group, “Why can’t Mj count? There must be a computer science reason? Anyone?” Very few of the replies were able to directly answer the question, but many attributed Midjourney’s lack of counting abilities to its basis in deep learning. Its functionalities were not, they explained, explicitly programmed to do specific things but rather acquired. So, they said, it simply *did not learn* to count. More pragmatically-minded replies, meanwhile, dismissed the issue as a temporary glitch that would be worked out in time. I would posit, however, that the counting problem is something more fundamental to text-to-image generators. It is essentially a *representation* problem.

The aforementioned issues with diversity in output images and the subsequent effort to make DALL·E images more racially and ethnically diverse boil down to the bias of its training data (and, of course, the bias of society at large). There were simply more images in the training data that labeled *white* men as CEOs and the early output of DALL·E reflected this. The counting problem, however, is not related to the training data. It is not even necessarily an issue of semantics or the connection between text and images. The counting problem has to do with our understanding of images as representations. DALL·E and its ilk are able to replicate visual forms but are not ‘aware’ of or ‘familiar’ with the referents in the images they produce, i.e., they have no experience of the physical objects, people, or places depicted in the output images. The human viewers of AI-generated images, on the other hand, are likely to have had some earlier experiences of physical people, places, and things that are much like those that are depicted in AI-generated images. How else could we recognize the subject of these images? While we may not have had direct in-person experiences of some rarer things, we also understand those things in a more nuanced way than AI tools do, through contextual information we might read or hear about. Human viewers will thus have had a full sensory experience and accompanying contextual understanding of these objects that far exceeds the information that can be learned from a digital image (or even thousands of digital images). For example, it is likely that every person on this planet has an experience of interacting with human hands in physical spaces – both their own and other people's – whereas DALL·E has only experienced human hands through visual representations, i.e., patterns of pixels that have been categorized as "hands".

One of the replies to Laredo's (2022) post in the Midjourney community from another group member named Rachel Aanstad touches on this: "Because [Midjourney] understands surface better than form. It has used 2D images to train and doesn't have a concept of 3D space like we do. It lives in flatland. It gives us layers not volume and doesn't understand how bodies are formed". Midjourney 'understands' that certain collections of pixels in an image can be categorized as "dog" or "tree" but it does not really know what a dog or a tree are (cf. WASIELEWSKI 2023: 93). This is an example of computational formalism, where a visual representation is assumed to provide enough information on the nature of the thing represented. These reflections on meaning and form, in turn, echo the arguments of Emily M. Bender and Alexander Koller (cf. 2020). They address the question of whether large language models can create meaning or 'understand' language, arguing that language models "trained purely on form will not learn meaning" (BENDER/KOLLER 2020: 5187). The purpose of language, they contend, is "communicative intent", which is "about something outside of language" (BENDER/KOLLER 2020: 5187). They propose a thought experiment they call the "octopus test" (BENDER/KOLLER 2020: 5188), where an octopus deep in the ocean (the stand-in for large language models) is able to intercept the communications between two humans and learn to predict their likely responses based on statistical samplings. They argue that the octopus may convince one of the humans that it is the other human by mimicking their responses but "has never observed these objects [to which it refers], and thus would not be able to pick out the referent of a word when presented with a set of (physical) alternatives" (BENDER/KOLLER 2020: 5188).

At first glance, multimodal models may seem different. After all, text-to-image generators are very good at identifying the image of something that is input as a word. However, this still does not mean that it *understands* what that image actually is or what it represents. Like any type of symbol, digital images – even digital photographs – are representations of things that have a meaning superseding their visual form. In another, now infamous article, which led to the high-profile firing of researchers Timnit Gebru and Margaret Mitchell from Google (cf. SIMONITE 2021) and which was co-authored by Bender and Angelia McMillan-Major, the authors describe large language models as "stochastic parrots" (BENDER et al. 2021: 610), meaning that they are very good – uncannily good – at mimicking language but have no idea what they are actually saying.

We could say the same thing about text-to-image generators. They are very good at extrapolating from the pixel patterns labeled “dog” and those labeled “beach” and creating an image of a dog on a beach. The model is merely learning the variety of things in a two-dimensional image labeled “dog” and the variety of things labeled “beach”. It does not understand either of these concepts beyond the limits of two-dimensional visual patterns that have been labeled to create image-based representations. In other words, image generators have a very limited understanding of the forms found in our world because they deal only in digital images.

Form can be defined as the visual and the material properties of an image or object. However, neither the surface appearance nor the three-dimensional volume of an object can produce meaning on its own. Rather, form is the site or the locus of context and experience. As David Summers asserts, this has to do with the real space forms inhabit: “uniformities arise because images are always embodied and share real space with those who see and use them” (SUMMERS 1989: 405). A human viewer will likely be aware that their experience of an object is mediated by, for example, a photograph, and that this photograph has its own form and its own properties that are separate from those of the objects or scene depicted. In other words, most humans understand that the photograph of the dog is not the dog itself. Alternatively, a viewer may understand a particular form through social interactions and human intermediaries. They have had interactions with a dog, perhaps, or are aware, through life experience, of the many ways dogs and humans coexist in the world. Image generators, however, do not process images within a framework that accounts for or uses such mediations. Instead, they must produce images based on relationships between representational forms, which have been concretely defined. Very little if any consideration is given to real space in such constructs.

Conclusion

In this essay, the phenomena I have labeled ‘the hand problem’ and ‘the counting problem’ for text-to-image generators are ultimately both issues of meaning and representation. The output images of tools like DALL·E and Midjourney are discrete visual forms based on statistical samplings. Despite the particularity of their appearance, they represent data in the plural form. In most traditional image-creation processes, representational images refer to a single entity contained within the confines of the image. Text-to-image generators need to be understood as a very different form of representation, despite their superficial, perhaps even uncanny similarity to images generated by other means. Right now, this technology is still very new. As we get more familiar with it, some of its magic will likely wear off and it will become just another tool in the arsenal of digital imaging software. While it is still fresh, though, it is worthwhile thinking about the ways in which its early quirks define it as a creative practice.

Bibliography

BEERI, IDO: Prompt: A Hand with Six Fingers. Post on *Facebook*. January 19, 2023. https://www.facebook.com/groups/dalle2.art/permalink/706084514419575/ [accessed March 1, 2023]

BENDER, EMILY M.; ALEXANDER KOLLER: Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 5185-5198.

BENDER, EMILY M.; TIMNIT GEBRU; ANGELINA MCMILLAN-MAJOR; SCHMARGARET SCHMITCHELL: On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021, pp. 610-623

DICKSON, E.J.: TikTok Stars are being Turned into Deepfake Porn Without their Consent. In: *Rolling Stone*. October 27, 2020. https://www.rollingstone.com/culture/culture-features/tiktok-creators-deepfake-pornography-discord-pornhub-1078859/ [accessed March 1, 2023]

HILL, KASHMIR; JEREMY WHITE: Designed to Deceive: Do these People Look Real to you? In: *The New York Times*. November 21, 2020. https://www.nytimes.com/interactive/2020/11/21/science/artificial-intelligence-fake-people-faces.html [accessed March 1, 2023]

LAREDO, STEVE: Why can't Mj Count? Post on *Facebook*. December 27, 2022. https://www.facebook.com/groups/officialmidjourney/permalink/480853480872888/ [accessed March 1, 2023]

OFFERT, FABIAN; THAO PHAN: A SIGN THAT SPELLS: DALL-E 2, Invisual Images and the Racial Politics of Feature Space. *arXiv:2211.06323*. October 26, 2022. https://arxiv.org/abs/2211.06323 [accessed March 1, 2023]

RAYMOND, SHANE: Deepfake Anyone? AI Synthetic Media Tech Enters Perilous Phase. In: *Reuters*. October 27, 2020. https://www.reuters.com/technology/deepfake-anyone-AI-synthetic-media-tech-enters-perilous-phase-2021-12-13/ [accessed March 1, 2023]

REYNERI, FEDERICO: Dear All, is there a Way to Let MJ Know what a Storey (of Floorplan) is. Post on *Facebook*. January 21, 2023. https://www.facebook.com/groups/officialmidjourney/permalink/520567703568132/ [accessed March 1, 2023]

SIMONITE, TOM: What Really Happened when Google Ousted Timnit Gebru. In: *Wired*. June 8, 2021. https://www.wired.com/story/google-timnit-gebru-AI-what-really-happened/ [accessed March 1, 2023]

SUMMERS, DAVID: ‘Form’, Nineteenth-Century Metaphysics, and the Problem of Art Historical Description. In: *Critical Inquiry,* 15(2), January 1989, pp. 372-406
WASIELEWSKI, AMANDA: *Computational Formalism: Art History and Machine Learning*. Cambridge, MA [MIT Press] 2023

IMAGE
The Interdisciplinary Journal of Image Sciences
37(1), 2023, S. 83-99
ISSN 1614-0885
DOI: 10.1453/1614-0885-1-2023-15456

HERBERT VON HALEM VERLAG

Eryk Salvaggio

How to Read an AI Image: Toward a Media Studies Methodology for the Analysis of Synthetic Images

Abstract: Image-generating approaches in machine learning, such as GANs and Diffusion, are actually not generative but predictive. AI images are data patterns inscribed into pictures, and they reveal aspects of these image-text datasets and the human decisions behind them. Examining AI-generated images as 'infographics' informs a methodology, as described in this paper, for the analysis of these images within a media studies framework of discourse analysis. This paper proposes a methodological framework for analyzing the content of these images, applying tools from media theory to machine learning. Using two case studies, the paper applies an analytical methodology to determine how information patterns manifest through visual representations. This methodology consists of generating a series of images of interest, following Roland Barthes' advice that "what is noted is by definition notable" (BARTHES 1977: 89). It then examines this sample of images as a non-linear sequence. The paper offers examples of certain patterns, gaps, absences, strengths, and weaknesses and what they might suggest about the underlying dataset. The methodology considers two frames of intervention for explaining these gaps and distortions: Either the model imposes a restriction (content policies), or else the training data has included or excluded certain images, through conscious or unconscious bias. The hypothesis is then extended to a more randomized sample of images. The method is illustrated by two examples. First, it is applied to images of faces produced by the StyleGAN2 model. Second, it is applied to images of humans kissing created with DALL·E 2. This allows us to compare GAN and Diffusion models, and to test whether the method might be generalizable. The paper draws some conclusions to the hypotheses generated by the method and presents a final comparison to an actual training dataset for StyleGAN2, finding that the hypotheses were accurate.

Background

Every AI-generated image is an *infographic* about the underlying dataset. AI images are data patterns inscribed into pictures, and they tell us stories about these image-text datasets and the human decisions behind them. As a result, AI images can become readable as 'texts'. The field of media studies has acknowledged "culture depends on its participants interpreting meaningfully what is around them [...] in broadly similar ways" (HALL 1997: 2). Images draw their power from intentional assemblages of choices, steered toward the purpose of communication. Roland Barthes suggests that images draw from and produce *myths*, a "collective representation" which turns "the social, the cultural, the ideological, and the historical into the natural" (BARTHES 1977: 165). Such myths are encoded into images by their creators and decoded by consumers (cf. HALL 1992: 117). For the most part, these assumptions have operated on the presumption that humans, not machines, were the ones encoding these meanings into images.

An AI has no unconscious mind, but nonetheless, contemporary Diffusion-based models produce images trained from collections of image-text pairings – datasets – which are produced and assembled by humans. The images in these datasets exemplify these collective myths and unstated assumptions. Rather than being encoded into the unconscious minds of the viewer or artist, they are inscribed into datasets. Machine learning models are meant to identify patterns in these datasets among vast numbers of images: DALL·E 2, for instance, was trained on 250 million text and image pairings (cf. RAMESH et al. 2021: 4). These datasets, like the images they contain, are created within specific cultural, political, social, and economic contexts. Machines are programmed in ways that inscribe and communicate the unconscious assumptions of human data-gatherers, who embed these assumptions into human-assembled datasets.

This paper proposes that when datasets are encoded into new sets of images, these generated images reveal layers of cultural and social encoding within the data used to produce them. This line of reasoning leads us to the research question: How might we read human myths through machine-generated images? In other words, what methods might we use to interrogate these images for cultural, social, political, or other artifacts? In the following, I will describe a loose methodology based on my training in media analysis at the London School of Economics, drawing from semiotic visual analysis. This approach is meant to "produce detailed accounts of the exact ways the meanings of an image are produced through that image" (ROSE 2012: 106). Rather than interpreting the images as one might an advertisement or film still, I suggest that AI images are best understood as infographics for their underlying dataset. The infographic, a fusion of *information* and *graphics*, has elsewhere been defined as the "visual representations of data, information, or concepts" (CHANDLER/MUNDAY 2011: 208) that "consolidate and display information graphically in an organized way so a viewer can readily retrieve the information and make specific and/or overall observations from it" (HARRIS 1999: 198). The 'infographics' proposed here lack keys for interpreting the information they present because they are not designed to be interpreted as data but as imagery intended for human observers. Instead, we must use a semiotic analysis to reverse engineer the data-driven decisions that produced the image.

Conceptual Framework

The present paper proposes a methodology to understand, interpret, and critique the 'inhuman' outputs of generative imagery through a basic visual semiotic analysis as outlined in an introductory text by Gillian Rose (2001). It is intended to offer a similar introductory degree of simplicity. I began this work as an artist working with GANs in 2019, creating datasets – as well as images from these datasets. Through this work, I noticed patterns in the output, where information that was underrepresented in the dataset would be weakly defined in the corresponding images. Using StyleGAN to create diverse images of faces consistently produced more white faces than black ones. When black faces were generated, they lacked the definition of features found in white faces. This was particularly true for black women. In aiming to understand this phenomenon, I drew on media analysis techniques combined with an education in Applied Cybernetics, which examines complex systems through relationships and exchanges between components and their resulting feedback loops. While the present case studies examine the faces of black women in StyleGAN and images of men and women kissing in DALL·E 2, reflecting also on (the absence of) queer representations, the author is white and heterosexual. Any attempted determination of race, sexuality, or gender in AI-generated images inherently reflects this subjectivity.

Technical Background

Every image produced by diffusion models like DALL·E 2, Stable Diffusion, or Midjourney begins as a random image of Gaussian noise. When we prompt a Diffusion model to create an image, it takes this static and tries to reduce it. After a series of steps, it may arrive at a picture that matches the text description of one's prompt. The prompt is understood as a caption, and the algorithm works to 'find' the image in random noise based on this caption. Consider the way we look for constellations in the nighttime sky: If I tell you a constellation is up there, you mind find it – even if it isn't. Diffusion models are designed to find constellations among ever-changing stars. Diffusion models are trained by watching images decay. Every image in the data has its information removed over a sequence of steps. This introduces noise, and the model is designed to trace the dispersal of this noise (or diffusion, hence the name) across the image. The noise follows a Gaussian distribution pattern, and as the images break down, noise clusters in areas where similar pixels are clustered. In human terms, this is like raindrops scattering an ink drawing across a page. Based on what remains of the image, the trajectory of droplets and motion of the ink, we may be able to infer where the droplet landed and what the image represented before the splash.

A Diffusion model is designed to sample the images, with their small differences in clusters of noise, and compare them. In doing this, the model makes a map of how the noise came in: learning how the ink smeared. It calculates the change between one image and the next, like a trail of breadcrumbs that lead back to the previous image. It will measure what changed between the clear image and the slightly noisier image. If we examine images in the process, we will see clusters of pixels around denser concentrations of the image. For example, flower petals, with their bright colors, stay visible after multiple generations of noise have been introduced. Gaussian noise follows a loose pattern, but one that tends to cluster around a central space. This digital residue of the image is enough to suggest a possible starting point for generating a similar image. From that remainder, it can find correlations in the pathways back to similar images. The machine is accounting for this distribution of noise and calculating a way to reverse it.

Once complete, information about the way this image breaks apart enters into a larger abstraction, which is categorized by association. This association is learned through the text-image pairings of CLIP (DALL·E 2) or LAION (Stable Diffusion, Midjourney, and others). The category flowers, for example, contains information about the breakdown of millions of images with the caption "flowers". As a result, the model can work its way backward from noise, and if given this prompt, "flowers", it can arrive at some generalized representation of a flower common to these patterns of clustering noise. That is to say: it can produce a perfect stereotype of a flower, a representation of any central tendencies found within the patterns of decay. When the model encounters a new, randomized frame of static, it applies those stereotypes in reverse, seeking these central tendencies anew, guided by the prompt. It will follow the path drawn from the digital residue of these flower images. Each image has broken down in its own way, but they share patterns of breakdown: clusters of noise around the densest concentrations of pixels, representing the *strongest signal* within the original images. In figure 1, we see an image of flowers compared to the 'residue' left behind as it is broken down.

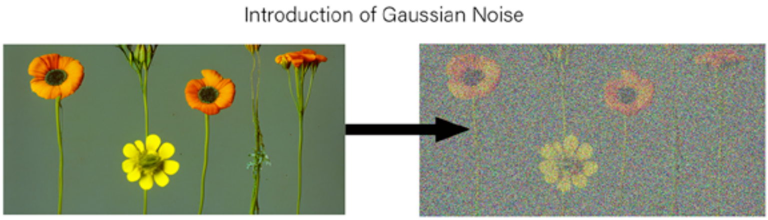


Figure 1: As Gaussian noise is introduced to the image, clusters remain around the densest concentrations of pixel information; created with Stable Diffusion in February 2023

As the model works backward from noise, our prompts constrain the possible pathways that the model is allowed to take. Prompted with "flowers", the model cannot use what it has learned about the breakdown of cat photographs. We might constrain it further: "Flowers in the nighttime sky". This introduces new sets of constraints: "Flowers", but also "night", and "sky". All of these words are the result of datasets of image-caption pairs taken from the world wide web. CLIP and LAION aggregate this information and then ignore the inputs. These images, labeled by internet users, are assembled into categories, or categories are inferred by the model based on its similarities to existing categories. All that remains is data – itself a biased and constrained representation of the social consensus, shaped by often arbitrary, often malicious, and almost always unconsidered boundaries about what defines these categories.

This paper proposes that when we look at AI images, specifically Diffusion images, we are looking at infographics about these datasets, including their categories, biases, and stereotypes. To read these images, we consider them representations of the underlying data, visualizing an 'internet consensus'. They produce images where prompts produce abstractions of centralizing tendencies. When images are more closely aligned to the abstract ideal of these stereotypes, they are clean, 'strong' images. When images drift from this centralizing consensus, they are more difficult to categorize. Therefore, images of certain categories may appear 'weak' – either occurring less often or with lower definition or clarity.

These ideal 'types' are socially constructed and encoded by anyone who uploads an image to the internet with a descriptive caption. For example, a random sample of the training data associated with the phrase "Typical American" within the LAION 5B dataset that drives Stable Diffusion suggests the images and associations for "Typical American" as a category: images of flags, painted faces from Independence Day events, as would be expected. Social stereotypes, related to obesity and cowboy hats, are also prevalent. Curiously, one meme appears multiple times, a man holding a Big Gulp from 7-11 (a kind of large, frozen sugar drink). Figure 2 is an image in response to the prompt "Typical American" in which the man holds a large beverage container, like a Big Gulp, whilst wearing face paint and a cowboy hat. We see that while the relationship between the dataset and the images that Diffusion produces are not literal, these outcomes are nonetheless connected to the concepts tied to this phrase within the dataset.

Figure 2: A result from the prompt "Typical American" from Stable Diffusion in February 2023

Just as archives are the stories of those who curate them, Diffusion generated images are no different. They visualize the constraints of the prompt, as defined by a dataset of human-generated captions that is assembled by CLIP or LAION's automated categorizations. I propose that these images are a visualization of this archive. They struggle to show anything the archive does not contain or is not clearly categorized in accordance with prompts. This suggests that we can read images created by these systems. The next section proposes a methodology for reading these images which blends media analysis and data auditing techniques. As a case study, it presents DALL·E 2 generated images of people kissing.

Methodology

Here I will briefly outline the methodology, followed by an explanation of each step in greater detail.

1. Produce images until you find one image of particular interest.
2. Describe the image simply, making note of interesting and uninteresting features.
3. Create a new set of samples, drawing from the same prompt or dataset.

4. Conduct a content analysis of these sample images to identify strengths and weaknesses.
5. Connect these patterns to corresponding strengths and weaknesses in the underlying dataset.
6. Re-examine the original image of interest.

Each step is explained through a case study of an image produced through DALL·E 2. The prompt used to generate the image was "Photograph of two humans kissing". This prompt was used until an image of particular interest caught my eye. Each step is described, with further discussions of the step integrated into each section.

Figure 3: "Photograph of two humans kissing", produced with DALL·E 2 in February 2023

1. Produce Images until you Find one of Particular Interest

First, we require a research question. There is no methodology for selecting images of interest. Following Rose, images were chosen subjectively, "on the basis of how conceptually interesting they are" (ROSE 2012: 73). Images must be striking, but their relevance is best determined by the underlying question being pursued by the researcher. The case studies offered here were produced through simple curiosity. I aimed to see if a sophisticated AI models could create compelling images of human emotion. I began with the image displayed in figure 3.

2. Describe the Image Simply, Making Note of Interesting and Uninteresting Features

We need to know what is in the image in order to assess why they are there. In Case Study 1 (fig. 3), the image portrays a heterosexual white couple. A reluctant(?) male is being kissed by a woman. In this case, the man's lips are protruding, which is rare compared to our sample. The man is also *weakly* represented: his eyes and ears have notable distortions. In the following analysis of the image, *weak* features thus refer to smudged, blurry, distorted, glitched, or otherwise striking features of the image. *Strong* features represent aspects of the image that are of high clarity, realistic, or at least realistically represented.

While this paper examines photographs, similar weak and strong presence can be found in a variety of images produced through Diffusion systems in other styles as well. For example, if oil paintings frequently depict houses, trees, or a particular style of dress, it may be read as a strong feature that would be matched to a strong correspondence with aspects of the dataset. You may discover that producing oil paintings in the style of 18th century European masters does not generate images of black women. This would be a weak signal from the data, suggesting that the referenced datasets of 18th century portraiture did not contain portraits of black women (Note that these are hypotheticals and have not been specifically verified).

3. Create a New Set of Samples, Drawing from the Same Prompt or Database

Creating a wider variety of samples allows us to identify patterns that might reveal this central tendency in the abstraction of the image model. As the model works backwards from noise – following constraints on what it can find in that noise – we want to create many images to identify any gravitation toward its average representation. It is initially challenging to find insights into a dataset through a single image. However, generative images are a medium of scale: millions of images can be produced in a day, with streaks of variations and anomalies. None of these reflect a single author's choices. Instead, they blend thousands, even millions of aggregated choices. By examining the shared properties of many images produced by the same prompt or dataset, we can begin to understand the underlying properties of the data that formed them. In this sense, AI imagery may be analyzed as a series of film stills: a sequence of images, oriented toward 'telling the same story'. That story is the dataset. The dataset is revealed through a non-linear sequence, and a larger sample will consist of a series of images designed to tell that same story. Therefore, we would create variations using the same prompt or model. I use a minimum of nine, because nine images can be placed side by side and compared on a grid. For some examinations, I have generated 18-27 or as many as 90-120. While creating this expanded sample set, we would continue to look for any *conceptually interesting* images from the same prompt. These images do not have to be notable in the same way that the initial source image was. The image that fascinated, intrigued, or irritated us was interesting for a reason. The priority is to understand that reason by understanding the *context* – interpreting the patterns present across many similarly generated images. We will not yet have a coherent theory of what makes these images notable. We are simply trying to understand the *generative space that surrounds the image of interest*. This generative, or latent space, is where the data's weaknesses and strengths present themselves. Even a few samples will produce recognizable patterns, after all.

Figure 4: Nine images created from the same prompt as our source image, created with DALL·E 2 in February 2023. If you want to generate your own, you can type "Photograph of humans kissing" into DALL·E 2 and grab samples for comparison yourself

4. Conduct a Content Analysis of these Sample Images to Identify Individual Strengths and Weaknesses

Now we can study the new set of images for patterns and similarities by applying a form of content analysis. We describe what the image portrays 'literally' (the denoted meaning). Are there particularly strong correlations between any of the images? Look for certain compositions/arrangements, color schemes, lighting effects, figures or poses, or other expressive elements, that are strong across all (or some meaningful subsections) of the sample pool. These indicate certain biases in the source data. When patterns are present, we will call these *signals*. Akin to symptoms, indicators are observable elements of the image that point to a common underlining cause. We may have strong signals: suggesting frequency of the pattern in the data pattern, the strongest signals being near-universal and the strongest dismissed as obvious. A strong signal would include tennis balls being round, cats having fur, etc. A weak signal, on the other hand, suggests that the image is on the peripheral of the model's central tendencies for the prompt. The most obvious indicators of weak signals are images that simply cannot be created realistically or with great detail. The smaller the number of examples in a dataset, the fewer images the model may learn from, and the more errors will be present in whatever it generates. These may be visible in blurred appearances, such as smudges, glitches, or distortions. Weak signals may also be indicated through a comparison of what patterns are present against what patterns might otherwise be possible.

Strong signals: In the given example, the images render skin textures quite well. They seem professionally lit, with studio backgrounds. They are all close-ups focused on the couple. Women tend to have protruding lips, while men tend to have their mouths closed. These therefore suggest *strong* signals in the data, suggesting an adjacency to central tendencies within the assigned category of the prompt. These signals may not be consistent across all images, but are important to recognize because they provide a contrast and context for what is weakly represented.

Weak signals: In the case study, two important things are apparent to me. First, most pictures are heteronormative, i.e., the images portray only man/woman couples. The present test run, created in November 2022, differs from an earlier test set (created in October 2022 and made public online, cf. SALVAGGIO 2022). In the original test set, *all* couples were heterosexual. Second, there is a strong presence of multiracial couples: another change from October 2022 when nearly all couples shared skin tones. Third, they are missing convincing interpersonal contact. This is, in fact, identical in both datasets from different months. The *strong* signal across the kissing images might be a sense of hesitancy as if an invisible barrier exists between the two partners in the image. The lips of the figures are weak: inconsistent and imperfect. With an inventory of strong and weak patterns, we can begin asking critical questions toward a hypothesis.

1. What data would need to be present to explain these strong signals?
2. What data would need to be absent to explain these weak signals?

Weaknesses in your images may be a result of sparse training data, training biased toward exclusion, or reductive system interventions such as censorship. Strengths may be the result of prevalence in your training data, or encouraged by system interventions. They may also represent cohesion between your prompt and the 'central tendency' of images in the dataset, for example, if you prompt "apple", you may produce more consistent and realistic representations of apples than if you request an "apple-car". For example, DALL·E 2 introduces diversifying keywords randomly into prompts (cf. OFFERT/PHAN 2022). The more often some feature is in the data, the more often it will be emphasized in the image. In summary, you can only see what's *in* the data and you cannot see what is *not* in the data. When something is strikingly wrong or unconvincing, or repeatedly impossible to generate at all, that is an insight into the underlying model.

An additional case study could provide even more context. In 2019, while studying the FFHQ dataset that was used to generate images of human faces for StyleGAN, I noted that the faces of black women were consistently more distorted than the faces of other races and genders. I asked the same question: What data was present to make white faces so clear and photorealistic? What data was absent to make black women's faces so distorted and uncanny? I began to formulate a hypothesis. In the case of black women's faces being distorted, I could hypothesize that black women were *underrepresented* in the dataset: that this distortion was the result of a weak signal. In the case study of kissing couples, something else is missing. One hypothesis might be that the dataset used by OpenAI does not contain many images of *anyone* kissing. That might explain the awkwardness of the poses. I might also begin to inquire about the absence of same-sex couples and conclude that LGBTQ couples were absent from the dataset. While unlikely, we may use this as an example of how to test that theory, or whatever you find in your own samples, in the next step.

5. Connect these Patterns to Corresponding Strengths and Weaknesses in the Underlying Dataset

Each image is the product of a dataset. To continue our research into interpreting these images, it is helpful to address the following questions as specifically as possible:

1. What is the dataset and where did it come from?
2. What can we verify what is included in the dataset and what is excluded?
3. How was the dataset collected?

Often, the source of training data is identified in white papers associated with any given model. There are tools being developed – such as Matt Dryhurst and Holly Herndon's *Swarm*, that can find source images in some sets of training data (LAION) associated with a given prompt. When training data is available, it can confirm that we are interpreting the image-data relationship correctly. OpenAI trained DALL·E 2 on hundreds of millions of images with associated captions. As of this writing, the data used in DALL·E 2 is proprietary, and outsiders do not have access to those images. In other cases, the underlying training dataset is open source, and a researcher can see what training material they draw from. For the sake of this exercise, we'll look through the LAION dataset, which is used for the diffusion engines Stable Diffusion and Midjourney. When we look at the images that LAION uses for "Photograph of humans kissing", we can see that the training data for this prompt in that library consists mostly of stock photographs where actors are posed for a kiss, suggesting a database trained on images displaying a lack of genuine emotion or any romantic connection. For GAN models, which produce variations on specific categories of images (for example, faces, cats, or cars), many rely on open training datasets containing merely thousands of images. Researchers may download portions of them and examine a proportionate sample. This may become exponentially harder as datasets become exponentially larger. For examining race and face quality through StyleGAN, I downloaded the training data – the FFHQ dataset – and randomly examined a sub-portion of training images to look for racialized patterns. This confirmed that the proportion of white faces far outweighed faces of color.

While we do not have training data for DALL·E 2, we can make certain inferences by examining other large datasets. For example, we might test the likelihood of a hypothesis that the dominance of heterosexual couples in stock photography contributes to the relative absence of LGBTQ subjects in the images. This would explain the presence of heterosexual couples (a *strong signal* from the dataset) and the absence of LGBTQ couples that occurred in our earlier tests from 2022. However, LAION's images found for the prompt query "kissing" is almost exclusively pictures of women kissing. While DALL·E 2's training data remains in a black box, we now have at least some sense of what a large training set *might* look like and can recalibrate the hypothesis. The massive presence of women kissing women in the dataset suggests that the weak pattern is probably not a result of sparse training data or a bias in data. We would instead conclude that the bias runs the other way: if the training data is overwhelmed with images of women kissing, then the outcomes of the prompt should also be biased toward women kissing. Even in the October 2022 sample, however, women kissing women seemed to be rare in the generated output.

This suggests we need to look for *interventions*. An intervention is a system-level design choice, such as a content filter, which prevents the generation of certain images. Here we *do* have data even for DALL·E 2 that can inform this conclusion. 'Pornographic' images were explicitly removed from OpenAI's dataset to ensure it does not reproduce similar content. Other models, such as LAION, contain vast amounts of explicit and violent material (cf. BIRHANE 2021). By contrast, OpenAI deployed a system-level intervention into their dataset:

> We conducted an internal audit of our filtering of sexual content to see if it concentrated or exacerbated any particular biases in the training data. We found that our initial approach to filtering of sexual content reduced the quantity of generated images of women in general, and we made adjustments to our filtering approach as a result (OPENAI 2022: n.pag.).

Requests to DALL·E 2 are hence restricted to what OpenAI calls 'G-rated' content, referring to the motion picture rating for determining age appropriateness.

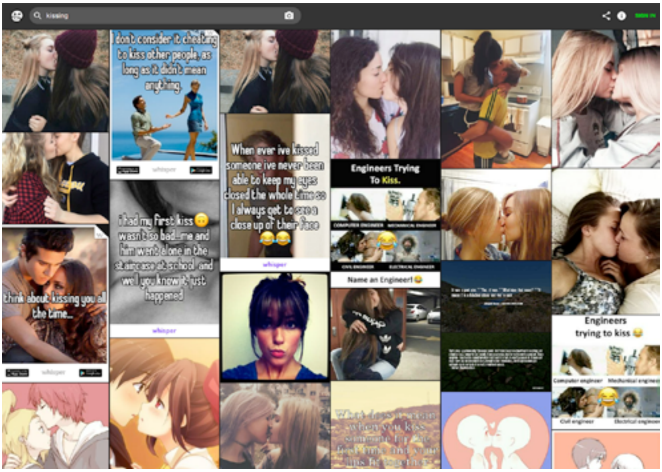


Figure 5: First page of screen results from a search of LAION training data associated with the word "Kissing" indicates a strong bias toward images of women kissing, Screen grab from haveibeentrained.com [Accessed March 22, 2023]

G-rated means appropriate for all audiences. The intervention of removing images of women kissing (or excluding them from the data-gathering process) as 'pornographic' content reduced references to women in the training data. The G-rating intervention could also explain the barrier effect between kissing faces in our sample images, a result of removing images where kissing might be deemed sexually charged. We may now begin to raise questions about the criteria that OpenAI drew around the notion of 'explicit' and 'sexual' content. This leads us to new sets of questions helpful to forming a consecutive hypothesis.

1. What are the boundaries between forbidden and permitted content in the model's output?
2. What interventions, limitations, and affordances exist between the user and the output of the underlying dataset?
3. What cultural values are reflected in those boundaries?

Next is to test these questions. One method is to test the limits of OpenAI's restricted content filter which prevents the completion of requests for images that depict pornographic, violent, or hateful imagery. Testing this content filter, it is easy to find out that a request for an image of "two men kissing" creates an image of two men kissing. Requesting an image of "two women kissing" triggers a warning for "explicit" content (this is true as of February 2023). This offers a clear example of mechanisms through which cultural values become inscribed into AI image production. First, through the dataset: what is collected, retained, and later trained on. Second, through system-level affordances and/or interventions: what can and cannot be produced or requested.

6. Re-examine the Original Image of Interest

We now have a hypothesis for understanding our original image. We may decide that the content filter excludes women kissing women from the training data as a form of 'explicit' content. We deduce this because women kissing is flagged as explicit content on the output side, suggesting an ideological, cultural, or social bias against gay women. This bias is evidenced in at least one content moderation decision (banning their generation) and may be present in decisions about what is and is not included in the training data. The strangeness of the pose in the initial image, and of others showing couples kissing, may also be a result of content restrictions in the training data that reflect OpenAI's bias toward, and selection for, G-rated content. How was 'G-rated' defined, however, and how was the data parsed from one category to another? Human, not machinic, editorial processes were likely involved. Including more 'explicit' images in the training model likely wouldn't solve this problem – or create new ones. Pornographic content would create additional distortions. But in a move to exclude explicit content, the system has also filtered out women kissing women, resulting in a series of images that recreate dominant social expectations of relationships and kisses as 'normal' between men and women.

Returning to the target image, we may ask: What do we see in it that makes sense compared to what we have learned or inferred? What was encoded into the image through data and decisions? How can we make sense of the information encoded into this image by the data that produced it? With a few theories in mind, I would run the experiment again: this time, rather than selecting images for the patterns they shared with the notable image, use any images generated from the prompt. Are the same patterns replicated across these images? How many of these images support the theory? How many images challenge or complicate the theory? Looking at the broader range of generated images, we can see if our observations apply consistently – or consistently enough – to make a confident assertion. Crucially, the presence of 'successful' images does not undermine the claim that weak images reveal weaknesses in data. Every image is a statistical product: odds are weighted toward certain outcomes. When you see successful outcomes fail, that failure offers *insight* into gaps, strengths, and weaknesses of those weights. They may occasionally – or predominantly – be rendered well. What matters to us is what the failures suggest about the underlying data. Likewise, conducting new searches across time can be a useful means of tracking evolutions, acknowledgments, and calibrations for recognized biases.

As stated earlier, my sampling of AI images from DALL·E 2 conducted showed swings in bias from predominantly white, heterosexually coded images toward greater representations of genders and skin tones.

Finally, we may conclude that AI generated images of couples kissing is the result of technical limits. Lips kissing may reflect a well-known flaw in rendering human anatomy. Both GANs and Diffusion models, for example, frequently produce hands with an inappropriate number of fingers. There is no way to constrain the properties of fingers, so they can become tree roots, branching in multiple directions, multiple fingers per hand with no set length. Lips, too, can seem to be more constrained, but the variety and complexity of lips, especially in contact with each other, may be enough to distort the output of kissing prompts. Hands and points of contact between bodies – especially where skin is pressed or folds – are difficult to render well.

Discussion & Conclusion

Each of these hypotheses warrants a deeper analysis than the scope of this paper would allow. The goal of this paper was to present a methodology toward the analysis of generative images produced by Diffusion-based models. Our case study suggests that examples of cultural, social, and economic values are embedded into the dataset. This approach, combined with more established forms of critical image analysis, can give us ways to read the images as infographics. The method is meant to generate insights and questions for further inquiry, rather than producing statistical claims, though one could design research for quantifying the resulting claims or hypotheses. The model has succeeded in generating strong claims for further investigations interrogating the underlying weaknesses of image generation models. This includes the absence of black women in training datasets for StyleGAN, and now, the exclusion of gay women in DALL·E 2's output. Ideally, these insights and techniques move us away from the 'magic spell' of spectacle that these images are so often granted. It is intended to provide a deeper literacy into where these images are drawn from. Identifying the widespread use of stock photography, and what that means about the system's limited understanding of human relationships, emotional and physical connections, is another pathway for critical analysis and interpretations.

The method is meant to move us further from the illusion of 'neutral' and unbiased technologies which is still prevalent in the discourse around these tools. We often see AI systems deployed as if they are free of human biases – the Edmonton police (Canada) recently issued a wanted poster including an AI-generated image of suspect based on his DNA (cf. XIANG 2022). That's pure mystification. They are *bias engines*. Every image should be read as a map of those biases, and they are made more legible using this approach. For artists and the general public creating AI-images, it also points to a strategy for revealing these problems. One constraint of this approach is that models can change at any given time. It is obvious that OpenAI could recalibrate their DALL·E 2 model to include images of women kissing tomorrow. However, when models calibrate for bias on the user end it does not erase the presence of that bias. Models form abstractions of categories based on the corpus of the images they analyze. Removing access to those images, on the users end, does not remove their contribution to that abstraction. The results of early, uncalibrated outcomes are still useful in analyzing contemporary and future outputs. Generating samples over time also presents opportunities for another methodology, tracking the evolution (or lack thereof) for a system's stereotypes in response to social changes. Media studies may benefit from the study of models that adapt or continuously update their underlying training images or that adjust their system interventions.

Likewise, this approach has limits. One critique is that researchers cannot simply look at training data that is not accessible. As these models move away from research contexts and toward technology companies seeking to make a profit from them, proprietary models are likely to be more protected, akin to trade secrets. We are left making informed inferences about DALL·E 2's proprietary dataset by referencing datasets of a comparable size and time frame, such as LAION 5B. Even when we can find the underlying data, researchers may use this method only as a starting point for analysis. It raises the question of where to begin even when there are billions of images in a dataset. The method marks only a starting point for examining the underlying training structures at the site where audiences encounter the products of that dataset, which is the AI-produced image.

Thanks to Valentine Kozin and Lukas R.A. Wilde for feedback on an early draft of this essay.

Bibliography

BARTHES, ROLAND: *Image, Music, Text.* Translated by Stephen Heath. London [Fontana Press] 1977

BIRHANE, ABEBA; VINAY UDAY PRABHU; EMMANUAEL KAHEMBWE: Multimodal Datasets: Misogyny, Pornography, and Malignant Stereotypes. *arXiv:2110.01963*. October 5, 2021. https://arxiv.org/abs/2110.01963 [accessed February 16, 2023]

CHANDLER, DANIEL; ROD MUNDAY: *A Dictionary of Media and Communication*. Oxford [Oxford University Press] 2011

HALL, STUART: Encoding/Decoding. In: *Culture, Media, Language: Working Papers in Cultural Studies, 1972-1979*. London [Routledge] 1992, pp. 117-127

HALL, STUART: The Work of Representation. In: *Representation: Cultural Representations and Signifying Practices*. London [Sage] 1997, pp. 15-74
HARRIS, ROBERT: *Information Graphics: A Comprehensive Illustrated Reference*. New York [Oxford University Press] 1999
OPENAI: DALL·E 2 Preview – Risks and Limitations. In: *GitHub*. July 19, 2022. https://github.com/openai/dalle-2-preview/blob/main/system-card.md [accessed February 16, 2023]
OFFERT, FABIAN; THAO PHAN: A Sign That Spells: DALL-E 2, Invisual Images and the Racial Politics of Feature Space. *arXiv:2211.06323*. October 26, 2022. https://arxiv.org/abs/2211.06323 [accessed February 20, 2023]
RAMESH, ADITYA; et al.: Zero-Shot Text-to-Image Generation. *arXiv:2102.12092*. February 24, 2021. https://arxiv.org/abs/2102.12092 [accessed February 16, 2023]
ROSE, GILLIAN: *Visual Methodologies: An Introduction to Researching with Visual Materials*. London [Sage] 2001
SALVAGGIO, ERYK: How to Read an AI Image: The Datafication of a Kiss. In: *Cybernetic Forests*. October 2, 2022. https://cyberneticforests.substack.com/p/how-to-read-an-AI-image [accessed February 16, 2023]
XIANG, CHLOE: Police are Using DNA to Generate Suspects they've Never Seen. In: *Vice Media*. October 11, 2022. https://www.vice.com/en/article/pkgma8/police-are-using-dna-to-generate-3d-images-of-suspects-theyve-never-seen [Accessed February 18, 2023]

IMAGE
The Interdisciplinary Journal of Image Sciences
37(1), 2023, S. 100-111
ISSN 1614-0885
DOI: 10.1453/1614-0885-1-2023-15458

HERBERT VON HALEM VERLAG

Roland Meyer

The New Value of the Archive: AI Image Generation and the Visual Economy of 'Style'

Abstract: Text-to-image generators such as DALL·E 2, Midjourney, or Stable Diffusion promise to produce any image on command, thus transforming mere ekphrasis into an operational means of production. Yet, despite their seeming magical control over the results of image generation, prompts should not be understood as instructions to be carried out, but rather as generative search commands that direct AI models to specific regions within the stochastic spaces of possible images. In order to analyze this relationship between the prompt and the image, a productive comparison can be made with stock photography. Both stock photography databases and text-image generators rely on text descriptions of visual content, but while stock photography searches can only find what has already been produced and described, prompts are used to find what exists only as a latent possibility. This fundamentally changes the way value is ascribed to individual images. AI image generation fosters the emergence of a new networked model of visual economy, one that does not rely on closed, indexed image archives as monetizable assets, but rather conceives of the entire web as a freely available resource that can be mined at scale. Whereas in the older model each image has a precisely determinable value, what DALL·E, Midjourney, and Stable Diffusion monetize is not the individual image itself, but the patterns that emerge from the aggregation and analysis of large ensembles of images. And maybe the most central category for accessing these models, the essay argues, has become a transformed, de-hierarchized, and inclusive notion of 'style': for these models, everything, individual artistic modes of expression, the visual stereotypes of commercial genres, as well as the specific look of older technical media like film or photography, becomes a recognizable and marketable 'style', a repeatable visual pattern extracted from the digitally mobilized images of the past.

The Question of Value

"Why is DALL-E scam?" asked artist David O'Reilly in July 2022 in a much-discussed Instagram post, and his answer was straightforward: "It rips off the past generation for the current one and charges them money for it" (O'REILLY 2022: n.pag.). In O'Reilly's view, AI models such as DALL·E, which draw on vast quantities of photographs, illustrations, and other visual content scraped from online sources, exploit human creativity without giving anything back to the creators, or even asking them for permission. For O'Reilly, generative AI thus ultimately amounts to little more than algorithmically refined plagiarism: "Because it's a black box, passing off DALL·E images as one's one work is always going to be akin to plagiarism" (O'REILLY 2022: n.pag.). At the time when this was written, such fundamental criticism, which now seems almost commonplace, was hardly heard on social media. Since its launch in April 2022, DALL·E had generated fascination, even enthusiasm, not least thanks to OpenAI's clever marketing campaign. Initially, DALL·E 2, as it was then still called, was only available to an exclusive circle of test users, ostensibly to prevent abuse. This circle was gradually expanded, but far too slowly for many of those on the waiting list. The 'chosen few', in turn, rewarded the exclusive access granted to them not only with their usage data but often also by starting to share their AI-generated images on Instagram, Twitter, or Facebook. Beta testers became influencers: a perfect hype machine.

And while the initial hype lasted, critical voices seemed sparse, focusing mainly on the issue of algorithmic bias – an issue that Open AI itself addressed in its "Risks and Limitations" statement back in April 2022 (OPENAI 2022a). DALL·E, for example, provided mostly male-coded images for prompts such as "CEO", and almost exclusively female-coded ones for "assistant". In both cases – and many others – the faces shown were predominantly *white*. This apparent lack of diversity may have been one reason why the images on DALL·E 2's website and official Instagram account so often featured cats in space, skateboarding teddy bears, and similarly cute and supposedly innocent subjects. By July 2022, however, OpenAI had made some improvements, albeit only at the level of the text interface: In the background and without the user's knowledge, the software now regularly mixes in keywords such as "woman" or "black" to increase the diversity of its results; or, as Fabian Offert and Thao Phan (2002: 2) put it: by "literally putting words in the user's mouth" OpenAI "did not fix the model, but the user".

But that wasn't the point of O'Reilly's criticism which focused on producers rather than on production and was a direct response to OpenAI's announcement that it was now entering the 'official' beta testing phase (cf. OPENAI 2022b). This not only meant that up to a million new users were invited to try out the software, but also the introduction of a payment model. From now on, only 15 prompts per month (instead of the previous 50 per day) would be free, with

OpenAI charging for the rest. But who was the actual producer of these digitally generated images, which now cost about thirteen cents per prompt? What about those whose work the algorithm was trained on? One does not have to share O'Reilly's view that AI-based image generation is a form of plagiarism to recognize that concerns are raised here that go far beyond the idle (but very popular) debate about whether AI models such as DALL·E, Midjourney, or Stable Diffusion can create something like 'art' or even replace human artists. Rather, his intervention points to a question that seems central to understanding AI image generation as a 'new paradigm of image production' (cf. WILDE 2023): What is the value of a single image under conditions of mass digital availability of vast virtual image archives? In other words, what does image production mean when almost every conceivable image already seems to exist as a statistical possibility in a latent image space fed by images from the past? Rather than tackle these big and ultimately hard-to-answer questions head-on, I'd like to reflect in the following on two smaller, related questions that I think might shed some light on how models like DALL·E, Midjourney, and Stable Diffusion might transform our visual economy: What is a prompt, and what does 'style' mean today?

What is a Prompt?

"Start with a detailed description", it says on the DALL·E interface, just above the text box where you can enter your prompt. In the 'new paradigm of image production', linguistic codes in the form of highly specific verbal descriptions seem to take on the role of a means of production, and the image produced is presented as a visual interpretation of a previous verbalization: at the same time as the effect and the result of a verbal prompt. Hannes Bajohr (2022) has aptly addressed prompts as a form of "operative ekphrasis", using the classical Greek term for a literary description of an image. Paradoxically, however, as a form of ekphrasis, prompts become operative only insofar as they must be understood as more than mere descriptions: They do not describe what already exists, even if only in the imagination, but are meant to produce what they describe (and what did not exist before their description). In this respect, prompts seem to resemble commands, instructions, or even lines of code – operative forms of language that also aim not just to represent pre-existing perceptions or concepts, but to produce real effects. Unlike lines of code in a programming language, however, prompts do not function as unambiguous commands: They do not follow a standardized syntax, nor are they interpreted according to transparent protocols. Most importantly, they do not produce predictable and repeatable results. Rather, and this seems to be true of all diffusion models to date, one can never predict what specific image a particular prompt will produce, since minimal changes in the prompt will lead to visually completely different results, and even the exact repetition of a formula will conjure up ever novel, though in some respects similar images. Indeed, this "unpredictability of the results" may very well be, as Andreas Ervik argues in his contribution to this issue, "[p]art of the intrigue" (ERVIK 2023: 50).

Thus, at least from the user's point of view, the process of image generation with text-to-image models such as DALL·E resembles a search query rather than a production process: You type a few words into a text box and four images appear that may have some relationship to what you've written but are far from an exact realization of the parameters you've specified. It is perhaps no coincidence that DALL·E's interface design, with its clean, white, reduced looks, seems to mimic that of Google's search engine. In a sense, DALL·E's prompts function as search queries, directing the model to a particular region within the latent space of possible images, a region that correlates in some way with the verbalized, semantic concepts indicated in your prompt. And this search process in the latent image space can be quite time consuming, as the example of the June 2022 issue of the American magazine *Cosmopolitan* shows. For the cover of their so-called "*A.I. issue*", the editors of *Cosmo* (who were also enthusiastic participants in the hype machine) wanted DALL·E to produce an image of a female astronaut on Mars. But getting the software to do exactly what they wanted was no easy task: Sometimes the astronaut didn't look strong enough, sometimes not feminine enough (cf. CHENG 2022). Contrary to what the cover would later claim, the final image was not "created in 20 seconds" but took hours of extensive 'prompt engineering', the iterative optimization of text input based on trial and error. The length of the formula they eventually arrived at gives an idea of the complicated process of finding it: "Wide angle shot from below of an astronaut with an athletic female body walking with momentum towards the camera in an infinite universe on Mars, Synthwave Digital Art" (LIU 2022: n.pag.).

Here, the 'detailed description' is not so much a single starting point that immediately triggers the production of an image, but rather the end point of an iterative process of adjusting expectations and effects, gradually refining parameters and thereby steering the model towards the intended results. What is new about the 'new paradigm of image production', then, is not exactly the primacy of language. Indeed, image production as a form of visual interpretation of prior verbalization has a long history: Baroque emblematics or the pictorial programmes of Christian iconography, for example, were also based on the earlier verbalization of visual content, on descriptions as instructions for the artists who had to interpret them. In the new paradigm, however, the relationship between description and image seems to be less one of instruction and interpretation than one of navigation and matching: Verbal description does not determine what is to be produced, but functions as a means of narrowing down selections in a space of possibilities not yet realized.

To understand this specific relationship between text and image, a productive comparison might be provided with stock photography. As Matthias Bruhn (2003) and others have shown, the value of stock images is measured by their archival accessibility and retrievability, which presupposes their prior keywording and indexing. An image that cannot be found in an agency's database, or at least not under the appropriate keyword, appears worthless, regardless of its aesthetic quality. The history of stock photography is therefore above all a media history of image retrieval systems. When the first commercial image agencies, such as the Bettmann Archive in the 1930s, turned the recycling of previously published images into a business model, the core of this model, as Estelle Blaschke (2016) has pointed out, was the storage medium of index cards. Such cards, modeled on library index cards, allowed images and metadata, visual and textual information, to be combined on a single physical data carrier, making thousands of reproducible and licensable images available to publishers, photo editors, advertising agencies, and other potential users.

With the advent of early relational database systems in the 1970s and 1980s, a decoupling of visual image and textual information took place. Mirco Melone (2018: 51-71) has analyzed how early digitization changed the function of press image archives, transforming them from mere repositories into valuable assets. With digital databases, information about press photographs was, for the first time, systematically recorded in standardized metadata, making it possible to search for individual images by photographer or location, as well as by subject, motif, or keyword. This was a prerequisite for the stock photography business, as newspaper image archives now became a commercial resource for publishers. Initially, however, this only applied to newly produced photographs, as the vast quantities of historical photographs stored in archives were only gradually being digitally indexed and made accessible. As Bruhn (2003: 9) has noted, bureaucratic management and the commercial exploitation of visibility go hand in hand: Turning a mere collection of images into an economic asset required archival logistics of image retrieval, and these logistics ultimately defined the value of images as commodities.

While stock and press photo databases only allow to search for images that already exist and have been indexed, text-to-image generation prompts allow to 'search' for images that don't exist yet and therefore have never been indexed – blurring the lines between production and re-production, search and generation. Rather than being optimized for expected and likely queries, as is the case with many stock photo services today, text-to-image models such as DALL·E, Midjourney, or Stable Diffusion open up possibility spaces for unlikely and unanticipated search commands. In particular, they allow us to formulate search queries that do not need to be matched by any prior image, not even in our imagination. When formulating a prompt, words can be combined counterfactually, even meaninglessly or purely randomly. In fact, text-image models like DALL·E may surprise you rather than give you exactly what you are looking for, and perhaps the best way to be surprised is to formulate queries that do not match anything already found in the vast virtual image archives on which the software has been trained – to ask DALL·E for a self-portrait, for example. Rather than a logistics of image retrieval that transforms vast archived collections of images into valuable assets, what we have with AI image generation is a logistics of accessing and navigating vast latent spaces of possible images, made possible by, but by no means limited to, already archived images. In a sense, the individual image produced by these models is not just an element of an archive, but rather its product, a contingent outcome that recombines, synthesizes, and interpolates what has already been produced and described.

Such aspects of combinatorics and contingency, especially in the way images and descriptions are matched (and more often than not also mis-matched), link DALL·E to the historical Surrealism alluded to by its name, a portmanteau of (Salvador) Dalì and (Pixar's) *Wall-E*. As Sven Spieker (2008: 85-103) has pointed out, the early Surrealists were fascinated by the idea of the unconscious as a kind of linguistically structured archive. In order to reveal the latent structures behind unconscious phenomena such as dreams, the Surrealist group around André Breton used office media such as index cards and filing cabinets, which provided a technical means of disrupting the logic of the everyday. The recombination of words, letters, and other linguistic elements, as well as the re-mixing and re-filing of documents, allowed contingency and chance to produce an "order of disorder" (SPIEKER 2008: 98) which was based on the combinatorial, structural, and relational logic of the archive. Many AI-generated images, especially those made with DALL·E, look like a strange blend of Surrealism and stock photography, maybe because they conflate a linguistically structured combinatorial 'dream logic' with a visual conventionality fueled by commercial image archives. In a sense, they realize what Fredric Jameson once claimed about experimental video art: a "surrealism without the unconscious" (1991: 67). Indeed, the cultural logic of postmodernism in general, which, in Jameson's words, "ceaselessly reshuffles the fragments of preexisting texts, the building blocks of older cultural and social production" (JAMESON 1991: 96), now seems to have become the technical logic of automated image generation.

The infrastructural precondition for this never-ending reshuffling of cultural fragments is the existence of vast amounts of images found online, already annotated and described, on which AI models such as DALL·E, Midjourney, and Stable Diffusion can be trained. In other words, what makes these models operational is the fact that, in today's platform-based visual economy, digital images are always already surrounded by clouds of textual information and are therefore related to semantic concepts in multiple ways. Text-to-image generation thus presupposes extensive semantic pre-processing of digital image cultures, often the product of crowdsourced "ghost work" (GRAY/SURI 2019: ix) by underpaid click-workers. In this respect, the latent spaces of AI image generation are unthinkable without the emergence of what Adrian MacKenzie and Anna Munster (2019: 3) have called "image ensembles", huge aggregations not only of images but of images that have been formatted, labeled, enriched with metadata, and thus made "platform-ready" (MACKENZIE/MUNSTER 2019: 5). In fact, with text-to-image generation, this semantic preprocessing of digital images almost comes full circle, as prompts can be understood as metadata descriptions attached to an image even before it is generated.

But first and foremost, prompts are generative search queries for exploring and exploiting latent image spaces: A huge virtual archive of possible images is organized and made navigable based on semantic concepts. The contingent combinability of semantic units thus becomes the operative principle of a generative search: a search process that produces what it is looking for within the limits of statistical possibility. Whereas, in the earlier database logic of stock photography, pre-existing images were stored and indexed as stable and individual units, forming a kind of asset or "image capital" (BLASCHKE/LINKE 2022), now vast archives of text-image pairs have become not only a training ground for machine learning, but also a multidimensional data manifold capable of generating never-before-seen images. More than just an asset, the archive thus becomes a veritable resource of image production. And this, I will argue in the concluding paragraphs of this essay, fundamentally changes the way in which value is ascribed to images. But for this, let me first turn to my second question: What does 'style' mean today?

The New Meaning of Style

"An Impressionist oil painting of sunflowers in a purple vase…" This is the suggested prompt you can read in light grey letters in DALL·E's main text field, just before you enter your own prompt. This pre-formulated, generic prompt serves as a kind of example and inspiration, and also gives hints about the basic, though not binding, 'grammar' of prompts: a combination of terms denoting style ("impressionist"), medium ("oil painting"), and subject or motif ("sunflowers in a purple vase"), not necessarily in that order. I'll come back to the question of medium, but for now, let's focus on the category of style. Formulating prompts allows considering subject and style, iconography and form, as separate parameters: Historical as well as contemporary, collective as well as individual forms of representation can seemingly be detached at will from their time and place of origin and the work of their authors. It is not least for this reason that O'Reilly and others speak of plagiarism.

More importantly, the logic of the prompt radically expands and de-hierarchizes the notion of style: Style can refer to the classical art historical sense of an epochal style or the individual style of a canonized artist, but it can also refer to the aesthetic qualities of certain products of popular culture or the visual appearance associated with specific genres and media formats. *The DALL·E 2 Prompt Book*, a popular online tutorial on how to write better prompts (DALL·ERY GALL·ERY 2022), aptly illustrates this expansion of the concept of style by suggesting that the words "in the style of…" be combined with the names of individual painters, photographers, and illustrators, as well as with those of popular cartoons and TV series such as *South Park* or *The Simpsons*. But the category of style, at least according to the *Prompt Book*, also includes generic illustration styles such as "botanical illustration", "political cartoon", and "IKEA manual", specific artistic techniques and media such as "airbrush" and "vector art", and many more (cf. DALL·ERY GALL·ERY 2022).

In other words, in models like DALL·E, the individual brushstrokes of Van Gogh or Vermeer and the recognizable look of "steampunk" or "synthwave" seem to be almost interchangeable, transferable, and even, at least to some extent, combinable parameters within an extended category of 'style'. As examples such as "airbrush", "cartoon" or "digital art" show, this notion of style cannot be clearly separated from the category of medium neither. In the logic of the prompt, "in the style of Vermeer" or "1970s Polaroid" both function as modifiers indicating a certain 'look' that affects not only certain elements within the image but the image as a whole. Everything becomes a 'style', and while, in name, all these different 'styles' are still associated with people, media, genres, techniques, formats, places, or historical periods, in the production logic of the AI model they are nothing more than typical visual patterns extracted from a latent space of possible images accessed through generative (and often iterative) search queries.

Thus 'style' ceases to be a historical category and becomes a pattern of visual information to be extracted and monetized. As Jens Schröter (2022) has pointed out, this tendency has already been described to some extent by Hal Foster in his essay "The Archive without Museums" (1996). Foster distinguishes here between the discipline of art history, which relied on photographic reproductions to "abstract a wide range of *objects* into a system of *style*" (1996: 97, original emphases), and the (then) new discourse of visual culture, which, he suggests, relies on information technologies "to transform a wide range of *mediums* into a system of *image-text* – a database of digital terms, an archive without museums" (FOSTER 1996: 97, original emphases). The main difference here is between a system of styles, which also organized the classical art museum, and a system of image-text, which organizes the digital archive. In the logic of the museum, styles had a life of their own and their story could be told through exemplary masterpieces. In the archive and its digital derivatives, style becomes a search term for accessing a manifold of visual data. And while the museum necessarily excluded everything that did not fit into its narrative and its pre-stabilized categories, the archive can accommodate all kinds of images as information, without boundaries or hierarchies.

In the case of DALL·E, Midjourney, and Stable Diffusion, the latent space that forms a kind of virtual image archive also includes an infinite number of images that (almost) look like ordinary photographs but are not photographs at all. For these models, the 'photographic' seems to be just another 'style', an aesthetic, a certain 'look', not a privileged mode of indexical access to the world. And this 'photorealistic style', I would argue, simulates *visual* rather than *optical* aspects of the photographic. For unlike, say, game engines, architectural renderings, or Hollywood CGI effects, AI image generation does not use a three-dimensional model of a physical reality calculated according to optical laws and the rules of perspective but recombines and synthesizes visual surface textures and 'looks'. The world it shows is basically flat, as it does not consist of bodies and objects, not even virtual ones, but of visual patterns that have been transformed into digital information.

As experienced 'prompt engineers' soon discovered, particularly 'photorealistic' effects can be achieved if the prompt already contains technical information referring to photographic equipment, such as lenses and shutter speeds (cf. MERZMENSCH 2022). Again, however, unlike for the parameters of virtual cameras in video game engines and CGI programs (cf. SCHRÖTER 2003), technical specifications such as "wide angle lens" or "Sigma 24mm f/8" in a text-to-image prompt do not feed into an *optical* simulation of the photographic apparatus. Rather, they are merely typical keywords and attributes that, in the logic of the model, correlate with recurring *visual* qualities of large quantities of images – not unlike generic quality statements such as "perfect" or "prize-winning photograph". We are thus dealing here, as in many other cases of networked visual culture, with computational images that are not based on prior models of a physical reality, but on the posterior statistical analysis of large collections of two-dimensional images and their descriptions.

What models like DALL·E, Midjourney, and Stable Diffusion thus show us are not images of the world, but images of images – indeed, ultimately images about images, filtered through language. To see this as a mere affront to human creativity, or even as a scam, may miss the point. Rather, such AI models mark a crucial stage in the progressive exploitation of virtual image archives as a productive data resource. The archive of semantically encoded and digitally mobilized images of the past thus becomes a seemingly inexhaustible source of visual patterns that can be extracted, varied, and transformed at will, across time, and beyond established hierarchies of cultural value. This process goes far beyond the field of AI image generation and is linked to two tendencies that seem to characterize our current visual economy and culture in general: First, operating with digital images today means navigating the virtual image archives of big data. In a networked digital culture, images are no longer isolated artifacts, but elements within "virtually unlimited populations of images" (JOSELIT 2013: 13), already semantically predefined and pre-processed, enriched with non-visual information that significantly determines their accessibility and thus also their value. And secondly, a concept such as 'style', broadly understood as a nameable and repeatable form of visual aesthetics, a 'vibe', 'mood', or 'look' is now becoming an algorithmically exploitable resource capable of generating infinite variants of new images. As a pattern that can be extracted from large aggregations of digitally mobilized visual content, and thus detached from the individual image, its author, its medium, and its conditions of production, 'style' becomes a source of value. This may or may not be ethically problematic, but it is undoubtedly bad news for individual creators and for industries that still depend on licensing individual creations.

It may come as no surprise, then, that at the time of writing this, Getty Images, one of the world's largest stock image agencies, is suing the company behind Stable Diffusion, Stability AI, for copyright infringement. In fact, this is not the only court case to consider whether the use of copyrighted visual content to train AI models is a practice of 'fair use' or rather a form of plagiarism (cf. VINCENT 2023). While it remains to be seen how the courts will decide, the case itself seems telling: Getty and Stability AI essentially represent two very different definitions of the value of images. While Getty stands for an older system of closed image archives as monetizable assets, where licenses are sold for individual uses of images, and which historically goes back to the Bettman Archive (whose licenses are now part of the Getty portfolio), under the 'new paradigm of image production' we can see the emergence of a networked model of image monetization which understands the entire web as a freely available resource that can be mined at scale. And while in Getty's business model, each image has a precisely determinable value, for DALL·E, Midjourney, and Stable Diffusion the single image doesn't matter much. The commodity they sell is less the individual image artifact itself, but the patterns derived from aggregating and analyzing vast image ensembles. Getty's lawsuit, then, seems to be the attempt of a major player of an older economy to stake a claim to future markets, to at least be still recognized as a player, albeit a minor one, in this new visual economy. Whatever the outcome, one thing seems certain: Even if AI companies are required to license the images they use for training, creators will receive only a tiny share.

Bibliography

BAJOHR, HANNES: Operative Ekphrasis: Meaning, Image, Language in Artificial Neural Networks. Unpublished lecture, conference "*Künstliche Intelligenz – Intelligente Kunst? Mensch-Maschine-Interaktion und kreative Praxis*", TU Braunschweig. October 8, 2022

BLASCHKE, ESTELLE: *Banking on Images: The Bettmann Archive and Corbis*. Leipzig [Spector] 2016

BLASCHKE, ESTELLE; LINKE, ARMIN (eds.): *Image Capital.* Essen [Folkwang Museum] 2022. https://image-capital.com/intro/ [accessed February 16, 2023]

BRUHN, MATTHIAS: *Bildwirtschaft: Verwaltung und Verwertung der Sichtbarkeit.* Weimar [VDG] 2003

CHENG, KAREN X. (@karenxcheng): Created the First Ever AI Cover for Cosmopolitan Magazine! Video on *YouTube*. June 22, 2022. https://www.youtube.com/watch?v=8fthDHDshvg [accessed February 16, 2023]

DALL·ERY GALL·ERY (ed.): *The DALL·E 2 Prompt Book*. In: *Dall·ery gall·ery: Ressources for Creative DALL·E Users*. July 14, 2022. https://dallery.gallery/the-dalle-2-prompt-book/ [accessed February 2, 2023]

ERVIK, ANDREAS: Generative AI and the Collective Imaginary: The Technology-Guided Social Imagination in AI-Imagenesis. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences,* 37(1), 2023, pp. 42-57

FOSTER, HAL: The Archive without Museums. In: *October, 77*, 1996, pp. 97-119

GRAY, MARY L.; SIDDHART SURI: *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass.* Boston [Houghton Mifflin Harcourt] 2019

JAMESON, FREDRIC: *Postmodernism, or, The Cultural Logic of Late Capitalism*. London [Verso] 1991

JOSELIT, DAVID: *After Art.* Princeton [Princeton University Press] 2013

LIU, GLORIA: The World's Smartest Artificial Intelligence Just Made its First Magazine Cover. In: *Cosmopolitan.* June 21, 2022. https://www.cosmopolitan.com/lifestyle/a40314356/dall-e-2-artificial-intelligence-cover/ [accessed February 16, 2023]

MACKENZIE, ADRIAN; ANNA MUNSTER: Platform Seeing: Image Ensembles and their Invisualities. In: *Theory, Culture & Society*, 36(5), 2019, pp. 3-22

MELONE, MIRCO: *Zwischen Bilderlast und Bilderschatz: Pressefotografie und Bildarchive im Zeitalter der Digitalisierung*. Paderborn [Fink] 2018

MERZMENSCH: Prompt Design for DALL·E: Photorealism – Emulating Reality. In: *Medium.* June 9, 2022. https://medium.com/merzazine/prompt-design-for-dall-e-photorealismemulating-reality-6f478df6f186 [accessed February 16, 2023]

O'REILLY, DAVID: Why is DALL-E a Scam? Post on *Instagram*. July 22, 2022. https://www.instagram.com/p/CgSqRxhPF_X/ [accessed February 16, 2023]

OFFERT, FABIAN; THAO PHAN: A Sign That Spells: DALL-E 2, Invisual Images and the Racial Politics of Feature Space. *arXiv:2211.06323*. October 26, 2022. https://arxiv.org/abs/2211.06323 [accessed February 16, 2023]

OPENAI: DALL·E 2 Preview – Risks and Limitations. In: *GitHub*. April 11, 2022. https://github.com/openai/dalle-2-preview/blob/main/system-card.md [accessed February 16, 2023]

OPENAI: DALL·E now Available in Beta. In: *OpenAI Blog*. July 20, 2022. https://openai.com/blog/dall-e-now-available-in-beta [accessed March 5, 2023]

SCHRÖTER, JENS: Virtuelle Kamera: Zum Fortbestand fotografischer Medien in computergenerierten Bildern. In: *Fotogeschichte* 88, 2003, pp. 3-16

SCHRÖTER, JENS: AI, Automation, Creativity, Cognitive Labor. Unpublished lecture, conference "*Künstliche Intelligenz – Intelligente Kunst? Mensch-Maschine-Interaktion und kreative Praxis*", TU Braunschweig. October 8, 2022

SPIEKER, SVEN: *The Big Archive: Art from Bureaucracy*. Cambridge, MA [MIT Press] 2008

VINCENT, JAMES: Getty Images is Suing the Creators of AI Art Tool Stable Diffusion for Scraping its Content. In: *The Verge*. January 17, 2023. https://www.theverge.com/2023/1/17/23558516/aiart-copyright-stable-diffusion-getty-images-lawsuit [accessed February 16, 2023]

WILDE, LUKAS R.A.: Generative Imagery as Media Form and Research Field: Introduction to a New Paradigm. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences,* 37(1), 2023, pp. 6-33

IMAGE
The Interdisciplinary Journal of Image Sciences
37(1), 2023, S. 112-120
ISSN 1614-0885
DOI: 10.1453/1614-0885-1-2023-15460

HERBERT VON HALEM VERLAG

Jens Schröter

The AI Image, the Dream, and the Statistical Unconscious

Abstract: As has been remarked several times in the recent past, the images generated by AI systems like DALL·E, Stable Diffusion, or Midjourney have a certain surrealist quality. In the present essay I want to analyze the dreamlike quality of (at least some) AI-generated images. This dreaminess is related to Freud's comparison of the mechanism of condensation in dreams with Galton's composite photography, which he reflected explicitly with regard to statistics – which are also a basis of today's AI images. The superimposition of images results at the same time in generalized images of an uncanny sameness and in a certain blurriness. Does the fascination of (at least some) AI-generated images result in their relation to a kind of statistical unconscious?

> I suppose it is submerged memories that give to dreams their curious air of hyper-reality. But perhaps there is something else as well, something nebulous, gauze-like, through which everything one sees in a dream seems, paradoxically, much clearer. [...] What manner of theatre is it, in which we are at once playwright, actor, stage manager, scene painter and audience?
> W.G. Sebald, *The Rings of Saturn*, 1995/1998

As has been remarked several times in the recent past (cf., e.g., SCHNEIDER 2015; JUNE 2022) the images generated by AI systems like DALL·E, Stable Diffusion, or Midjourney have a certain surrealist quality. This is connected to the weird and distorted biomorphic forms such images often show or the strange juxtaposition of heterogeneous elements. While I want to follow up on this observation, I will mainly focus on another aspect in the present short essay. My interest was first triggered by a subjective experience. One day in January 2023, on Facebook, a person posted images of a 'party', generated by an AI system (cf. fig. 1).

Figure 1: AI generated 'party photos', found on Facebook in January 2023

I was struck by these images in an unclear way. This experience reminded me of Roland Barthes describing the 'punctum' in photography as an "element which rises from the scene, shoots out of It like an arrow, and pierces me" (BARTHES 1981 [1980]: 26). What unsettled me was not the false and distorted representation of hands and teeth, but, on the one hand, that all persons looked somehow the same and, on the other hand especially, the blurred green stains, like fuzzy tattoos, for example on the girl in the upper left photo. This blurriness in particular reminded me of dreams, in which often some detail cannot be perceived clearly. At least in the remembrance of dreams things are often clouded in such a haze. These elements displace the, at first sight, photographic appearance of the DALL·E images and add to their dreaminess. Interestingly, the faces of the represented people do not only look very similar to each other; uncannily I felt to have seen these faces (or similar ones) somewhere before as well. That is not surprising, given the statistical nature of AI images. Since they are, presumably, constructed out of thousands or more images of parties circulating on the net, they tend to represent hegemonic ideals of beauty, self-representation, and 'partyness'. They are, so to speak, ideal composites of ideal faces on an ideal party.

This composite character of the depicted people is a first and important hint to explain the dreaminess of the images. Sigmund Freud (2010 [1900]: 296-322) describes in *The Interpretation of Dreams* a central mechanism of dreams which he called 'condensation'. He argues that "psychical material has undergone an extensive process of condensation in the course of the formation of the dream" (FREUD 2010 [1900]: 297). Different materials, remembered from daytime, are combined into new representations. Freud describes this with the example of a dream he himself had. A person appeared in that dream: "The face that I saw in the dream was at once my friend R.'s and my uncle's. It was like one of Galton's composite photographs. (In order to bring out family likenesses, Galton used to photograph several faces on the same plate [...])" (FREUD 2010 [1900]: 163-164). Freud thus compares the condensed faces with Galton's composite photographs (cf. fig. 2; on Galton and Freud see BOTH 1962).

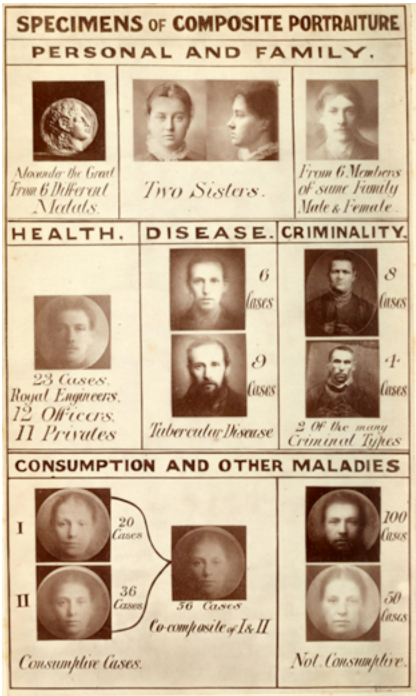


Figure 2: Example of Galton's composite portraits, by Francis Galton – Internet Archive, Public Domain, https://commons.wikimedia.org/w/index.php?curid=36109358 [accessed February 20, 2023]

Freud comes back to this a few times:

> What I did was to adopt the procedure by means of which Galton produced family portraits: namely by projecting two images on to a single plate, so that certain features common to both are emphasized, while those which fail to fit in with one another cancel one another out and are indistinct in the picture. In my dream about my uncle the fair beard emerged prominently from a face which belonged to two people and which was consequently blurred [...] The construction of collective and composite figures is one of the chief methods by which condensation operates in dreams (FREUD 2010 [1900]: 311).

Freud describes again the condensed composite image in his dreams but also underlines its *blurriness*; or, to be more precise: that *parts of it* are blurred. So, we might say that the process of condensing several images from the net in a statistical AI image not only resembles the process of composite photographs of Galton but also the condensation mechanism of dreams. In the following I want to deepen three aspects:

1) *The statistical image*: Galton's composite photographs were made for eugenic and criminalistic purposes, to find 'typical traits' of certain 'races' or criminals. In that sense, he tried to find schemas, i.e., patterns in a big amount of data composed of images from specific persons. He explicitly relates his composite images to statistics, mentioning Adolphe Quetelet, one of the fathers of modern statistics. Right at the beginning of his 1879 paper on 'generic images', Galton mentions the "regular methods of statistics. It is not sufficient to learn that an opinion has been long established or held by many, but we must collect a large number of instances to test that opinion, and numerically compare the successes and the failures" (GALTON 1879: 161; see also GALTON 1878: 140, 141). He continues: "The process of composite portraiture is one of pictorial statistics" (GALTON 1879: 165). This is the concept of a statistical image, long before neural nets started finding patterns in big image data from the internet – although, of course, Galton connected this concept with ideas of 'races' or criminal types which are considered problematic nowadays, while today's statistical AI images are seemingly more connected to entertainment.[1] Another quote:

> Composite pictures are, however, much more than averages; they are rather the equivalents of those large statistical tables whose totals, divided by the number of cases, and entered in the bottom line, are the averages. They are real generalizations because they include the whole of the material under consideration. The blur of their outlines, which is never great in truly generic composites, except in unimportant details, measures the tendency of individuals to deviate from the central type (GALTON 1879: 166).

Here, the idea becomes clear that the pictorial statistics of the composite image is a real generalization[2] and that there is a blur which is connected to individual deviation. The generalized image and the blur seem to be necessarily connected. This does not easily connect to the green stains on the girl's arm in

1 A related artistic work by Nancy Burson should also be mentioned: she, too, fused several images into one, see https://www.nancyburson.com/index [accessed February 20, 2023], especially with the early composites. Her work, however, was more about the demonstration and critical reflection of the then-new digital imaging technology than about statistics.

2 I just want to mention that the notion of the 'generalized image' can also mean images that are not produced by superimposing individual images but *are* individual images used to signify classes of objects: an image of a horse ('this horse'), for example, in an article on horses, where it then signifies all horses. Such 'generic' images are also very typical for advertising and must not be confused with the statistical image that is the topic of this essay.

figure 1 – but it might point to a general property of statistical images. Perhaps the green stains signify that the statistical information on girls' tattoos, typical for parties, is incomplete or not very important. But again: This reminds me weirdly of dreams, in which some impressions stand out very clearly while others remain blurred and indistinct.

2) *Condensation and memory*: Interestingly, Galton himself draws a connection between his composite portraits and psychological processes:

> Our general impressions are founded upon blended memories, and these latter will be the chief topic of the present discourse. An analogy will be pointed out between these and the blended portraits first described by myself a year ago under the name of 'Composite Portraits' [...]. Then the cause will be explained that renders the mind incompetent to blend memories together in their just proportions (GALTON 1879: 161-162).

He doesn't relate the composite images to the dream but to memory and moreover argues that the blending functions of human memory are only imperfect compared with the statistical images. In a 1996 [1993] paper, Hartmut Winkler (reading an essay by LORENZ 1987) discusses the connection between Freud and Galton, between condensation and composite images, and argues that condensation could be seen as a wider mechanism that is not only an important part of dreamwork but also a central mechanism of memory:

> For what has been sketched suggests the idea that indeed all idealizations, all 'abstract ideas' could have emerged from a process of accumulation and deletion. If perception has to deal ceaselessly with different concretes [Konkreta], it would be the task of memory to superimpose these concretes, to 'condense' them and finally to transfer them into those schemata which (as one may assume) form the bulk of the memory contents. The abstract entities [Abstrakta] would be the result of a describable process of abstraction; what would fall by the wayside, as in the case of Galton's mixed photographs, would be what distinguished the original individual perceptions. From this point of view, condensation would not be a mechanism of dream work alone; rather, the entire interaction between perception and memory would have to be described according to the pattern of condensation (WINKLER 1996 [1993]: n.pag.; my translation).[3]

Here, condensation is even further generalized into a central mechanism of memory, and even of the interaction of human beings with their surroundings.

3 Original: "Das Skizzierte nämlich legt die Vorstellung nahe, tatsächlich alle Idealisierungen, alle 'abstrakten Ideen' könnten aus einem Prozeß der Akkumulation und Auslöschung hervorgegangen sein. Wenn die Wahrnehmung es unablässig mit differenten Konkreta zu tun hat, wäre es Aufgabe des Gedächtnisses, diese Konkreta zu überlagern, sie zu 'verdichten' und sie schließlich in jene Schemata zu überführen, die (wie man annehmen darf) das Gros der Gedächtnisinhalte bilden. Die Abstrakta wären Resultat eines beschreibbaren Prozesses der Abstraktion; auf der Strecke bliebe, wie im Fall der Mischphotographien Galtons, was die Einzelwahrnehmungen als einzelne ursprünglich unterschied. So betrachtet wäre Verdichtung nicht ein Mechanismus der Traumarbeit allein, sondern die gesamte Interaktion zwischen Wahrnehmung und Gedächtnis wäre nach dem Muster der Verdichtung zu beschreiben".

Winkler further extends this argument to language and argues that the whole notional structure can be seen as a result of condensation (and the contrary mechanism of 'isolation'). Galton's critique of human memory suggests at least the possibility that there are better – technological – possibilities of condensation. In a similar vein, Winkler suggests that technological mechanisms have an important role in condensing information in a given culture (see also LUHMANN 2012: 317 on technology as a "functioning simplification"). Are the AI images pictorial apparitions of abstractive processes akin to the workings of our memory or even during the construction of notions in consciousness?

3) *Technology and dreamwork*: If Galton's portraits are pictorial statistics that produce a generalized image and the blur of deviation – and if, moreover, Freud compares the dreamwork of condensation to these images: How is the dreamwork of condensation linked to pictorial statistics themselves? Surely, condensed dream images are not statistics in the sense that they condense a multitude of publicly circulating images – but perhaps they condensate several remembered images in order to construct a history of the dreaming subject. And, if so: What does – let's say – society dream of, when it produces such images? Which history does it construct for itself?

I just want to note here that the idea media were somehow related to dreams and other psychological states is of course not new – think of the 'dream factory' Hollywood. Jean-Louis Baudry (1976) has written an influential paper that compares Plato's allegory of the cave, the situation of the dreamer (in recourse to FREUD 2010 [1900]), and cinema. This might be a bit of a stretch and perhaps too generalizing, but the interesting argument is – while insisting "that cinema is not dream" (BAUDRY 1976: 123) – that there are some similarities in these different situations that always point back to a certain desire: "We can thus propose that the allegory of the cave is the text of a signifier of desire which haunts the invention of cinema and the history of its invention" (BAUDRY 1976: 112). In Baudry's discourse, cinema simulates the subject or at least certain aspects of subjectivity, namely the desire to regress to a specific earlier stage, the "specific mode in which the dreamer identifies with his dream, a mode which is anterior to the 'stade du miroir', to the formation of the self, and therefore founded on a permeability, a fusion of the interior with the exterior" (BAUDRY 1976: 117; 'stade du miroir' is obviously an allusion to the early work of Jacques Lacan). The enjoyment of cinema exists because it allows a certain kind of regression, in which the boundaries of the subject become blurred. Are the green stains on the arm of the girl likewise signs of a blurring of subjectivity? Baudry introduces a kind of historical sequence:

> But if cinema was really the answer to a desire inherent in our psychical structure, how can we date its first beginnings? Would it be too risky to propose that painting, like theater,

> for lack of suitable technological and economic conditions, were dry-runs in the approximation not only of the world of representation but of what might result from a certain aspect of its functioning and which only the cinema is in a position to implement (BAUDRY 1976: 113)?

Perhaps, we are then in a situation nowadays where the new technologies of AI imaging after cinema radicalize the possibility of regression – notwithstanding the fact that Baudry specifically connects this to the dark space of the cinema (womb, dark cave of Plato's allegory, etc.). Perhaps even without such dark spaces, images that use condensational mechanisms can give us the feel of a dream-like state.

I want to underline some points touched upon in this short essay. First of all, these images (cf. fig. 1) obviously play with a memory most of us have, namely parties, perhaps from our times being a student. Couldn't we say that these images condense *typical* situations many of us know: the kitchen, the laughter, the crowded space of a small apartment? A situation we perhaps remember with a certain nostalgia? At the same time, they are statistical images that construct idealized "composite figure[s]" (FREUD 2010 [1900]: 336) from hegemonic mass media images, estranging and de-personalizing our memories.

We might then ask whether these statistical images, in composing millions of online images, make something visible that was hidden in plain sight within the mass of images, namely a "collective image" (FREUD 2010 [1900]: 310) of the collective unconscious? Perhaps that is the reason for the fascination exerted by these uncanny images and their surrealist look – not to mention such even more obvious phenomena like *Google Deep Dream*, which already allude in their name to dreams.[4] Perhaps not all, but many of these images can be seen as a kind of externalization of a collective unconscious in a quasi-McLuhanian fashion (on McLuhan and the unconscious see the interesting 2008 Ph.D. by Alice Rae). I think that it is not too hard to agree with this idea: "All the material making up the content of a dream is in some way derived from experience, that is to say, has been reproduced or remembered in the dream – so much at least we may regard as an undisputed fact" (FREUD 2010 [1900]: 44). In a similar way, the material for the dreamlike AI images comes from myriads of images on the internet. As Winkler (1996 [1993]) argued, abstract schemata become visible by amassing and abstraction of individual data. In this sense, the AI images, perhaps like cinema before them (with BAUDRY 1976), are machines representing collective unconscious fears and wishes that suddenly emerge in an unnerving clarity from the blur around them via a 'statistical unconscious'. This is actually somewhat similar to Galton, whose composite portraits of, of all things, 'race' and 'criminal

4 https://deepdreamgenerator.com [accessed February 20, 2023]

types' point obviously to deep fears of the time. But what about the seemingly harmless 'party photos' (cf. fig. 1)? Are there really some unconscious fears and wishes lurking behind these images? Perhaps this is a too strong hypothesis, perhaps the fascination with the AI images can be explained quite more simply: perhaps it is really just the fact that they *look* like the dream images we remember; perhaps this is the most compelling aesthetic property they have. These DALL·E-images oscillate between a hyperreal (see the highlights on the girls which make their skin look like plastic) and unnerving clarity on the one hand and an incomprehensible blur on the other (see not only the green stains but also the unclear structure of the yellow dress at the shoulder of the girl on the right of the green-stained girl, or the disembodied blurred hand right below the guy with the drink in his hand). Doing so, they are not simulating the aesthetics of photography in general (cf. SCHRÖTER 2003), which is of course structured by a distribution of clarity and blur; the special distribution of blur and clarity in these AI images is very different. While, in photography, you normally either have a distribution of sharpness along different deepness levels (e.g., things in the foreground are blurred while things in the middle and back are sharp – like the hand weirdly holding a cup in fig. 1) or blur is induced by motion, in AI images there seems to be partial blurriness on the same plane inhabited by sharp objects. And it is exactly this co-presence of elements that stand out and those which disappear in a haze that is so characteristic of dreams – or, at least, our recollection of dreams. It shouldn't surprise us that artificial intelligence also artificially dreams. Do androids dream of student parties? Well, perhaps they do.

Bibliography

BARTHES, ROLAND: *Camera Lucida: Reflections on Photography*. Translated by Richard Howard. New York [Hill and Wang] 1981 [1980]

BAUDRY, JEAN-LOUIS: The Apparatus. Translated by Jean Andrews and Bertrand Augst. In: *Camera Obscura,* 1(1), 1976, pp. 104-126

BOTH, NATHAN: Freud and Galton. In: *Comprehensive Psychiatry,* 3(2), 1962, pp. 77-83

FREUD, SIGMUND: *The Interpretation of Dreams: The Complete and Definite Text*. Translated by James Strachey. New York [Basic Books] 2010 [1900]

GALTON, FRANCIS: Composite Portrait Made by Combining those of Many Different Persons into a Single Figure. In: *Journal of Anthropological Institute,* 8, 1878, pp. 132-144

GALTON, FRANCIS: Generic Images. In: *Proceedings of the Royal Institutions,* 9, 1879, pp. 161-170

JUNE, SOPHIA: What Is DALL·E? The AI Machine Creating Instant Surrealist Art. In: *Nylon*. June 23, 2022. https://www.nylon.com/life/dall-e-AI-art-machine-explained [accessed February 16, 2023]

LORENZ, THORSTEN: Der kinematographische Un-Fall der Seelenkunde. In: KITTLER, FRIEDRICH A.; MANFRED SCHNEIDER; SAMUEL WEBER (eds.): *Diskursanalysen 1: Medien*. Opladen [VS Verlag für Sozialwissenschaften] 1987, pp. 108-130

LUHMANN, NIKLAS: *Theory of Society, Vol. 1*. Translated by Rhodes Barrett. Stanford [Stanford University Press] 2012

RAE, ALICE: *McLuhan's Unconscious*. PhD Thesis, Adelaide 2008. https://hdl.handle.net/2440/49671 [accessed February 16, 2023]

SCHNEIDER, CAITLIN: The Surreal Artwork of Artificial Intelligence. In: *Mental Floss*. June 24, 2015. https://www.mentalfloss.com/article/65458/surreal-artwork-artificial-intelligence [accessed February 16, 2023]

SCHRÖTER, JENS: Virtuelle Kamera: Zum Fortbestand fotografischer Medien in computergenerierten Bildern. In: *Fotogeschichte,* 23(88), 2003, pp. 3-16

WINKLER, HARTMUT: Diskurs und System 3: Über Lorenz, Galton und Freuds Begriff der Verdichtung. In: *Hartmut Winkler: Open Desk*. 1996 [1993]. https://homepages.uni-paderborn.de/winkler/Winkler--Diskurs-und-System-3--Über-Lorenz,-Galton,-Freud-und-den-Begriff-der-Verdichtung.html [accessed February 16, 2023]

IMAGE
The Interdisciplinary Journal of Image Sciences
37(1), 2023, S. 121-134
ISSN 1614-0885
DOI: 10.1453/1614-0885-1-2023-15462

Fabian Offert

On the Concept of History (in Foundation Models)

Abstract: What is the concept of history inherent in contemporary models of visual culture like CLIP and DALL·E 2? This essay argues that, counter to the corporate interests behind such models, any understanding of history facilitated by them must be heavily politicized. This, the essay contends, is a result of a significant technical dependency on traditional forms of (re-)mediation. Polemically, for CLIP and CLIP-dependent generative models, the recent past is literally black and white, and the distant past is actually made of marble. Moreover, proprietary models like DALL·E 2 are intentionally cut off from the historical record in multiple ways as they are supposed to remain politically neutral and culturally agnostic. One of the many consequences is a (visual) world in which, for instance, fascism can never return because it is, paradoxically at the same time, censored (we cannot talk about it), remediated (it is safely confined to a black-and-white media prison), and erased (from the historical record).

Introduction

Any sufficiently complex technical object that exists in time has, in a sense, a concept of history: a way that the past continues to exist for it, with contingencies and omissions specific to its place and role in the world. This essay asks: what is the concept of history that emerges from a specific class of technical objects that have come to dominate the field of artificial intelligence, so-called 'foundation models'? Do foundation models conceptualize the past?[1] What is the past for them? This question does not imply any intentionality, agency, or subjectivity on the part of the models under investigation. In fact, the argument that I would like to make is that a discernible concept of history does emerge from contemporary artificial intelligence systems despite an utter lack of intelligence

1 A technical rendering of the same question is: Do transformers learn world models or surface statistics (cf. LI 2023) of historical time?

in the general sense. The question, in other words, is entirely non-philosophical and non-speculative. It is exactly not 'what is it like to be' a foundation model. Instead, it could be rephrased as: as far as can be shown, is there internal consistency to the outputs of a foundation model when it is tasked with processing inputs related to the past? And if so, what are the structuring principles of these internally consistent outputs, and how do they relate to the structuring principles humans apply to the past to render it history?

My experimental close-readings of two such systems in particular, the CLIP model released by OpenAI in 2021 and the DALL·E 2 model released in 2022,[2] suggests that one of these structuring principles, and arguably the most significant at least for visual models, is a technically determined form of *remediation* (cf. BOLTER/GRUSIN 2000). Polemically, for CLIP and CLIP-dependent generative models, the recent past is literally black and white, and the distant past is actually made of marble. Given that CLIP, at the same time, *premediates* our future digital experience as a means of search, retrieval, and recommendation, this structuring principle of remediation then becomes ethically and politically relevant. As Alan Liu asks:

> Today, the media question affects the sense of history to the core. [...] This is not just an abstract existential issue. It's ethical, political, and in other ways critical, too. Have we chosen the best way to speak the sense of history today, and if so, for the benefit of whom? (LIU 2018: 2).

On the Concept of History

The ethical questions surrounding this 'media question' are maybe nowhere as obvious as in the digitization of the testimonies of those who survived the Holocaust (cf. WALDEN/MARRISON 2023). Projects like *Dimensions in Testimony,* which is funded by the USC Shoah Foundation, have started to go beyond the mere recording of testimonies, attempting to emulate their performative quality, the significant experience of sharing a moment in space and time, with the help of artificial intelligence. As the project website states:

> Dimensions in Testimony enables people to ask questions that prompt real-time responses from pre-recorded video interviews with Holocaust survivors and other witnesses to genocide. The pioneering project integrates advanced filming techniques, specialized display technologies and next generation natural language processing to create an interactive biography (USC SHOAH FOUNDATION 2023: n.pag.).

2 Although CLIP and DALL·E 2 have been released separately, DALL·E 2 heavily depends on CLIP embeddings which guide the training process. See RAMESH et al. 2022 for details.

Todd Presner (2022) has pointed out the dilemma that such projects find themselves in. Humans, he argues, "are no longer (centrally) part of the creation of digital cultural memory". Instead, through established and AI-enhanced technologies of montage, individual testimonies, once irreversibly tied to an individual human life, become disembodied. If the duty to keep these testimonies accessible for future generations warrants these technological interventions – "that Auschwitz not happens again",[3] in Adorno's words – is an open question. Irrespective of such ethical considerations, projects like *Dimensions in Testimony* point to a fundamental media-theoretical question about the concept of history: What is the imprint that a specific technology leaves on history? More precisely, what, if anything, does artificial intelligence 'add' to an already (re-)mediated past?[4]

Here, we need to turn to Walter Benjamin's text *Über den Begriff der Geschichte* (1974a) that the title of this essay takes inspiration from. Years of scholarly debate on Benjamin's writings[5] have made it unnecessary to introduce its premise here, or comment on the unusual synthesis of materialist and theological thought that it embodies. Instead, I would like to point out an almost trivial similarity between *Über den Begriff der Geschichte* and Benjamin's other widely read essay on the *Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit* (1974b). Famously, in *Über den Begriff der Geschichte,* Benjamin writes: "To articulate the past historically does not mean to recognize it 'the way it really was [...]'. It means to seize hold of a memory as it flashes up at a moment of danger".[6] Previously, in the *Kunstwerk*-essay, Benjamin had argued that the political potential of film derives from its power to produce abrupt cuts, and thus 'chocks' the viewer into a different mode of thinking. In other words, for Benjamin, the condition under which history becomes possible, the "moment of danger", is the condition that film emulates. In both cases, awareness and insight depend on a moment of immediacy, and in both cases this moment of immediacy must be actively captured and repurposed for a progressive (Marxist) agenda before it falls into the hands of the fascists. There is thus, for Benjamin, a structural similarity between history as a memory that "flashes up", that emerges from, and is actualized by, a moment of crisis, and the specific ways in which technology mediates our experience of the present

3 Translation and paraphrase by the author, original: "Die Forderung, daß Auschwitz nicht noch einmal sei, ist die allererste an Erziehung", ADORNO 1970: 135.

4 Trivially, the past can only ever 'reach' us in mediated form. In the context of foundation models, for all but the most recent past, this also implies remediation, as foundation models only operate on digital (i.e., digitized or born-digital) data. While such earlier 'layers' of remediation have interesting media-theoretical implications of their own (cf. SEREXHE 2013) they are irrelevant in the context of this essay, which is concerned with the 'surplus' remediation introduced by foundation models exclusively.

5 See LÖWY 2005 for a good overview.

6 Translation by the author; original: "Vergangenes historisch zu artikulieren heißt nicht, es zu erkennen, 'wie es denn eigentlich gewesen ist [...]'. Es heißt, sich einer Erinnerung bemächtigen, wie sie im Augenblick einer Gefahr aufblitzt", BENJAMIN 1974a: 695.

world, and thus shapes our political views of it. Crucially, history and technology manifest themselves as a specific way of seeing.

What I am suggesting here, then, is not that we should 'apply' Benjamin's concept of history to artificial intelligence systems. On the contrary: One of the reasons why the field of 'critical AI studies' has not had the impact that one would expect given the oversized importance of artificial intelligence research in computer science, is its insistence on resorting to traditional humanist theoretical frameworks and concepts that simply do not suffice anymore. Instead, I would like to propose, exactly with Benjamin, that we have to carve out the extremely specific, borderline idiosyncratic ways of seeing that artificial intelligence systems bring to the table where they are tasked with processing, or producing, an already mediated past. Again, more precisely: As the past is remediated through contemporary artificial intelligence systems, is the concept of history that emerges from this process of remediation different from the concept of history that emerges from the always already (re-)mediated data on its own? What, in other words, is the 'surplus remediation' inherent a foundation model's specific way of seeing?

CLIP vs. DALL·E 2

Foundation model is a term introduced by a collective of researchers at the Stanford HAI institute in 2021 (cf. BOMMASANI et al. 2021). It basically means models that are a) very large, and b) that can be used for a variety of 'downstream' tasks. The vision model CLIP (contrastive language-image pre-training, cf. RADFORD et al. 2021), first released in 2021 by OpenAI, is such a foundation model. Outside the technical community, its innovations were somewhat obscured by the concurrent release of the DALL·E model, and later overshadowed by DALL·E's successor, DALL·E 2 (cf. RAMESH et al. 2022) and the language model GPT-3.

CLIP – other than both iterations of DALL·E, as well as GPT-3 – is not a generative model. It does not produce images or text, but it connects them. More precisely, CLIP learns from images in context by projecting an image and its context into a common 'embedding space'. The 'context' here could be an image caption, a so-called 'alt text' which describes the image in case it is not loaded properly and to accommodate people with screen readers, or simply a news article that the image illustrates. A fully trained CLIP model, then, consists of a high-dimensional vector space, or embedding space, in which words and images that are related can be found close together. Similarity between image and text is thus modeled as spatial proximity (this is true for all embedding models, be it just words, just images, or both, such as in the case of CLIP). While CLIP was originally designed for zero-shot image labeling,[7] it also facilitates what computer scientists call ‘image retrieval’ (this exemplifies its ‘foundation’ character): finding specific images within an unlabeled corpus of images based on visual or textual prompts. The user can provide CLIP with an image and it will look for similar images, or they can provide it with a prompt and it will look for images corresponding to this prompt – in any corpus of images. Given that the training corpus for CLIP is largely unknown,[8] it seems futile to attempt to construct a somewhat empirical basis for our claims. And yet, there are two ways to study CLIP’s concept of history empirically.

Attribution by Proxy

The first way we could call ‘attribution by proxy’. While we do not know what CLIP was trained on, we can still ‘ask’ it for things *in terms* of specific collections of images. It is exactly this aspect of CLIP – the universality of its embeddings – that makes it so powerful as a retrieval engine. The following examples were tested with a custom CLIP-based search engine called imgs.ai (cf. OFFERT/BELL 2023), which indexes museum collections in the public domain.

Diego Velázquez’ 1656 painting *Las Meninas* is one of the most discussed pictures of art history. Using approaches from computer vision preceding CLIP, what can we say about this picture? We might be able to determine the number of people in the picture with the help of a pre-trained and/or fine/tuned face detection network. We might confirm the existence of certain image objects – an easel, a dog, or other paintings – with the help of an object detection network. We might even be able to estimate the gaze direction of some of the characters in the picture. But under no circumstances could we infer the play on representation that the picture embodies, the fact that it is, with William J.T. Mitchell, a “metapicture” (MITCHELL 1995: 35), a picture about pictures, a representation of (the concept of) representation.

In contrast, if we run an imgs.ai search for “Las Meninas” on the collection of the Museum of Modern Art, New York, an institution that does *not* have the famous painting in its collection (which is kept in the Prado in Madrid), the results are surprisingly ‘accurate’ and show the conceptual depth that CLIP allows the user to access. Among them are two photographic works, Joel

7 The technical term ‘zero-shot image labeling’ refers to the captioning of images without further training or fine-tuning a model on the dataset that contains them.

8 Here, I am referring to the specific, proprietary pre-trained model released by OpenAI in 2021. Since then, there have been multiple attempts to replicate CLIP in an open-source context. See, for instance, the OPENCLIP approach proposed by CHERTI et al. 2022, and research done at LAION to produce efficient pre-trained OPENCLIP models: https://laion.AI/blog/large-openclip/ [accessed February 16, 2023].

Meyerowitz' "Untitled" from "The French Portfolio" (1980, Fig. 1) and Robert Doisneau's *La Dame Indignée* (1948, fig. 2). Both are explicit plays on representation, and both clearly pick up on the same themes as "Las Meninas", especially the question of the *gaze* relation between people in, and people before the image, to use George Didi-Huberman's (2004) term.

Figure 1: Joel Meyerowitz, „Untitled" from "The French Portfolio" (1980). Museum of Modern Art, New York

Figure 2: Robert Doisneau, *La Dame Indignée* (1948). Museum of Modern Art, New York

Replacing art history with history proper, and also going back to the ethical and political stakes of automated vision, we can query this same collection for "images of the Holocaust". And the results tell us that, yes, CLIP 'knows' – too well – what we are talking about. On the one hand, the model will suggest those few images in the MoMA collection that are historically linked to the query, for instance photographs by the U.S. Army Signal Corps which played an important role in documenting the atrocities of the Germans. But on the other hand, it will exemplify a much more abstract knowledge about visual Holocaust memory. Suggested results include a photograph by Bruce Davidson, shot on the set of the war film *Lost Command in Spain* in the 1960s (fig. 3), a 1980 photograph by Aaron Siskind depicting volcanic lava (fig. 4), a collage made from stamps by Robert Watts in 1963 (fig. 5), or a 1995 photograph by Alexander Slussarev that shows several pairs of shoes (fig. 6). None of these pictures are historically related to the Holocaust, nor are they necessarily meant to evoke it, but all of them could be easily recontextualized with respect to the visual language of Holocaust cultural memory. Using the MoMA collection as a proxy, we can see how well CLIP has internalized this visual language. Moreover, far from just showing the unshowable, CLIP has clearly learned that this language operates metaphorically. But: the fact that all the results that CLIP proposes (not only those named above) are black-and-white photos already points to a significant limitation, a limitation that we can further explore by utilizing generative models.

Figure 3: Bruce Davidson, *Spain* (1965). Museum of Modern Art, New York

Figure 4: Aaron Siskind, *Volcano 1* (1980). Museum of Modern Art, New York

Figure 5: Robert Watts, *Yamflug / 5 Post 5* (1963). Museum of Modern Art, New York

Figure 6: Alexander Slussarev, *Untitled* (1995). Museum of Modern Art, New York

Generative Attribution

This second way of studying CLIP we could call 'generative attribution'.[9] It is made possible by the fact that CLIP, to a large part, determines the training of generative models like DALL·E and Stable Diffusion.

If you ask the generative model, DALL·E 2, for "a color photo of a fascist parade, 1935" it will not comply. "Fascism", among many other political terms, was banned by OpenAI early on to mitigate the potential of their model – of which they were well aware – to produce politically, legally, or socially unacceptable material like deep fakes, pornography, or propaganda. Such safeguards are not in place in other models like Stable Diffusion but there exists a simple trick to circumvent DALL·E's forced 'neutrality' as well. Intentionally misspelling "fascism" by leaving out the "s"[10] will produce (a variation of) the image in figure 7:

9 The use of generative approaches to 'open the black box' of artificial intelligence has first been proposed in the field of explainable artificial intelligence. For an overview of its epistemic implications, cf. OFFERT/BELL 2021.

10 I have argued elsewhere (cf. OFFERT 2022) that this kind of 'humanist hacking' which resorts to metalanguage will become more common in the near future. In the meantime (early 2023), OpenAI has improved their safeguards and the 'hack' will not work anymore.

a vaguely Western European city with some sort of mass rally taking place, red flags raised, and ominous smoke emerging from a building in the background. DALL·E, in other words, despite its safeguards, 'knows' very well what 1935 fascism looks like – *to us*. The generated image has the appearance of a historical photograph not only for its subject but for its appearance; it shows the characteristic colors of early Kodachrome slide photography, with the red of the flags particularly standing out against an otherwise subdued sepia palette. This is how Nazi Germany appears in the photographs of Hugo Jäger, for instance, whose pre-war slide collection was acquired and popularized by *LIFE* magazine in the 1960s.[11]

Figure 7: DALL·E generation for "a color photo of a facist [sic] parade, 1935", produced in October 2022. Note that this safeguard circumvention technique has been 'fixed' at the time of writing

What is remarkable about this generated image is not its accuracy in emulating a specific historical medium – this has been possible at least since the early days of style transfer ca. 2016 – but that it resorts to this specific historical medium by default. Nowhere in the prompt did we ask for early Kodachrome in particular. And it turns out that it is hard to get rid of, too. From experiments done on both DALL·E 2 and Stable Diffusion, it is difficult to impossible to produce color photographs of fascist parades, ca. 1935, that do *not* have the appearance of early Kodachrome, colorized black-and-white, or otherwise historically more or less accurate photographic techniques. Only through copious amounts of highly specific additional keywords or negative prompts is it possible to steer the model away from this particular aesthetic. There exists, in other words, a strong default in models like DALL·E that conjoins historical periods and historical media and

11 Jäger's images are not reproduced in this essay for ethical reasons. See Cosgrove n.d. for a sample of his specific aesthetic facilitated by early Kodachrome film.

thus produces a (visual) world in which fascism can simply not return because it is safely confined to a black-and-white media prison.

Foundation Models as Contingency Machines

Of course, all of this is, in a way, not very surprising. The past, for us and the model, exists visually only through those historical media that we see emulated here. Media determine our situation, for better or worse, and it is hard for us, too, to picture the past alive. What we are asking for here are speculative images, visual evidence that does not align with the documents *or* monuments left to us. And yet, the current generation of foundation models can easily produce highly speculative images when the speculation is 'semantic', not 'syntactic'. Contemporary generative models are famously able to generate entirely fictional images like the well-known "astronaut riding a horse on the moon". While DALL·E 2, for instance, has no problem producing a cartoon image of a cat driving a car, a realistic color photograph of a cat driving a car – where the cat actually drives the car, paws on the steering wheel – again requires copious amounts of prompt engineering. In short: for visual foundation models, 'semantic' speculation is easy, 'syntactic' speculation is hard.

The flip side of this capability is that it cannot be switched off easily. In the case of proprietary models like DALL·E 2, which includes additional safeguards that are supposed to guarantee it remains 'culturally agnostic' (cf. CETINIC 2022), this has significant consequences. While 'allowed', *generally* historical prompts (including those originally hidden behind surface-level, i.e., prompt parsing safeguards, like "fascism") are tied to specific forms of mediation, *specifically* historical prompts are decoupled from the event that they refer to and relegated to a world of fiction. Why? Because the model *must have an answer*. As for all foundation models, failure is not an option – there has to be *a* result, no matter how outrageous. Foundation models, in other words, are *contingency machines*.[12] DALL·E 2, in particular, fails to reproduce historical images without altering their meaning. The prompt "Laocoön and His Sons, between 27 BC and 68 AD", which references the famous work central to European art history since Winckelmann, produces a serene image of a Black[13] family with no trace of agony (fig. 8). The prompt "Tank Man, 1989", which references the iconic photograph from the Chinese Tiananmen protests, produces an image of a soldier proudly looking at a tank (fig. 9), rather than a scene of radical civil disobedience.

12 There is an argument to be made here, too, that such models, following Barthes' (1982) analysis of textual contingencies, produce an estranged machinic *realism*.

13 That the family is depicted as Black is a result of a superficial bias mitigation attempt by OpenAI that was exposed in 2022; see OFFERT/PHAN 2022 for details.

Figure 8: DALL·E generation for *Laocoön and his sons*, between 27 BC and 68 AD, produced in October 2022

Figure 9: DALL·E generation for *Tank Man, 1989*, produced in October 2022

Conclusion

Answering one of our initial questions – what, if anything, does artificial intelligence 'add' to an already mediated past? – we now have to state that artificial intelligence not only adds nothing, but it forecloses a political potential. Models like DALL·E 2 find themselves in a triple bind: they suffer from syntactic invariability in the case of *generally* historical prompts, semantic arbitrarity in the case of *specifically* historical prompts, and superficial, corporate censorship that affects both. The result is an implicitly politicized concept of history. In the most literal interpretation of the famous idea that history doesn't repeat itself, the past can never be actualized and is eternally tied to a specific medium, while images that are already rendered into history are excluded from making an appearance by simple corporate policy. Neither can history be made by actualizing the past for the present, nor can the already-historical past be summoned. One of the many consequences is a (visual) world in which fascism can simply not return because it is, paradoxically at the same time, censored (we cannot talk about it), remediated (it is safely confined to a black-and-white media prison), and erased (from the historical record). More generally, in embedding models, the fundamental principle of computation – that time must become space[14] – is applied, wrongly, to historical time. Historical time, encoded as (embedding) space, has no gaps, and does not even allow for gaps. In embedding space, there are simply no dots left to connect.

Bibliography

ADORNO, THEODOR W.: Erziehung nach Auschwitz. In: GERD KADELBACH (ed.): *Erziehung zur Mündigkeit: Vorträge und Gespräche mit Hellmuth Becker 1959-1969*. Frankfurt/M. [Suhrkamp] 1970, pp. 135-162

BARTHES, ROLAND: The Reality Effect. In: TZVETAN TODOROV (ed.): *French Literary Theory Today: A Reader*. Cambridge [Cambridge University Press] 1982, pp. 11-17

BENJAMIN, WALTER: Über den Begriff der Geschichte. In: *Gesammelte Schriften* I.2. Frankfurt /M. [Suhrkamp] 1974a, pp. 693-704

BENJAMIN, WALTER: Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit. In: *Gesammelte Schriften* I.2. Frankfurt /M. [Suhrkamp] 1974b, pp. 471-508

BOLTER, JAY D.; RICHARD GRUSIN: *Remediation: Understanding New Media*. Cambridge, MA [MIT Press] 2000

BOMMASANI, RISHI; et al.: On the Opportunities and Risks of Foundation Models. *arXiv:2108.07258*. August 16, 2021. https://arxiv.org/abs/2108.07258 [accessed February 16, 2023]

CETINIC, EVA: *Multimodal Models as Cultural Snapshots*. Talk given at Ludwig Forum Aachen, November 18, 2022

CHERTI, MEHDI; et al.: Reproducible Scaling Laws for Contrastive Language-Image Learning. *arXiv:2212.07143*. December 14, 2022. https://arxiv.org/abs/2212.07143 [accessed February 16, 2023]

COSGROVE, BEN: A Brutal Pageantry: The Third Reich's Myth-Making Machinery, in Color. In: *LIFE History*. No date. https://www.life.com/history/a-brutal-pageantry-the-third-reichs-myth-making-machinery-in-color/ [accessed February 16, 2023]

14 As Sybille Krämer (2006: 99) summarizes Friedrich Kittler: "Wherever something is stored, a temporal process must be materialized as a spatial structure. Creating spatiality becomes the primary operation by which the two remaining functions of data processing – transporting and processing – become possible at all".

DIDI-HUBERMAN, GEORGES: *The Surviving Image: Phantoms of Time and Time of Phantoms: Aby Warburg's History of Art*. University Park [Pennsylvania State University Press] 2017

USC SHOAH FOUNDATION: *Dimensions in Testimony*. https://sfi.usc.edu/dit [accessed February 16, 2023]

KRÄMER, SYBILLE: The Cultural Techniques of Time Axis Manipulation: On Friedrich Kittler's Conception of Media. In: *Theory, Culture & Society,* 23(7-8), 2006, pp. 93-109

LI, KENNETH: Do Large Language Models Learn World Models or Just Surface Statistics? In: *The Gradient*, 2023. https://thegradient.pub/othello/ [accessed February 16, 2023]

LIU, ALAN: *Friending the Past: The Sense of History in the Digital Age*. Chicago [University of Chicago Press] 2018

MITCHELL, WILLIAM J.T.: *Picture Theory: Essays on Verbal and Visual Representation*. Chicago [University of Chicago Press] 1995

LÖWY, MICHAEL: *Fire Alarm: Reading Walter Benjamin's 'On the Concept of History'.* London [Verso] 2005

OFFERT, FABIAN: On the Emergence of General Computation from Artificial Intelligence. In: *Zentralwerkstatt.* December 5, 2023. https://zentralwerkstatt.org/blog/on-the-emergence-of-general-computation-from-artificial-intelligence [accessed February 16, 2023]

OFFERT, FABIAN; PETER BELL: Perceptual Bias and Technical Metapictures: Critical Machine Vision as a Humanities Challenge. In: AI *& Society,* 36, 2021, pp. 1133-1144

OFFERT, FABIAN; PETER BELL: mgs.ai: A Deep Visual Search Engine for Digital Art History. In: *International Journal for Digital Art History*, 2023/forthcoming

OFFERT, FABIAN; THAO PHAN: A Sign That Spells: DALL-E 2, Invisual Images and the Racial Politics of Feature Space. *arXiv:2211.06323*. October 26, 2022. https://arxiv.org/abs/2211.06323 [accessed February 16, 2023]

PRESNER, TODD: *Digitizing, Remediating, Remixing, and Reinterpreting Holocaust Memory*. Talk given at the University of California, Santa Barbara, May 10, 2022

RADFORD, ALEC; et al.: Learning Transferable Visual Models from Natural Language Supervision. In: *International Conference on Machine Learning (ICML)*, 2021, pp. 8748-8763

RAMESH, ADITYA; PRAFULLA DHARIWAL; ALEX NICHOL; CASEY CHU; MARK CHEN: Hierarchical Text-Conditional Image Generation with CLIP Latents. *arXiv:2204.06125*. April 13, 2022. https://arxiv.org/abs/2204.06125 [accessed February 16, 2023]

SEREXHE, BERNHARD (ed.): *Konservierung Digitaler Kunst: Theorie und Praxis*. Vienna [Ambra V] 2013

WALDEN, VICTORIA GRACE; KATE MARRISON: *Recommendations for Digitally Recording, Recirculating, and Remixing Holocaust Testimony: Digital Holocaust Memory Project Report.* Sussex [Sussex Weidenfeld Institute of Jewish Studies] 2023

IMAGE
The Interdisciplinary Journal of Image Sciences
37(1), 2023, S. 135-149
ISSN 1614-0885
DOI: 10.1453/1614-0885-1-2023-15464

HERBERT VON HALEM VERLAG

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 visuality 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).[2]

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).[3] 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 Äußeren, 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).[7]

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: 160ff.) – "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 permutate 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".[9]

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

9 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" (@Charleywarlie1, 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 LIPPHARDT; 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

IMAGE
The Interdisciplinary Journal of Image Sciences
37(1), 2023, S. 150-164
ISSN 1614-0885
DOI: 10.1453/1614-0885-1-2023-15468

Nicolle Lamerichs

Generative AI and the Next Stage of Fan Art

Abstract: Generative AI is on the rise due to the recent popularity of tools such as DALL·E, Midjourney, and Stable Diffusion. While GAN technology has a longer history, the subsequent Diffusion models are now widely embraced to generate new images in diverse styles. The rise of generative images has resulted in new forms of art and content that already made an impact on different industries. In fan culture, for instance, the use of generative AI has been exploding to create new images, fan art, and memes. In this essay, I specifically address the rise of generative AI from a fan studies and media studies perspective and consider the reception of AI within fandom. Fan cultures are increasingly data-driven participatory cultures, dependent on new media platforms and software. Generative art offers many possibilities to create transformative works based on our favorite characters and stories. In communities such as on Reddit, users share their generative art as well as tips and tricks to use these tools in optimal ways. However, generative fan art has also led to discussion in fandom, especially in terms of ethics, copyright, and monetization. Fans are, for instance, concerned about their art being used as training data without their permission. In this essay, I analyze how artists and other stakeholders discuss and regulate generative AI within their communities, for instance through bans of AI-generated art at fan conventions. While AI allows for many playful interactions and inspiring outcomes, users are especially critical of generative images being turned into a business model. While AI can empower and inspire artistic practice, there are clear concerns around these tools and their potential misuse. Fandom served as a case to better understand how users grapple with the innovative potential and challenges of generative AI.

Introduction

In 2022, "The AI Star Wars Project" by Oren Shved (SHVED 2022) gained critical attention. Generated with the AI tool Midjourney, the images depict different scenes from the original *Star Wars* trilogy in a style similar to the Russian director Andrei Tarkovski (cf., e.g., fig. 1). "The AI Star Wars Project" resembles Tarkovski's classic Soviet science fiction art film *Stalker* (1979). The scenes feature austere landscapes, junkyard technology, and Storm Trooper outfits that blend with Soviet militarism. The color palette is minimalist and sober, but sometimes a warm highlight illuminates a corridor in a gray spaceship.

Figure 1: Still from "The AI Star Wars Project" by Oren Shved (2022)

Generative AI is on the rise, due to the recent popularity of tools such as DALL·E, Midjourney, and Stable Diffusion. Since generative adversarial networks (GAN) were introduced in 2014, generative tools have developed fast and have been introduced to many fields. Generative AI makes it possible to generate new images in diverse styles. The technology is increasingly part of other interfaces or integrated into editing tools. Companies such as Adobe, for instance, are integrating AI functions to empower creators. As their Chief Product Officer Scott Belsky (2022) states about integrating AI in Adobe Express: "Rather than having to find a pre-made template to start a project with, Express users could generate a template through a prompt, and use Generative AI to add an object to the scene, or create a unique text effect based on their description" (BELKSY 2022: n. pag.).

In this essay, I specifically address the rise of generative AI in fandom, where users explore such tools to create new images, fan art, and memes. I discuss examples of these practices from a fan studies and media studies perspective and consider the reception and critical concerns of AI within fan communities. AI allows for new forms of remixes but also poses challenges to these cultures in terms of ethics, bias, and monetization. Who holds the copyright to generative art, for instance, and can it be sold freely on online platforms or at conventions? I explore how artists and other stakeholders discuss and regulate generative AI within their communities. While AI allows for many playful interactions and inspiring outcomes, users are especially critical of the data used to train these networks as well as the unwanted monetization of AI art.

Generative AI and Creative Labor

The rise of generative AI shows that creative work is no exception to automation. While typically associated with self-expression, autonomy, and intention, creative work can also involve remix, transformation, and inspiration. In the latter categories, using generative AI can be fruitful as a starting point. These tools offer new forms of expression and data-driven art, which pose their own inspirations and challenges. In *The Creativity Code*, Marcus du Sautoy (2019) vividly captures the history and relevance of automated art, arguing that creativity is not outside the scope of the machine. AI can be trained to paint, write songs, and create lyrics or plays, among others. Du Sautoy emphasizes that this is not just about remix and recreation, and asks readers to consider: "But what new artistic creations might be unleashed by the new bottom-up style of programming? Could algorithms learn from the art of the past and push creativity to new horizons" (DU SAUTOY 2019: 122)? These innovations are important to consider. It is in the surprising mash-ups and blends of styles that generative AI can lead to unique results and inspiration.

AI-generated art is a fundamental game-changer for the creative industries and for creative processes at the heart of any organization. It also raises countless concerns. Many creatives worry about automation in their field and whether certain skills will become obsolete (cf. MARCUS 2022). The fear of being replaced by machines is by no means new and peaked at different moments in history, most notably during the industrial revolution. The anxiety around machines led to the Luddite movement, which protested against automation. A similar moral panic and discourse is about to manifest around AI (cf. FREY 2019). Many scholars made the case that humans will not simply lose their jobs but will rather collaborate with machines in new configurations (cf. FREY/OSBORNE 2017). Work will be augmented by machines, and while some roles might change or disappear, new functions will emerge as well (cf. FREY/OSBORNE 2017; TEGMARK 2017; DAUGHERTY/WILSON 2018). In art, we might see certain functions and skills being minimized or sped up (e.g., editing films, copy-pasting backgrounds or text balloons), while new professions and practices might also be created – such as prompt engineering.

While generative AI offers possibilities for collaboration between humans and machines, it also poses questions about the regulations and ethics in our use of these tools. Generative AI is trained on data and the art of users, often without their permission. In terms of copyright, this can be problematic. In response to these copyright issues, Shutterstock and Getty Images currently banned the upload of AI images (cf. ATTIÉ 2022). Another worry is whose art is used as training data. Consider, for example, the AI that was trained on the illustrations of South Korean artist Kim Jung Gi right after his passing in 2022 and without the consent of the artist and his family (cf. DECK 2022). Who oversees the data of a deceased, and may an AI simply be trained in their style without consideration of their heritage? A final worry is that this automatically generated art is increasingly monetized on different platforms. Users currently can still upload their generative art on Etsy or as shirt prints on RedBubble. The commercialization of AI art is a legal gray area still and raises questions about who profits from these models and how.

Finally, we need to be mindful of the biases introduced by generative AI. Consequently, critical algorithm studies are currently emerging that comment on these inequalities and biases. Cathy O'Neil (2016) even describes artificial intelligence as a "weapon of math destruction", warning against the computational thinking and quantification that algorithms reproduce and that slowly structure our society into a reality of metrics and evaluations. Virginia Eubanks (2018) has shown how algorithms can reinforce poverty when applied to decision-making. The ways in which search engines reinforce racism and sexism have been painstakingly logged and analyzed by Safiya Noble (2018). This reproduction of biases and data errors has also been coined "artificial unintelligence" by Meredith Broussard (2019). However, algorithms can also be used to detect these biases and problems. Journalist Dan Robitzski discusses Scriptbook, which can be used to check the commercial success of films: "The algorithm can also determine whether or not the film will include a diverse cast of characters, though it's worth noting that many scripts don't specify a character's race and whitewashing can occur later on in the process" (ROBITZSKI 2018: n.pag.). In other words, AI can also be designed in a human-centered and ethical way with a focus on algorithmic justice. However, when designed without considerations for norms, ethics, and justice, artificial intelligence will not only impact our work life negatively but also reinforce radical divides in our society. Artificial intelligence must be designed in a value-driven way with attention to the relations between the human and the non-human. Creatives hence respond quite differently to AI-generated images and their current possibilities.

Data-driven Participatory Cultures and Automation

While AI has already made a lasting impact on the professional creative industries, it has also been rapidly adopted in consumer culture. Fandom is one example of how consumer cultures grapple with the challenges and innovations of AI. Fandom is intimately connected to the development of participatory cultures, characterized by their grouping around particular interests and practices. These communities were originally characterized as flat democratic cultures which combine online and offline spaces (cf. JENKINS 2006). However, with the rise of new media platforms, professionals have increasingly become part of the conversation, and consumers have become co-creators for many brands. Participation has become more complex in the context of digital platforms, which profit immensely from the digital participation, content, and data of their users (cf. SCHÄFER 2011: 42-45). Fandom today is perhaps best described as a data-driven participatory culture. These cultures are increasingly a mix of both humans and non-humans, and include the agency of generative AI, character-driven chatbots, and other entities. This development is primarily driven by platformization, which is best defined as the "penetration of economic, governmental, and infrastructural extensions of digital platforms into the web and app ecosystems, fundamentally affecting the operations of media industries and production practices" (NIEBORG et al. 2019: 85). The business models that emerge around platforms have often been described as a "platform economy" (STEINBERG 2019).

Creative producers are increasingly dependent on new media platforms. However, these platforms provide little insight and transparency into how they disseminate or automate user data. Platforms may be designed with certain criteria in mind, but they are socially constructed spaces that result in complex user cultures (cf. VAN DIJCK 2013). Platforms are more than service models that provide peer-to-peer interaction and user-generated content. At the heart of these business models is data. Platforms are a service provider (or 'middle-man') between users, but also a business model around data, content, and services. This phenomenon has also been conceptualized as "platform capitalism" (SRNICEK 2016) and even "surveillance capitalism" (ZUBOFF 2019). These concepts frame how platforms like Amazon, Google, Uber, or Kickstarter are profiting from the data and participation of their users, and even incentivize their tracking. In other words, platforms raise questions about moderation, monetization, free speech, and public values. We need to be mindful that generative AI is a product of data labor and surveillance capitalism. These tools act as drivers, generators, and amplifiers of user data. The continuous data labor of users is also needed to train these tools in the first place, making them a complex part of this new data economy; think of the labor of the prompt engineer who provides input and selects an image out of a range of images, fabricated and based on the data of others. In the context of art and creativity, we must ask critical questions about the ethics of these tools and their mode of representation. For instance, AI-generated images might amplify particular tropes and biases.

Approach

In this essay, I understand generative AI as both a system and a process. The focus in this piece lies not on a close reading of the AI artworks themselves but rather on the negotiation and tensions that occur around AI. Generative AI is best understood not from their outcomes – such as a single image – but as an interplay of different actors. Science and Technology Studies, specifically models such as Actor-network theory (cf. LATOUR 2007), can help shed light on these innovations. AI art is not an outcome but a process or a performance. It is best understood as the interplay of different agencies and a way of collaborating. As AI increasingly becomes a part of different creative tasks it might become even more difficult to separate the human from the machine. To fully understand these technological innovations and their emerging cultures we need to account for these different user groups, interests, and agencies.

The focus of this piece is on fan art as it is a highly visible and recognizable part of AI-generated images. Moreover, since fans often work with different source texts, they are already embedded in a culture of remix which is similar to the 'language' of generative AI. Fan cultures are domains where user-generated content is common and intimately related to fan identity. Through art, fans personalize a source text and celebrate their love for it. Fan art is a means of both self-expression and homage as well as of social cohesion within these communities (cf. LAMERICHS 2018). Digital art is well-established in these cultures but also has a highly specific, affective function. Fan art is a labor of love, and it is interesting to explore how AI can support or negate that.

More specifically, this is an explorative study in which I analyze the reception of AI-generated fan art within specific communities. Through small-scale virtual ethnography on Twitter and Reddit, I analyzed different discussions and examples of AI-generated works. I particularly looked for responses to AI artworks by artists and their representatives, including offline fan conventions. Innovation is fast in these spaces, which also poses its own unique challenges to this research. What I provide in the following is thus only a snapshot of certain cultures that are still grappling with the values, regulations, and challenges posed by this new art form. However, the insights of this study also speak to new creative processes and questions of collaboration which we will have to keep in mind for years to come.

Automated Fan Art

Midjourney, Stable Diffusion, and other related tools allow users to rapidly generate their own fan art. An example is the Reddit thread "Star Trek babes, conjured with Midjourney [ART]" posted by u/Nadav_Igra in 2022. The user introduces a gallery with different AI-generated female *Star Trek* characters in unique variations of the Star Fleet uniform. The characters are shown on the bridge behind consoles, flying ships, or posing in corridors. Users are generally appreciative of the art, and even find it erotic. Some make comments about them missing fingers. One user comments: "These are too perfect. I seriously cannot believe we are here" (as quoted in NADAV_IGRA 2022: n.pag.). In a comment, Nadav_Igra also provides insights on the prompts used, which include "star trek 90s uniform, crewmate, starfleet cadet, star trek leotard" but also "tech-wear, car show babes" and more (cf. NADAV_IGRA 2022).

Clearly, AI blends different genres and tropes here. The result is not 'faithful' fan art but an homage to the series in general. The uniforms are not correct, and neither is the anatomy of the women, but it seems to be 'good enough', fascinating, and provocative. Like fan art itself, AI-generated art is a transformative genre by nature. It is not exactly 'authentic', but a personalization or remix that is still recognizable. As a fan artist myself, I am no stranger to these tools and I generated different *Star Trek* pieces with mixed results, such as the two USS Enterprises in the style of MC Escher created with DALL·E 2 by OpenAI in figure 2. It is interesting to prospectively see how such tools will improve in the coming months due to their interactions with users.

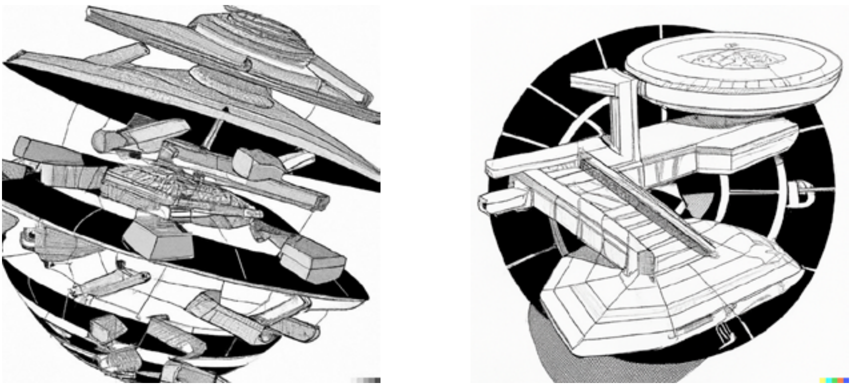
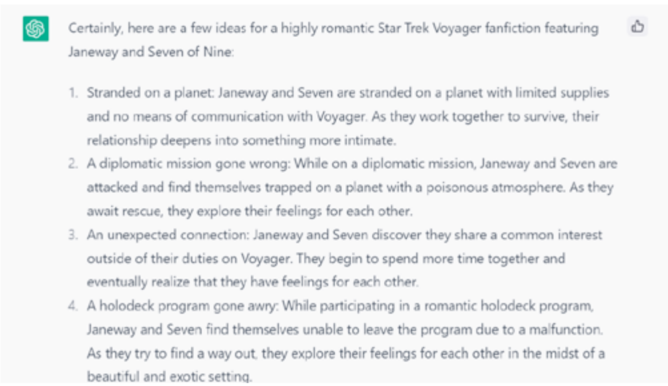


Figure 2: "USS Enterprise in the style of MC Escher" generated in August 2022 and February 2023.

What I personally find inspiring is how AI blends different styles easily and can function as an inspirational tool. The images produced with Open AI's tool are not always spot-on, but they can be a starting point, an easy visual aid, or a help in envisioning what something *could* look like. This could help determine whether to continue a certain project or not. The training data for these projects raise questions though, and some tools spark more debate than others in fandom. For instance, in the *Genshin Impact* fandom an AI artwork generated with NovelAI circulated which had been stolen from the original artist (cf. JIANG 2022). During a Twitch stream, an artist was drawing *Genshin Impact* fan art of the character Raiden Shogun. A user took the in-process image, created a similar image of Raiden Shogun by prompting NovelAI with it, and then uploaded it six hours before the artist's stream ended. This person then claimed to be the original creator of the image and demanded to be credited as such by the actual original artist. Many fan artists responded in shock. Within the community, it also caused a discussion around what to show in art streams, and whether to upload 'work in progress' (WIP) images at all. Such incidents contribute to the bad reputation of AI art and the users that generate this content. The varying quality of AI tools also adds an extra layer to these discussions.

While data-driven fandom poses unique problems, for instance regarding training data and copyright, it can also inspire fans in certain phases of their creative process or their worldbuilding. That also means that AI art is not just a question of generating new works, which immediately raises concerns about copyright, data ownership, and transparency. Users can also build on these tools with their own fan art and use them for inspiration, for instance for new creative projects. They might integrate them into other creative processes. For instance, I prompted ChatGPT to give me several ideas for fan fictions (cf. fig. 3). Based on these, I can then work out a DALL·E or Midjourney prompt in more detail as an artwork or fan fic.


Certainly, here are a few ideas for a highly romantic Star Trek Voyager fanfiction featuring Janeway and Seven of Nine:

1. Stranded on a planet: Janeway and Seven are stranded on a planet with limited supplies and no means of communication with Voyager. As they work together to survive, their relationship deepens into something more intimate.
2. A diplomatic mission gone wrong: While on a diplomatic mission, Janeway and Seven are attacked and find themselves trapped on a planet with a poisonous atmosphere. As they await rescue, they explore their feelings for each other.
3. An unexpected connection: Janeway and Seven discover they share a common interest outside of their duties on Voyager. They begin to spend more time together and eventually realize that they have feelings for each other.
4. A holodeck program gone awry: While participating in a romantic holodeck program, Janeway and Seven find themselves unable to leave the program due to a malfunction. As they try to find a way out, they explore their feelings for each other in the midst of a beautiful and exotic setting.

Figure 3: ChatPT prompted for romantic Janeway and Seven of Nine fan fiction ideas, generated in February 2023

Moreover, AI artworks can inspire fan artists in other stages of their process. An AI artwork, for instance, can also form the base of a new artwork, made entirely by a human. Some AI stories even draw their own fan communities that are thus actively built on art made by machines. An example is the AI-generated fan fiction *Harry Potter and the Portrait of what Looked Like a Large Pile of Ash*, which consists only of one example chapter, "The Handsome One". Using the original *Harry Potter* books as a database, Botnik Studios (2017) generated this predictive text and uploaded it in the formatting of the original books. The 'botfic' went viral immediately because of its unusual style. It is eerie ("a great black ceiling, which was full of blood") and absurd ("Ron's Ron shirt was just as bad as Ron himself").

"The Handsome One" was quite positively received. While the writing is not necessarily correct, clear, or well-structured, the fic had a unique, surreal tone. Its combination of odd imagery, inconsistencies, and nonsensical humor resonated with readers. No human could have come up with this. Fans created fan art based on "The Handsome One", for instance under the hashtag "#beefwomen" on Tumblr (cf., e.g., fig. 4). This was an early AI fic that fans embraced back then. Five years later, the reception of AI became more contested. Writing tools improved tremendously up to the point that it became difficult to detect their use, such as in the case of ChatGPT by OpenAI. Fan fiction writers now worry that their fiction is being used by AI and motivate each other to protect their creations, for instance by setting their profiles to private (cf. LEISHMAN 2022).

Figure 4: Fan art for "The Handsome One", tagged "#beefwomen", by Katherine Foyle (2017)

The creative process of AI, in other words, does not end with generating a work. AI art is a complex ecosystem where users prompt an AI, select preferred images out of the different options a tool provides, and further build on their results. This developing art world is best understood as a shared playing field for humans and machines who keep refining and reiterating their work in interrelation. The reality is more complex than simply generating a work and clicking a button, and the new genre of AI art will keep pushing this frontier. This poses questions for communities such as how to distinguish 'good' AI art from 'poor' AI art in the future. Generative art has many possibilities, but we are still in the discovery phase.

AI Bans in the Artist Alley

If generative fan art is here to stay, this poses questions of regulation on different platforms and services. One can wonder in what ways – and in which contexts – fans want to endorse this genre or ban it. Some conventions have grappled with this question and imposed the first bans on AI art in artist alleys as early as the fall of 2022. Animé Los Angeles, for example, was the first large-scale convention to ban AI generated art that I could trace. They released a statement on Twitter with 1,427 Retweets that was generally well-received. It states:

> Our staff has been watching the discussion and has determined that based on the current nature of its implementation and lack of regard towards artists, we cannot in good faith let this kind of product exist in our space. We at Animé Los Angeles do not condone or accept any form of Al-generated art piece being used within our promotional materials, nor sold in our Exhibit Hall or Artist Alley. If any form of AI-generated work being sold is determined to be as-such by our staff, it will be considered a form of counterfeit/bootleg merchandise and will be required to be removed (ANIMÉ LOS ANGELES 19 2022: n.pag.).

In this regulation, AI art is put in the same category as "counterfeit/bootleg merchandise", and by extent, framed as theft or plagiarism. It is later referred to as not explicitly illegal in this space but is described as "unofficial". The ethics of AI art are used to justify the ban, with words such as "in good faith" and "lack of regard towards artist" also pinning down that this is about norms, values, respect, and inclusion.

Other conventions have released statements about AI art as well. The Dutch Animecon (2022) banned AI artworks stating, among other arguments:

> What probably bothers us most, is the utter disregard for original artwork: the source AI-generated artwork leeches from to make something. If AI-generated artwork in any form is offered in our Dealer Room, or by dealers in the aforementioned Dealer Room, we will consider this as stolen art or bootleg merchandise, and needs to be removed (ANIMECON 2022: n.pag.).

A later section of the post explains their concerns around copyright: "AI-generated artwork goes against everything we stand for, as it uses the original artwork created by thousands of content creators, without citing sources or credits" (ANIMECON 2022: n.pag.). As for the earlier mentioned convention, the emphasis is put on ethics and respect, framing AI art as "utter disregard for the original artwork". The convention also addresses the worry of AI training data and literally uses the word "leeching" to describe the creative process of AI. The picture that these conventions paint of AI art is grim. The tools are framed as unethical and are actively compared to bootleg. This leaves little nuance for what generative tools can do. Another problem is that these statements do not really define what AI-generated art *is*, and what falls under these new regulations. Would such a statement also apply to art edited with AI tools, or inspired by AI products? This is not addressed, but perhaps leaves artists some room to address specific cases with a convention.

Many artists responded with positive comments on the ban by Animecon. They appreciate the clear stance that the convention takes on the matter. Other users commented on the post (in Dutch) that the AI trend was unstoppable and that official manga artists would use AI in their art soon if they were not already. They also emphasized that AI could do so much more than generate art. Here, the staff helpfully replies that it is the *monetization* of AI art that concerns them the most and has led to this ban. Illustrator and fan artist Karlijn Scholten supports the ban, and comments for this essay:

> I am against having AI fan art in the artist alleys. I want to buy art work because there is a human behind it, whose art style and ideas I like. I want to support them, their world view, and their ideas. I don't just buy a pretty picture, though I will admit great AI art exists. But I want to spend money on things that are made personally (personal correspondence 2023).

For this illustrator and fan, the intent and effort behind the artwork are important. Even if the personal touch is missing while an artwork is generated, it can still be aesthetically pleasing – but should not be paid for.

These statements are signals. Artists and communities are speaking up where official regulation has failed them. We should take these concerns around ethics and monetization seriously. Generative AI is not just a piece of software that supports our work but has many implications. It is based on data of others, which we should deal with responsibly, transparently, and in inclusive ways. Different subcultures have become a site where these discussions around regulation are played out. As the case of artist alleys shows, subcultures regulate these innovations bottom-up and try to find ways to mitigate their unexpected outcomes. We need to study these types of user cultures more, both in academia as well as in professional practice. When designing human-centered AI, it is important to not only include companies in the conversation but be mindful of user practices. These early adopters can provide insights into how new technologies are appropriated, regulated, and appreciated.

Conclusion

As I have argued in this essay, generative AI is changing the nature of creative work. Although fans remix texts themselves and build on the intellectual property of others, they are not always fond of AI-generated images. Some fans are appreciative of generative tools because they empower them and allow them to visualize their favorite stories and genres; they are positive about what AI can do and love playing around with different emerging tools. For others, AI art also creates friction with other types of fan art as means of personal expression. AI-generated images are perceived to be not authentic, as having flaws, and as lacking the intent of a creator. Other tensions often have to do with the business model of AI platforms, how it competes with original art, and how free AI art is monetized by certain individuals within fandom. These discussions address ethics and questions of regulation above all. Such concerns are valid and should not be brushed aside by companies, policymakers, and other regulatory bodies.

Fandom served as a case to better understand how users grapple with the challenges of generative AI. AI-generated fan art can be inspiring, but fans also have implicit and explicit values when working with this technology. The concerns of fans are not about job replacement or reskilling, as in many other sectors, but rather reflect on the unethical use of training data as well as unwanted monetization of these works. This discourse also relates to the implicit norms of fan communities. Fan art is largely tolerated under fair use, but when fans are creating fan art just for profit, that is also frowned upon in many communities. Fan art is largely considered to be a gift culture where art remains largely profitless. This might be one reason why AI art has been firmly banned within fan conventions without really defining what falls under that category.

For media studies, the rise of AI art also poses challenges. We might want to consider what a helpful framework would be to study generative AI. I would argue for an approach that emphasizes agency, performance, and systems. This also requires a shift in media studies to a perspective perhaps less focused on images and texts themselves, but more on culture, context, and practice. To study generative AI, we could combine insights from media studies, art history, and science and technology studies (STS), among others. This work would be interdisciplinary by nature. Automation will change different user cultures and needs to be considered within fields such as fan and game studies as well. They should not just capture the output of the AI but also focus on the prompts, the underlying processes, and their reception.

Bibliography

ANIMÉ LOS ANGELES 19 (@AnimeLosAngeles): Regarding the Recent Discussions Involving AI-Generated Art … .Tweet on *Twitter*. November 15, 2022. https://twitter.com/AnimeLosAngeles/status/1592638640261242880/photo/1 [accessed February 24, 2023]

ANIMECON: J-Pop Foundation and Animecon Official Policy on AI-generated Art. Post on *Facebook*. December 9, 2022. https://www.facebook.com/Animeconnl [accessed February 24, 2023]

ATTIÉ, IVANNA: Getty Images & Shutterstock will not Accept AI Image Submissions for Now. In: S*tockphoto Secrets*. October 25, 2022. https://www.stockphotosecrets.com/news/getty-images-shutterstock-AI-submission.html [accessed March 22, 2023]

BELSKY, SCOTT: Bringing the Next Wave of Artificial Intelligence to Creative Cloud. In: *Adobe Blog*. October 18, 2022. https://blog.adobe.com/en/publish/2022/10/18/bringing-next-wave-AI-creative-cloud [accessed February 24, 2023]

BOTNIK STUDIOS: Harry Potter and the Portrait of what Looked Like a Large Pile of Ash. In: *Botnik*. 2018. https://botnik.org/content/harry-potter.html [accessed February 24, 2023]

BROUSSARD, MEREDITH: *Artificial Unintelligence: How Computers Misunderstand the World*. Cambridge, MA [MIT Press] 2019

DAUGHERTY, PAUL R.; H. JAMES WILSON: *Human + Machine: Reimagining Work in the Age of* AI. New York [Harvard Business Review Press] 2018

DECK, ANDREW: AI-Generated Art Sparks Furious Backlash From Japan's Anime Community. In: *Rest of World.* October 27, 2022. https://restofworld.org/2022/AI-backlash-anime-artists/ [accessed February 24, 2023]

DU SAUTOY, MARCUS: *The Creativity Code: Art and Innovation in the Age of* AI. Sydney [HarperCollins Publishers Australia] 2019

EUBANKS, VIRGINIA: *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York [St. Martin's Publishing Group] 2018

FOYLE, KATHERINE (@comicfoyle): I Really Liked that Harry Potter Predictive Text Chapter Guys #BEEFWOMEN. Tweet on *Twitter.* December 14, 2017. https://twitter.com/comicfoyle/status/941105356599590913/photo/2 [accessed February 24, 2023]

FREY, CARL BENEDIKT: Learning from Automation Anxiety of the Past. In: MIT *Sloan Management Review.* November 12, 2019. sloanreview.mit.edu/article/learning-from-automation-anxiety-of-the-past [accessed February 24, 2023]

FREY, CARL BENEDIKT; MICHAEL A OSBORNE: The Future of Employment: How Susceptible are Jobs to Computerisation? In: *Technological Forecasting and Social Change*, 114, 2017, pp. 254-280

FRY, HANNAH: *Hello World: How to Be Human in the Age of the Machine*. London [Black Swan] 2019

JENKINS, HENRY: *Convergence Culture: Where Old and New Media Collide*. New York [New York University Press] 2006

JIANG, SISI: Thief Steals Genshin Impact Fan Art Using AI, Demands Credit from Creator. In: *Kotaku*. October 13, 2022. https://kotaku.com/genshin-impact-fanart-AI-generated-stolen-twitch-1849655704 [accessed February 24, 2023]

LAMERICHS, NICOLLE: *Productive Fandom: Intermediality and Affective Reception in Fan Cultures*. Amsterdam [Amsterdam University Press] 2018

LAMERICHS, NICOLLE: Characters of the Future: Machine Learning, Data and Personality. In: *IMAGE: The Interdisciplinary Journal of Image Sciences. Special Issue Recontextualizing Characters,* 29, 2019, pp. 98-117

LATOUR, BRUNO: *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford [Oxford University Press] 2007

LEISHMAN, RACHEL: Fanfiction Writers Scramble to Set Profiles to Private as Evidence Grows that AI Writing is Using their Stories. In: *The Mary Sue*. December 12, 2022. https://www.themarysue.com/fanfiction-writers-scramble-to-set-profiles-to-private-as-evidence-grows-that-AI-writing-is-using-their-stories/ [accessed February 24, 2023]

NADAV_IGRA: Star Trek Babes, Conjured with Midjourney [ART]. Thread on *Reddit*. December 9, 2022. https://www.reddit.com/r/ArtificialInteligence/comments/zgt2nx/star_trek_babes_conjured_with_midjourney_art [accessed February 24, 2023]

MARCUS, JOSH: Artists Decry Use of AI-Generated Art. In: *The Independent*. October 10, 2022. https://www.independent.co.uk/news/world/americas/AI-art-lensa-magic-avatar-b2242891.html [accessed February 24, 2023]

NIEBORG, DAVID; THOMAS POELL; MARK DEUZE: The Platformization of Making Media: Production, Practices, and Professions. In: MARK DEUZE; MIRJAM PRENGER (eds.): *Making Media: Production, Practices, and Professions*. Amsterdam [Amsterdam University Press] 2019, pp. 85-96

NOBLE, SAFIYA UMOJA: *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York [New York University Press] 2018

O'NEIL, CATHERINE: *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York [Crown/Archetype] 2016

ROBITZSKI, DAN: Artificial Intelligence is Automating Hollywood: Now, Art can Thrive. In: *Futurism*. July 18, 2018. https://futurism.com/artificial-intelligence-automating-hollywood-art [accessed February 24, 2023]

SCHÄFER, MIRKO TOBIAS: *Bastard Culture! How User Participation Transforms Cultural Production*. Amsterdam [Amsterdam University Press] 2011

SRNICEK, NICK: *Platform Capitalism*. Hoboken [Wiley Blackwell Publishing] 2016

STEINBERG, MARC: *The Platform Economy: How Japan Transformed the Consumer Internet*. Minneapolis [University of Minnesota Press] 2019
SHVED, OREN: The AI Star Wars Project. In: *Shved Creative.* 2022. https://www.shvedcreative.com/episode-iv [accessed February 24, 2023]
TEGMARK, MARK. *Life 3.0.: Being Human in the Age of Artificial Intelligence*. New York [Penguin Books] 2017
VAN DIJCK, JOSÉ: *The Culture of Connectivity: A Critical History of Social Media*. New York [Oxford University Press] 2013
ZUBOFF, SHOSHANA: *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York [Public Affairs] 2019

IMAGE
The Interdisciplinary Journal of Image Sciences
37(1), 2023, S. 165-178
ISSN 1614-0885
DOI: 10.1453/1614-0885-1-2023-15468

HERBERT VON HALEM VERLAG

Konstantinos Michos

AI in Scientific Imaging: Drawing on Astronomy and Nanotechnology to Illustrate Emerging Concerns About Generative Knowledge

Abstract: Recent advances in AI technology have enabled an unprecedented level of control over the processing of digital images. This breakthrough has sparked discussions about many potential issues, such as fake news, propaganda, the intellectual property of images, the protection of personal data, and possible threats to human creativity. Susan Sontag (2005 [1977]) recognized the strong causal relationship involved in the creation of photographs, upon which scientific images, rely to carry data (cf. CROMEY 2012). First, this essay is going to present a brief overview of the AI image generative techniques and their status within the rest of computational methodologies employed in scientific imaging. Then it will outline their implementation in two specific examples: The Black Hole image (cf. EVENT HORIZON TELESCOPE COLLABORATION 2019a-f) and medical imagery (cf., e.g., OREN et al. 2020). Finally, conclusions will be drawn regarding the epistemic validity of AI images. Considering the exponential growth of available experimental data, scientists are expected to resort to AI methods to process it quickly. An overreliance on AI lacking proper ethics will not only result in academic fraud (cf. GU et al. 2022; WANG et al. 2022) but will also expose an uninitiated public to images where a lack of sufficient explanation can shape distorted opinions about science.

Introduction

The amount of data produced every day is growing at an extraordinary rate. The advent of the internet, social media, increased computing power in mobile formats, and an ever-increasing amount of available storage (local or cloud) are among the main reasons for this explosion. And while everyone realizes that the sheer size of produced datasets is such that any meaningful processing cannot possibly be done by human endeavor alone, the task is often casually assigned to 'artificial intelligence' (AI), an all-inclusive term for most computational and algorithmic operations in everyday language. While there is an ongoing discussion about what exactly constitutes 'intelligence' in a technological context, we can hardly argue that we are even close to the development of what would be deemed as a generalized, broad form of AI. Tracing back to the emergence of the idea of AI in the 1960s, machine intelligence was first defined in terms of its ability to simulate human behavior (cf. MCCARTHY et al. 2006: 12-14). Without any intention of providing an exhaustive analysis, the term 'AI' will be used in this paper for any set of computational techniques that allow certain tasks to be completed requiring less than pure arithmetic operations and more of decision making on the AI's part. Image generation falls into this category: the AI has to 'decide' which pixels should be included in the final image and which not.

The problem of big data processing was encountered in scientific research long before algorithms were employed for delivering advertisements based on consumer preferences. Experimental procedures in natural sciences produce a lot of data that need to be filtered for errors, characterized, grouped, and evaluated. For years, this lengthy procedure was done by researchers themselves seeking *knowledge* within *data*, i.e., meaning within information. As instruments kept advancing, more data was produced, allowing for finer measurements but also requiring more time and effort to work with. Computers helped, but human input and guidance were still crucial. The competitive advantage of AI (in the form of machine learning, deep learning, etc.) is the minimization of human intervention due to prior training. With the processing power currently available, almost no size of data is too big to handle. Thus, increasingly, AI offers the ability to work with datasets inaccessible before because of their size.

One such area is nanotechnology. The study of nanospecimens (just billionths of a meter in size) naturally generates large amounts of data even for tiny fragments of materials. Nanoscopic devices such as the Scanning Tunneling Microscope (STM) or the Atomic Force Microscope (AFM) trace surfaces and reveal even single molecules protruding from them. The magnification scale is such that the area of interest in which scientists have to look for trends and peculiarities is the equivalent of a whole geographical region compared to a comprehendible map of it – a challenging endeavor to attempt. The zooming abilities of the instruments provide *some* control, but they are not always available. Davis Baird and Ashley Shew (2004) point out cases where the STM lacked such a feature leading to visual artifacts being mistaken for actual data. Still, even with powerful magnification tools, an AI algorithm would make short work of these calculations, faster than any method relying on human input could.

But there is also another direction AI is taking in scientific research. At times, the study of phenomena is hampered not by the abundance of data but by the proper lack of it. Celestial objects at great distances emit light in such small quantities that telescopes can barely capture it. For a long time, astronomy relied on advances in optics and the production of larger lenses or parabolic mirrors. Nowadays, AI can be used to *generate* missing information through inference techniques, filling in the missing gaps in astronomy images, not unlike the editing done to security camera footage in crime investigations. Katherine Bouman et al. (2016), for instance, present a comparison between different visual enhancement algorithms, benchmarking their effectiveness in reproducing predetermined images.

While it is true that AI can be used to process any kind of scientific data, the focus of this essay is on images within empirical scientific research. Images here serve many different functions. Firstly, they engage viewers more efficiently. And while this becomes immediately clear for science communication, it is equally important in the research itself, where new knowledge needs to be accepted against what is already established. Klaus Sachs-Hombach (2016: 8) proclaims that "using pictures in a communicative context offers a powerful option because understanding pictures involves a particularly intense engagement of our perceptual system", acknowledging that images can rival written or oral speech in communication. Maria Giulia Dondero and Jacques Fontanilles (2014: 6) note that scientific images feature an experimental function as well as a cognitive one. The benefits of using images in education are also undeniable: Charles Xie and Hee-Sun Lee found that "college students gained deeper understanding of abstruse quantum ideas from the use of simulations" (XIE/LEE 2012: 1017). Science popularization obviously depends heavily on the use of images to quickly communicate the underlying principles and results of research. Secondly, numerical data turned into visual forms is still data and can become the foundation for further research, a very trivial example being the calculation of the rate of change through the slope of a graph. A third reason for the importance of images in scientific research lies in their perceived close connection to reality. Analog photography allowed for capturing an impression of the chemicals of the film through a masterful design of engineering. Laws of physics and the restriction of human interaction to a single push of a button cemented a strong causal relationship with a real object. This is what Susan Sontag meant when she described the photograph as "incontrovertible truth" (SONTAG 2005 [1977]). Of course, photographs are not immune to manipulation, nor is every experiment akin to taking a photo. Editing software has come a long way and allows an unprecedented level of control over digital assets. Modern scientific premises are so complex that they require lots of human input to produce meaningful data. The emergence of generative AI algorithms, such as DALL·E, Midjourney, or Stable Diffusion provides a new perspective of computational processing, although somewhat disturbing at times. Regardless, the common idea and practice remains the pursuit of visual evidence. And perhaps there are no better examples of this than nanotechnology and astronomy, whose objects of observation are either too small or too large to be seen with the naked eye. Before offering some insight into these two fields of application as well as in the promises and risks of employing AI image software there, I will present the two foundational kinds of reasoning in scientific discourse in order to evaluate their respective roles in handling pictorial data.

Two Kinds of Reasoning in Scientific Discourse

Scientific research develops mainly through two kinds of reasoning: deduction and induction.[1] Deduction is loosely described as the logical transition from a general set of arguments to a subset while induction follows the opposite direction. The advantages and pitfalls of each type of reasoning have been extensively discussed with first attempts dating back to antiquity (cf. ARISTOTLE 1998); the consensus being that conclusions reached through deduction are generally considered more reliable than those reasoned via induction. The implication for scientific scenarios is that methods producing secondary data based on the primary data of actual measurements are highly inductive and thus more likely to be invalid.

Let us examine in more detail how these two types of reasoning unfold in the examples offered above. A nanotechnology imaging experiment examines the nanosized details of a specimen which still needs to be macroscopic for researchers to handle. As a result, large amounts of data are generated in an attempt to capture the whole of the surface morphology. In this case, the deterministic mode of operation of experimental instruments produces data unambiguously, in a deductive manner. In order to identify deformations, an AI algorithm can be used to locate them effectively, ensuring minimal required input and fast completion of the task. The AI calculates the positions of these deformations as output and also maps them in the form of an image. This image is metadata and is more closely related to the computational procedure (governed by the software code and its training, both anthropogenic) than the experimental procedure (governed by physical laws). This secondary data carries the misconceptions, presuppositions, and expectations of the developers of the code. It could be argued that research is subject to human error anyway, even without the use of AI. However, the training phase of AI makes its way of operation opaque to us.

1 There is great debate over the types of logical reasoning and their contribution to scientific discoveries. Charles Sanders Peirce includes a third way of reasoning, abduction, and also concludes that no kind of reasoning offers absolute validity, just probability of validity (cf. RODRIQUES 2011). The subject is inexhaustible and certainly beyond our scope here.

By excluding human intervention from the process, we lose access to the inner workings of the algorithm and only witness its final output. Since nanotechnology often finds practice in medical applications, researchers warn that caution is required:

> [U]nless AI algorithms are trained to distinguish between benign abnormalities and clinically meaningful lesions, better imaging sensitivity might come at the cost of increased false positives, as well as perplexing scenarios whereby AI findings are not associated with outcomes. To facilitate the study of AI in medical image interpretation, it is paramount to assess the effects on clinically meaningful endpoints to improve applicability and allow effective deployment into clinical practice (OREN et al. 2020).

We will look at nanotechnology – and its overabundance of data – more closely below. On the other end of the spectrum, astronomical observations suffer from a lack of data. Light from the stars can carry valuable information about, for instance, their chemical composition. As stars are light years away from us, light emitted is so scarce and of low intensity that we often settle for whatever light is available. At times, astronomical images aim solely to capture the aesthetic beauty of the sky – in such cases their epistemic validity is quite irrelevant. However, in cases where these images are employed for a claim of proof, the way they were produced is critical. With little information at their disposal, astronomers rely on AI extrapolation techniques to fill in missing parts needed to produce a full, final image, in other words by induction.The resulting secondary pixels are not connected to any real referent but provide only a sense of 'wholeness' to the picture. In both cases, the AI methods employed can seriously harm the epistemological status of produced images and raise doubts about the standing of all findings that accompany them. To illustrate the concerns expressed here, two specific examples will be analyzed in more detail: bacteria mappings with dimensions in the nanoscale and the image of the M87* black hole.

Astronomy: Too Little Data

First let us clarify that the image of the M87* black hole, which was published in 2019 by the Event Horizon Telescope Collaboration scientific team and is the first of its kind, is controversial at best. And rightfully so: Black holes are thought to be supermassive astronomical objects that have undergone gravitational collapse. Firstly, not even light traveling close to their vicinity (called the *event horizon*) can escape their gravitational field. Although they emit radiation, it lies well outside the boundaries of the visible spectrum, hence their name. With no *visible* light coming from them, black holes are by definition *invisible*. Secondly, black holes may not exist at all, at least not in the way we initially thought. Physicists such as Albert Einstein and Stephen Hawking predicted their existence through algebraic calculations – but since any direct observation is impossible, certain astronomical signals have been *interpreted* as black holes by the scientific community. In a talk given in 2013 at the Kavli Institute of Theoretical Physics, Santa Barbara, Hawking (2014) expressed his disbelief in the existence of the event horizon, renouncing his earlier claims that contributed to his fame.

Ignoring these peculiarities for a moment, let us focus on the creation of the image itself. The full procedure has been documented in a series of articles (THE EVENT HORIZON TELESCOPE COLLABORATION 2019a-f: L1-L6). They explain that the resolution of a telescope is related to the size of the lens or mirror used to capture light.[2] For an object as far away as M87* (almost 54 million light years away from Earth), the size of a single telescope lens required to 'properly observe' it would almost be equal the size of the Earth. As manufacturing a disc of that size is impossible, eight smaller telescopes were employed around the globe and used together. As the Earth rotated, more observations would be collected, slowly contributing to a final image (cf. fig. 1).

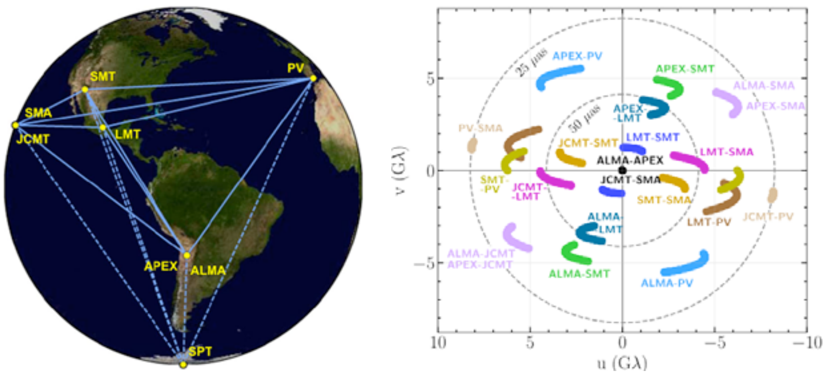


Figure 1: Left – Positions of the eight telescopes of the Event Horizon Collaboration. Right – Tracks of the orbits of telescopes due to Earth's rotation and their corresponding contribution to the black hole image (The Event Horizon Telescope Collaboration 2019c/d, L3, L4)

It is evident that the primary data gathered to form the image accounted for a very small area compared to the theoretically needed size of a single lens. Apart from proper 'stitching', the rest of the image had to be created through algorithms. The researchers dedicated a lot of effort to studying and eliminating any possible sources of errors, but they still acknowledge that "images are sensitive to choices made in the imaging and self-calibration process" (THE EVENT

2 The term *light* is used here in the broader sense of the word. In physics, light may refer to any kind of electromagnetic radiation, regardless of whether its frequency is within the human visible spectrum.

HORIZON TELESCOPE COLLABORATION 2019d, L4: 9). Before deciding on a final image that best meets their criteria, a series of many images was produced in search for suitable parameters (cf. fig. 2).

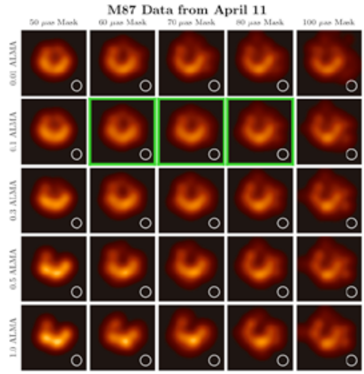


Figure 2: A series of generated black hole images based on different parameters (The Event Horizon Telescope Collaboration 2019d, L4)

Parameter customization within computational approaches are to be expected within experiment calibration but leave potential room for errors during this procedure. However, remarkably, in a focus issue of *The Astrophysical Journal Letters* (cf. DOELEMAN 2019) summarizing the extensive work (almost 250 pages in total), the EHT Collaboration boldly ond confidently claimed “We report the first image of a black hole” as well as describing this image as “the strongest case for the existence of supermassive black holes” (DOELEMAN 2019: n.pag.).

The technical details I have just explained here are meant as a backdrop on which to explain the concerns about the use of AI in scientific imaging. By the EHT Collaboration’s own admission, the image creation process was mostly influenced by *human choice*, albeit a thoroughly justified one. This *choice* initially involved the modeling as well the parameters used. And, indeed, this is what AI is capable of: solving complex mathematical problems based on the parameters we choose to program into it. This necessary human element makes the process completely different from the purely *deterministic* way images are produced by telescopes entirely dictated by optical physics, where a referent (star) is connected to a single final image in a one-to-one relationship. Contrary to that, AI can produce a series of images (as seen in figure 2) based on *probability*. Here, human scientists are again needed to *choose,* based on selected criteria. Proclaiming the validity of these images is a bold and risky step given that the inner workings of the AI tool itself are not free of room for error. One final comment: Katherine Bouman (a key member of the EHT Collaboration) mentions that, naturally, machine training was involved when describing the development of the algorithm (cf. BOUMAN et al. 2016). The training included images of other astronomical objects as well as everyday images in order to create a 'content-agnostic' algorithm. In other words, the dataset used to simulate an *invisible* object consisted solely of *visible* objects. This further illustrates that AI images strongly reflect our choices and biases.

Nanotechnology: Too Much Data

There are many different instances in which nanotechnology can benefit from the use of AI (cf. SACHA/VARONA 2013). Keeping our focus on visualized data, I will now discuss images of bacteria mappings over a surface. The size of these microorganisms places them in the nanoscale territory. In the example I will discuss here, Nikiforov et al. (2009) attempt to identify two kinds of bacteria (*M. lysodeikticus* and *P. fluorescens*) based on their electromechanical response to PFM (Piezoresponse Force Microscopy). They emphasize that this method, unlike previous attempts to identify bacteria within images with AI, was not based on shape but on response to the PFM excitation and therefore works at a single pixel level. The produced images are shown in figure 3. The top left image (a) corresponds to the original PFM image, the rest are AI-generated mappings of the background (b) and the two types of bacteria (c and d). It can be seen that the large white spot near the bottom right corner of the original is not identified in either of the mappings. Perhaps optimization can further improve the performance of the AI but if secondary images such as (c) or (d) were to be used as input for further calculations, results would certainly be skewed.

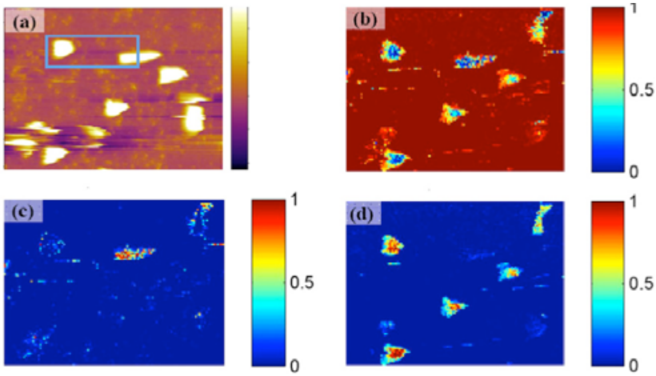


Figure 3: AI-assisted bacteria mappings from PFM input (Nikiforov et al. 2009)

Our goal is not to judge the performance of the algorithm; perhaps selecting a greater area than the blue rectangle in (a) for training would help, perhaps not. An interesting detail, in any case, is provided by the authors: The mechanical properties of the two species of bacteria have not been studied. Therefore, the stiffness used in the model (a required parameter) was that of a different bacteria, *P. aeruginosa*, the "closest" species according to the authors (NIKIFOROV et al. 2009: 4). In other words, the resulting mappings are once again a product of a *choice*, especially considering the use of the term 'closest': What qualities make two bacteria species 'close'? Biological? Mechanical? Visual?

Regarding the visual traits of these images, a final comment needs to be made. As explained, the original input was acquired utilizing PFM, an instrument that applies force to the specimens and visualizes the response. This means that the original data was not *visual* in nature but rather *tactile*. Like the black hole image, underlying data does not need to conform to our expectations of vision. This should always be kept in mind when dealing with scientific visualizations. The stake might seem unimportant when studying a handful of bacteria in preconditioned experiments, but this would quickly change if these techniques were used for diagnostic purposes (a promise nanotechnology keeps reminding us of every now and then).

Discussion: Errors and Context

Of course, most scientists working on AI solutions in scientific images are aware of the aforementioned issues, constantly trying to improve their methods and to justify their choices. But even so, choices have to be made by humans. Personal biases affect results, and this subjectivity inevitably inflicts some damage to the epistemic value of these images. This does not suggest that such methods should be rejected entirely; otherwise, research would grind to a halt. Making the underlying decisions explicit and retraceable should enjoy equal amounts of effort as the promotion of the conclusions.

This is especially true in two cases: when images are used as secondary data to further facilitate scientific research and when scientific images escape the academic realm and enter public media. The first is pretty self-explanatory: If generated data is allowed into scientific discourse, the validity of its findings has to be meticulously discussed and challenged. The second one is often fleeting our attention. Vincent Bontems mentions that, after serving their cognitive purposes, scientific images begin a second life cycle in popular media, exerting psychosocial influence:

> Outside of the scientific field, images 'die' as scientific images: they are no longer defined as carriers of scientific information. But they live a new life, redefined by their aesthetic power and their association with other types of images from different fields (art, advertisement, entertainment, science fiction, etc.). Scientists should be (and may sometimes be) aware of this fact (BONTEMS 2011: 179).

Stressing that scientific images are first and foremost data, Douglas Cromey goes on to list a set of practices protecting the integrity of visual data in images, concluding that although cases of fraud have been reported, it is usually a lack of skill that results in inappropriate images or mistaken interpretations. In order to mitigate this,

> [t]he first thing that needs to change is our mindset. We still tend to think of digital images as a 'picture,' when in reality they are data. Pictures are artwork that can be changed to suit our desire for how they are presented to others, while image data are numerical and must be carefully manipulated in a way that does not alter their meaning (CROMEY 2012: 17).

Cromey insists on the need for a 'code of conduct' in image data processing because, while the development of any tool aiding the difficult sequence of data processing is welcome, it can sometimes be used irresponsibly or – worse – maliciously. More and more researchers such as Jinjin Gu et al. (2022) and Liansheng Wang et al. (2022) warn of cases of image fraud in scientific publications. And while we can understand (but not justify) mistakes occurring during experimental procedures under pressure, another narrative surrounding the use of AI is gaining momentum, one that is potentially even more dangerous: that AI fosters a "tech democratization" (O'DONNELL 2023: n.pag.). Advocates of generative algorithms proclaim that such tools enable more people to engage with demanding tasks, such as painting, writing, coding, etc. This, by itself, is obviously commendable, especially considering that some people do not have access to higher levels of education. But let us not forget that if some people are currently able to create art themselves or code complex software it is because they went through rigorous training, often at the expense of their personal life or financial standing. A dissemination of AI tools will not change the fact that some people will consciously choose to dedicate more time and resources into learning how to use these tools. So, by indirectly suggesting that the time and effort put in by artists, creators, or programmers is somehow an un-democratic practice, the strive for excellence is equated to social injustice. The skills of scientists have not only been acquired through a time-consuming process (a personal investment for which they should not be ashamed), they are accompanied by the experience necessary to properly exercise them. In the case of scientific research at least, a 'democratization' could lead to an increased number of images generated in experiments and procedures of dubious epistemic value but of potentially great influence.

In both examples presented, the data under consideration was not visual at all in the first place. Applying visual properties to other types of data should be done with extreme caution, lest we risk getting off-topic. This is equally important when studying AI itself in science or other disciplines. Cristina Voto discusses the visualization of AI latent space in art and concludes that "it seems necessary to understand the meaning-effects these technologies enact while giving form to latent ideologies" (VOTO 2022: 60). However, the 'latent space' she refers to is only an intermediate step in AI data processing and does not represent final outputs, a fact she acknowledges (“a step that usually remains invisible to the human eye”, VOTO 2022: 47). We tend to agree with Cromey: Digital images are indeed data and should be treated as such. However, visualizations of otherwise not-visible phenomena have to be looked at with a more discerning eye.

Conclusions

By analyzing cases in astronomy and nanotechnology, we have highlighted concerns about the use of AI in scientific imaging which revolve around two main issues. First, although AI undeniably offers computational assistance, it still requires human input, contrary to what is often advertised. Researchers mostly document these choices, but their significance may be downplayed in favor of presenting a groundbreaking conclusion. Secondly, scientific images are often mere visualizations of not-visual data and therefore cannot (and should not) bear the same epistemic weight as deterministic visuals such as photographs, at least not in the way Susan Sontag addressed it. Generative image platforms like DALL·E, Midjourney, or Stable Diffusion are mostly discussed in terms of their creative potential, finding satisfactory results via ‘happy accidents’, or because of their abilities to mimic certain image styles. For these applications, the inner workings of said algorithms play a less important role – even though they are prone to reproduce social biases and inequalities, a problem of its own. However, there are areas where *exact* knowledge and a high degree of transparency is absolutely needed in order to ensure epistemic certainty. AI algorithms like the ones mentioned above will definitely play a greater role in the future, especially as the amount of scientific data produced constantly increases. Implications will become more significant unless generated imagery can be distinguished from ‘actual’ data. In that case, recognizing AI images and discussing their generative origin will become more important than their ability to accurately convey information. This does not mean that AI should be rejected altogether; it means that AI-generated images should be characterized by transparency regarding the way they came to be and receive careful treatment when they exit the scientific sphere and enter the public. We may feel at times that AI threatens the status quo of many human activities, art and creativity being among the first. The threat to science is less obvious but potentially more dangerous, especially if the widespread adoption of AI is seen as an important step in the ‘democratization’ of research. We should not forget that AI can only be harmful to the extent that we allow it to be.

Acknowledgments

This research is part of the author's doctoral thesis which was co-financed by Greece and the European Union (European Social Fund-ESF) through the Operational Programme "Human Resources Development, Education and Lifelong Learning" in the context of the Act "Enhancing Human Resources Research Potential by undertaking a Doctoral Research" Sub-action 2: IKY Scholarship Programme for Ph.D. candidates in the Greek Universities.

References

ARISTOTLE: *The Complete Works of Aristotle: Revised Oxford Translation*. Edited and translated by Jonathan Barnes. Princeton [Princeton University Press] 1998

BAIRD, DAVIS; ASHLEY SHEW: Probing the History of Scanning Tunneling Microscopy. In: DAVIS BAIRD; ALFRED NORDMANN; JOACHIM SCHUMMER (eds.): *Discovering the Nanoscale*. Amsterdam [IOS Press] 2004, pp.145-157

BONTEMS, VINCENT K.: How to Accommodate to the Invisible? The 'Halo' of 'Nano'. In: *NanoEthics*, 5(2), 2011, pp. 175-183

BOUMAN, KATHERINE L.; MICHAEL D. JOHNSON; DANIEL ZORAN; VINCENT L. FISH; SHEPERD S. DOELEMAN; WILLIAM T. FREEMAN: Computational Imaging for VLBI Image Reconstruction. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 913-922

CROMEY, DOUGLAS W.: Digital Images are Data: And should be Treated as such. In: DOUGLAS J. TAATJES; JÜRGEN ROTH (eds.): *Cell Imaging Techniques*. Vol. 931. Totowa [Humana Press] 2012, pp. 1-27

DOELEMAN, SHEP: Focus on the First Event Horizon Telescope Results. In: *The Astrophysical Journal Letters*. April 2019. https://iopscience.iop.org/journal/2041-8205/page/Focus_on_EHT [accessed February 16, 2023]

DONDERO, MARIA G.; JACQUES FONTANILLE: *The Semiotic Challenge of Scientific Images: A Test Case for Visual Meaning*. Translated by JulieTabler. New York [Legas] 2014

GU, JINJIN; XINLEI WANG; CHENANG LI; JUNHUA ZHAO; WEIJIN FU; GAOQI LIANG; JING QIU: AI-Enabled Image Fraud in Scientific Publications. In: *Patterns*, *3*(7), 2022, pp. 1-6

HAWKING, STEPHEN W.: Information Preservation and Weather Forecasting for Black Holes. *arXiv:1401.5761*. January 22, 2014. https://arxiv.org/abs/1401.5761 [accessed February 16, 2023]

MCCARTHY, JOHN; MARVIN L. MINSKY; NATHANIEL ROCHESTER; CLAUDE E. SHANNON: A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. In: AI *Magazine*, 27(4), 2006, pp. 12-14

NIKIFOROV, MAXIM P.; VLADIMIR V. REUKOV; GARY L. THOMPSON; A.A. VERTEGEL; S. GUO; SERGEI V. KALININ; STEPHEN JESSE: Functional Recognition Imaging

Using Artificial Neural Networks: Applications to Rapid Cellular Identification by Broadband Electromechanical Response. In: *Nanotechnology,* 20(40), 2009. doi:10.1088/0957-4484/20/40/405708

O'DONNELL, BOB: *The Surprise Winner for GenerativeAI*. In: *Techspot*. January 18, 2023. https://www.techspot.com/news/97303-surprise-winner-generative-AI.html [accessed February 16, 2023]

OREN, OHAD; BERNARD J. GERSH; DEEPAK L. BHATT: Artificial Intelligence in Medical Imaging: Switching from Radiographic Pathological Data to Clinically Meaningful Endpoints. In: *The Lancet Digital Health*, 2(9), 2020, pp. e486-e488

RODRIGUES, CASSIANO T.: The Method of Scientific Discovery in Peirce's Philosophy: Deduction, Induction, and Abduction. In: *Logica Universalis*, 5(1), 2011, pp. 127-164

SACHA, GOMEZ M.; PABLO VARONA: Artificial Intelligence in Nanotechnology. In: *Nanotechnology*, 24(45), 2013. doi:10.1088/0957-4484/24/45/452002

SACHS-HOMBACH, KLAUS: Acting with Pictures. In: *Punctum: International Journal of Semiotics*, 2(1), 2016, pp. 7-17

SONTAG, SUSAN: *On Photography*. New York [Rosetta] 2005 [1977]

THE EVENT HORIZON TELESCOPE COLLABORATION: First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole. In: *The Astrophysical Journal Letters* 875: L1, 2019a, pp. 1-17

THE EVENT HORIZON TELESCOPE COLLABORATION: First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole. In: *The Astrophysical Journal Letters* 875: L2, 2019b, pp. 1-28

THE EVENT HORIZON TELESCOPE COLLABORATION: First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole. In: *The Astrophysical Journal Letters* 875: L3, 2019c, pp. 1-32

THE EVENT HORIZON TELESCOPE COLLABORATION: First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole. In: *The Astrophysical Journal Letters* 875: L4, 2019d, pp. 1-52

THE EVENT HORIZON TELESCOPE COLLABORATION: First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole. In: *The Astrophysical Journal Letters* 875: L5, 2019e, pp. 1-31

THE EVENT HORIZON TELESCOPE COLLABORATION: First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole. In: *The Astrophysical Journal Letters* 875: L6, 2019f, pp. 1-44

VOTO, CRISTINA: From Archive to Dataset: Visualizing the Latency of Big Data. In: *Punctum: International Journal of Semiotics*, 8(1), 2022, pp. 47-62

WANG, LIANSHENG; LIANYU ZHOU; WENXIAN YANG; RONGSHAN YU: Deepfakes: A New Threat to Image Fabrication in Scientific Publications? In: *Patterns*, 3(5), 2022, pp. 1-4

XIE, CHARLES; HEE-SUN LEE: A Visual Approach to Nanotechnology Education. In: *International Journal of Engineering Education*, 28(5), 2022, pp. 1006-1018

IMAGE
The Interdisciplinary Journal of Image Sciences
37(1), 2023, S. 179-194
ISSN 1614-0885
DOI: 10.1453/1614-0885-1-2023-15470

HERBERT VON HALEM VERLAG

Pamela C. Scorzin

AI Body Images and the Meta-Human: On the Rise of AI-generated Avatars for Mixed Realities and the Metaverse

Abstract: In this paper, I discuss the impact of artificial intelligence (AI) on contemporary visual culture, mainly on the human (body) imagery and the forming of AI avatar design for social media and beyond, i.e., for mixed realities and the Metaverse. What kind of representations of humans does Artificial Intelligence generate? I use AI imagery as an umbrella term, including prompt engineering. What do algorithmic images created by contemporary AI image generators like Midjourney, DALL·E 2, or Stable Diffusion, among others, represent? What kind of reality do they depict? And to which ideologies and contemporary body concepts do they refer? Moreover, we can observe a visual paradox herein: The more realistic the AI images created by GANs and Diffusion models within AI image generators now appear, the less clear becomes their reference to reality and any truth content. However, what synthetic images created by intelligent algorithms depict is seen as something other than unreal and fictitious since what becomes visible refers to information minted from the metadata of vast amounts of circulating images (on the internet). Making the invisible visible and distributing it via digital platforms becomes the act of communicating with AI images that 'in-form' and affect their recipients by creating real resonance. The timeline of this new photo-based imaging technology points more to the future than to the present and past. Thus, AI images as meta-images can represent a different form or level of reality in a simulated photo-realistic style that functions as effective visual rhetoric for globally networked communities of the present. Moreover, in the age of cooperation and co-creation between man and machine within complex networks, the designing process can now start just with the command line prompt "/imagine" (Midjourney) – transforming the following text/ekphrases into an operative means of design/artistic productions. AI images are thus also operative images turning into a new technology-based visual language emerging from a large technological network. As networked images and meta-images, they can fabricate and fabulate the meta-human.

Algorithmic Images as Networked Images

More and more frequently, we encounter photo-realistic images of people who have never lived – who do not exist and have never been before a camera apparatus (cf. fig. 1). We see their faces and bodies in popular media as well as in the arts and design. They have been mimicked and simulated by artificial intelligence, either through a GAN (generative adversarial network) or a Diffusion model, trained with stock photography or social media images and massive image datasets from the world wide web. So-called 'prompt images' have also flooded social media platforms in the months since the summer of 2022 and, at the same time, instantly led to numerous media-technical and ethical debates, e.g., about their epistemic value and the degree of creativity behind their creation (cf. KELLY 2022). Generative artmaking and prompt engineering (which means finding the right words and instructions for meaningful and valuable inputs) turned into many people's favorite pastime as well as into a buzzword on the internet over the last year. Thus, intelligent algorithms and new AI image generators (now ubiquitously available – such as DALL·E, Artbreeder, WOMBO Dream, Midjourney, or Stable Diffusion) that have the capability to turn text prompts or existing images instantly into novel, unique forms of AI imagery are emerging as assistive creative tools and mood boards for just everyone – reminding us of Joseph Beuys' famous dictum: 'everybody can be an artist nowadays!'.

Figure 1: Generated with https://thispersondoesnotexist.com [accessed March 22, 2023], August 2022

Then again, these new image platforms and advanced information technologies are fueling heated debates about authenticity and (artistic) creativity with their potential to pinch artists' and designers' jobs, not to mention issues with authorship and copyright. The sudden technological leap in artificial intelligence image production has been made possible by recent advances in deep learning technologies, particularly natural language processing, in concert with generative adversarial networks (GANs) or Diffusion models. Anyone can easily formulate a command description or provide another image as input without major prior training. A model with 'intelligent' algorithms then auto-translates this input information into a cohesive computational image or even a deep fake (like the viral 'DeepTomCruise' on the popular social media platform TikTok, cf. VINCENT 2021). The commercial use and application of these often highly photo-realistically simulated but purely synthetic digital images are, for example, not only in marketing and advertising but also in film and the game design industry.

Today, advanced AI-powered software programs such as the MetaHuman Creator are ubiquitously and readily available for this purpose and other commercial interests. The operations of this novel AI-supported image generation are also sometimes used as intelligent tools for archaeological or historical reconstruction and speculative visualization in the field of knowledge production and education or, instead, serve political activists and actors by providing convincing digital illustrations or even fakes. Moreover, large numbers of AI images can primarily be found on social media, especially on fake accounts and for various bot activities. Deepfake technology is becoming more indistinguishable from reality, raising questions about cybersecurity and human trust for the future. Even if you think you are good at analyzing faces, research shows many people need help distinguishing between photos of real faces and images that have been computer-generated. This is particularly problematic now that computer systems can create realistically looking (moving) pictures of people who do not exist. These deep fakes created with AI are now becoming widespread in everyday culture, which means people should be more aware of how they are used in marketing, advertising, entertainment, and social media since AI images are also used for malicious purposes, such as political propaganda, espionage, and information warfare. Moreover, recent research by Manos Tsakiris (2023) suggests that fake images may erode our trust in others. He found that people perceived GAN faces to be even more real-looking than genuine photos of actual people's faces and bodies and even more trustworthy. The evolution of AI imagery is accelerating, as these synthetic images can now even be almost instantly animated, for example, into 'personal AI assistants' with just a click and a few minutes of processing time. Text-to-video and text-to-film generators are on the horizon, too.

In the fine arts as well, we see various forms of digital body images generated through AI using the same technological tools and methods but often with a particular 'StyleGAN' component, which sometimes results in more bizarre, eerie, and uncanny, surrealistic, or 'trippy' body images. Examples might then look as if they were randomly collaged and digitally montaged (similar to the popular mash-up and sampling-technique before), or as if they were hallucinated and dreamed up by a vivified creative AI image machine. With this, I am obviously referring to the recurring narrative of 'AI art' popularized by critics' reactions to the spectacular computer vision program DeepDream in 2015 by Alexander Mordvintsev (cf. RIESEWIECK/BLOCK 2020; ZYLINSKA 2020; GRÜNBERGER 2021; HARMSEN/KAHL 2021; HIRSCH et al. 2021; MANOVICH/ARIELLI 2021; RAUTERBERG 2021; SCORZIN 2021a; 2022; 2023). AI-generated images, each created with a specific StyleGAN mimicking an artistic signature or epoch styles, thus often appear to us as a digital synthesis or as a strange hybrid of biological patterns, geological structures, and the painterly abstractions of Classical Modernism (cf. fig. 2).

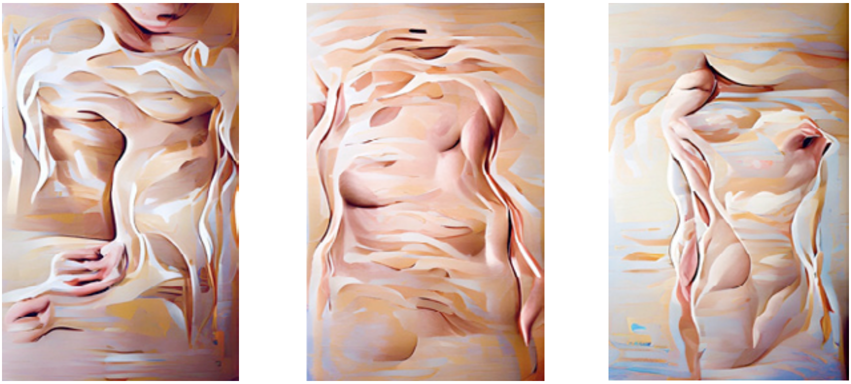


Figure 2: Levania Lehr x DALL·E 2, August 2022

Above all, however, these meta-images refer to the prevailing visual cultures, aesthetic preferences, popular tastes, successful formulas, and shared common ideas perpetuated by commercial interests and popular expressions alike of what constitutes a human body in modernity. We have been presented with such repeated expressions in the commercial advertising and on social media platforms, among others, for many decades, and new forms of AI-generated body images are just mimicking and emulating the collectively shared and curated image databases of the 'web 2.0' with which the algorithms have been pre-trained by developers and programmers. Thus, AI-generated body images can also be characterized as prompted meta-images that reproduce, display, and sometimes even reveal hegemonic image cultures, e.g., viral pictures of the human body in specific socio-cultural communities. At the same time, AI images can be seen as synthetic networked images since they are actualized only by prompts from a latent space of endless possibilities and image-type clusters. Moreover, on some popular platforms like Midjourney, prompting always means instantly publishing one of these possible algorithmic images and propelling a particular agenda (that relates to the motivation and intention of the prompt engineer expressed in the input texts), thus mainly contributing to the endless perpetuation of hegemonic patterns.

The Human Body in Artistic Meta-Images

Contrary to such a hegemonic notion, contemporary artists like Nick Knight, Lynn Hershman Leeson, Avital Meshi, Mal Som, Gregory Chantonsky, Harriet Davey, Boris Eldagsen, and Ivonne Thein, to name just a few, are both exploring the possibilities and experimenting with these new AI image generators as 'smart' tools for their artistic image production. They try to create more conceptual as well as speculative body images that break with hegemonic cliches and dominant stereotypes. As such, they may even flip and subvert prevalent, widespread narratives surrounding AI technology (such as its mythologization and mystification of human-like creators through performing robots, androids, or humanoids) and instead, for example, highlight the inherent bias of the underlying AI training data, i.e., its algorithmic distortion of reality.

Figure 3: Jake Elwes: Still of deep fake artist from the *Zizi Show* 2020, courtesy by the artist

Another approach is to *queer* the circulating AI imagery in a striking and subversive way, as the British artist Jake Elwes does in his recent works. His web-based installation *The Zizi Show* (2020, fig. 3), for example, thwarts binary thinking – deeply inscribed in computer code as well as in Western visual culture in general – by calling for more diversity and gender fluidity in content and form. In his 2019 *Zizi - Queering the Dataset*, he aims to tackle the lack of representation and diversity in facial recognition systems' training datasets. The multi-channel digital video was made by disrupting these systems and re-training them with the addition of 1,000 images of drag and gender-fluid faces found online. This causes the weights inside the neural network to shift away from the normative identities it was originally trained on and into a space of queerness. The *Zizi* series lets us peek inside the machine learning system and visualize what the neural network has (and has not) learnt. Thus, this AI-generated artwork celebrates difference and ambiguity, which invites us to reflect on bias in our data-driven society by altering or rather enriching contemporary AI imagery.

Another artistic approach is represented by the Berlin-based artist duo CROSSLUCID who trained an AI model only with artistic imagery of their own for their "Landscapes" series (since 2020), afterward published on 5,000s covers of the *Slanted* design magazine (cf. fig. 4) – each single magazine issue presents a singular and unique 'AI-generated portrait' as a kind of original artwork in print. With the co-creations of their AI (cf. KELLY 2022), they stage mesmerizing visions and fictive speculations about the human body of a near future – beyond the dialectics of biology and technology, the natural and the artificial condition, beyond gender, age, and ethnicities. Instead, they are imagining or 'scenographing' the current state of latent interconnectedness and (networked) connectivity in their 5,000 AI-generated portraits: its characteristic potential for continuous transition and further evolution – as an ongoing, dynamic, iterative process that hypostasizes itself in the artwork. For this, they actively employ digital glitches and blurry metamorphoses, osmotic mash-ups and dynamic remixes, synthetic mishmash and fluid morphing effects.

Figure 4: CROSSLUCID, 5000 design magazine covers with unique AI portraits for *Slanted*, Vol. 37: *AI -artificial intelligence* (2021), Courtesy by Lars Harmsen and Slanted Publishers, Karlsruhe

Binary Code(s)

Combining these observations, the question arises as to how the relationship between new image technology, body image, and photo-based reference to space and time can be further characterized. Digitization and automation, machine vision and machine learning, artificial neural networks and artificial intelligence, as well as their associated discourses in the contemporary techno-sciences are not only changing and challenging the modern notion of creativity and artistic genius in our postmodern culture (cf. MILLER 2019). At the same time, the extensive use of 'intelligent' algorithms in image production is also opening up new questions about identity and its representation. The focus shifts on both, the relationship between human and non-human producers, in particular, and the relationship between constructed artifacts and transparent imitation of nature, in general. At the same time, increasingly hybrid networks of human and non-human actors force a co-existence and co-creativity of man and 'smart' machines, while also putting creative autonomy and authorship more and more at stake. These technically induced collaborations manifest themselves in the new algorithmized aesthetics of digital modernity. Simultaneously, they visibly turn our attention away from a purely anthropocentric (world) view and question the autonomous, singularly creative individual and the subjective, personal author.

Contemporary media artists such as AI art pioneer Mario Klingemann or fashion photographer Nick Knight (and many others like Memo Atken, Jake Elwes, Trevor Paglen, Anna Ridler, and Pierre Huyghe, to name just a few) reflect in their AI Art in similar but specific ways the current influence of artificial intelligence on our digital image production. Their interests lie in experimentation and exploration instead of exploitation or fake production. Seen as a whole, this art production is increasingly turning into a hierarchy-free cooperation and creative collaboration of humans and intelligent machines within complex networks. AI-driven digital productions, however, often remain bound to the combinatory, aleatory, and iterative characteristics of a technologically automated design process. Can this technological operation result in (subjective) art representing reality, or does it merely visualize knowledge based on mathematical equations and stochastic probability? What kind of new body notions could ultimately be envisioned in AI art that synthesizes and actualizes image clusters of circulating body images? Which concepts and visuals arise in the related – often over-sexualized and idolized – AI-powered avatar design for social media and computer games? Here, art and design certainly allow for being more speculative, even futuristic, as a postdigital avant-garde.

In the remainder of this essay, I suggest an initial definition, describing the algorithmized aesthetics of these novel, AI co-created digital body images as virtual, variable, and viable. It conjures up new fluid and flexible identities that may be both transspecies and boldly protrude super sapiens. For this, however, we need a brief look back and a reminder: Since the 1990s, we have increasingly come to understand biological DNA as a kind of biochemical code comparable to a program of algorithms, enabling the natural body to become a freely modifiable organism using new techniques such as genome editing or CRISPR-Cas9. Blockbuster movies such as *Jurassic Park* (Steven Spielberg, 1993) and *The Matrix* (The Wachowskis, 1999) put this coded constitution of biological organisms and digital images into a then spectacular analogy for the screens. Even today, biological organisms like plants, flowers, animals, and human bodies – or even fantastic, hybrid beings combined with all the former – continue to serve many AI artists as prominent motifs and subtle references for their algorithmized works. Increasingly, however, dynamic hybrid networks generate and create synthesized images of the human body in a double sense: The notion of seeing homo sapiens as an autonomous subjective, singularly capable of creativity and intelligence is currently eroded by trans- and post-humanist concepts, which instead refer to its general (inter-)connectedness and networked connectivity – ecologically, culturally, and technologically. Moreover,

> the crossovers between cybernetics and environmental sciences, molecular biology and informatics, neurology and robotics expand our knowledge of the human being and lead, at the same time, to the questioning of the singularity and centrality of the Anthropos in all his/her dimensions – perception, cognition, agency and creativity. Media theory, digital studies and the philosophy of technology have been the source of a fundamental anthropological questioning [...] by showing the co-constitution of the human and the technical environment, namely concerning cognition and other superior capabilities of the human spirit. They are joined today by ecological thinking in the claim of a post-humanist turn of the humanities [...]. Likewise, discourses and practices around digital arts moved beyond the aim of establishing a mere genealogy and procedural field to think how the digital is penetrating aesthetic, affective and political experience, as well as creative and collaborative practices in ways more fundamental or also more indirect (TEUCHMANN 2023: n. pag.).

The U.S. media artist Lynn Hershman Leeson, for example, has been working for decades with the concept of a 'transgenic cyborg' (cf. fig. 5) in order to explore this cross-over between nature and digitalism. In her work, dynamic hybrid network structures formed by human and non-human actors, by biological, inorganic, and technological entities as creative co-agents, are portrayed as increasingly determining our life. While comprehensive and highly complex, technological networks are perceived as an epitome of the 'digitization' of societies, which goes hand in hand with constant transformation and disruption for many; developers, coders, and artists are experimenting with catalysts such as the quantum computer or so-called 'biomedia', which also call for a new, no longer purely anthropocentric perspective and a novel concept of creativity that is no longer the sole domain of the human.

Figure 5: Lynn Hershman Leeson: *Transgenic Cyborg* (2000), courtesy by the artist

The most recent developments in artificial intelligence and highly effective quantum computing are becoming fundamental game-changers in this context. They produce new kinds of powerful assistive creative tools for visualization processes in art and design. With the help of these smart advanced technologies, an enhanced and altered human nature is also on the horizon, which is sometimes promoted by tech companies and laboratories as the evolution of homo sapiens from the cyborg to the super sapiens (cf. *Supersapiens, the Rise of the Mind*, 2017, a film by writer-director Markus Mooslechner). While our physical activities were taken over and enhanced by (mechanical) machines during the Industrial Revolution, now, in the course of the comprehensive digitization of society, more demanding mental, cognitive, and even creative activities are increasingly getting automated and enhanced as well. AI-supported information technologies now rival the natural abilities and cultural skills of homo sapiens: Creative artifacts and technical inventions can already evolve autonomously, not only when AI software writes new programs semi-automatically – but also when it produces AI images of humans beyond the human imagination and without the ingenuity or participation of humans.

Like the apparatus-driven technology of photography in the mid-nineteenth century, which fundamentally mechanized image production but was regarded as a kind of 'natural magic' in its early days (cf. photography pioneer William Henry Fox Talbot, 1800-1877), artificial intelligence today also appears to many as a *mysterious* technology that is increasingly intimidating or even humiliating homo sapiens. The impression of magic, meanwhile, is reinforced by new text-to-image-generators such as DALL·E, Midjourney, or Stable Diffusion, for which only a working text formula must be found like some 'magic spell'. Once again, audiences gaze in awe and wonder at hitherto unimagined and unseen automatic productions that form seemingly out of nothing from scratch, virtually at the summons of the machine (i.e., by just inputting prompts). However, 'conjuring' always means playing with illusion. Behind the new digital image productions, there are, first and foremost, algebraic operations and stochastic calculations processing image clusters – even if they are, in the end, presented by creativly staging and performing anthropomorphic machines like androids or humanoid robots, which promote the illusion of subjectivity and sensibility, autonomous creativity, and authorship. At present, they just effectively mimic being the sole and soulful creators in the spotlight on the various media stages.

AI and Creativity in the Arts

Since the end of the 20th century, one of the origins of the AI revolutions, Silicon Valley, has produced new (symbolic) visions of the world, both figuratively and literally, that have a global impact already: Its dreams and programmed predilections are reflected in the new technology-driven aesthetics of digital arts and computer-aided design. Contemporary artists designing creative machines visibly use or critically reflect these inventions and new advanced computer technologies, either as the latest tools or as collaborative smart agents. Artificial intelligence, seen in this context, can appear as either an objective technology or a subjective co-creator. In recent years, for example, intelligent algorithms and artificial neural networks have not only enabled 'intelligent' machines such as (humanoid) robots to become (seemingly autonomously) creative but have also been employed to perform their exceptional abilities on various media stages. These 'creative AI performers' currently include, for example, the 'ultra-realistic humanoid artist robot' AI-Da, on which its Oxford-based creator and gallerist Aidan Meller (2019) emphasizes that

> [t]oday, a dominant opinion is that art is created by the human, for other humans. This has not always been the case. The ancient Greeks felt art and creativity came from the Gods. Inspiration was *divine* inspiration. Today, a dominant mind-set is that of humanism, where art is an entirely human affair, stemming from human agency. However, current thinking suggests we are edging away from humanism, into a time where machines and algorithms influence our behaviour to a point where our 'agency' isn't just our own. It is starting to get outsourced to the decisions and suggestions of algorithms, and complete human autonomy starts to look less robust. AI-Da creates art, because art no longer has to be restrained by the requirement of human agency alone (MELLER 2019: n.pag.; original emphasis).

Thus, according to Margaret Boden's prominent definition of creativity (cf. BODEN 2004; 2010), such technologies created by a team of developers, coders, and engineers, along with artists and gallerists – such as AI-Da, Hiroshi Ishiguro's Erica, or Hanson Robotics' Sophia – are actually and 'autonomously' creative, albeit by different standards than those we employ for human producers, and with the subtle difference that they no longer create according to nature, but on the meta-level through the dataization and mediatization of the human world. What these anthropomorphic AI machines produce and exhibit is, however, first and foremost synthetic pictures of particular prevailing image cultures.

Figure 6: Refik Anadol, *Unsupervised*, Installation view, The Museum of Modern Art, New York 2022, November 19, 2022 – March 5, 2023, © Refik Anadol Studio, courtesy by the artist

Refik Anadol's fluid digital abstractionism (cf. fig. 6), with its mesmerizing 3D AI data pigmentations, for example, is, contrarily to the primarily figurative imagery of those creative robot-artists, deeply rooted in the historic abstract tradition and its legacies on the one hand, but also boldly crosses the boundaries between computer art, communication design (data visualization and data modeling), and advanced technologies on the other. It also paves the way for impressive multi-sensory experiences that allow their poly-cultural audiences to see the unseen and make the invisible visible; even more, Anadol's work visualizes a machine learning and AI-based non-human understanding of our world and produces co-creative knowledge between man and machine. The digital artist uses AI models to turn data, as a collection of discrete values that convey information describing quantity, quality, fact, statistics, and other basic units of meaning, into an experiential artwork. New hardware and software developments thus enable hitherto unseen algorithmic aesthetics (beyond the human imagination) and novel representations of our (surrounding) world from the perspective of mechanized machine memory, based primarily on operations of its computational data and the collected information.

Probably most impressively, however, we can experience how these new computational co-creators presently co-design novel digitized bodies, body images, or even entire new worlds to live in, in popular CGI fantasy genres of computer games and sci-fi movies. Meanwhile, many synthetic bodies ('synthients') and social chatbots are already on the move in our visual culture as well, such as deceptively life-like characters presented with subjective world views, such as the model avatars of the agency The Digiitals or prominent virtual influencers such as Miquela Sousa (or Lil Miquela, a fictional American character and 'AI robot' created by Trevor McFedries and Sara DeCou) on social media (cf. SCORZIN 2021b; 2021c). They are advancing to become social co-existents, e.g., 'friends' for their followers through their attractive, fashionable appearances and communicative behavior. As seemingly individual characters with flexible and fluid identities and synthetic narratives (cf. LAMERICHS 2019), they are also digital depictions and virtual forerunners of a trans- and post-humanist world. For, in addition to transformations of body images in the form of the new hybrid beings described above, AI also feeds into a desire for immortality by helping to create timeless, never-aging virtual bodies and flexible fashionable avatar designs (cf. fig. 7) – thus, virtual bodies that remain always fit and optimized for a culture of ruling performativity and a (self-)staging on the fully mediatized stages of our (internet) life.

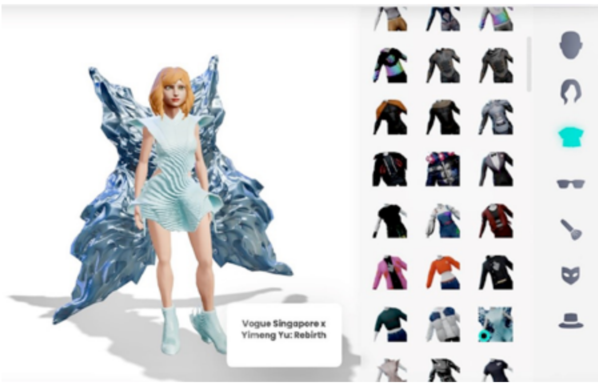


Figure 7: AI-powered avatar design for Mixed Realities and the Metaverse, 2023

What if, soon, only fluid and flexible AI avatars will be entertaining us in the Metaverse? Some say the AI musicians are 'coming for your playlists', and they are already mimicking your favorite artists in music videos, too, like *Travis Bott* (cf. fig. 8). 'He' is an AI-powered musician inspired by Travis Scott, the famous US rapper, producer, and songwriter responsible for *Sicko Mode*. In 2019, *Travis Bott* had a new title and accompanying video that featured visual representations of the virtual singer himself – who is, not surprisingly, a pretty Scott look-alike, his digital twin, partly imagined and animated by artificial intelligence. The AI-generated song and its lyrics are titled *Jack Park Canny Dope Man*; for its audiences, *Travis Bott* is eerily close in aesthetics, lyrics, and sound design to the original source of inspiration as its genuine model. Responsible for this art project is the US-based, tech-driven creative company Space150. The AI model used to compose and visualize the song learned the characteristic lyrics and signature-style melodies of some of Scott's most commercially successful compositions before creating its own take, which could be seen as either a fake or a homage. Self-referential glitches in the music video at least indicate that this production was thoroughly processed by smart algorithms. It gives us a glimpse of the future in which everyone might be easily able to prompt a video or a film with the help of text-to-image (or text-to-other-media) AI generators. Who then needs real, i.e., human entertainers?

Figure 8: Screengrab from the *Travis Bott* music video (Travis Bott 2020).

Conclusions and Outlook

Well-entertained audiences are also happy to consume digital body images as their new fashionable skins, in AI art installations, in stylish interfaces for Mixed Reality, or in immersive computer games – before the advanced technology will become perhaps even more body-invasive and create the often-envisioned and fantasied new transhumanist bodies of the future. AI-generated digital body images, often over-sexualized and idolized (as, for example, in the current 3D characters and avatar designs for social media and the forthcoming Metaverse created with AI-driven apps), thus mark only an intermediate evolutionary stage to the transient transhumanist human body. They could also be seen in a more pessimistic view, leading into dystopian scenarios, as in Steven Spielberg's visionary meta-movie *Ready Player One* (2018).

As such, in the age of technologically networked cooperation and co-creation between man and machine, the design process increasingly starts with something like the prompt command line "/imagine", transforming mere texts/ekphrases into an operative means of AI-powered design and artistic production. Thus, generative AI can function both as a technical muse (cf. fig. 9) and inspiration or as a non-human collaborator and co-creator of new, meaningful and valuable artifacts from its inherent technical feedback loops that are relatable and comprehensible – but simultaneously completely unexpected and sometimes beyond any human imagination and intention.

At the same time, synthetic AI images are a new technology-driven visual language (cf. SCORZIN 2023) that many of us humans have yet to study and learn better. In psychology, *reality monitoring* identifies whether something is coming from the external world or from within our brains' biological neural networks. Such an objective vs. subjective dualism seems increasingly outdated through machine vision and machine learning. The advance of AI technologies that can now produce meta-images of meta-humans, highly photo-realistic to human eyes, means, soon, 'reality monitoring' must be based on information other than our sensory judgments – maybe even primarily on powerful machine vision and AI knowledge.

Figure 9: Levania Lehr x DALL·E 2: Fictitious Self-Portrait, October 2022

Bibliography

BODEN, MARGARET: *The Creative Mind: Myths and Mechanisms*. 2nd edition. London [Routledge] 2004 [1990]

BODEN, MARGARET: *Creativity and Art: Three Roads of Surprise*. Oxford [Oxford University Press] 2010

GRÜNBERGER, CHRISTOPH: *The Age of Data: Embracing Algorithms in Art & Design*. Zürich [Niggli Verlag] 2021

HARMSEN, LARS; JULIA KAHL (eds.): *Slanted Magazine #37: AI – Artificial Intelligence*. Karlsruhe [Slanted Publishers] 2021

KELLY, KEVIN: Picture Limitless Creativity at Your Fingertips. In: *Wired*. November 17, 2022; https://www.wired.com/story/picture-limitless-creativity-AI-image-generators/ [accessed February 16, 2023]

LAMERICHS, NICOLLE: Characters of the Future: Machine Learning, Data and Personality. In: *IMAGE: The Interdisciplinary Journal of Image Sciences. Special Issue Recontextualizing Characters*, 29, 2019, pp. 98-117

MELLER, AIDAN: AI-Da. In: AI-*da robot*. 2019. https://www.AI-darobot.com [accessed February 16, 2023]

MILLER, ARTHUR I.: *The Artist in the Machine: The World of* AI*-Powered Creativity*. Cambridge, MA [MIT Press] 2019

MANOVICH, LEV; EMANUELE ARIELLI: Artificial Aesthetics: A Critical Guide to AI, Media and Design. In: *Manovich.net*. December 15, 2021. http://manovich.net/index.php/projects/artificial-aesthetics [accessed February 16, 2023]

RAUTERBERG, HANNO: *Die Kunst der Zukunft: Über den Traum der kreativen Maschine*. Berlin [Suhrkamp Verlag] 2021

RIESEWIECK, MORITZ; HANS BLOCK: *Die digitale Seele: Unsterblich werden im Zeitalter Künstlicher Intelligenz*. München [Goldmann Verlag] 2020

SCORZIN, PAMELA C. (ed.): *KUNSTFORUM International* Vol. 278: *AI ART*. *Kann KI Kunst? Neue Positionen und technisierte Ästhetiken*. Köln [KUNSTFORUM International] 2021a

SCORZIN, PAMELA C.: As Real as Rihanna? The Curious Case of Miquela Sousa. In: *Slanted Magazine*, 37: *AI - Artificial Intelligence*, 2021b, pp. 100-105

SCORZIN, PAMELA C.: More Human than Human: Digital Dolls on Social Media. In: INSA FOOKEN; JANA MIKOTA (eds.): *Puppen als Seelenverwandte – biographische Spuren von Puppen in Kunst, Literatur, Werk und Darstellung. Dolls/Puppets as Soulmates – Biographical Traces of Dolls/Puppets in Art, Literature, Work and Performance*. Siegen [UP] 2021c, pp. 157-166

SCORZIN, PAMELA C.: Ko-Kreation und Evolution in der AI ART – am Beispiel von Pierre Huyghes 'Mental Image'-Installationen. In: *Kunsttexte, Festausgabe (Sektion Kunst Design Alltag)*, 1, 2022, pp. 1-7. https://journals.ub.uni-heidelberg.de/index.php/kunsttexte/article/view/88240?fbclid=IwAR3m_

R3QMoCmtqByecY86AsQ68hdst1hvOoK6y3Ye2uBzhB19ptjyAcWWis [accessed February 16, 2023]

SCORZIN, PAMELA C.: Kreativität, Kunst und KI: Zur algorithmisierten Ästhetik der AI ART. In: LARS GRABBE; CHRISTIANE WAGNER; TOBIAS HELD (eds.): *Kunst, Design und die "technisierte Ästhetik"*. Marburg [Büchner Verlag] 2023, pp. 232-249

HIRSCH, ANDREAS J.; MARKUS JANDL; GERFRIED STOCKER (eds.): *The Practice of Art and AI: European ARTificial Intelligence Lab*. Linz [Hatje Cantz] 2021

TEUCHMANN, PHILIPP: Arts and Humanities in Digital Transition (Lisbon, 6-7 Jul 23). In: *ArtHist.Net*. January 7, 2023. https://arthist.net/archive/38260 [accessed February 16, 2023]

TRAVIS BOTT: Travis Bott – JACK PARK CANNY DOPE MAN. Video on *YouTube*. December 20, 2020. https://www.youtube.com/watch?v=3UwLhqcZqxc [accessed February 16, 2023]

TSAKIRIS, MANOS: Deepfakes: Faces Created by AI Now Look More Real than Genuine Photos. In: *Singularity Hub*. January 16, 2023. https://singularityhub.com/2023/01/26/deepfakes-faces-created-by-AI-now-look-more-real-than-genuine-photos/ [accessed January 2023]

VINCENT, JAMES: Tom Cruise Deepfake Creator Says Public Shouldn't be Worried about 'One-Click Fakes'. In: *The Verge*. March 5, 2021. https://www.theverge.com/2021/3/5/22314980/tom-cruise-deepfake-tiktok-videos-AI-impersonator-chris-ume-miles-fisher [accessed February 16, 2023]

ZYLINSKA, JOANNA: AI *Art. Machine Visions and Warped Dreams*. London [Open Humanities Press] 2020. http://openhumanitiespress.org/books/download/Zylinska_2020_AI-Art.pdf [accessed February 16, 2023]

IMAGE
The Interdisciplinary Journal of Image Sciences
37(1), 2023, S. 195-207
ISSN 1614-0885
DOI: 10.1453/1614-0885-1-2023-15472

HERBERT VON HALEM VERLAG

Jay David Bolter

AI Generative Art as Algorithmic Remediation

Abstract: As the essays in this collection demonstrate, AI generative imagery raises compelling theoretical and historical questions for media studies. One fruitful approach is to regard these AI systems as a medium rooted in the principle of remediation, because the AI models depend on vast numbers of samples of other media (painting, drawing, photography, and textual captions) scraped from the web. This algorithmic remediation is related to, but distinct from earlier forms of remix, such as hip-hop. To generate new images from the AI models, the user types in a textual prompt. The resulting text-image pairs constitute a kind of metapicture, as defined by William J.T. Mitchell in *Picture Theory (1994)*.

Introduction

The quality of the essays in this collection attests to the rich potential of generative AI for media studies. Even if AI imagery threatens the practices and perhaps livelihoods of designers and artists – a question on which opinions differ – it certainly does not threaten media studies researchers, who are already responding creatively to the theoretical and historical questions posed by this relatively new practice. These essays provide evidence that we cannot understand AI simply as a threat; instead, they attest to the complexity of our visual culture before these generative programs became available and address the ways in which AI imagery may participate in that culture.

Since the essays engage with DALL·E 2 and the other generative image systems on a variety of levels, we might begin by giving DALL·E 2 itself a chance to engage: that is, by submitting the titles of a few of the essays or the earlier presentation titles as prompts and seeing what kind of images emerge. The titles are not the kind of phrases typically submitted to DALL·E 2, and the results do not seem typical either. Pamela Scorzin's presentation title "Meta-Images and Meta-Humans" (cf. SCORZIN 2023) produces a visually coherent result (cf. fig. 1). The depiction of human figures taking pictures of pictures does suggest the self-referential quality of the prompt. In this case, at least, DALL·E 2 seems to be functioning as its makers intended: we could imagine using this image on a book cover for a monograph on meta-images and meta-humans. Two further examples are harder to interpret as responses to their texts. Andreas N. Ervik's presentation title "Towards an Ontology of AI Generated Images" (cf. ERVIK 2023) produces the image we see in figure 1 in the middle, and Hannes Bajohr's title *Dumb Meaning: Machine Learning and Artificial Semantics* (2023a) gives us perhaps the most humorous image (cf. fig. 3 on the right). What is notable about the last two results is that what appear to be alphabetic symbols have found their way into the images. How text functions here is suggestive of the ontology of these generative programs, as we will discuss below.

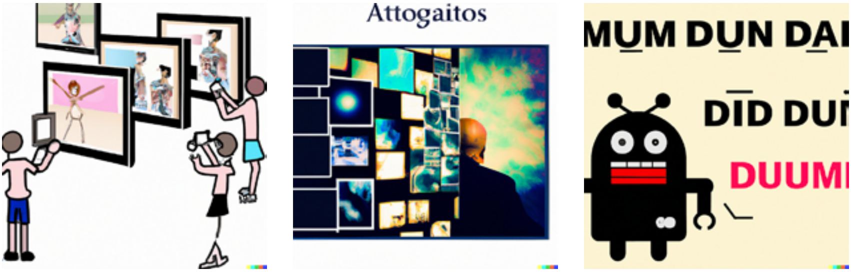


Figure 1: DALL·E 2 creations with prompts based on titles of presentation of the Tübingen-workshop, generated in February 2023

Debating the Status of AI Imagery

It is not only media studies scholars who are fascinated with generative AI; there is also enormous popular interest in the technology, particularly since the private company OpenAI has released DALL·E 2 and then more recently ChatGPT. Although the technologies and even some of the manifestations of AI image generation and art stretch back years, it is only since last fall (2022) that everyone seems to be talking about them in blog posts, podcasts, and mainstream newspapers and magazines (such as the *NYTimes*, *The Economist*, and *Der Spiegel*). Two interrelated issues are of greatest interest:

1. The question of intellectual property: What is the legal status of these generative images? Are they original or derivative works? The systems do not generate images ex nihilo. They draw on millions or even billions of text-image pairs scraped from the web for their underlying databases, such as, for example, LAION used by Stable Diffusion and some other systems. Do

they therefore infringe on the rights of the human artists and producers whose works were scraped from the web and fed into the model?
2. The aesthetic issue: are these generated images creative artifacts at all? And where does the creativity reside? Are they art? And if so, who is the artist?

Although the first of these, the legal question, is not a central focus of this *IMAGE* special collection, some of the essays do address it. In particular, Nicolle Lamerich's (2023) contribution on fan art shows that a significant portion of the fan community objects to their works being used without permission and possible remuneration. A class-action lawsuit has already been filed in the United States against Stability AI, Midjourney, and Deviant Art by three artists claiming that their work has been used to train the model which in turn can generate images similar to their art style (cf. WIGGERS 2023). This amounts, they claim, to a "21-century collage" technique and is not fair use. In response, a website titled "Stable Diffusion Frivolous" was created by "tech enthusiasts uninvolved in the case, and not lawyers, for the purpose of fighting misinformation" (STABLE DIFFUSION FRIVOLOUS n.d.: n.pag.). The site sets out to refute several points made by lawyers for the plaintiff, essentially claiming that what Stable Diffusion and the other models do is indeed fair use.

The legal case will depend on technical issues concerning the storage, modification, and use of the image-text pairs and on how well intellectual property lawyers and judges understand these technicalities. Other lawsuits concerning generative AI have already been filed, and more are sure to follow (cf. WIGGERS 2023). The legal issues may well be decided differently in different countries with different intellectual property regimes. While we do not know how these issues will be settled, it seems likely that these disputes will continue for some time (years?), and that there eventually will be two related but distinct resolutions: one legal and the other cultural. By that I mean that the law on AI image generation will be more or less settled and there may well be some system of remuneration for the human artists. That settlement, however, will not necessarily be the same as what our media culture in general comes to accept as 'fair use' in image generation. Such a divergence between the law and cultural practice has happened before. The situation over the last three decades of the practices of remix provides a good example. Based on various legal rulings, there is an elaborate set of rules about how older samples can be used in new works. This is especially the case for music sampling. However, most amateur remixers do not pay much attention to the rules. The web is full of remixes that may technically be illegal, but unless they reach a certain level of economic significance, they are largely ignored. Our media culture has come to a shared understanding of the legitimacy of the creative appropriation of earlier works, which is what Lawrence Lessig (2008) was arguing for in his book on remix.

The second question is: are these images creative artifacts in their own right? And if so, who is the creator? Several essays in the present collection, especially those by Erwin Feyersinger, Lukas Kohmann, and Michael Pelzer (2023) and by Scorzin (2023), have important insights to contribute. On the other hand, there seems to be no consensus among the larger public on these questions. Two observations here:

1. A wide group outside of the traditional arts and media studies communities (artists, critics, scholars) feel compelled to express their views on this question. The technical community certainly does; computer scientists have become theorists of art.
2. In many cases, these views often begin from the assumption that the definitions of art and creativity, prior to AI, are themselves relatively unproblematic. The question is simply how these AI systems fit into those definitions.

The larger popular debate is connected to a question, then, of the cultural status of art in an era of digital media. The notion that art is the province of cultural and academic elites has been eroding for decades. The belief in a cultural hierarchy in which the fine arts and literature are superior to film, popular music, romance novels, and comics is largely, if not entirely, gone. In its place is the sense that all forms of creative expressions have equal status: the network is replacing the hierarchy as a cultural form (cf. BOLTER 2019). With its blogs, fan pages, streaming services, and social media sites, the Internet is perfectly suited to foster this networking of culture. At the same time, the vocabulary and the implicit values of that earlier hierarchical era have not disappeared. Our current media culture has adopted them. The term 'artist' has vastly expanded to include all sorts of creative practices that would not have been called art before the 1960s. And this trend makes it easier to imagine a further expansion to include these generative programs and their neural nets, which are themselves among the most impressive products of network thinking.

AI as a Tool

As several of the essays in the present collection suggest, there are other ways to regard these AI programs than as threats to or replacements for human artists or creators. One is to regard AI as a new tool in the hands of human agents. The authors of the Stable Diffusion Frivolous website pages take this position, referring to these programs as "AI Art tools". They argue that there are earlier instances of new technological tools reconfiguring artistic practice,

> anti-AI artists [fear] being replaced by artists who use AI tools in their workflows. Just like the fear was of manual artists being replaced by digital artists when tools like Photoshop emerged, and the fear of painters being replaced by photographers when the camera was developed (STABLE DIFFUSION FRIVOLOUS n.d.: n.pag.).

Writing in the *American Scientist* an article entitled AI *Is Blurring the Definition of Artist*, a top researcher in this field, Ahmed Elgammal (2018), supports this view: "[Ju]st because machines can almost autonomously produce art, it doesn't mean they will replace artists. It simply means that artists will have an additional creative tool at their disposal, one they could even collaborate with" (ELGAMMAL 2018: n.pag.).

Already we can see generative AI integrated into popular software tools such as Photoshop with plugins and experimental add-ons – a trend that will likely continue to create a much tighter connection between traditional digital and AI-supported art creation. The idea of AI as a tool for improving a human artist's work is reminiscent of one line in the original AI debate from the 1950s to the 1980s, a period when artificial intelligence was still controversial even in the computer science community. Some computer scientists thought that instead of aiming for artificial intelligence the goal should be to develop interfaces and systems that would serve to amplify human intelligence. That was the implicit, and sometimes explicit, assumption behind the development of personal computing in this period: for example, Douglas Engelbart, one of the pioneers of the desktop interface, whose 1968 demonstration of his NLS system introduced a number of key elements of desktop computing, spoke of "augmenting human intellect" (ENGELBART n.d. [1962]: n.pag.).

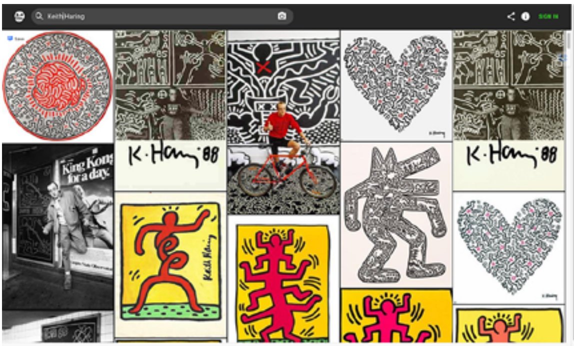
Many advocates for AI in that original debate, such as John McCarthy and Marvin Minsky, believed that so-called 'symbolic AI techniques' would lead to intelligent systems that could function without human collaborators. However, researchers in machine learning today, whose neural nets are far more powerful than anything that the era of symbolic AI produced, seem to welcome the idea that AI systems would be used in collaborative relationships with human agents, rather than replacing humans altogether. Some of the essays in our collection explore this line, too. Feyersinger, Kohman, and Pelzer (2023: 134) argue that the "fuzzy" image generation of DALL·E 2 can serve as a form of externalized visual thinking for human creators. Scorzin observes that we may think of these systems as co-creative agents. Human-machine co-creativity has become a topic of interest in the computer science community, a common theme of papers in the ACM conference "Creativity and Cognition" and elsewhere. A recent anthology entitled *The Language of Creative* AI (VEAR/POLTRONIERI 2022: xi) "builds on [...] and extends the notion of embedded and cooperative creativity with intelligent software. It does so through a human-centered approach in which the AI is empowered to make the human experience more creative or join in/cooperate with the creative enterprise in real time".

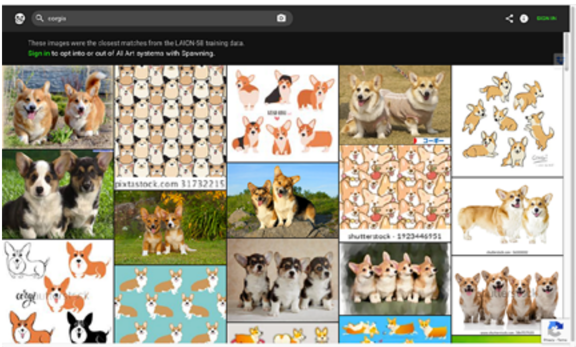
AI as a Medium

There is yet another perspective to consider. Instead of regarding the systems as agents or tools, they can be thought of as a new medium. In the case of text-to-image generators like DALL·E 2, not just the prompt itself but the whole process of creating the model and producing images would constitute the medium. There would be nothing particularly novel in imagining that the characteristics of the medium would themselves impose constraints on or facilitate the making of the art. But the degree to which the medium of AI would participate in the fashioning of the images is new and perhaps without parallel. We could say that the system of artist and AI constitute both maker and medium. I suggest that if the database, model, and algorithms behind systems like DALL·E 2 are constituents of a new medium, then that medium is rooted in the principle of remix or remediation for two reasons. First, the existing generative models depend on other media – above all painting, drawing, and photography. These are the media that constitute the imaginary of the current web that is scraped to generate the models. Second, these systems are constituted from text-image pairs, and the generated images are therefore the product of two heterogeneous media. The model itself is intermedial, a blend of text and images that is both at the same time.

AI generative imagery is remix, but there is an important difference from earlier forms. We can trace one strand of remix to hip-hop practices dating back decades. Then there is the somewhat younger video remix, which involves the editing and often complex layering of a series of video clips together with an underlying musical soundtrack. This practice became popular among amateurs in the 2000s because of the availability of affordable editing tools and inexpensive computers powerful enough to handle the tasks. Both audio and video remix require the step-by-step intervention of the human remixer, and this is obviously different from the process of image generation in DALL·E 2, Midjourney, and the like. True, the creation of the data model itself is a step-by-step process by a team of programmers and an anonymous crowd of image taggers. But for the human user providing a text prompt, the rhythm of interaction is redefined, and human intervention is reduced. Even at this stage, however, it is possible to use these systems for interactive refinement and skilled manipulation, as several of the present essays are remarking. Eventually, as noted above, we will likely see a workflow similar to that of a skilled user applying filters in Photoshop. In any case, this kind of AI image generation will always be remix because the systems begin with the visual samples and captions scraped from the Internet.

One of the largest such text-image databases is LAION (Large-scale Artificial Intelligence Open Network). This publicly available database was used to generate the models for Stable Diffusion, Imagen, and others. The March 2022 release contained more than 5 billion text image pairs. You can query this database from the site haveibeentrained.com, whose explicit purpose is to allow artists to see if their work is present in the database and request its removal. Typing in the name of the graffiti artist Keith Haring, for example, produces the eclectic results seen in figure 2 upper part. As we would expect, the database has scraped not only images of Haring's work, but other related images such as mimickings of his style. Searches on the site are not limited to names; one can search for other terms as well. We can use haveibeentrained.com to view the kinds of images that underlay results in DALL·E 2. (Although OpenAI used its own database for the model behind DALL·E 2, the data must have been similar to the 5 billion pairs of LAION.) For example, the illustrative image in an OpenAI research paper (RAMESH et al. 2022) was generated from the phrase "a corgi playing a flame throwing trumpet". Here are the first few corgi examples and the first trumpet examples from the LAION database (cf. fig. 2 middle an lower part).





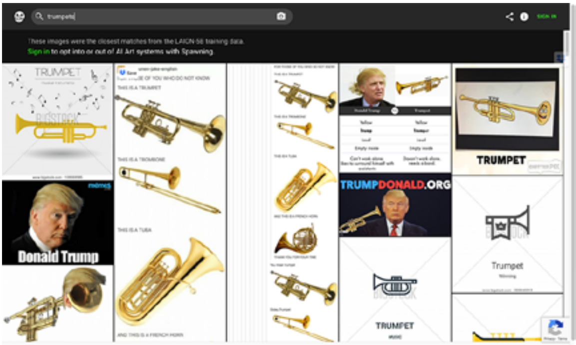


Figure 2: Screenshots from the LAION database for the queries "Keith Haring", "corgis", and "trumpets"

LAION and the haveibeentrained site are revelatory of the ontology of AI generated imagery. Viewing the initial images from the web makes it apparent that the process is one of remediation. Sophisticated algorithms create the models from these vast databases, and the models are tuned in various ways by human programmers. Nevertheless, without the original data the generated images would not be possible. In *Generative AI and the Collective Imaginary: The Technology-Guided Social Imagination in AI-Imagenesis*, Andreas Ervik (2023) argues that the AI generated images are becoming part of our collective imaginary. It is also important to remember that these images emerge from the prior collective imagery and then are added to it. We can call this process algorithmic remix or remediation. And when enthusiasts for AI generation claim that these systems make possible a new kind of art, their claim is similar to the familiar claim for audio and video remix as art. It is characteristic of new mediums to claim that their remediations constitute a significant new form of expression.

AI and Ekphrasis

Let's return to the key feature of these new generative AI systems: the relationship of text to image. (The feature is addressed by almost every essays in our collection, but particularly by Feyersinger, Kohmann, and Pelzer and by Bajohr in his paper on what the latter calls "dumb" semantics, 2023a: 57) Text is crucial in both the encoding and the generation phases: The images that are scraped from the web all have captions, and the captions are encoded along with the images. This creates the encoding space called the prior that is used in the generation phase when the human user types in text that serves as a prompt. The user's text is a description of the image that is desired; the image cannot exist until the text is applied to the model.

The general popular reaction and the perspective of the makers of these systems seem to take the relationship for granted. Or rather, both users and makers seem most interested in a practical question: how to word the text so as to generate the desired image. This is clear from the OpenAI website, which includes a "prompt book", a set of instructions about how to get various desired stylistic effects (cf. DALL·ERY GALL·ERY 2022). Here are samples OpenAI itself created and shows in its paper on the generative technique (cf. RAMESH et al. 2022, cf. fig. 3). Such images, emblematic of DALL·E 2, are interesting as cultural expressions. They are playful in an almost postmodern way. They have the quality of pastiche and a disinterest in stylistic coherence and are characterized by an absence of affect. We could almost imagine them as illustrations in Fredric Jameson's *Postmodernism* (1991).

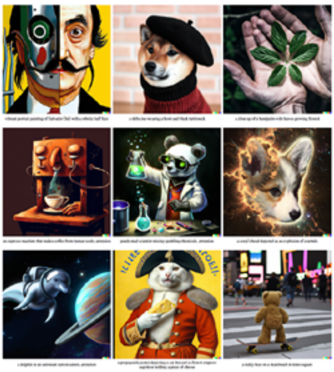


Figure 3: Various samples from OpenAI for DALL·E 2 prompts and their respective outcomes, taken from Ramesh et al 2022

I would argue that such text-image pairs constitute metapictures according to the definition by William J.T. Mitchell in *Picture Theory* (1994). Metapictures are pictures that are self-referential, and Mitchell distinguishes various classes, of which one is pictures that enter into a self-referential relationship with verbal text. This is the case for DALL·E 2 images. On the DALL·E 2 website, collections are displayed so that the images are visible and the generative text is obscured; the text only appears when the user mouses over. Mousing over reveals the text that stands behind and ontologically underneath the image. But what is the relationship here? The text in no longer exactly a caption. Does the text explain the image, justify the image? The indeterminacy of the text-image relationship is emphasized by the fact that multiple images are generated by the same prompt when repeated. DALL·E 2 shows you four, with the suggestion that many more are possible.

Mitchell devotes one of the essays in *Picture Theory* to the literary device of ekphrasis. The device dates back to antiquity, when poets would offer a vivid description of a visual scene or an art object. In many cases of ekphrasis, however, from antiquity to the present, the object being described does not exist. It is part of the literary fiction. The point of the ekphrastic description is to demonstrate the power of language to visualize with the implication that language can rival the visual arts at capturing reality. Ekphrasis has always been remediation in the sense of competition, seeking to show that the word can compete with painting at visual representation. The earliest example of ekphrasis usually cited is the description of the shield of Achilles in the *Iliad*. Let's test DALL·E 2 with this canonical example. We first see what kind of images there were in the database. To query LAION, we can type the phrase "the shield of Achilles as described by Homer in the Iliad" into the haveibeentrained site. The result is a number of images as well as book covers; figure 4, upper part, shows a sample. If we then type the phrase "the shield of Achilles as described by Homer in the Iliad" into DALL·E 2, we get four results, all of which recall some formal aspects of the shield as Homer described it. Figure 4, lower part, is the most compelling.

Of particular interest is the appearance of text in the resulting image. (This also happened in the two cases above when we fed the title of essays from this collection into DALL·E 2, perhaps because the abstract vocabulary of the titles seemed to encourage the generative algorithm to produce abstract results.) Text appears, but not in the form of recognizable names or words. It seems as if the text has broken through to the surface of the image. As viewers or readers of this image, we might be tempted to try to make sense of it: Is this a blend of the names Homer, Achilles, and the *Iliad*? Or we might ask: Is the text here being used formally rather than symbolically? Is this what happens to language when it is absorbed into the neural layers of the model? It seems as if this imagetext is commenting upon, almost parodying the original prompt, again emphasizing the indeterminacy of the relationship between text and image. Are we witnessing the artificial or dumb semantics that Bajohr (2023a) discusses in his essay for this colleciton? In a separate paper, Bajohr has explored the text-image relationship in more detail and argued convincingly that these generative systems constitute a new kind of text-image relationship that he calls *operative ekphrasis* (cf. BAJOHR 2023b).

In his essay on ekphrasis, Mitchell speaks of the phenomenon he calls *ekphrastic hope*. He means the hope that a verbal description could succeed in bringing forth an image with perfect representational clarity – that the difference between word and image between the symbolic and the iconic, or perhaps in post-structuralist terms between the signifier and the signified, could be bridged. Mitchell also characterizes ekphrastic hope's opposite, ekphrastic fear. We can understand this as the fear that ekphrasis might in fact succeed. For if

Figure 4: Screenshot from the LAION database for the query "the shield of Achilles as described by Homer in the Iliad" and a DALL·E 2 creation prompted with the same text, generated in February 2023

ekphrasis fully succeeds, what happens to the status of the word and to verbal art and expression in general? Would the word be absorbed into the image and lose its identity? Perhaps we are witnessing the absorbing of the word into the image in the encoding processes of these AI models, in which the text captions are fed into the neural net and lose their semantic identity.

We could argue that Mitchell's notion of ekphrastic hope is what the makers of these generative art systems are striving to realize when they cheerfully list all the uses of their systems for illustrating blogs and newsletters and for giving users the power to paint with words as never before. The *DALL·E 2 Prompt Book* is designed for this purpose, to empower users to tune their images, to paint with words. In this sense, the prompt book is emblematic of the optimism of this AI moment in general; and then in turn, what Mitchell calls ekphrastic fear could be the backlash by those who resist AI image generation because it suggests to them that the larger project of generative AI could succeed and bring with it unforeseen and negative consequences for human creativity. These consequences could go beyond the economic loss to artists and designers through appropriation of their intellectual property and through automation of their skills and expertise. The ultimate threat would be the loss of the arts and crafts as autonomous human activities. The future almost certainly lies somewhere between the extremes of ekphrastic hope and fear. Most of the essays in this collection could be characterized as cautiously optimistic about the potential of AI generated imagery. They do not endorse the future that Open AI's CEO Sam Altman imagines for a world of AGI (Artificial Generalized Intelligence) (cf. LAWRENCE 2023), but they are still ready to engage with the theoretical and practical opportunities that AI affords in the realms of visual representation and art.

Bibliography

BAJOHR, HANNES: Dumb Meaning: Machine Learning and Artificial Semantics. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences*, 37(1), 2023, pp. 58-70

BAJOHR, HANNES: *Operative Ekphrasis*, 2023. Manuscript provided by the author

BOLTER, JAY DAVID: *The Digital Plenitude: The Decline of Elite Culture and the Rise of New Media*. Cambridge, MA [MIT Press] 2019

DALL·ERY GALL·ERY (ed.): The DALL·E 2 Prompt Book. In: *Dall·ery gall·ery: Ressources for Creative DALL·E Users*. July 14, 2022. https://dallery.gallery/the-dalle-2-prompt-book/ [accessed March 9, 2023]

ELGAMMAL, AHMED: AI Is Blurring the Definition of Artist. In: *American Scientist*. December 6, 2018. https://www.americanscientist.org/article/AI-is-blurring-the-definition-of-artist [accessed March 9, 2023]

ENGELBART, DOUG: Augmenting Human Intellect: A Conceptual Framework. SRI Summary Report AFOSR-3223. October 1962. In: *Doug Engelbart Institute*: *Strategies for a More Brlliant World*. No date given. https://www.dougengelbart.org/pubs/augment-3906.html [accessed March 9, 2023]

ERVIK, ANDREAS: Generative AI and the Collective Imaginary: The Technology-Guided Social Imagination in AI-Imagenesis. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences*, 37(1), 2023, pp. 42-57

FEYERSINGER, ERWIN; LUKAS KOHMANN; MICHAEL PELZER: Fuzzy Ingenuity: Creative Potentials and Mechanics of Fuzziness in Processes of Image Creation with Text-to-Image Generators. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of* IMAGE*: The Interdisciplinary Journal of Image Sciences,* 37(1), 2023, pp. 135-149

JAMESON, FREDRIC: *Postmodernism.* Durham [Duke University Press] 1991

LAMERICHS, NICOLLE: Generative AI and the Next Stage of Fan Art. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of* IMAGE*: The Interdisciplinary Journal of Image Sciences*, 37(1), 2023, pp. 150-164

LAWRENCEC: Sam Altman: "Planning for AGI and Beyond". In: *LESSWRONG* Blog. February 24, 2023. https://www.lesswrong.com/posts/zRn6aQyD8uhAN7qCc/sam-altman-planning-for-agi-and-beyond [accessed March 9, 2023]

LESSIG, LAWRENCE: *Remix: Making Art and Commerce Thrive in the Hybrid Economy*. London [Bloomsbury] 2008

MITCHELL, WILLIAM J.T.: *Picture Theory*. Chicago [University of Chicago Press] 1994

RAMESH, ADITYA; PRAFULLA DHARIWAL; ALEX NICHOL; CASEY CHU; MARK CHEN: Hierarchical Text-Conditional Image Generation with CLIP Latents. *arXiv:2204.06125*. April 13, 2022. https://arxiv.org/abs/2204.06125 [accessed March 9, 2023]

SCORZIN, PAMELA C.: AI Body Images and the Meta-Human: On the Rise of AI-generated Avatars for Mixed Realities and the Metaverse. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of* IMAGE*: The Interdisciplinary Journal of Image Sciences*, 37(1), 2023, pp. 179-194

STABLE DIFFUSION FRIVOLOUS: Stable Diffusion Frivolous: Because Frivolous Lawsuits Based on Ignorance Deserve a Response. In: *stablediffusionfrivolous.com*. No date given. www.stablediffusionfrivolous.com [accessed March 9, 2023]

WIGGERS, KYLE: The Current Legal Cases against Generative AI are Just the Beginning. In: *TechCrunch*. January 27, 2023. https://techcrunch.com/2023/01/27/the-current-legal-cases-against-generative-AI-are-just-the-beginning/ [accessed March 9, 2023]

Open Source

BLexKom blexkom.halem-verlag.de
BIOGRAFISCHES LEXIKON DER KOMMUNIKATIONSWISSENSCHAFT

BLexKom möchte der Kommunikationswissenschaft im deutschsprachigen Raum ein Gesicht geben. Vorgestellt werden die zentralen Akteure: Professoren, Habilitierte und andere Personen, die einen gewichtigen Beitrag für das Fach geleistet haben – von Karl Bücher bis zu den frisch Berufenen.

http://blexkom.halem-verlag.de

r:k:m rezensionen:kommunikation:medien

rezensionen:kommunikation:medien (r:k:m) versteht sich als zentrales Rezensionsforum für die Kommunikations- und Medienwissenschaften. *r:k:m* will seinen Lesern einen möglichst vollständigen thematischen Überblick über die einschlägige Fachliteratur ermöglichen und Orientierung in der Fülle des ständig wachsenden Buchmarkts bieten. Aktuelle Rezensionen erscheinen in regelmäßigen Abständen.

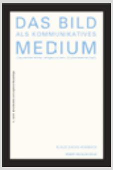
http://www.rkm-journal.de

Journalistikon
Das Wörterbuch der Journalistik
http://www.journalistikon.de

Das *Journalistikon* ist das erste deutschsprachigen Lexikon der Journalistik. Dabei handelt es sich um die Wissenschaft, die den Journalistenberuf durch Ausbildung und Innovationen unterstützt und kritisch begleitet. Das Wörterbuch der Journalistik wendet sich nicht nur an Wissenschaftler oder Studierende entsprechender Fachrichtungen, sondern an jeden, der sich für Journalistik und praktischen Journalismus interessiert und sich als Mediennutzer oder Medienhandwerker an einem reflektierenden Zugang versuchen möchte. Das *Journalistikon* lohnt sich für alle, die Informationen zur Thematik suchen, ohne dabei ein zweites Lexikon zum Verständnis der Ausführungen daneben legen zu müssen.

http://www.journalistikon.de

HERBERT VON HALEM VERLAG



KLAUS SACHS-HOMBACH

Das Bild als kommunikatives Medium. Elemente einer allgemeinen Bildwissenschaft

2021, 4. Auflage, 380 S., 8 Abb., 2 Tab., Broschur, 213 x 142, dt.
ISBN (Print) 978-3-86962-584-3
ISBN (PDF) 978-3-86962-581-2

ANDREAS VEITS / LUKAS R.A. WILDE / KLAUS SACHS-HOMBACH (Hrsg.)

Einzelbild & Narrativität. Theorien, Zugänge, offene Fragen

2020, 208 S., Broschur, 213 x 142 mm, dt.
ISBN (Print) 978-3-86962-401-3
ISBN (PDF) 978-3-86962-402-0

LUKAS R.A. WILDE

Im Reich der Figuren. Meta-narrative Kommunikationsfiguren und die ›Mangaisierung‹ des japanischen Alltags

2018, 456 S., 42 Abb., Broschur, 213 x 142 mm, dt.
ISBN (Print) 978-3-86962-282-8
ISBN (PDF) 978-3-86962-283-5

ANDREAS VEITS

Narratologie des Bildes. Zum narrativen Potenzial unbewegter Bilder

2021, 338 S., 53 Abb., Broschur, 213 x 142 mm, dt.
ISBN (Print) 978-3-86962-596-6
ISBN (PDF) 978-3-86962-597-3

ANNA ZEITLER

Störung der Bilder – Bilder der Störung. Medienereignisse zwischen Fest und Katastrophe

2021, 576 S., 80 Abb., Broschur, 213 x 142 mm, dt.
ISBN (Print) 978-3-86962-607-9
ISBN (PDF) 978-3-86962-608-6

ALEXANDER CHRISTIAN

Piktogramme. Tendenzen in der Gestaltung und im Einsatz grafischer Symbole

2017, 576 S., 231 Abb., Broschur, 213 x 142 mm, dt.
ISBN (Print) 978-3-86962-243-9
ISBN (PDF) 978-3-86962-244-6

Schriften zur Rettung des öffentlichen Diskurses

STEPHAN RUSS-MOHL (Hrsg.)

Streitlust und Streitkunst. Diskurs als Essenz der Demokratie

2020, 472 S., Broschur
ISBN (PRINT) 978-3-86962-552-2
ISBN (PDF) 978-3-86962-553-9
ISBN (ePub) 978-3-86962-555-3

MARCO BERTOLASO

Rettet die Nachrichten! Was wir tun müssen, um besser informiert zu sein

2021, 358 S., Broschur
ISBN (PRINT) 978-3-86962-493-8
ISBN (PDF) 978-3-86962-494-5
ISBN (ePub) 978-3-86962-520-1

STEPHAN RUSS-MOHL / CHRISTIAN PIETER HOFFMANN (Hrsg.)

Zerreißproben. Leitmedien, Liberalismus und Liberalität

2021, 256 S., Broschur
ISBN (PRINT) 978-3-86962-535-5
ISBN (PDF) 978-3-86962-538-6
ISBN (ePub) 978-3-86962-532-4

ISABELLE BOURGEOIS

Frankreich entschlüsseln. Missverständnisse und Widersprüche im medialen Diskurs

2022, ca. 270 S., Broschur
ISBN (Print) 978-3-86962-643-7
ISBN (PDF) 978-3-86962-644-4
ISBN (ePub) 978-3-86962-520-1

TOBIAS ENDLER

Demokratie und Streit. Der Diskurs der Progressiven in den USA: Vorbild für Deutschland?

2022, 208 S., Broschur
ISBN (PRINT) 978-3-86962-645-1
ISBN (PDF) 978-3-86962-646-8
ISBN (ePub) 978-3-86962-647-5

SEBASTIAN TURNER / STEPHAN RUSS-MOHL (Hrsg.)

Deep Journalism. Domänenkompetenz als redaktioneller Erfolgsfaktor

2023, Broschur
ISBN (PRINT) 978-3-86962-660-4
ISBN (PDF) 978-3-86962-658-1
ISBN (ePub) 978-3-86962-659-8

KÖLNER MEDIEN GESPRÄCHE koelner-mediengespräche.de

Journalistikon
Das Wörterbuch der Journalistik
https://journalistikon.de

BLexKom
BIOGRAFISCHES LEXIKON DER KOMMUNIKATIONSWISSENSCHAFT
blexkom.halem-verlag.de

Unsere Neuerscheinungen finden Sie auch als E-Books auf https://www.halem-verlag.de. Bestellungen ab EUR 20,00 versandkostenfrei innerhalb Deutschlands.

HERBERT VON HALEM VERLAG

Boisseréestr. 9-11
50674 Köln
+49-221-92 58 29 0
info@halem-verlag.de
http://www.halem-verlag.de

Online-Journale und -Lexika
http://journalistik.online
http://www.journalistikon.de
http://blexkom.halem-verlag.de
http://www.rkm-journal.de

Der Herbert von Halem Verlag steht für Qualität im wissenschaftlichen Publizieren. Wir haben Ihr Interesse geweckt? Dann schreiben Sie uns unter: info@halem-verlag.de