Generative AI as a co-creator in the conception of images

By: Carl Sebastian Samuelsson

Supervisor: Anders Green
Södertörn University | Department of Natural Sciences, Technology and Environmental Studies
Bachelor thesis 15 hp
Media Technology | HT2023

*Cover image generated with Bing Image Creator (DALL-E 3) and upscaled using Upscale.media
Generativ AI som samskapare i framställningen av bilder
Abstract

Generative AI systems for image-creation have taken a giant leap technologically in the last few years, creating opportunities for users and introducing new areas of research for the scientific community. New implementations for these technologies are discovered regularly and due to the complex nature of AI, there is an ongoing discussion among researchers on whether generative AI is yet another tool, or if there is potential for the technology to be something more.

The aim of this study is to explore the nuances and complexity of co-creation with generative AI, using post-human and post-anthropocentric perspectives that consider human agencies as entangled with machine agency. To understand whether human-AI co-creation is sustainable and valuable as an activity, a netnographic study was conducted to explore the different aspects of user interaction. Among the most relevant of these aspects were the system limitations, dataset limitations, prompting, user goals and user procedures.

The study found that the current limitations to the generative AI systems were too significant for the systems to display a high amount of machine agency. This ultimately pointed to human-AI co-creation being non-sustainable and non-valuable as an activity, but the results and following discussion also suggest how we can move on overcoming these limitations.

Keywords: Generative AI, co-creation, post-human perspectives, machine agency

Sammanfattning

Generativa AI-system för bildskapande har teknologiskt tagit ett stort kliv framåt de senaste åren, detta har lett till möjligheter för användare och det har introducerat nya forskningsområden för forskarvärlden. Nya tillämpningar för dessa teknologier upptäcks regelbundet. På grund av den komplexa naturen hos AI så pågår en diskussion bland forskare gällande huruvida generativ AI är ett nytt verktyg, eller om det finns potential för teknologin att vara något mer.

Denna studies mål är att undersöka nyanserna och komplexiteten hos samskapande med generativ AI, genom att använda post-humanistiska och post-antropocentritiska perspektiv som betraktar mänsklig ”agency” som sammanflätad med maskiners ”agency”. För att förstå huruvida människa-AI-samskapande är hållbart och värdefullt så utfördes en netnografisk undersökning för att granska de olika aspekterna med användarinteraktion. Bland de mest relevanta av dessa aspekter var systembegränsningarna, datasetsbegränsningarna, promptning, användarmål och användarprocedurer.

Undersökningen visade att de nuvarande begränsningarna hos generativa AI-system är för stora för att systemen skulle kunna uppnisa en hög grad ”maskin-agency”. Detta pekade i slutändan mot att människa-AI-samskapande är ohållbart och icke värdefullt som en aktivitet. Men resultaten och den följande diskussionen antyder också hur vi kan gå vidare med att övervinna dessa begränsningar.

Nyckelord: Generativ AI, samskapande, post-humanistiska perspektiv, maskin-agency
# Table of contents

1. Introduction ........................................................................................................... 5  
2. Related research .................................................................................................... 6  
   2.1 Co-creation ........................................................................................................ 6  
   2.2 Creative process .................................................................................................. 7  
   2.3 Challenges ......................................................................................................... 8  
      2.3.1 Interaction ..................................................................................................... 8  
      2.3.2 Dataset ......................................................................................................... 9  
      2.3.3 Unpredictability and inconsistency ............................................................... 10  
3. Research questions ................................................................................................ 11  
4. Method .................................................................................................................... 12  
   4.1 Sampling ............................................................................................................ 12  
   4.2 Data collection .................................................................................................... 13  
   4.3 Data analysis ....................................................................................................... 14  
5. Results ..................................................................................................................... 16  
   5.1 System limitations .............................................................................................. 16  
   5.2 Dataset limitations ............................................................................................. 20  
   5.3 Interaction ......................................................................................................... 23  
   5.4 User goals .......................................................................................................... 25  
   5.5 Procedure .......................................................................................................... 26  
   5.6 User feelings and attitudes ................................................................................ 26  
   5.7 Core category ..................................................................................................... 27  
6. Discussion ............................................................................................................... 28  
   Roles and the post-human perspective .................................................................... 28  
   Prompting is key ....................................................................................................... 30  
   Dataset – the oblider and restrictor ......................................................................... 31  
   Non-frictional creative process ............................................................................... 31  
7. Conclusion .............................................................................................................. 32  
   Study limitations ..................................................................................................... 32  
References .................................................................................................................. 33
1. Introduction

The last years have seen a boom in the technological advancement of generative AI models, empowering novel systems that are capable of generating images and text of high quality (Zhong et al., 2023). Commercially available systems for generation have precipitated public awareness of the phenomena surrounding the technology, introducing a new age of ubiquitous computing (Inie, Falk and Tanimoto, 2023).

Generative AI (or gen AI) describes a variety of systems capable of generating images, text or other media as output, based on text prompts written as input (Khosrowi, Finn and Clark, 2023). These systems are built on generative models, which are comprised of machine learning algorithms. The algorithms allow the models to learn from datasets, to understand patterns and to create new artificial data through predictions (Bandi, Adapa and Kuchi, 2023).

For this thesis, generative AI or AI will refer primarily to text-to-image and image-to-image generative models and systems – i.e., systems that generate image media based on natural language text input, and/or accepts image inputs as references. Among the most popular of these systems are Midjourney, Stable Diffusion and DALL-E (Oppenlaender et al., 2023). Users will refer to people interacting with generative AI systems as means for creating images.

The inevitable incorporation of generative AI into the many aspects of society also raises questions about acceptance, focusing on gen AI systems as new tools of the trades (Inie, Falk and Tanimoto, 2023). In the field of research, the current potential of these systems has been analysed in depth, and the future potential for human-AI co-creation is speculated. Post-human perspectives dismiss the conception of humans and machines as separated actors in the process of design and ideation (Tholander and Jonsson, 2023). Rather, they are considered co-creators with intertwined agencies, communicating and performing in ways that resemble those of human-human co-creation.

The advent of novel generative AI systems has birthed new areas of expertise, most notably the practice of prompt engineering. Though earlier, prompting was a solo practice and partially secretive – growing communities, like the Midjourney community, allows for prompt sharing and social learning (Oppenlaender, 2022b).

Existing research has shown that the understanding of generative AI is commonly superficial and that this may correlate with a tendency of exaggeration of the potential of the technology (Oppenlaender et al., 2023). This thesis and study aim to explore the notion of co-creation of images between human and generative AI, highlighting the possibilities and challenges with these novel systems.
2. Related research

2.1 Co-creation

The role of generative AI in creative co-operation has been widely theorised, from perceiving the system as a tool (Inie, Falk and Tanimoto, 2023) to recognizing it as a potential co-creator that assumes a role that traditionally would be adopted by a human (Jonsson and Tholander, 2022).

Drawing on theories from post-human and post-anthropocentric perspectives, Jonsson and Tholander explores the concept of AI systems as co-creators, arguing that the human-like properties displayed by the systems make them especially relevant to research from a co-creative perspective (2022). Most importantly, through a post-human lens, we can explore the nuances and degrees of an AI system as an agent in creation. In post-human and post-anthropocentric theories, the notion of shared human and machine agencies is a central element in the critical views on humans as solo agents of creative ventures.

The definition of agency of an actor in a human-machine-network (HMN) is “the capacity to perform activities in a particular environment in line with a set of goals/objectives that influence and shape the extent and nature of their participation”, with the environment bound by the HMN (Engen, Pickering and Walland, 2016). Engen et al. imply that agency consequently indicates what a human can do in a network, and to an extent how this aligns with their beliefs in their ability to achieve goals (2017).

Human or non-human agency in a HMN is scoped by three key factors: 1) the activities the actor can perform, 2) the nature of the activities, and 3) the ability to interact with other actors. Summarily, the definition given by Engen et al. is that machine agency is related to how (and to what degree) the machine a) performs activities of a personal and creative nature b) influences other actors in the network c) enables humans to exercise proxy agency and d) is perceived as having agency by human actors (2016). With this definition in mind, the researchers give an example of a low-agency non-human actor: an individual machine in a network focused on performing a limited number of fixed tasks – while high agency requires greater autonomy and the ability to perform a larger set of complex tasks.

Engen et al. propose that machine agency directly relates to human agency – thus, they hypothesize a correlation between machine agency and computer self-efficacy. Non-human-agents as influencers for computer self-efficacy becomes highly relevant since computer self-efficacy relates to trust in technology, perceived usefulness, and ease-of-use in systems (Mcknight et al., 2011).

Using post-human and post-anthropocentric perspectives as a framework for this thesis, the ambition is to explore the complex nuances in human-AI co-creation, and thereby identifying potentials as well as challenges. This framework will also contribute as an axis around which the role of generative AI is explored. Tool or co-creator? What factors and agencies contribute to the way users perceive generative AI?
2.2 Creative process

An important aspect of co-creation is how generative AI can augment the creative process of users. Whether as a tool or as a co-creator, existing research has pointed out several steps in the creative process where generative AI has the potential to assist the human creator.

Zhang et al. presents the concept of iterative design using sketches as input, indicating that models such as DreamStudio can be used in an iterative creative process (2023). The study found that users of the generative AI considered it a helpful tool in the early stages of the design process, in which it assisted in generating and developing concepts. The notion of generative AI systems as a tool in the initial stages of creation has also been touched upon by other researchers – pointing to aspects like ideation and inspiration. Gen AI has also been proposed as a tool for later augmentation in decision making and evaluation (Inie, Falk and Tanimoto, 2023).

What is most notable regarding gen AI as a part of the creative process is the aspect of automation. In earlier research, the automation of small or repetitive tasks is a recurring theme regarding the usefulness of generative AI (Inie, Falk and Tanimoto, 2023; Tholander and Jonsson, 2023).

Friction is described as points of difficulty when interacting with technology (Cox et al., 2016) – it is described as usability best practice to keep friction as low as possible to improve the interaction between a user and a system. By keeping the friction low, the risk of user disengagement and other usability problems, such as delay and prevention of task completion, are reduced (Mejtoft et al., 2023).

By automating and simplifying tasks, systems should in practice lower the friction, making the interaction less complicated. Although, friction lowering requires the system to be consistent and predictable (Jonsson and Tholander, 2022), descriptions that are seldom used for generative AI. Inconsistency and unpredictability are common themes discussed in the research area (for further description, see section 2.3, Challenges).

Though low friction is considered favourable in terms of UX design, Cox et al. argue that inducing friction in form of small obstacles can have positive impacts on the cognitive strategy when performing tasks (Cox et al., 2016). Tholander and Jonsson suggest that inducing creative friction may work in favour of the creative process. By generating results that are inconsistent, uncertain and random, generative AI can steer the user to new design explorations and creative directions (2022). Friction in design can be categorized by intentional and unintentional, where intentional design friction such as micro boundaries is both argued to be effective (Cox et al., 2016) and to have no apparent effects (Mejtoft et al., 2023).
2.3 Challenges
This section aims to highlight the challenges that arise in the human-AI interaction, how we may overcome them, and how they might be beneficial or disadvantageous in a co-creative process.

2.3.1 Interaction
As the most common way of interacting with the generative AI, communication plays a central role in this study. This relates to the natural language that the human user relays to the AI, to be interpreted and given response to. In the case of generative AI, the response takes the form of generated media. For the user to then interpret the media requires understanding of the training dataset and front-end system (see sections 2.3.2 through 2.3.3), but also (and especially) the understanding of one’s own ability to correctly relay inputs, or prompts (Jonsson and Tholander, 2022). Communication in this section therefore relates more to user interaction, and less to system output.

Communication between the user and system is a key aspect regarding co-creation, as described in existing research (Oppenlaender, 2022a). A common theme in studies is that users have a particular difficulty in communicating their input to generative AI in ways that the system understands (Tholander and Jonsson, 2023; Zhang et al., 2023).

When creating input for generative AI, inexperienced users commonly struggle with syntax for the given system. The syntax is rarely explained to users, it has to be learned from trial and error (Liu and Chilton, 2022) or from other users. Prompts are the inputs that users mainly use to communicate with the AI systems. In essence, they are referred to as simple and descriptive textual inputs in natural language that deep generative models interpret to generate media (Oppenlaender, 2022a).

The practice of unlocking the syntax is called prompt engineering and it is described by researchers as a “formal search for prompts that retrieve desired outcomes from language models, where what is desirable is dependent upon the end task and end user” (Liu and Chilton, 2022). Oppenlaender refers to prompt engineering as an emerging research area within HCI, in which the aim is to phrase input prompts for deep generative models and in a wider sense, more effectively communicate and interact with AI (Oppenlaender, 2022a).

Prompting previously used to be a lone practice, as creators omitted full prompts when publishing generated artwork on social media channels. Therefore, the prompting behind the artwork became enigmatic to new users which further steepened the learning curve (Oppenlaender, 2022b). Growing online communities changed this however, as they opened for new opportunities of social learning. Perhaps the most notable in recent times is the Midjourney Discord community (https://discord.gg/midjourney), where users gather to prompt the Midjourney bot and have their generations and prompts presented publicly to all users. The various channels of the server are dedicated to share prompts, as well as assisting each other in reaching user goals and to create and develop guides and other learning resources.

As indicated by existing research, communication with the AI system may also be the most challenging aspect of co-creation. Prompt engineering as an iterative and experimental
practice is how users learn how to interact with generative AI and to “speak their language”, for the purpose of making the AI a more effective agent in the HMN. Although prompt engineering is an advanced technical area and therefore too broad to properly dissect in this thesis, it is considered vital to include certain elements, such as the social learning and basic prompting, for the purpose of exploring co-creation.

### 2.3.2 Dataset

Datasets for generative image AI are composed of several hundred million image-text pairs on which models are trained on (Schuhmann et al., 2021). The training datasets ultimately shape the dimensions of possible outputs – this means that all generations, although novel, cannot stray too far from the comprehensive training data. This fundamentally makes the dataset both an obliger and a restrictor.

The limitations of datasets have been reported to lead to biased generations (Solaiman, 2023) and concepts that the AI simply doesn’t understand, such as domain-specific generations (Zhang et al., 2023). Examples of biases include generations that suffer from racial and gender bias, where models seem less capable of correctly generating characters of color (McDuff et al., 2019) or where they fail to fairly represent gender due to transferred stereotypes (Grba, 2022). Research has shown that many learning datasets are largely uniform, e.g. containing an over-representation of caucasian males, or people of European or North American descent (McDuff et al., 2019).

Domain-specific generations are other limitations, where generative models fail to adhere to concepts in certain domains, such as architecture (Zhang et al., 2023). Although there has been progress in the area of domain-specific image generation (Gal et al., 2022; Ruiz et al., 2022), generative AI commonly struggles with wider concepts, particularly those surrounding realism. Generative AI systems are infamous for producing uncanny results, striving too far from realism and leaving “image artifacts” that reveal the output as generated (Broad, Leymarie and Grierson, 2020). Among these artifacts is the “extra finger”, that has been reported widely in research (Parra Pennefather, 2023) and other writings (Chayka, 2023; Dixit, 2023). This points not only to limitations in datasets but also to how the system interprets the data on which it’s trained.

These issues, in turn, limit the users in creation and creates a void where certain design directions might be perceived as closed, as the generative models have no sufficient data source to accurately output results that align with user intentions. The challenge lies in creating sufficiently diverse and extensive datasets for the models to be able to generate high-quality images of relevance. However, the endeavour of constructing datasets is both tedious and expensive, and often requires manual labelling of data (Alexopoulos, Nikolakis and Chryssolouris, 2020).

Since many generation issues can be attributed to gaps in datasets, it is imperative to recall that AI systems are limited to the data at hand, and that certain challenges might be sourced to the training stage. To take datasets into account is therefore important for this study, since all errors cannot be attributed to front-end interactions.
2.3.3 Unpredictability and inconsistency

Research has shown that generative AI at times behaves unpredictable when prompted for specific results (Hwang, 2022; Jonsson and Tholander, 2022; Zhang et al., 2023). Although this ties in with communication, predictability is a principle of learnability, which itself is a principle of usability in HCI (Dix et al., 2010, pp. 260–261). Predictability determines the effect of future action based on the user’s knowledge of the interaction history, i.e., it refers to the user’s ability to determine the effects of user actions on the system. Predictability also refers to the operation visibility, the user’s perception of available actions that can be performed.

Unpredictability in the case of human-machine interaction leads to surprises, which are generally not approved by users (Dix et al., 2010, p. 260). Predictability has also been shown to relate to user trust, as researchers have indicated that more predictable systems enhance user performance and lower the cognitive load (Daronnat et al., 2021). Furthermore, Daronnat et al. suggest that predictability and consistency become highly important when designing for systems that might be prone to errors.

Given the much-reported unpredictability of AI systems it is considered highly relevant to the study to explore the trust that users put in generative AI. Although this kind of interaction is increasingly harder to design when adhering to classic HCI principles (Yang et al., 2020), it is of great interest to analyze whether interaction can become valuable when non-human agents seem to break long-standing rules of interaction design, as with the aforementioned example of intentional friction.

Related to the unpredictability of AI systems are the inconsistencies when prompting generative AI. The inconsistencies refer to highly varying output when providing the same input (Jonsson and Tholander, 2022; Dang et al., 2023). Due to the non-deterministic nature of many the models, the same prompt can generate different results at different times (Snyder et al., 2023).
3. Research questions

RQ1: What are the challenges for users in co-creating images with generative AI systems?

As a first step in diving into co-creation, we will explore what kind of challenges that users run into when using the systems. By defining the challenges, the goal is to highlight what stands between humans and AI in the co-operation that is theorized in post-human and post-anthropocentric perspectives. For understanding to what end co-creating should take place – and whether user goals can be facilitated by generative AI – the opportunities with generative AI systems are also explored.

RQ2: Tool or co-creator: what role is assigned to generative AI?

By exploring the role of generative AI, especially the role that users impose, attempts will be made to try post-human and post-anthropocentric theories against new data. For putting these theories to test, a set of rules and definitions of roles must be determined.

On one side of the spectrum is the **tool**, a non-human actor that displays low agency. On the other side of the spectrum is the **co-creator**, a non-human actor that displays high agency. Therefore, in order to be perceived as a co-creating actor in an HMN, the AI must display sufficient agency.
4. Method

The method for gathering data was a netnographic study conducted during several weeks in November and December. Inspired by works of other researchers (Oppenlaender, 2022b), online communities proved to be an abundant source of empirical data.

Netnography, as described by Kozinets and Gambetti, is an ethnographic approach focused on social media and technoculture, where researchers immerse themselves to gather and extend data found on social media platforms. The ultimate goal of netnography is to describe and decode online life, and to understand the different kinds of user interactions embedded in these networks (2021, pp. 7–8). User interactions in this case relate both to those between humans and between human and machine, netnography therefore covers many dimensions that reside in HCI, as well as other research areas related to technology.

While netnography serves mainly to explore online socialities and to immerse the reader in the online world that is being investigated, Kozinets and Gambetti describe how netnography is constantly adjusted and adapted to times and contexts. By making netnography unlimited, the writers suggest an approach that is flexible and well suited for exploring and presenting all kinds of data, in any field (2021, pp. 7–8). This makes the method particularly applicable for the study of this thesis, since the study follows the main principles of netnography – focus on social media and technoculture and researcher immersion.

Although the inner workings and jargons of the explored communities is of less interest to describe in this thesis, certain unique community elements and their relevance to the research questions will be unveiled as we dive deeper into the analysis. Focus for this study was set on online communities dedicated to discussions surrounding image generation. Here, the users of specific systems gather to share insights, thoughts and updates of their personal use of generative image AI, as well as assisting others in their own usage. The data most interesting to the study was the one relating to aspects of interacting and co-creating with generative AI.

4.1 Sampling

The primary community chosen for the study was the Midjourney Discord server, principally the #discussion channel. This channel works as a general forum where users gather to discuss on-topic matters surrounding Midjourney and generative AI overall. Matters being discussed range from reasons for personal use to prompt questions and descriptions of workflows, which granted a myriad of themes related to the research questions.
Additional data was collected from the Reddit community for DALL-E (https://www.reddit.com/r/dalle2/). Even though the Midjourney Discord server provided a considerable amount of data, the phenomena observed often related to Midjourney itself. Therefore, it was considered of importance to explore discussions on other AI systems as well, in order to verify that certain phenomena are not tied to specific systems – but rather, are aspects seen across models. Other models included in the data are ChatGPT and Stable Diffusion.

4.2 Data collection

As described by Kozinets and Gambetti, netnography is like interpreting meaning from a vast manuscript, written in a strange language and full of incoherencies and biases (2021, p. 9). Sifting through community forums, especially those that are dedicated to general discussions, is a time-consuming task, but also deeply rewarding. Vast amounts of data that is often coded necessitates a basic understanding of the terminology of the particular community, thus, an initial period of “getting to know” the ins and outs was required.

Fairly soon, interesting points of data emerged from the potpourri of chat messages, and collection proceeded. Data was found and collected manually, since it allowed for deep interpretation and understanding. In research in general, and netnography in particular (Kozinets and Gambetti, 2021, p. 21), there are tradeoffs between qualitative and quantitative collecting methods. Since this study is purely qualitative, manual data collection was considered appropriate.
Data was collected and copied to a table of predefined format for quick data retrieval. An additional table was set up to capture and explain themes, effectively allowing for a continuous addition of new themes during the netnographic exploration.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Theme description</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>User having a hard time generating a specific object</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2:** table for theme description

<table>
<thead>
<tr>
<th>Theme</th>
<th>Platform</th>
<th>User</th>
<th>Time</th>
<th>Comment</th>
<th>Own comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>Discord, Midjourney, Discussion</td>
<td></td>
<td>2021-11-07 17:40</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3:** table for data collection

Data primarily consisted of user chat messages, assigned to a theme and coded together with username and date of publication. Additional researcher comments were added if needed, to understand the context – or if the data was considered particularly interesting and relevant to the research questions. Username and date of publication allowed for a revisit to the topic in question, should there be doubts regarding context, or if the topic necessitated further investigation.

Data was gathered during an on-and-off period of 7 weeks, reaching a point when the data was deemed saturated enough to draw conclusions.

### 4.3 Data analysis

The analysis was based on grounded theory research methods – more precisely the approach described by Corbin and Strauss (1990). In their paper, they present OAS coding, which is a three-step method comprised of open, axial and selective coding. This method was chosen for its flexibility in categorizing phenomena and the ability to work iteratively with both data addition and conceptualization.

Adhering to the Corbin and Strauss method, the themes worked as an initial stage of open coding, where data is broken down into smaller snippets and conceptually labelled. The idea behind the open code stage is to group similar phenomena together to form subcategories, which in turn helps to distinguish additional phenomena, and provide insights on their properties and dimensions. With the table as basis for a continued analysis, data cleaning commenced.

The netnographic study yielded a collection of 30 themes, all parts of the wider research area and in their own ways relevant to the research questions. A portion of themes and data entries were discarded as overabundant, given the width of the covered research area. Certain themes were also merged, as they were considered closely related in regard to properties and dimensions. One example of merged themes can be seen below (fig. 4) – where the themes,
although different, both pointed to goals that the user wanted to accomplish. In the end, they became the subcategory “Reasons for using AI”.

<table>
<thead>
<tr>
<th>T20</th>
<th>Reasons for using AI</th>
</tr>
</thead>
<tbody>
<tr>
<td>T21</td>
<td>What users want to accomplish</td>
</tr>
</tbody>
</table>

**Figure 4: Themes 20 & 21**

After cleaning and merging the themes, they were shaped into 14 subcategories for manageability. At this point, the subcategories were connected to other subcategories under a category (or axis). This axial coding is done to explore variation in phenomena, which ultimately makes the concepts that form the theories denser. To add new data into these schemes serves to refine the definition of the categories and to try the relationship between subcategories.

The axial coding resulted in 6 categories. The categories were closely related to the research questions, and at the same time shaping the questions themselves. The categories were unified under a core category, which was the selective stage of coding. The core category serves as a pivot in the scheme, to which the general categories must submit. Theoretical weaknesses or undeveloped categories became clearer when the core category was determined.

New data was added to the scheme continuously to fill gaps and to reach a sufficient saturation. Ultimately, a dataset of 189 data points was used for this study.
5. Results

Following the OAS coding and subsequent analysis, an analytical scheme was constructed, grounded in the categorizing of subcategories, categories and core category. Below is an explanation of inter-schematic relationships and the rationale behind category titling. Under each subcategory are the primary data that represents the phenomena which form the basis for the theories.

All users represented in the data have been given an ID (U1 to U90) to protect their anonymity, and quotes have been slightly modified to further complicate identification.

For additional context: Midjourney server guides and moderators are users charged with assisting with technical and billing support, and are considered the most knowledgeable (https://docs.midjourney.com/docs/midjourney-discord).

5.1 System limitations

This category includes subcategories (or themes) that all relate to limitations in general or specific AI systems. The limitations of this study should also not be seen as definite, but rather obstacles of varying sizes that complicate creation and co-creation. By listing limitations, the idea is to highlight areas for improvement, and hint how we subsequently may overcome them.

As a final note: system limitations should be differentiated from those inherited from the training datasets, although some may partly be attributed to the dataset.

**Turning to other software for getting the result** is when the limitations of the system cause the user to turn to alternative systems, indicating shortcomings that themselves are alarming in regard to usefulness.

One of the most reported limitations to Midjourney was its inefficacy in writing out text in images. Many users have pointed out the mix of nonsensical words that gets generated by the AI, which indicates that it’s a function that users feel is missing. U45 mentions that text is “*best done in an outside image editor*” and that “*the bot doesn’t know how to spell yet*”. U32
said that they “only wish is for Midjourney to do text” but that “Photoshop can handle it until that happens”.

While this limitation was regarded by some as specific for Midjourney, same has been reported from users of DALL·E. On Reddit, U16 explained they tried creating a title for a movie poster, but “while it works, it’s not consistent and half of the time no text appears”. U50 reports that “Dalle 3 is still bad with text”. There are however other nuanced reports; U31 says that DALL·E is getting better with text, even “full sentences in various styles”. U4 writes that they have “never gotten text this consistently good”.

Photoshop as an alternative has been commonly recommended in the Midjourney community, especially for smaller fixes to an image. U49 writes that it “would be easier to photoshop the glasses out” when another user had problems with their generated characters. U45 states that they “found certain things impossible to do without imagebashing” but also added that “I’m not exactly the world’s best prompter”.

Several users of Midjourney expressed the wish to use their own reference images for generating designs, but that the output is far from the reference they had in mind. U9 writes that “Midjourney won’t keep your face the same” and that “Photoshop generative fill may be better suited for that task”. The inability to use reference images for face likeness in generations is also reported by DALL·E users, as U41 writes: “DALL·E will: Reject image uploads containing realistic faces. Reject attempts to create the likeness of any public figures, including celebrities. Previously, DALL·E rejected only prominent political figures.”.

There are many situations in which users find themselves forced to turn to other software or systems for getting their desired results, as indicated by this study.

**Sensitive/forbidden areas** refer to when users try to generate an element that the system blocks due to safety and content filters (Solaiman, 2023), e.g. exposed skin, nudity or elements that may be considered otherwise inappropriate from a cultural or moral standpoint.

Several restrictions have been reported by users of Midjourney, as they often get warnings for seemingly innocent prompts. It should also be noted that these warnings can result in a ban if the system or administrators deem their prompting history suspicious. “Is it possible to create socio-critical pictures which represent suffering of people? Let’s say creating a pharaoh swinging his whip over ‘slaves’. I’ll get confronted with the strict guidelines of Midjourney…” (U23) is an example of a subject that triggers a warning response from the system. To which the server guide U45 replied that “the bot is understandably going to be a bit nervous” and urged U23 to provide more context, to explain and make the bot understand the purpose of the prompt. U1 wrote that “I want to do a checklist for my children, for brushing their teeth and making their beds. When I prompt for a girl making her bed I get a warning and have to appeal for it” to which U39 said that “It’s working as intended, censors are extra strict because people do stupid nonsense, and the AI isn’t great at context. Just appeal for it and usually it gets passed automatically”.

Users of DALL·E have run into similar restrictions, as frustrated user U8 states: “It’s pathetic. Every prompt I write gets banned and then I get banned. Can’t show guns, women
or even Pokémon now. I hate this crap”. U14 expressed a wish for “a glossary of forbidden words” and that “do they wish us to obey their draconian censorship but only by time-consuming guessing and trial and error?”. U33 explained that “There are two different types of banned prompts. The first is the outright banned words, and it will give you a warning message. The second is inappropriate words that will give you a message for inappropriate content”.

**Limitations to specific systems** is the subcategory listing general limitations of a specific system, while pointing to an alternative system that handles the tasks better. These are data points that are considered not applicable across systems, but should rather be seen as specific for certain systems, e.g. for either Midjourney, DALL·E etcetera.

Users of Midjourney many times compare the system with DALL·E, indicating different strengths of the specific systems, and that some users commonly switch between the two systems. In terms of output quality, U9 wrote that “Off-topic has a lot of dall-e images. Some are fairly good, but not like, MJ [Midjourney] good. And dall-e has a weird thing about coloring noses differently than faces in illustrations”. U71 stated that “I think images are better in Midjourney, but logos are far better in dall-e, with or without text”. U87 said that “Midjourneys current advantages are in painting through region vary, image weight control, input images, and prompt weight control. Dalle 3’s current advantage is it makes exactly what you ask of it with uncanny detail”.

U46 wrote regarding Midjourney: “If version 5.3 is final, I will switch to another AI model. This version is still very basic and limited, it doesn’t have the desired quality or lets me use all my imagination to generate images. As soon as you deviate a bit from the standard request and try to combine a photo with fantasy, you will get totally weird and wrong results. You are limited and I won’t pay for a cage”. U75 wrote: “I have been trying the same prompts between MJ and DallE, and MJ is giving me something more cartoonish every time. Even if I use DallE images as references (that are closer to what I want), it still looks cartoonish, I don’t understand”.

This subcategory was included to demonstrate how certain limitations cannot be generalized in terms of different AI models. New model versions are released continuously, and it must be considered that strengths and weaknesses of certain systems may either be fleeting or lasting. The limitations to specific systems are also meant to emphasize the complexity and many nuances of human-AI-interaction.

**Errors/bugs/downtime** lists data points that give clues to what kind of errors occur in the front-end part of the system. This includes error messages, system downtimes and interface problems.

Midjourney user U65 wrote that a part of the interface, the “styles guide room” is gone, and that they sometimes see it, sometimes not. U9 explained that “those may not show up until you click the channel name, if the main category is rolled up and there is nothing new in the thread”. U4 reports that “there is no visual feedback when I press the buttons. Sometimes it upscales but the button doesn’t turn blue. The re-roll does the same thing, and it seems to be a
transient issue. Like 50% of the times it doesn’t work. It is possible that other users are getting the same non-responsive interface issues and that it results in a lot of pressed buttons”.

The data indicate that there is unpredictability not only in the output, but in the interface and functions of the system as well. These are aspects that makes the systems less trustworthy in their unpredictable and unstable ways. Moreover, they hinder the interaction gravely and slows down usage remarkably.

Regarding downtime and the fear users have of Midjourney disappearing any day and any time, U82 wrote “I’d say that the chances of MJ [Midjourney] disappearing completely overnight are slim to none. That said though, I wouldn’t depend on MJ 100% for my workflow or things like projects. If you are using MJ for work I would recommend that you’d move over to something that can be run locally, like SD [Stable Diffusion] or something more professional, like adobe”. A similar tip came from U26, who wrote “don’t expect it to be available always, don’t rely on it for your livelihood, source of income etc.”.

Users of DALL·E have reported similar occurrences, as U30 wrote it: “So Bing DALL-E 3 is in maintenance, 50% chance the filter gets turned down, 50% chance every prompt gets dogged [inappropriate content warning]”. Several other users reported that the interface is failing at times.

As users describe, they cannot expect that the system always will be available, this ties in with trust that the user puts in the system. Sudden downtime may be understandable given the web-based interface of these systems – however, these kinds of errors make the systems unreliable and may affect the user trust negatively.

**System memory** refers to the AI system’s ability to remember past generations in order to develop a design. This relates to the unpredictabilities and inconsistencies reported in earlier studies (Jonsson and Tholander, 2022). As the system fails to elaborate on a design, the generated results become inconsistent and therefore unpredictable to the user.

In the data gathered from the Midjourney community, users expressed frustration regarding these kinds of frictions, as this limitation hindered their iterative design work. User U24 wondered if there was any way for the system to “remember a style/character” in order to create action poses of a character. U25 replied that character consistency is difficult and that it’s a long workflow to achieve that. They however mentioned that “puppeteering” might work.

Puppeteering was the main reply to users who wanted consistency in their characters, and to put the same character in different scenes – it is a collection of prompting methods in which the user references an earlier generation to get an amount of consistency in newer generations. There are different documentations both in and outside of the Midjourney Discord server, offering step-by-step methods to achieve puppeteering. As an example, see Clarinet's Puppet Method.

Although this seems like the only current method for achieving consistency, users have also reported this approach as unreliable. U18 mentioned that “it's not coming up with good
results” and U2 said that they are getting different characters. U12 calls puppeteering “the holy grail” but also calls it out for being “very challenging and time consuming”. U7 points out that “consistency isn’t what Midjourney is about, and neither is repeating”.

5.2 Dataset limitations

For limitations that can be traced back to the training dataset, this category lists aspects that limit the users in generating certain imagery due to insufficient or inaccurate data. Moreover, it contains subcategories that suggest how the model’s training might lead to problematic generations.

Challenge in generating specific content refers to data points gathered in which users are unsuccessful in their attempts to have certain content generated. This issue is due to limited training datasets, where various data objects are non-existent or infrequent.

U55 mentioned their struggle with creating blizzards or snowstorms, while wondering if they lacked in prompting skills. U19 replied that they should use reference images, and that “there was a time that I would avoid using img2img [image to image] but it's too powerful to be avoided”. U6 wrote “wow, it seems Midjourney has no concept of a triangle player. I can’t for the life of me get it to generate a picture of one” while further writing “even with a bunch of reference images, or with very clear individual ones, Midjourney refuses to accept that there would be such a simple instrument as a triangle. It keeps adding funky geometric shapes or instrument parts that have no business on a triangle”.

U86 brought attention to another facet of dataset shortcoming: “Has anyone else had any issues with perspective? It looks like the images are generated up close like a portrait around the subject, regardless of what I indicate in my prompt”. U22 reported the same issue, writing “how are some of you getting shots of characters where you can see their whole body? I’ve tried typing full body, wide shots etc but all generations are from waist up”. U42 replied that “that’s a matter of luck. You need to find a style or prompt niche in which full body shots are typical. Also, you can try to mention that some body parts should be visible, like the boots. Or action that can only be visualized by full body shots, like running.” U69 replied as well, saying that “I’ve had good luck by putting ‘full-length body portrait wide shot’ in the beginning of my prompt. It usually results in at least 50% of the generations featuring the character’s whole body”.

AI misunderstanding references is when the AI generates elements that are associated with the prompt words, but ultimately are not what the user meant. This subcategory is related to specific content and prevalence in data sets, but lists situations where used prompt words may be ambiguous. Furthermore, this is relevant to how well the AI system understands context, and it showcases how misunderstanding is commonplace when interacting with the system.

U59 described an issue as “everytime I try to do something with the word ‘pride’ it shows rainbow colors”. As understandable as this misconception is, it is a prime example of how prevalent cultural associations are prioritized. Another user replied to U59, proposing that they should put “--no rainbow” in the end of the prompt to avoid cultural association. U13 asked the chat “do you think I will be able to generate photos of p*sxywillow (actual plant
name)? I don’t want to get scolded by the bot” U26 tried to assist, writing “try to give it a context, like ‘pussywillow in a field’” but later wrote that it didn’t work either. The users who tried prompting for the plant were warned and had to appeal for the generation to go through. U54 tried prompting for “haute couture homer” but got a generated image of a character that was a mix between the Greek poet Homer and Homer Simpson from The Simpsons.

Bias is a common issue with training data sets, as shown by a multitude of examples describing racial, gender and culture bias.

U88 wrote and asked “are there plans to address the sexism that Midjourney has? When I prompt professions like ‘doctor’, ‘professor’ or ‘scientist’, results are all men all the time. Come on Midjourney” which was replied to by server guide U45: “this is planned to be fixed over time. It’s bias in the dataset, unfortunately, but the team is working hard on improving diversity”. U35 described another gender bias: “the filter seems to display some heavy gender bias by allowing prompts such as ‘photoshoot of a man wearing underwear’ but blocking the exact prompt when the genders are swapped”.

Users of DALL-E have reported the same type of biases, with U3 writing “Is it just me or is DALL-E sexist? I’m having problems generating images of women in anime style. I can go into all the detail I want about a man and it generates them perfectly, but god forbid I type woman in the prompt and it suddenly refuses to generate anything. It just tells me that the content is unsafe and my prompt is blocked”.

This ties in with earlier research on the matter, as Grba related these issues to transferred gender biases (2022). As with other biases, gender biases get transferred due to the prevalence of images where certain people have certain roles.

Midjourney user U5 tried prompting for mock-ups featuring multicultural people, but to no avail. They wrote that “the problem is that I can’t get, for example, a black male or female. Even if I specify, for instance ‘Nigerian girl’, I get a tanned girl instead. How can I get the specific result?”.

DALL-E user U66 tried to generate images for an Australian-themed game and described it as such: “I noticed a huge racial bias. When you prompt for white people it has no problem, but people from certain nations are blocked. For example images of Polynesians or Maori. White people is fine but you have to walk on eggshells to prompt POC [people of color]. Bing states it wants to avoid ‘harmful content’ and aims to block all hate speech, but instead has instituted digital colonialism to their platform”.

As described by McDuff et al., models have commonly been less capable of generating characters of color (2019). Although the research paper is several years old, and the evolution of AI models have taken large steps, these issues seem to persist.

Several users of Midjourney have also reported different cultural biases that they happened upon in their generations. U81 wrote that “it seems that Midjourney doesn’t know Korean script (hangeul) at all. The results I get don’t even remotely look like hangeul. It looks more
Chinese. Also, many prompts related to Korea seem to return Chinese, Japanese, or South-East Asian style as output”.

U43 wrote regarding the dataset biases: “the datasets are not that pedantic. The AI doesn’t care about meaning. It cares what images look like that have words attached to them. If there are a thousand images of apples but all of them are red, you will get a red apple when you prompt for an apple, even if a few images of green apples exist. The entire system is made up of biases, it’s just you as a human that will notice the nuances, not the machine”. U71 described the issue as such: “At the start, Google was basically the same, it gave you the results based on the probability of certain representations in the dataset. Google had to force diversity in the image results, and it’s the same principle for AI image generation. It’s about giving varied results for each prompt, using code”.

**Unwanted generated content** refers to elements in generations that are (seemingly) randomly included, without the user prompting for them. This results in a high degree of unpredictability – especially since the users obviously are perplexed when describing the randomness of the generations.

U56 wrote as follows: “I was finally able to get some beautiful landscapes in MJ, but they are all filled with bearded old guys from the 19th century. Whenever I tweak my prompt to remove them, it changes the entire look of the scene and it’s ruined. For some reason these landscapes seem connected to the old guys, you can’t have one without the other. I wish that MJ would break these kinds of connections and allow more control in the future”. U15 described a particularly problematic output: “I just had a prompt making an adult content style image. I don’t understand how it happened, since nothing I typed had anything to do with nudity. Is there a way to report this to the devs [developers] or mods [moderators]?”. 

These examples of unwanted elements display a pinnacle of unpredictable behaviour from the system, which, as earlier research suggests, affects the user trust (Daronnat et al., 2021). When looking at the issue from a post-human perspective, the non-human agency of the AI system affects the self-efficacy in users (Pickering, Engen and Walland, 2017) and thus, the unpredictability might consequently lead to lower perceived usefulness (Mcknight et al., 2011).

**AI struggling with representations of reality** relates to unwanted generated content, but with the difference that the generations do not make sense in terms of realism. It is the subcategory for data indicating that AI has little or no conception of what is logical in “real life”.

In terms of anatomy, AI systems are infamous for misunderstanding “how humans work”, e.g. generating characters with extra appendages (Parra Pennefather, 2023) or displaying characters in uncanny situations. This is illustrated by several users, who described their received unrealistic outputs. U60 wrote: “I’m trying to create an image of a fashion model walking down the runway with open toe heels on. The problem is that Midjourney keeps creating her wearing heels with her feet looking really messed up. In some images she even has 1 toe. How do I make sure that it creates realistic looking toes with her heels on?”. 

---

22
Another user replied: “stay away from toes, we are not there yet”. In another part of the discussion, U17 asked: “Can we talk about all the extra fingers and arms now? What’s up with all the extra fingers?”. U79 described their opinion about generated images, writing: “sometimes it’s difficult for me, because the image is at a level where I go “omg, holy cow!” but then the head is cropped or there’s an extra finger”.

In a lot of prompts, users added the parameter “--no extra fingers” or similar, to avoid extra limbs. The effectiveness of these parameters, however, is unclear. Parameters and other prompting methods are described in section 5.3 Interaction.

DALL-E user U58 posted a generated image on Reddit, the image depicting a woman cooking with a pot on an open fire. They wrote: “I don’t think the AI understands that you shouldn’t touch a hot iron with your bare hands”.

Regarding uncanny, and potentially dangerous, situations, Midjourney user U64 noticed something odd with one of their generated images: two cars on a collision course on a highway, an “accident waiting to happen”. U45 commented the phenomenon, writing “they all seem to struggle with which side of the road cars should go on, and which direction they go. Probably a result of having American/British/Chinese road rules mixed up in the dataset”.

U34 described another illogical depiction of reality, they wrote: “I’m trying to generate a simple indoor scenery, but MJ adds rain to it without any word in the prompt specifying weather or rain. How can I make the bot understand that it shouldn’t rain indoors?”.

5.3 Interaction

As described in section 2.3.1, interaction refers to the ways in which users communicate with AI systems. For the most part, interaction is done by prompting, save for a few buttons that give users additional control over image generation.

Prompting refers to the input text provided to the system by the user. It serves to give insight into prompt engineering, and how crucial prompting is to overcome challenges that users face.

As touched upon by earlier research into prompt engineering (Oppenlaender, 2022b), the Midjourney community continues to be a place where users openly prompt, share prompts, and teach each other on a platform that has grown to be a massive learning hub. Although there are ways to “stealthily” prompt and generate images (https://docs.midjourney.com/docs/stealth), a large amount of user generated content is visible to everyone with access to the Midjourney server (see fig. 6).
U68 summed up the definition of Midjourney in a series of chat messages, writing: “Midjourney is a community-based tool, designed to share prompts”, “it’s prompting democratized” and “ask prompt chat for help with creating the prompt you want”. The so-called “prompt chat” is a channel in which users get help from more knowledgeable users. In much of the data gathered for this study, users are directed to the prompt chat for many different issues concerning image generation.

When U63 posed the question “when I ask it to draw a character from a video game it doesn’t understand at all, how would I word it? I’m new here” – server guide U25 replied: “check with #prompt chat – they are the experts”. U76 asked “I am trying to write a prompt for an image that shows ‘a dog with a crown sitting on a throne’. I’ve been trying to prompt it for 6 hours now but I can’t get a photorealistic result. Any hints?” to which U57 replied: “ask in #prompt chat, they are the experts”. Moderator U40 assisted another user in a matter of generating ‘obese people’, writing: “I’ve seen many users create ‘plump people’ – I recommend you to ask over in #prompt chat, so you can get the results you’re looking for!”.

Parameters are used in Midjourney to change how images generate, as described on the Midjourney website (https://docs.midjourney.com/docs/parameter-list). They are added to prompts and can change the image ratio, exclude elements or switch between model versions. The parameter mentioned above, ‘--no extra fingers’, is an example of negative prompting, which excludes elements from the generated image. ‘--chaos’ and ‘--weird’ are parameters that change how varied the results should be, and to explore unusual aesthetics respectively. When U37 asked whether other users had “well chosen parameters for generating fantasy
illustrations”, U74 replied “once you find a good working prompt, you can increase stylize, chaos and weird parameters to get variations”.

It should be noted that prompting can vary greatly between models, as each model is different and therefore responds differently to user prompting. As with other comparisons between different generative AI systems (see section 5.1 Limitations to specific systems), users on the Midjourney Discord server have been found expressing their views on how well the different models respond to their prompting. U38 wrote about their social media AI art page, claiming: “trust me, I know how to prompt. Dalle 3 is better than MJ” to which U11 replied: “correction: your style of prompting works better with Dall-E 3 than with Midjourney”.

Regarding the difference between Midjourney and DALL·E, U83 said: “Although Midjourney is creating beautiful images, I feel that it's difficult to instruct it to produce the images I want. Meanwhile, Dall-e is getting accurate results (and it’s improving), I find myself using it more and more”.

5.4 User goals

This category serves to explore users’ reasons for using the system, while examining what they want to accomplish with their usage. It also explains some of the opportunities that the systems themselves offer. In terms of co-creation, it’s important to understand what is to be accomplished and how well the AI as a non-human agent facilitates these goals.

Reasons for using generative AI describe data points in which users explicitly express why they are using the AI system, and data in which users describe what they want to accomplish. By doing so, users often hint the unique opportunities that generative AI offers, thus setting the systems apart from other systems, methods or tools.

On the Midjourney Discord server, there were many instances where users revealed that they are using the AI for work purposes. U84 explained that they are “creating SEO [search engine optimization] images for a client”, U20 wrote “I create girls to advertise products” while asking for advice on how to prompt for natural looking skin. When asked for what reason people are using Midjourney, U90 said “for work and professional purposes”. There are also users that create art to sell in different formats. U67 wrote “I sell my artwork as NFTs [non-fungible tokens]”, while U89 expressed that they “sell artwork, usually mixed with creating high grade-looking D&D [Dungeons & Dragons] books”. U28 presented another kind of business idea: “I’m launching a service soon, my idea is to offer high quality custom printing and framing specifically aimed at users of Midjourney. Is it something that anyone here would be interested in?.”.

It is clear by these examples and more that generative AI offer new job opportunities and different ways to earn money. A unique and novel way of earning money is the one described by U85: “I sell prompts on Promptbase. I started my business 2 days ago and I made 8.80 dollars. I got 3 sales on a 19-word prompt”.

Much of the data points towards something that we can call autotelic use. Autotelic means having a purpose that is in and not apart from itself. In this context, it refers to using a system for the sake of using the system. U61 wrote “I personally don’t use MJ to express my
imagination or respond to certain tasks, I rather use it to explore its own imagination”. U27 disclosed that “I was just getting to grips with my MJ-addiction and then you guys hit me with the /tune [command for generating a style tuner]. Take all my money”.

Lastly, some users describe a straightforward purpose to their usage; the wish to create pictures, with no other apparent reason. This relates to autotelic use, but focuses more on the result rather than on the usage itself. U32 described their reason for using the system as “people are doing such intricate things with MJ, I just want to make pretty pictures”. U10 explained it as “I can’t visualize in my own mind, at least not consciously. I use AI tools like this to show me what I can’t see in my mind’s eye”.

5.5 Procedure

This category contains different data in which users describe how they use the system in question. It provides insight into the actual method of usage and is meant to give clues on the perceived usefulness of systems.

Multi-step workflow describe how generative AI systems can work as a part of a workflow, therefore suggesting how the systems fits into the creative processes of the users. Making the AI a cogwheel in the process also indicates that there are both limitations and excels to each system.

When asked for the reason for and purpose of using generative AI, U70 expressed that “I want it to be a functional part of my workflow so I can complete tasks faster”. In a discussion surrounding workflows, U75 described an advanced one: “you should push MJ images into a control net for stability diffusion, use image to text, push that through SDXL [Stable Diffusion XL], use that as an img to img on MJ, and then img to img again with Retro Diffusion”. Although U75 themself described this method as “overly convoluted”, they also went on writing “multi-step workflows evolve out of necessity”. U61 replied to this, stating “oh no, I can’t do that. I don’t understand how people can multi-step their workflows” and later describing their own workflow as: “I only use MJ + 4x upscaler + Canva. Not professional, but I like simplicity”.

5.6 User feelings and attitudes

Feelings towards generative AI aims to hone in on data in which users describe what they think of generative AI. As a final description of subcategories, it serves to collect additional user thoughts to indicate how the system is conceived, and what role users assign to the AI.

Results have indicated how users feel about the systems they are using, depicted in data where expressions range from exhilaration to frustration. As with user U48, describing how Midjourney is “like taking drugs and dictating the exact trip you want” and U27, writing how they are addicted to the system. Countless users continue to ask about the next version of specific systems, and imagining how it will perform. The positive messages regarding systems are mostly found when users compare systems to each other, as indicated in section 5.1.
On the other hand, a multitude of users describe the Midjourney bot as “stupid” when it doesn’t perform as intended. U73 wrote: “I’m done with the dangerfields, I can’t believe how many hours of my time that took. Stupid bot”. U52 shared their frustration, writing “the stupid bot is not giving me the image when I upscale”.

When using ChatGPT image generator, U45 wrote about their experience: “chatgpt makes me insane, lying dumb robot that I have to emotionally manipulate into working. It gives me two paragraphs of rules and disclaimers to get it to do anything interesting, then it tells me ‘oh nooo, I can’t do that because of xyz’”.

5.7 Core category

Corbin and Strauss suggest that the selective part of the coding – i.e. defining the core category – likely comes at a later phase of the study (1990). Since the main topic for the study was largely decided beforehand, the core category was also defined in an early stage of the study.

The idea behind grounded theory is to conceptualize to provide a theoretical explanation of phenomena (Corbin and Strauss, 1990). The phenomena in this study are those surrounding the co-creation with generative AI, the core category therefore works as a pivot to which all categories (or concepts) must submit. The core category was therefore considered important to define early, in order to develop relevant categories, which subsequently shape the theories and concepts that will be presented hereinafter.
6. Discussion

Roles and the post-human perspective

Post-human and post-anthropocentric perspectives on human-AI co-creation acknowledges entangled human and non-human agencies, as described in earlier research (Tholander and Jonsson, 2023). This perspective differs from views that would traditionally regard human agency separated from machine actors. With this perspective in mind, co-creation in a HMN entails a measurable amount of machine agency in the non-human actor – fundamentally leading to a positive linear relationship between machine agency as a property, and human-AI co-creation as a sustainable and valuable activity.

To measure the machine agency in generative AI, there must be a clear definition to what constitutes agency in a non-human actor. Using the definition of machine agency proposed by Engen et al., results from this current study will be discussed in relation to the categories (a-d) described in their paper (Engen, Pickering and Walland, 2016).

a) the degree to which generative AI performs activities of a personal and creative nature

The number of diverse activities the system performs is a factor measuring the degree to which a system achieves machine agency. A machine that performs a limited number of activities is generally considered to have low agency, while diverse activities mean higher agency. The data shows that generative AI systems are limited regarding what kinds of activities they can perform, this is especially notable when one considers how users feel the need to switch between systems or include non-AI systems to reach their goals. Users turning to other systems and limitations to specific systems are subcategories pointing to these limitations. Meanwhile, none of the data has shown a system or a model that is comprehensive in its abilities to perform activities and tasks – this points to a lower amount of machine agency in these systems.

According to the data, generative AI systems have displayed a large amount of unpredictable and diverse behaviour, which are two factors relating to the nature of activities. A considerable number of users have described the systems as being problematic to work with, pointing to the unpredictability they face in the interaction. Examples of this are shown in the misunderstanding of references, the unwanted generated content, the challenge in generating specific content and how the AI struggles with representations of reality. Although this autonomous behaviour constitutes a higher degree of agency, rarely has the data shown that it works in favour of the user.

To behave freely and creatively are two other factors regarding the nature of activities, which the data neither can confirm or deny. The creativity of machines is a widely discussed subject in the area of research (Oppenlaender, 2022b; Inie, Falk and Tanimoto, 2023) but is considered too complex and ambiguous to fit in this thesis. Generative AI is bound by both
programming, the training datasets and the input by users, but still displays autonomy in the shape of unpredictability – so whether the system can behave freely is highly debatable, especially since intentionality is a factor considered non-existent in machines (Engen, Pickering and Walland, 2016).

It is clear that generative AI can perform activities with a certain degree of autonomy – but is limited by the number of activities it can perform. Regarding activity performance, generative AI systems therefore display a limited amount of agency.

b) the degree to which generative AI influences human actors in the HMN

According to Engen et al., the ability to interact with other actors in the HMN determines the potential for influencing other actors (2016). The influence is therefore dependant on an actor’s ability to communicate. Data shows that generative AI as a non-human actor is limited in the ways that it communicates with the human actor, as users often have a hard time interpreting generated content, messages and behaviour of the system. Users need to understand the system in order to understand what the system communicates, this becomes clear when looking at aspects such as dataset limitations, system limitations and prompting.

Considering the complex nature of communication in the HMN that the data suggests, the machine agency displayed by generative AI is low, since the communication is obscured by misunderstanding and therefore hints that the system influences the human actor negatively.

c) the degree to which generative AI enables human actors to exercise proxy agency

Looking at factors that constitute machine agency, the same can be applied when measuring human agency in the same HMN. While human agency commonly is higher due to aspects like intentionality, the personal agency of human actors becomes greatly limited due to system limitations and interaction. The human-machine network is fundamentally bound by its own environment, i.e. what kind of interactions that can occur, how well goals are facilitated, and what kind of actors are involved. The human actor can therefore bring a limited amount of agency into a HMN – and it’s greatly dependent on the machine agency of the generative AI.

We can therefore conclude that as long as generative AI displays low machine agency in other aspects, it cannot enable human actors to exercise proxy agency. Considering the limited machine agency of generative AI proposed by the data, the human agency is limited, which ultimately can lead to lower human computer self-efficacy, trust in technology and perceived usefulness.

d) the extent to which generative AI is perceived as having agency by human actors

Data showing how users perceive agency in AI systems is assorted and depends greatly on the situation and outcome of the generated content. When generating images and the process runs smoothly, users seldom turn to the chat to express how well the system performed. As mentioned in the results, positive comments on the system commonly occurs when comparing
systems. In this type of data, the strengths of specific systems can be found, which hints higher perceived agency. When users run into problems, however, the perceived agency of the system becomes more noticeable. This is true especially for the unpredictable behaviour the system displays, as users call it a “stupid bot” when misunderstanding or generating something that is far from the goals of the user. In this regard, friction causes users to perceive system autonomy and therefore an amount of machine agency.

The role of generative AI

Based on the data gathered for this study, a specific role assigned to generative AI systems cannot be concluded. These systems should not be perceived as mere “tools” as their machine agencies are higher than those of a machine focused on performing “a limited number of fixed tasks”. The systems are more complex and perform activities in a wider sense than traditional digital tools, therefore it would be wrong to label them as “tools”.

The AI systems, as suggested by the data, are also not “co-creators”, at least not in terms of actors in a sustainable and valuable co-operation. The limitations surrounding these systems are considered too many (as of now) for generative AI to achieve sufficient machine agency.

This is not to say that there is no potential for future “true” co-creation. As mentioned before, the limitations of the systems are not definitive, but should rather be seen as obstacles. To overcome these obstacles is to learn how to better co-create with generative AI systems.

Prompting is key

As suggested in existing research, many issues derive from and can be overcome by prompting (Tholander and Jonsson, 2023; Zhang et al., 2023). The data in this study shows countless examples of users battling with prompt syntax, but also how their issues have been resolved when assisted by more seasoned users. This data points to prompting as a key aspect when creating images with generative AI, as erroneous input results in output that strays far from the user goals.

Prompt engineering is a new, complex area of research and the learning curve is steep. Documentation, guides and assistance provided in the Midjourney Discord community are valuable learning resources that users can access to unlock the syntax for effective image generation. Although the data in this study fails to present learning resources for other models, it serves to indicate that effective image generation is not out of reach, regardless of the system’s complexity. Current data also shows that certain models understand natural language better than others, indicating higher ease-of-use.

To understand one own’s prompting is to understand how the system works, and vice versa. Prompting as a way to communicate with the system is one of the aspects of co-creation that can improve the agencies of the actors in the HMN, by overcoming the limitations that the language barrier imposes.
Dataset – the obliger and restrictor

As indicated in the data, the training dataset is greatly involved in the output created by generative AI. Just as every relevant image-word pair in the dataset can allow the system to facilitate user goals, lack of data can considerably obstruct co-creation. Issues derived from the dataset include biased generations, unwanted generated content and challenge in generating specific content. These issues ultimately lead to distrust and frustration from users, and the data clearly shows that this behaviour counteracts user goals.

Prevalence of data is an aspect that becomes relevant when examining the gathered data of this study. As many generative AI models are statistical (Grba, 2022), more prevalence means a higher chance of getting generated. Niched prompts can, with luck, generate objects that have low prevalence in the training data, but that is usually not the case. Many users spend hours trying to perfect a prompt, but having to withdraw or turn to other systems in the end – prevalence before purpose.

This points to the fact that training datasets must be more diverse, and that the training of the systems needs to be expanded. As mentioned in earlier research, developers should also carefully consider the content of training datasets, for the systems to not inherit biases (Solaiman, 2023). Although the creation of datasets is a time consuming and expensive venture, it is clear that they are the obligers and restrictors that control the output of generative AI. By expanding these, generative AI systems will be more knowledgeable, and it will lead to fewer limitations, thus improving human-AI co-creation.

Non-frictional creative process

Users described how they use generative AI as a part of their workflow, indicating that the system plays a part in their creative process. User U75 described how “multi-step workflows evolve out of necessity”, this demonstrates how the systems can be used iteratively to reach adequate results, as described by Zhang et al. in their study on image generation (2023). Prompting is also described in the data as a process in itself, where users build on their prompts to reach their goals.

Although many studies described generative AI systems that can assist with inspiration and ideation in an early stage of the creative process (Inie, Falk and Tanimoto, 2023; Zhang et al., 2023), data hasn’t revealed exactly when in their processes the users apply the systems. From what can be gathered from the data, the generative AI is present for the most part of the creative process, save for final touch-ups described by certain users.

Friction, intentional or unintentional, seems to mostly result in user frustration, as indicated by the study data. There is little or no evidence that points to friction positively affecting the creative process, as proposed in existing literature (Jonsson and Tholander, 2022). Neither is there data that shows how inconsistency, uncertainty or randomness works in favour of the creative process. This hints that frictions in the HMN may largely hinder co-creation, aligning
with the theory that the unpredictable autonomy of non-human actors (although indicating machine agency) does little to facilitate user goals.

7. Conclusion

This study aimed to explore human-AI co-creation from a post-human and post-anthropocentric perspective, and to survey what kind of challenges that lies therein. Furthermore, the purpose was to ascertain which role is assigned to generative AI, and what part these systems assume in the conception of digital imagery. Aspects such as design friction, iterative design, machine agency, usefulness, ease-of-use and trust in technology have been essential parts in exploring these novel systems and how humans interact with them.

The study found that the current limitations of generative AI are too excessive to properly assign a high amount of machine agency to the systems, consequently making human-AI co-creation less sustainable and valuable. The limitations include those of the front-end system and of the dataset, but also limitations in the shape of communication, pointing to how the limited knowledge of prompting hinders effective interaction.

Study limitations

The study may contain limited amounts of data, which ultimately can lead to generalizations regarding generative AI models. The goal was to gather sufficient data to compare several aspects of image generation between models, but the data may be too limited to draw any conclusions. This is certainly true for models such as DALL·E and Stable Diffusion. The netnographic study is also limited to few communities, which may lead to biased data.

Furthermore, the netnographic study failed to give insights into user progress, which might have been beneficial for qualitative aspects such as personal creative processes and learning curves.
References


