# CONCEPT FORMATION IN COMPUTATIONAL CREATIVITY

GIOVANNI LION

PhD

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University

School of Design

# Concept Formation in Computational Creativity

Giovanni Lion

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

April 2023

# Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgment has been made in the text.

(signed)

Giovanni Lion (Name of the student)

I dedicate this thesis to my family and friends whose love, support, and encouragement have been essential in shaping my academic journey.

Thesis advisor: Prof. dr. dr. Johan F. Hoorn Giovanni Lion

# Concept Formation in Computational Creativity


## Abstract

This thesis investigates the relationship between creativity and machines through the lens of mediation theory. The research provides a thorough review of existing Theories of Concepts (TOCs) and examines their influence on the evolution of rule-based and data-driven computational approaches, in the context of creative practices. The research objective is to establish a theoretical framework able to describe the two approaches and utilize it to identify and validate critical factors of concept formation that affect the creative process, with particular focus on the observed trend towards data-driven technologies.

An extension of mediation theory is proposed, distinguishing between two computational approaches and their associated TOCs. The first two studies presented are aimed exploring the practice of concept representation using data-driven tools, with the objective of identifying the process' critical aspects. The first study is a collaboration with a music composer aimed at training a model to generate music in her unique style. The second study explores the reflective potential of dataset curation using generative adversarial networks, in partnership with a fellow researcher and photographer. The third study investigates the relationship between critical factors identified in the first two studies, within the context of text-to-image generation using StableDiffusion.

The findings highlight the significance of dataset curation for artists and designers adopting data-driven tools. From the studies, language emerges as a powerful interface for concepts, with potential implications on human and non-human creativity as large language models advance. The research indicates a possible shift in focus from product to process in creative practices, emphasizing the need for adaptation and skill development in the age of abundant content generation.

# Contents


1 Introduction 1
- 1.1 Technology and creativity . . . . . 1
- 1.2 Concepts and machines . . . . . 5
- 1.3 Two approaches to creative machines . . . . . 7
- 1.4 Societal impact . . . . . 8
- 1.5 Thesis summary . . . . . 10

2 Literature review 12
- 2.1 Concepts . . . . . 13
- 2.2 Technology and creativity . . . . . 31
- 2.3 Summary . . . . . 62

3 Methodology 63
- 3.1 Technological mediation . . . . . 64
- 3.2 Creativity theories . . . . . 71
- 3.3 Practitioner research . . . . . 82
- 3.4 Studies overview . . . . . 85

4 Study: Music 87
- 4.1 Scope . . . . . 88
- 4.2 Collaboration in practice . . . . . 91
- 4.3 Method . . . . . 93
- 4.4 Results and discussion . . . . . 95

5 Study: Images 99
- 5.1 Scope . . . . . 100
- 5.2 Collaboration in practice . . . . . 102
- 5.3 Method . . . . . 103
- 5.4 Results and discussion . . . . . 108

6 Study: Text-to-image 111
- 6.1 Scope . . . . . 112
- 6.2 A clash of paradigms . . . . . 115
- 6.3 Stable Diffusion . . . . . 117
- 6.4 Method . . . . . 118

6.5 Results and discussion . . . . . 124
6.6 Limitations and further research . . . . . 129

7 Discussion **131**
7.1 Datasets curation . . . . . 132
7.2 Language as interface for concepts . . . . . 136
7.3 Blurring human and non-human creativity . . . . . 139

8 Conclusions **142**
8.1 Research summary . . . . . 143
8.2 Post-phenomenology of creativity . . . . . 144
8.3 Limitations . . . . . 146
8.4 Further research . . . . . 148

Appendix A Supplement to the literature review **151**
A.1 Scopus search retrieval . . . . . 151
A.2 GPT categorization prompt . . . . . 155
A.3 Top cited paper and their research question . . . . . 157

Appendix B Questionnaire items and details **162**

References **164**

# Listing of figures


1.1 Stable Diffusion and ControlNet for image transformation . . . . . . . . . 4
1.2 A fork in the road . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

2.1 Physical representation of Turing's machine . . . . . . . . . . . . . . . . 39
2.2 The Zuse Graphomat Z64 plotter . . . . . . . . . . . . . . . . . . . . . . 43
2.3 Four computer-generated artworks . . . . . . . . . . . . . . . . . . . . . 44
2.4 Rosenblatt's first model of artificial neural network . . . . . . . . . . . . 46
2.5 Gärdenfors's conceptual space of colors . . . . . . . . . . . . . . . . . . 49
2.6 Systematic literature review: distribution of *Scope* . . . . . . . . . . . . . 57
2.7 Systematic literature review: distribution of *Medium* . . . . . . . . . . . . 58
2.8 Systematic literature review: breakdown of *Scope* by *Medium* . . . . . . . . 59
2.9 Systematic literature review: breakdown of *Approach* by *Year* . . . . . . . . 60
2.10 Systematic literature review: breakdown of *Scope* by *Year* . . . . . . . . . . 61
2.11 Systematic literature review: breakdown of *Approach* by *Scope* . . . . . . . 61

3.1 Methodology . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 64

4.1 A sample of Vicky's produced Canto-pop hits . . . . . . . . . . . . . . . 92
4.2 Generation pipeline of the first iteration . . . . . . . . . . . . . . . . . . 93

5.1 Storefront in Sham Shui Po as inspiration for dataset . . . . . . . . . . . . 104
5.2 Screenshot of the training process nearing completion . . . . . . . . . . . 105
5.3 Samples generated from the first iteration . . . . . . . . . . . . . . . . . . 106
5.4 Samples generated from the second iteration . . . . . . . . . . . . . . . . 107

6.1 Generated images using Stable Diffusion and the prompt "two circles" . . . 114
6.2 Facebook post expressing frustration about AI-generated content . . . . . 115
6.3 The explosion of AI-generated images . . . . . . . . . . . . . . . . . . . . 116
6.4 Example of the `/image` slashcommand interface . . . . . . . . . . . . . . 120
6.5 Example of the output of the `/image` command . . . . . . . . . . . . . . 121
6.6 Initial user-uploaded image for `/riff` . . . . . . . . . . . . . . . . . . . . 122
6.7 Output of `/riff` command . . . . . . . . . . . . . . . . . . . . . . . . . 122
6.8 Output of `/interpolate` command . . . . . . . . . . . . . . . . . . . . . . 123
6.9 Summary of distributions . . . . . . . . . . . . . . . . . . . . . . . . . . 125
6.10 Q-Q plots for tolerance for ambiguity . . . . . . . . . . . . . . . . . . . . 126
6.11 Q-Q plots for number of generated images per user . . . . . . . . . . . . 126

6.12 Linear regression on EA with TA and EHC . . . . . . . . . . . . . . . . . . . 127

# Listing of tables

2.1 Summary of theories of concepts . . . . . . . . . . . . . . . . . . . . . . . 31

3.1 Summary of human-technology relations . . . . . . . . . . . . . . . . . . . 66
3.2 ACASIA modules contextualizations . . . . . . . . . . . . . . . . . . . . . 81

6.1 Overview of workshops . . . . . . . . . . . . . . . . . . . . . . . . . . . . 119

A.1 25 most cited papers in the systematic literature review . . . . . . . . . . . 161

# Listing of Abbreviations

| | |
|---|---|
| **AI** | Artificial Intelligence |
| **AGI** | Artificial General Intelligence |
| **CA** | Conceptual Atomism |
| **CC** | Computational Creativity |
| **CLIP** | Contrastive Language-Image Pre-training |
| **CT** | Classical Theory |
| **DL** | Deep Learning |
| **DNN** | Deep Neural Network |
| **EA** | Expectations of Alignment |
| **EHC** | Expectations of Human Compatibility |
| **ET** | Exemplar Theory |
| **GAN** | Generative Adversarial Network |
| **GDL** | Generative Deep Learning |
| **GOFAI** | Good Old-Fashioned Artificial Intelligence |
| **GPT** | Generative Pre-trained Transformer |
| **GPU** | Graphics Processing Unit |
| **LDM** | Large Diffusion Model |
| **LLM** | Large Language Model |
| **LoRA** | Low Rank Adaptation |
| **ML** | Machine Learning |
| **NLP** | Natural Language Processing |
| **NT** | Neoclassical Theory |
| **PR** | Practitioner Research |
| **PT** | Prototype Theory |
| **SD** | Stable Diffusion |
| **TA** | Tolerance for Ambiguity |
| **TOC** | Theory Of Concepts |
| **TT** | Theory-Theory |
| **VAE** | Variational AutoEncoder |

# Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisor, Johan Hoorn, for his unwavering support and guidance throughout my graduate studies. His expertise, dedication, and encouragement have been invaluable to me.

I would also like to thank the members of my dissertation committee for their time, feedback, and insights. Their constructive comments and suggestions have greatly improved the quality of my work.

Special recognition goes to all the individuals who have contributed directly to this study. I would like to extend my heartfelt thanks to Vicky and Marty, for their collaboration on the studies that are part of this thesis. Their collaborative efforts have not only enriched the quality of this work but also made the research journey more rewarding and enjoyable. Additionally, I am grateful to the participants who generously shared their experiences and perspectives, as well as mentors and colleagues whose invaluable insights and expertise have enriched my research.

I would also like to extend my gratitude to Derek Lomas and Willem van der Maden from TU Delft for their generosity in providing me access to GPT. Their support has enabled me to carry out my research and complete my thesis successfully.

Special thanks go to Mikey and Tesfa for their invaluable help in proofreading my thesis. Their careful attention to detail and suggestions have helped improve the clarity and coherence of my writing.

In addition to my family's love which has sustained me during this journey, I am grateful to my friends who have been my cheerleaders and confidants throughout the process. A special mention goes to Kay and Luci for taking care of my basic necessities such as food, coffee, and for reminding me to sleep when work consumed me. Their encouragement, thoughtfulness, and camaraderie have made this experience more meaningful and enjoyable.

Lastly, I would like to express my appreciation for everyone else who has played a role in my academic journey—your support, guidance, and belief in me have made a significant impact on my life. Thank you all for your contributions and support.

*Do not follow where the path may lead, go instead where there is no path and make a trail.*

Ralph Waldo Emerson

# 1 Introduction

## 1.1 Technology and creativity

It is generally agreed that human technology acts on nature for a purpose: an axe can chop a tree to make firewood. *Prima facie*, the essence of technology may be superficially conceptualized as instrumental. While it is true that the first stone tools were used to help with the physical tasks of hunting and gathering, it is also true that the same tools were used to engrave caves with non-utilitarian abstract patterns (Rodríguez-Vidal et al., 2014). The first writing was used at first to keep records of crops (Hallo, 1971), but it quickly became a defining tool of our society, its social laws, religious beliefs and cultural identity. These artifacts tell us who we are, how we see the world and what our place in it is (Ihde, 1990; Rosenberger & Verbeek, 2015). Technology and creativity work hand-in-hand to create the fabric of human society. If our existence and perception of the world is indeed so heavily influenced by the technological environment around us, it would be limiting to see technology just as means to an end (Feenberg, 1991; Heidegger, 1977; Ihde, 1990; Winner, 1980).

Throughout history, we have examples of how technology has disrupted how we produce, consume and evaluate creative artifacts. The invention of the printing press in the 15th century changed the way books were produced and distributed. Likewise, the invention of the camera in the 19th century transformed the way we capture and share images. The invention of the computer in the 20th century revolutionized the way we store and process any type of information. Arguably, generative tools based on Machine Learning (ML) are on the verge of redefining the creative industry. Today, "producing large amounts of quality text, images and music is much less labor intensive than it ever was" (Colton et al., 2020, p. 1).

For example, in a not-so-distant future it may be possible to generate narratives about any given topic by training a ML algorithm on a large corpus of text. This is already possible to some extent: the current state of the art is such that it is possible to generate relatively complex stories starting from a simple *prompt*. Here is an example from GPT3.5 (Brown et al., 2020):

> PROMPT: Write a short story with a moral, with protagonist Jack, Mary and their pet dragon George who is very fond of pizza.
>
> ---
>
> Once upon a time, there was a boy named Jack and a girl named Mary who owned a pet dragon named George. George was an unusual dragon because he was very fond of pizza. In fact, he would do anything for a slice of pizza. Jack and Mary loved George very much, so they would always order pizza for him whenever they could. One day, Jack and Mary noticed that George was getting bigger and bigger. They realized that he was eating too much pizza and not exercising enough. They knew

they had to do something about it. So, they decided to take George for a walk every day and limit his pizza intake.
At first, George was unhappy about the change. He missed his favorite food and didn't want to exercise. But Jack and Mary were patient and persistent. They knew that it was for his own good. Slowly but surely, George started to enjoy his walks and even started trying new foods that were healthier for him.
One day, while on a walk, they came across a group of people who were littering the park. Jack and Mary were upset and wanted to do something about it. George, who had now become fit and strong, decided to take action. He picked up the litter and put it in the trash bin. Everyone was amazed by what he did and started applauding him.
From that day on, George realized that he didn't need only pizza to be happy. He found joy in helping others and doing good deeds. Jack and Mary were proud of their pet dragon and were grateful for what they had learned. They learned that sometimes, the things we think we need the most may not be what's best for us. It's important to find balance in life and to take care of ourselves and our surroundings.
The moral of the story is that we should not let our desires control us. We need to learn to make healthy choices and take care of ourselves. We also need to take care of the world around us and help others when we can. With determination and persistence, we can achieve anything we set our minds to.

In a similar way, generating images from text descriptions has become a trivial matter of seconds. We are now able to generate photo-realistic images, digital illustrations, sketches and much more from a few keywords. The impact that this technology might have in the workflow of graphic illustrators, designers, architects in the next ten years is quite substantial. For example, Large Diffusion Models (LDM) such as Stable Diffusion (Rombach et al., 2021) make it possible to transform basic sketches into full fledged images (see Figure 1.1).

The sound domain is also filled with new tools and applications that make use of large public datasets. Generative tools can be found in the field of speech recognition and synthe-

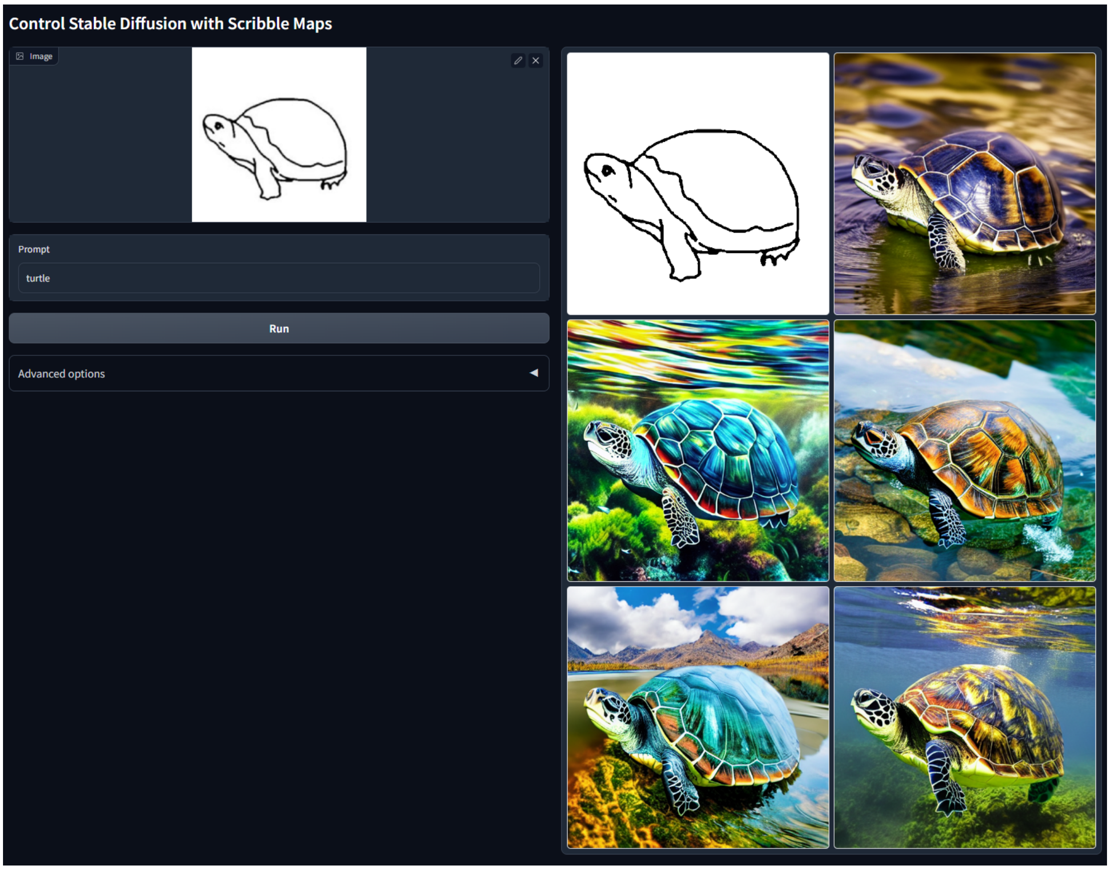


**Figure 1.1:** Illustration of Stable Diffusion and ControlNet, a technique for transforming sketches into full images by Zhang and Agrawala (2023). The image demonstrates the effectiveness of the method in generating clear and accurate images from rough sketches, highlighting the potential applications of this technology for creative tasks.

sis, for example OpenAI's whisper (Radford et al., 2022) and Google's audioLM (Borsos et al., 2022). For music composition and generation, it is now also possible to generate rich and complex audio clips based on a text description or even an image (Agostinelli et al., 2023; Q. Huang et al., 2022).

All of this could mean that, in the not-so-distant future, the creative industry will be transformed in a radical way. Technology will allow us to generate large amounts of personalized content with little effort. In the long term, it is likely that this will have a profound impact on the way we consume content, and it also could have substantial impact on the way we value creativity. Within this context, it becomes increasingly evident that these computational tools will become intertwined with our way of thinking creatively.

## 1.2 Concepts and machines

Concept formation is a central part of creative tasks such as problem solving, design, and scientific discovery. The overarching goal of this thesis is to investigate how computational methods affect different components of the creative process, such as making associations or combinations of concepts, performing abstractions, evaluating and selecting concepts (Hoorn, 2014; 2023). The automation of analytical reasoning and the advent of computers has made it possible to build machines that can help us come up with new ideas or refine existing ones. But the nature of human concepts and their origin is still not well understood, so these automated implementations only reflect our hypothesis about how concepts might work.

It follows that, in order to discuss concept formation for machines, it is necessary to have a mental model of how human concepts work, fail to work, combine with each other and are applied to objects, events, people or experiences. In cognitive psychology this phenomenon is known as categorization. The simplest and most traditional way of thinking about conceptual categorization (i.e. the process of assigning object instances to the appropriate category) is to use definitions. Scholars refer to this definitional framework as Classical Theory (CT). According to this view, conceptual categorization involves checking whether the necessary defined properties are present or not in the object instance in question. Albeit simple, this method can accurately model some aspects of how concepts work (Laurence & Margolis, 1999). CT also makes it possible to formalize and mechanize some parts of the creative process. Philosophers and researchers have highlighted the many limitations about this approach to concepts. Prototype Theory is an alternative theory attempting to deal with the problems of definition-based approaches (Medin, 1989; Murphy & Medin, 1985; Rosch, 1978; Rosch & Mervis, 1975; E. E. Smith & Medin, 1981). In PT, each concept is represented by a prototype, which is an idealized example of the concept. PT focuses on the properties that are *typical* for the category in question, as opposed to the properties that are *necessary*. This means that in PT, an instance need not have all the typical properties to be assigned to a category, it simply needs to be more similar to that category's prototype than any other (Wittgenstein, 1953). Compared to CT, PT is more attuned to the way human categorization works, but also has its shortcomings (Laurence & Margolis, 1999; Osherson & Smith, 1981). For example, it is not obvious what the prototype is if a category has many or lacks typical instances, such as New species or Objects that weigh more than one gram (Laurence & Margolis, 1999).

PT is well suited for explaining and mechanizing conceptual categorization because it allows for flexible representations of concepts that can accommodate the inherent *fuzziness* of some categories[1]. However, the notion of similarity upon which they rely on to assign instances to their categories is not so easy to formalize. Some implementations of these theories adopt geometrical models to measure similarity, such as euclidean or cosine distances, while others rely on presence and absence of features, such as Tversky's *contrast model* (Tversky, 1977) which is the most commonly used in psychology.

Another way of formalizing categorization is to use probabilistic models. This approach has been gaining popularity in recent years, as it has been shown to be more successful in dealing with the problems of definition-based approaches (Hüllermeier & Waegeman, 2021; Linardatos et al., 2020; Xu et al., 2021). Probabilistic models of categorization are essentially statistical models that estimate the probability that a given object instance belongs to a given category. The probability is calculated based on the observed properties

[1] Medin (1989) for example asks: are carpets part of the furniture?

and their relations. Statistical models typically require a lot of training data in order to achieve human-level accuracy and of course they will still make wrong predictions. Nevertheless, the non-deterministic nature of the probabilistic approach allows models to exhibit inherent creativity through misclassification and errors (Hoorn, 2023).

This thesis is concerned with how these different accounts of conceptual categorization affect the creative process. It investigates how different aspects of creativity are affected by a technology's embedded assumptions about what concepts and categories are.

## 1.3 Two approaches to creative machines

When we use of machines and other computational tools for creative purposes, the choice of technology has impact on the kinds of tasks required for the creative process. Imagine for a moment that you are an artist or designer working on a digital artwork around the theme *two circles*. You have the option of using a programming language or a traditional program to produce your design, or you might decide to use as text-to-image generator like StableDiffusion or Midjourney. How does this choice affect your creative process? On one hand, for the code or program to work, you might have to define every aspect of your design, the size of the canvas, the positioning of the circles, etc... On the other, you could simply type "two circles" as a prompt and the model will generate many images of circles, which may or may not contain exactly two circles (see Figure 1.2). The result and the process are both quite different.

Indeed, these two approaches have opposing theoretical origins, which will be discussed in the literature review. On one hand the rule-based approach is firmly grounded in the idea of intelligence as analytical thinking and computation (Turing, 1950). This approach is tightly connected with CT and has been the dominant paradigm for over five decades. On the other hand, the data-driven approach that has been gaining popularity in the last

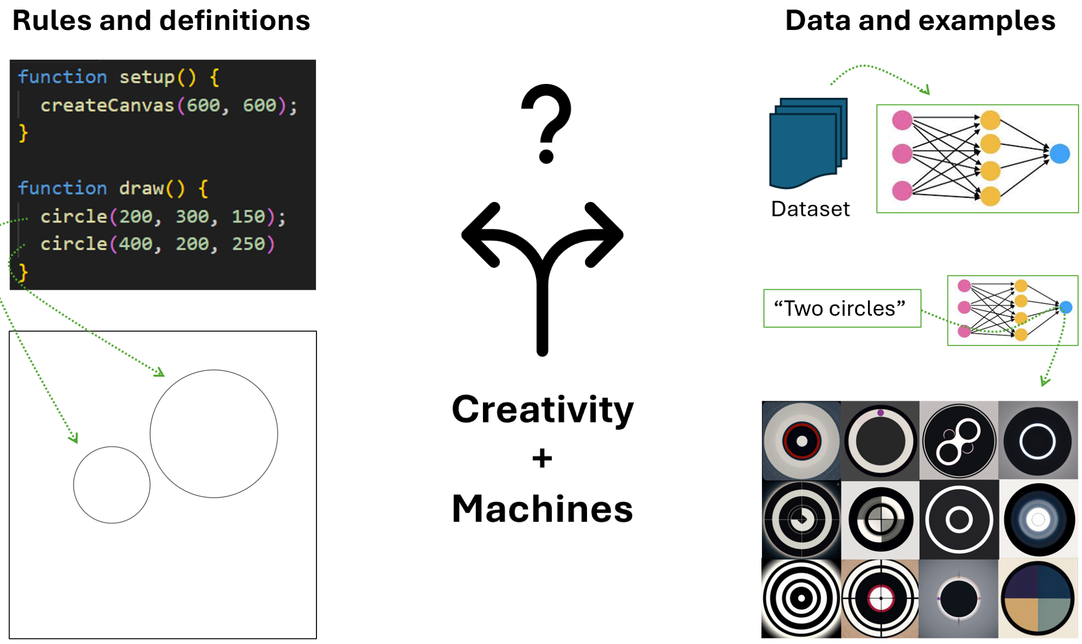


**Figure 1.2:** The diagram distinguishes between two different approaches to machine creativity.

two decades is instead grounded in PT, as it subscribes to a probabilistic representation of concepts.

This thesis is set out to investigate the differences between these two approaches and the impact that this choice has on the creative process. The objective is to provide a framework that can describe these two approaches and their assumptions about what concepts are.

## 1.4 Societal impact

In the last 15 years, the corpus of papers in the field of Artificial Intelligence (AI) and ML has been growing at an exponential rate (Krenn et al., 2022). A handful of new companies, such as OpenAI and Stability.ai, alongside few industry giants, such as Google, Meta, Microsoft and Nvidia, are competing fiercely to offer products and services that leverage ML.

As these technologies become largely available and more accessible, not much attention is directed towards gaining a better understanding of the long-term implications that these data-driven tools have on the way creativity is valued by authors and audiences.

Moreover, state-of-the-art computational methods for media generation also expose how the technological front has evolved much faster and far beyond our theoretical, ethical and legal frameworks. Acclaimed projects, such as GPT3 (Brown et al., 2020), StyleGAN 2 (Karras et al., 2019), Magenta (C.-Z. A. Huang et al., 2018; Roberts et al., 2018), build upon large datasets of text, images and music, often obtained without the consent of the people involved. So is the case of Stable Diffusion (Rombach et al., 2021), an open-source model trained on the 5 billion images in LAION-5B dataset (Schuhmann et al., 2022). Several artists have found that "the systems create art in their styles when their names are used as prompts, and that users have been creating works that are *indistinguishable* from theirs." (Brittain, 2023). This discovery led them to file lawsuits against the Stability.ai, the company that trained the model. This is a unprecedented situation, which poses theoretical and ethical questions as well as copyright headaches.

Deep Learning (DL) relies on probabilistic representations of concepts obtained from large datasets. This means that the generated output is optimized to match the bias in the training data. For this reason, DL and, more specifically, large pre-trained models that learn from data gathered on the internet, stand in a direct relationship with the cultural and social norms that influenced the training datasets.

The relationship between society and technology has been discussed at length by philosophers such as Heidegger (1977), Ihde (1990), Latour (1990), Flusser (2000) and Verbeek (2011). The impact that different forms of technology have on the creative process has also been addressed by the academic efforts of the Computational Creativity (CC) community. Scholars in this field, have thoroughly discussed whether non-human agents can be considered creative (Wiggins, 2006a; 2006b), thus focusing primarily on how the evaluation of artifacts produced by autonomous systems should be conducted and which criteria should be used (Agrawal, 2019; Boden, 2010; Colton, 2008; 2012; Jordanous, 2009; Lamb, 2018). In parallel, the Human Computer Interaction (HCI) literature stream addresses interactions with computers as a design problem. A few authors attempt to discuss how human perception is affected by the interaction with new technologies (Algarni, 2020; Hoorn, 2023; Ragot, 2020).

This thesis also discusses the impact of this data-driven trend on society at large. It attempts to identify critical factors of data-driven technologies and the effect they have on the creative process.

## 1.5 Thesis summary

Chapter 2 is dedicated to the literature review and discusses two main topics. The first part (Section 2.1) addresses various TOCs introducing main authors, supporting evidence and issues related to each theory (Laurence & Margolis, 1999). The second part (Section 2.2) presents a review of philosophy of technology introducing phenomenology and post-phenomenology (Heidegger, 1977; Ihde, 1993; Latour, 1990; 1993; Verbeek, 2011) and discusses the evolution of different computational approaches within the context of creative practices. The second part ends with a systematic literature review of computational creativity papers, highlighting the trend of data-driven tools and how it is affecting the research community.

Chapter 3 in this thesis reviews existing theoretical frameworks addressing creativity (Boden, 2003; Hoorn, 2014; Rhodes, 1961) and presents an extension of mediation theory developed by Ihde (1990, 1993). The chapter introduces the distinction between two approaches to computation and their implied TOCs, as discussed in Chapter 2. This theoretical foundation is then used in the studies and in the final discussion chapter. Chapter 3 also discusses matters related to Practitioner Research (PR) which was adopted for the first two studies in this doctoral thesis.

Three distinct studies in the creative fields of music, images, and text-to-image are covered in Chapters 4, 5, and 6, respectively. Chapter 4 presents a collaboration with a music composer in which we attempted to train a model to generate music in her style using ML (Dinculescu et al., 2019; C.-Z. A. Huang et al., 2018). The study in Chapter 5 examines the process of dataset curation as *reflective practice* (Schön, 1983) and was carried out in partnership with a fellow PhD student who is also a photographer. Chapter 6 presents a study involving 76 participants which investigates how users of a Stable Diffusion (SD) Discord bot use text-to-image technology in relation to their Tolerance for Ambiguity (TA) personality trait (Furnham & Marks, 2013; Herman et al., 2010; Norton, 1975; Zenasni et al., 2008) and user expectations.

Chapter 7 combines the insights collected from each study and discusses three main points which emerge from the results. First, it addresses the central role of dataset curation within the ACASIA framework (Hoorn, 2023), highlighting how custom models trained by the community might represent an novel form of creative product. It then addresses the central role of text-based model conditioning in relation to Ihde's *hermenentic intentionality* (Ihde, 1993), speculating on the role of language as interface for concepts. Finally it discusses how data-driven tools are blurring the boundaries of authorship and ownership between human and non-human and the implications of this ambiguity.

The last chapter in this thesis summarizes once again the purpose of the research, the gaps that have been addressed and the main findings. It then discusses limitations of the studies and suggests several directions for further research.

*No man ever steps in the same river twice, for it's not the same river and he's not the same man.*

Heraclitus

# 2

# Literature review

THE GOAL OF THIS LITERATURE REVIEW IS TO FRAME THE EXISTING DISCOURSE AROUND CATEGORIZATION AND CONCEPT FORMATION together with the literature stream discussing technology in relation to human cognition and creativity. In the first part, it reviews the existing hypotheses about what concepts are and how these concept ontologies model our creative process. In the second part, it addresses how different technologies embed different Theories of Concepts (TOCs) and how this affects the creative process. From this view point, it will be possible to formulate a methodology of inquiry to understand where the current trend towards a probabilistic interpretation of concepts may lead us to.

Because of its trans-disciplinary nature, this literature review spans several fields. The review of concept ontologies lies at the intersection between philosophy, psychology and cognitive science, while the discussion about the relationship between technology and creativity touches upon design, philosophy of technology and creativity literature.

## 2.1 Concepts

There is no consensus on a single theory of concepts as of today, in spite of the thousands of years of philosophical discussions about the nature of ideas that have been ongoing since the pre-Socratics. The two fundamental questions that a TOC must provide an answer to are: (1) how do concepts form in our mind (2) how they relate to one another. In this section several competing hypotheses are presented as discussed by Laurence and Margolis (1999):

- **The Classical Theory of concepts (CT)**. CT is based on definitions. It relies on presence/absence of given properties to establish reference to the world. CT has been criticized for its rigidity and for its inability to explain evidence emerging from psychological experiments.
- **Prototype Theory (PT)**. Rosch (1978) and E. E. Smith and Medin (1981) developed PT to explain the results of their experiments involving typicality effects. PT adopts probability and fuzzy logic to formalize the categorization process.
- **Theory-Theory (TT)**. Murphy and Medin (1985) and Carey (1991, 2009) developed TT based on the intuition that concepts rely on theories about the world that we formulate, assimilating concept formation to the scientific method.
- **Neo-classical Theory (NT)**. Jackendoff (1989) proposed that concepts have partial definitions which are necessary to identify their extension. NT is however primarily

concerned with lexical concepts.

- **Conceptual Atomism (CA)**. Fodor (2008) proposes that concepts have no structure and highlights compositionality as a necessary component of concepts.

Each theory is discussed in a separate subsection, highlighting along the way the potential fallacies and weaknesses. Some of these issues are shared across more than one theory and serve as guide for the theoretical framework. The summary descriptions for each theory are taken verbatim from Laurence and Margolis (1999).

### 2.1.1 Classical Theory of Concepts

> Most concepts are structured mental representations that encode a set of necessary and sufficient conditions for their application, if possible, in sensory or perceptual terms.

Classical Theory (CT) holds that most concepts have definitional structure. According to this view, the concept BIRD might be composed of a set of properties an object must have in order to count as a bird, such as HAS WINGS, CAN FLY, LAYS EGGS and so on. This implies that concepts must have a hierarchical structure, in that concepts are effectively composed by other structurally simpler representations. Following this hierarchical structure, new concepts can be formed using existing concepts combined in a new definition.

Albeit its simplistic approach and numerous problems, CT has been around since antiquity and stood undisputed until the 1950s because of its explanatory power. Here is a list of different aspects of concepts and their corresponding explanation according to CT Laurence and Margolis (1999):

- **Concept Acquisition**. Learning a concept is just a matter of learning the simpler individual components that form its definition.

- **Categorization**. The process of applying a concept to a particular instance is as trivial as checking the properties of that object.
- **Epistemic Justification**. We can justify a belief we have about the world by determining whether its defining properties are satisfied.
- **Analyticity and Analytical Inferences**. The definitional structure of concepts guarantees that certain statements can be inferred from others without empirical evidence.
- **Reference Determination**. Concepts are semantically evaluable by their definition, which means that every statement can either be true or false.

In spite of the obvious merits of CT, numerous critiques were raised by philosophers and psychology scholars over the years. In the forthcoming paragraphs are highlighted six major ones.

### Plato's problem

In Meno, Plato provides one of the first critiques to CT. In this dialogue, Meno begins by asking Socrates how "virtue" (in Greek ἀρετή) is acquired and Socrates replies that he does not know how to define it and asks Meno to help. Meno at first suggests that virtuous actions depend on the person's age, gender, role in society and so on, but Socrates is looking for some quality that is common to all, something more general. Meno then suggest that there are some common quality to all virtuous men such as justice and temperance, but Socrates is not satisfied because Meno did not provide a full list and he does not know what is common among these qualities. Meno is understandably confused and frustrated. The exchange continues:

> **Meno:** But in what way will you look for it, Socrates, this thing that you don't know at all what it is? What sort of thing, among the things you don't know, will you propose to look for? Or even if you should meet right up against it, how will you know that this is the thing you didn't know?
>
> **Socrates:** Do you see what a contentious debater's argument you're bringing up—that it seems impossible for a person to seek either what he knows or what he doesn't know? He couldn't seek what he knows, because he knows it, and there's no need for him to seek it. Nor could he seek what he doesn't know, because he doesn't know what to look for.
>
> (Plato, trans. 1998, Meno, 80d-e)

Plato is suggesting here through Socrates' words that if we subscribe to a definitional structure of concepts, we end up in a paradox as it seems that for most concepts a definition is very hard to find. It also becomes very problematic to justify how we recognize something we have no definition for. Very few examples of defined concepts exist, mostly mathematical, such as Prime Number. Other rather fundamental concepts like Event, Object or Cause do not have clear definitions, yet we use them all the time. This is a fundamental problem of CT as there seems to be no need for a definition to exist in order to apply concepts. For this reason, ascribing definitional structure to concepts creates the epistemological paradox highlighted by Plato.

### The problem of analyticity

The idea that some statements are true by virtue of meaning alone has fascinated many, but we owe a great deal to Kant (1998) for discussing at length the distinction between analytic and synthetic propositions in his "Critique of Pure Reason". Analytic judgments are those deemed to be true only by virtue of definition, a typical example would be: "all bachelors are unmarried". Synthetic judgments, on the other hand, are statements that must be verified through experience, for example: "all bachelors are wealthy". The essential difference, according to Kant, is that the truth value of analytic statements does not depend on the state of the world, while the opposite is true for synthetic propositions.

This line that Kant drew has been subject of debate for centuries. On one hand, the distinction has been the foundation of the empiricist paradigm, which, in an attempt to get rid of metaphysics altogether, radically stated that all knowledge comes from experience. Problematic knowledge for this agenda such as mathematics and logic could be placed in the box of analytic *a priori*.

On the other hand, the empiricist view has been criticized for being too reductionist, as it implies that knowledge is limited to what can be determined through the senses and that all knowledge must come from experience. This view has been challenged by other more metaphysical approaches such as idealism, which hold that there is knowledge that cannot be acquired by experience, and that some knowledge can be accessed through pure thought (*a priori*).

Kant's cleavage between analytic *a priori* and synthetic *a posteriori* is still relevant today and is the basis of many debates in philosophy, which continue to this day, with no clear consensus on the correct approach. Among the critics, Quine questioned the notion of analyticity itself, arguing that self-contradiction and analyticity are the "two sides of a single dubious coin" (1951). Quine suggested that such distinction "is an unempirical dogma of empiricists, a metaphysical article of faith" (1951), primarily by claiming that there is no such thing as analytic statements. Firstly because there is no definition of similarity or analyticity that is *analytic* and second, because individual statements are never confirmed in isolation.

### The problem of typicality effects

The typicality effect is the finding that people are quicker to make category judgments about typical members of a category than they are to make such judgments about atypical members. For example, they are more quickly able to judge that a dog is a mammal than they are able to judge that a whale is a mammal. The study conducted by Rosch and Mervis (1975), highlights one of the most influential arguments against CT. According to CT all instances of a concept should be equally good examples as long as they match the definition, but prototypical judgments provide evidence against this claim as people seem to rank certain members of a category as more representative than others. The studies on typicality effect demanded for a revision of CT to match the empirical data, which pushed the development of alternative theories of concepts able to accommodate for typicality effects, as well as some of the other fundamental issues of CT discussed so far.

### The problem of psychological reality

The advent of experimental psychology led to many of the critiques to CT. One particular implication about the hierarchical structure has been shown to be incompatible with the definitional hypothesis. If concepts have complex structure, then one would expect the definitional complexity to affect the psychological complexity, but the evidence collected by Foss (1969) demonstrates that "lexical concepts show no effects of definitional structure in psychological experiments" (Laurence & Margolis, 1999, p. 27). There seems to be no significant difference in cognitive effort when applying concepts that are more complex than others, such as in the case of BELIEVE and CONVINCE (which is defined as the combination of CAUSE TO and BELIEVE). This line of research suggests that all concepts are equally *expensive* to apply, a finding that does not support the hypothesis of a hierarchical structure of concepts, a foundational aspect CT.

#### The problem of ignorance and error

It is possible to have a concept in spite of massive ignorance and/or error, so concept possession cannot be a matter of knowing the definition (Kripke, 1980; Putnam, 1970; 1975). For example, the concept associated with the disease SMALLPOX may have been erroneously defined in the past as divine punishment, whereas now we understand it as something rather different. This does not imply the concept of SMALLPOX refers to a different disease now, yet there is substantial difference in its definition. It seems that it is indeed "possible to possess a concept without representing necessary conditions for its application" (Laurence & Margolis, 1999, p. 21).

#### The problem of conceptual fuzziness

Medin (1989) suggests that "concepts have determinate extensions and categorization judgments should also yield determinate answers, yet concepts and categorization both admit of a certain amount of indeterminacy" (Laurence & Margolis, 1999, p. 27). There are abundant examples of this phenomenon. Medin asks, are carpets part of the furniture? This problem can be reduced to the first issue, that is most concepts lack of a definition.

### 2.1.2 Prototype Theory

> Most concepts are structured mental representation that encode the properties that objects in their extension tend to possess

In Prototype Theory (PT), concepts have no definitions. According to PT, a concept "should encode the distribution of statistically prominent properties in a category" (Laurence & Margolis, 1999, p.29), following the idea of family resemblance by Wittgenstein (1953). PT is developed as an alternative to fix the shortcomings of CT exposed by Psychology in the 1970's, primarily the evidence of typicality effect discussed earlier in this chapter. For Wittgenstein (1953), as for Rosch and Mervis (1975), "a concept like GAME isn't governed by definitions but rather by a possibly open-ended set of properties which may occur in different arrangements" (Laurence & Margolis, 1999, p. 29). The whole set of properties that overlap in games forms a similarity space: "what makes something a GAME is that it falls within the boundaries of this space" (Laurence & Margolis, 1999, p. 29).

PT is also able to effectively describe concept acquisition: acquiring a concept is as simple as assembling its features. Effectively, this mechanism embodies a statistical procedure, rather than a logical one as described by CT. Following from the categorization model, this procedure consists of identifying a new region in the similarity space. Because similarity is such a central aspect in the theory, PT advocates developed a number of psychological measures for it. Tversky (1977) proposed a similarity measure known as *Contrast Principle* which is widely used in psychology, which is based on the presence and absence of features. Another popular method is to define properties as geometrical dimensions and use measures of distance within this space to describe similarity between two items (see Section 2.2.6). Furthermore, PT is well suited to sidestep conceptual fuzziness as it allows to formally describe ambiguity using the mathematical construct of fuzzy sets. There are, however, significant shortcomings also for PT.

#### THE PROBLEM OF PROTOTYPICAL PRIMES

Experimental psychology has shown that typicality effects occur even in well-defined concepts, i.e. concepts that people can immediately provide a definition for, such as PRIME NUMBER or FEMALE (Armstrong et al., 1983). Interestingly, this study would be considered controversial today given that the binary notion of gender has been obsoleted. However, the argument for mathematically defined concepts still stands, so the results of this study suggest that typicality effects cannot be counted as strong evidence supporting PT.

### The problem of ignorance and error

Ignorance and error is as much a problem for PT as it is for CT. Indeed, "the problem is considerably worse for PT, since concepts with prototype structure fail to cover highly atypical instances and incorrectly include non-instances" (Laurence & Margolis, 1999, p. 44). There is essentially no way to account for misapplied concepts in PT. The Swan concept would encode that the prototypical instance is white based on the most frequent occurrences, but black swans are still swans. Penguins lack one of the most typical properties of birds, yet they are considered birds.

### The missing prototypes problem

Just as many concepts lack definitions, many concepts also lack prototypes. One can easily construct such concepts as Objects longer than 1 centimeter or 41st century technology, that are either too broad or simply have an empty extension and therefore cannot have prototypical instances. This aspect of concept combination making is where CT seems to have an edge over PT. Crisp concepts with formal rules of interaction among them can generate new concepts that may or may not have an extension.

### The problem of compositionality

PT does not have an adequate account for compositionality, since the properties of complex concepts are not generally a function of the prototype of their constituent concepts. For example, the prototype for pet fish would have properties such as small, colorful and LIVING IN BOWLS OR TANKS which hardly relate to the qualities of neither a prototypical PET, a dog or a cat would be FURRY and AFFECTIONATE, nor a prototypical FISH, a trout or a sea bass would be GRAY and LIVING IN THE WILD (Osherson & Smith, 1981).

The role of compositionality is central to language and meaning. This idea is already present in Frege's work, "Über Sinn und Bedeutung" (On Sense and Reference), where he argued that the meaning of complex expressions is determined by the meanings of its parts and the rules of its composition (Frege, 1892). It is echoed by Fodor in "The Language of Thought", where Fodor argues against definitions (critiquing CT), exposing them as too vague and general to be of any use (Fodor, 1975). Fodor pointed out that instead of relying on definitions, we should focus on understanding the context in which words are used, and more specifically in understanding the compositionality of language (Fodor, 2008). Without a proper account for how compositionality is achieved in PT, the framework cannot explain how concepts relate to one another in order to form new meanings.

### 2.1.3 Theory-Theory

> Concepts are representations whose structure consists in their relations to other concepts as specified by a mental theory

According to Theory-Theory (TT) cognition is similar to theory construction and scientific reasoning. The main appeal of TT is that it provides an explanation for "conceptual change along the lines of theory change in science" (Laurence & Margolis, 1999, p. 45), where a theory is replaced with a better one to explain the same data. TT provides an account of how conceptual change is driven by a combination of rational and empirical processes, allowing for the development of more sophisticated concepts that are better suited to the environment. TT also provides a unified explanation of the development of core cognitive abilities such as language and counting (Carey, 1991; 2009).

TT suggests that cognition is driven by a process of hypothesizing, experimentation, and revision. As new information is gathered and existing concepts are modified, more sophisticated concepts emerge. This process is driven by the desire to make sense of the world and to predict future events. As a result, TT can account for both conceptual development and the transfer of knowledge across domains (Murphy & Medin, 1985).

TT also provides an account of how knowledge and conceptual understanding are acquired and retained. Knowledge is acquired by actively engaging with the environment, gathering information, and testing hypotheses. This process allows for the development of more sophisticated concepts as knowledge is accumulated and understood. In addition, TT suggests that knowledge is retained through a process of encoding and recall. This means that knowledge is stored in a form that can be retrieved when needed.

In spite of the appeal that this framework might have there are several unanswered questions summarized below.

### The problem of ignorance and error

It is possible to have a concept in spite of its being tied up with a deficient or erroneous mental theory, but according to TT concepts do not inform us about the properties of the objects in their extension. The SMALLPOX argument still holds against TT as different mental theories do not imply a different disease.

### The problem of stability

"The content of a concept cannot remain invariant across changes in its mental theory" (Laurence & Margolis, 1999, p. 51). Imagine two people have slightly different theories about the concept ANIMAL, for example person A believes that all animals are physical entities and person B believes instead that some species also have a non-physical soul. These theories are similar, but they are not the same, so TT fails to explain how "people with different beliefs systems have concepts with the same or similar content." (Laurence & Margolis, 1999, p. 51)

#### The "mysteries of science" problem

"The mechanisms that are responsible for the emergence of new scientific theories and for the shift from one theory to another are poorly understood" (Laurence & Margolis, 1999, p. 53). There is no account for how scientific theories transition other than the "the mysterious logic of discovery" (Gopnik & Meltzoff, 1997, p. 40) so claiming the similarity to the scientific method does not provide any additional insight about how concepts shift.

### 2.1.4 Neoclassical Theory

> Concepts have partial definitions in that their structure encodes a set of necessary conditions that must be satisfied by things in their extension.

According to Neo-classical Theory (NT), most concepts are structured mental representations "that encode partial definitions, i.e., necessary conditions for their application" (Laurence & Margolis, 1999, p. 55). NT claims that certain aspects of linguistic phenomena can be explained by conceptual structure. The fact that causal constructs have a clear distributional pattern serves as Jackendoff's starting point:

1. $x$ killed $y$ $\rightarrow$ $y$ died
2. $x$ lifted $y$ $\rightarrow$ $y$ rose
3. $x$ gave $z$ to $y$ $\rightarrow$ $y$ received $z$
4. $x$ persuaded $y$ that $\mathcal{P}$ $\rightarrow$ $y$ came to believe that $\mathcal{P}$

All of these inferences might be viewed as being unrelated to one another, "but they are strikingly similar, and this suggests that they have a common explanation" (Laurence & Margolis, 1999, p. 55). According to Jackendoff, the definition of a causative implicates the occurrence of a specific event. On the basis of this supposition, a single rule that applies to all of these circumstances may be introduced to explain this pattern:

- $\mathcal{X}$ cause $\mathcal{E}$ to occur $\rightarrow$ $\mathcal{E}$ occur

For example, (1) could be analyzed as "$x$ cause [$y$ died] to occur", and similarly for (2),(3) and (4). Causatives are only one of the aspects in which NT finds supports, others include polysemy, syntactic alterations, and lexical acquisition.

Jackendoff tackles the issue of compositionality, which he refers to as the "creativity of language", assessing that concepts cannot just be the encoding of all the encountered instances. Therefore he deduces that there must be a "potential degree of indeterminacy either in the lexical concept itself, or in the procedure for comparing it with mental representations of novel objects, or in both" (Jackendoff, 1989). He suggests that the words and ideas we hold in our vocabulary are built from a natural set of potential concepts, influenced by both language and non-language experiences.

NT does not intend to explain everything about concepts and is more focused on the semantic and lexical patterns that emerge in language. For this reason, some of the issues found in previous theories of concepts are not addressed and remain unresolved.

### The problem of completing definitions

Definitions are still problematic for NT. In fact, if incomplete definitions are expanded into complete definitions then NT has all the problems that are associated with CT. If, instead, they are left incomplete, then NT has no account of how concepts are applied to their instances. NT is not concerned with how lexical concepts are applied to the entities they refer to, so this aspect of concepts is simply not explained.

### The problem of ignorance and error

The introduction of partial definitions does not help with the problem of ignorance and error, much like in PT. Because NT effectively does not provide a theory for reference determination, it is still unclear how we can have a concept in spite of having erroneous information about its definitions (Small pox argument still applies). Jackendoff proposes that for lexical concepts that refer to physical objects incorporate a 3D model of them. However, it is also possible that an animal that closely resembles a duck is not actually one, and vice versa, that a duck may not appear to be one for whatever reason.

### The regress problem for semantic fields

Neoclassical structure cannot explain how a word retains aspects of its meaning across different *semantic fields*. Either its conceptual constituents must themselves have neoclassical structure, and so on, or else no structure is needed at all. To understand this methodological objection pushed vigorously by Fodor (1998, p . 50), here is an example:

- Harry kept the bird in the cage
- Sam kept the crowd happy

In these two sentences, Jackendoff would argue that on one hand there is the intuition that the same word is used (Keep), on the other that the sense of Keep is different in the two cases. The definition of Keep as "causation of a state that endures over time" would account for the feeling that Keep is univocal, while the differences are explained by the different *semantic fields*, each of which has its own particular inferential patterns (Jackendoff, 1995). Fodor objects to this explanation by questioning the assumption that the constituent of the definition (i.e. CAUSE, STATE, TIME, ENDURE) must also themselves be univocal. He suggests that Jackendoff logic runs into a paradox when attempting to assess if CAUSE is polysemic. If it is, then the definition of KEEP is no longer univocal and the argument in favor of definitions is lost. If it is not, then what explains the univocality of CAUSE across *semantic fields*? Fodor suggests that there are only dead ends from here. Either a new univocal concept X is needed for the definition of CAUSE, which would lead to an infinite regress (what guarantees the univocality of X?), or just accept that CAUSE is univical because it always means *cause*, in which case then the same could be said for KEEP and no theory is needed at all.

### 2.1.5 CONCEPTUAL ATOMISM

> Lexical concepts have no hierarchy or structure, thus they cannot be broken down

Conceptual Atomism (CA) is different from previous theories because, rather than arguing a particular structure, it questions the fundamental assumption of conceptual structure itself. Because CA assumes no structure, it is able to sidestep most problems discussed for other theories. The theory mainly provides an account for how the concept references are determined, that is Asymmetric Dependance Theory. There is an asymmetric dependency between laws such as "$Y_1$ CAUSES $\hat{X}$", "$Y_2$ CAUSES $\hat{X}$",' etc., and the law "X CAUSES $\hat{X}$". This is because the latter law does not depend on any of the $Y_1$, $Y_2$, ..., $Y_n$ laws in the same way. The intuition here is that $\hat{X}$ will only be caused by the question "What kind of animal is called Fido?" because dogs (X) cause instances of $\hat{X}$. Instances of foxes causing instances of $\hat{X}$ are only due to them being mistaken for dogs and dogs causing instances of $\hat{X}$.

Two important implications of this theory are the rejection of mental images and the rejection of context-sensitive meaning. First, while other theories might argue that a particular concept is represented by a mental image, CA rejects this idea and instead suggests that the concept is merely a reference to an external object or situation. Second, while other theories might allow for context-sensitive meaning (i.e., different meanings of words depending on the context), CA rejects this idea as well and instead suggests that all meanings of atoms are fixed and will not vary based on context.

Overall, CA offers an interesting alternative to traditional theories of conceptual structure, since it does away with many of their complexities and assumptions about mental images and context-sensitive meanings. While it obviously has its critics (who argue that it fails to account for more complex concepts), it offers an interesting look at how language might work.

There are however several issues with CA, discussed below.

### The problem of radical nativism

According to CA, "most lexical concepts turn out to be innate, including such unlikely candidates as Carburetor and Xylophone" (Laurence & Margolis, 1999, p. 75). According to Fodor, there is only one way that cognitive science can explain how an idea is learned, and that is by speculating on a mechanism by which a brand-new, complex concept is constructed from its constituent parts. For example one can learn the concept of Mother by combining the concepts Female and Parent. This process assumes that one already possesses the concepts Female and Parent, so when we ask about how these concepts were acquired, the answer might be that they are themselves composed of simpler concepts, but eventually this has to stop. So if there are no other explanations about how concepts are learned, one must conclude that there are some primitive concepts that are *innate*.

### The problem of explanatory impotence

CA cannot explain psychological phenomena such as categorization. If concepts lack structure, atomists have no way to make sense of the empirical evidence about typicality effects, documented in psychology by (Rosch, 1978; Rosch & Mervis, 1975). Although Fodor acknowledges the significance of prototypes, he disputes their role in the semantic structure of concepts.

### The problem of analytic data

CA lacks for "an adequate explanation of why people have intuitions of analyticity" (Laurence & Margolis, 1999, p. 75). Rey (1993) has put together a case against conceptual atomism based on the fact that NT's partial definitions provide an explanation about analytic data. He asserts that, regardless of whether there are any analytical truths, individuals undoubtedly have intuitions about what is analytical. According to Rey, these intuitions emerge from the relationships that are formed among concepts. We therefore have an argument against CA and an argument in favor of NT: as Rey points out, there is no plausible atomistic alternative.

### The problem of compositionality

In a sense CA has no issue with combining concepts, but primarily because the theory is centered around lexical concepts. If we extend CA to a comprehensive theory of concepts, then some familiar issues still arise. In this context, "atomistic theories of concepts have as much difficulty with conceptual combination as PT" (Laurence & Margolis, 1999, p. 75). For example consider the concept Grandfathers whose granddaughters are friends with politicians. It is unlikely that this concept stands in a lawful dependency relation with the property of being "a grandfather whose granddaughters are friends with politicians". In other words, just like in PT and the PET FISH example, the asymmetric dependence relations of complex concepts are not a function of the asymmetric dependence relations of their constituents.

### 2.1.6 Summary

The theories in this section, summarized in Table 2.1, provide a foundation for understanding the relationship between technology and creativity, which is explored further in the next section. The assumption is that when these concept theories are mechanized, their advantages and problems impact the creative interaction. Thus, examining how different concept ontologies in various implementations affect the creative process becomes valuable.

| Theory | Description | Problems |
|---|---|---|
| Classical Theory | "Most concepts are structured mental representations that encode a set of necessary and sufficient conditions for their application, if possible, in sensory or perceptual terms." (Laurence & Margolis, 1999, p. 9) | 1. Plato's problem<br>2. The problem of analyticity<br>3. The problem of typicality effect<br>4. The problem of psychological reality<br>5. The problem of ignorance and error<br>6. The problem of conceptual fuzziness |
| Prototype Theory | "Most concepts are structured mental representation that encode the properties that objects in their extension tend to possess" (Laurence & Margolis, 1999, p. 28) | 1. The problem of prototypical primes<br>2. The problem of ignorance and error<br>3. The missing prototypes problem<br>4. The problem of compositionality |

| | | |
|---|---|---|
| Theory-Theory | "Concepts are representations whose structure consists in their relations to other concepts as specified by a mental theory" (Laurence & Margolis, 1999, p. 47) | 1. The problem of ignorance and error<br>2. The problem of stability<br>3. "Mysteries of Science" Problem |
| Neoclassical Theory | "Concepts have partial definitions in that their structure encodes a set of necessary conditions that must be satisfied by things in their extension." (Laurence & Margolis, 1999, p. 55) | 1. The problem of completing definitions<br>2. The problem of ignorance and error<br>3. The regress problem for semantic fields |
| Conceptual Atomism | "Lexical concepts are primitive, they have no structure" (Laurence & Margolis, 1999, p. 63) | 1. The problem of radical nativism<br>2. The problem of explanatory impotence<br>3. The problem of analytic data<br>4. The problem of compositionality |

**Table 2.1:** This table summarizes the theories of concepts discussed in this section.

## 2.2 Technology and creativity

This section reviews some key literature streams discussing the relationship between humans, technology and creativity. The first two subsections address the perspective of philosophy of technology, which will serve as a starting point for the inquiry into the history of computation and creativity. In the following subsections an overview of a series of milestones that link computing technology and creativity are presented in parallel. The purpose of this arrangement is to highlight the trans-disciplinary spillover of ideas that interconnects the theories of concepts discussed in the previous section with research in artificial intelligence as well as art and design. The objective of this review is to explore the possibility of framing computational creativity and the current trend of data-driven AI within the post-phenomenological view of technology, which constitutes the starting point of this thesis' methodology discussed in Chapter 3.

### 2.2.1 Philosophy of technology

Philosophy of technology is an interdisciplinary field that combines insights from philosophy, sociology, history, and cultural studies to investigate the complex relationship between humans and technology (Ihde, 1993). It raises questions about the nature, purpose, and impact of technology on human existence, ethics, and values. By examining technology's historical development, the philosophy of technology seeks to understand how technology has transformed human lives and societies. Three main themes can be found in literature:

- **Technicity:** The concept of technicity refers to the idea that technology is not merely a tool but a condition of human existence (Feenberg, 1991). Technicity highlights the inseparable relationship between humans and technology, emphasizing that technology is not just an external object but an integral part of human experience.
- **Technological Determinism:** Technological determinism is the belief that technology drives social change, often leading to unintended consequences (M. R. Smith & Marx, 1994). This perspective suggests that technological innovations have their own logic and inevitability, which can lead to both positive and negative outcomes.
- **Technological Mediation:** Technological mediation refers to the role of technology as a mediator between humans and the world (Verbeek, 2005). This concept emphasizes that technology does not simply represent an external reality but shapes human perceptions and experiences.

Technology is a relatively young topic in Philosophy, yet most authors start their inquiry by looking at how Greek philosophers addressed it in relation to nature. For example, Aristotle deemed technē to be deeply interrelated with the notion of physis. In Physics, he states: "action for an end is present in things which come to be and are by nature" (Atwill, 2009) and therefore technē is the perfect example of action for a purpose. According to Aristotle, what sets physis and technē apart is that physis is itself its own efficient cause, whereas technē requires an external cause to be set in motion. In other words, physis is self-realizing towards its final cause, just as much as technē is, and the only difference is that physis is set in motion by itself, whereas technē requires an external driving force. According to Aristotle, technē is not understood as imitation of physis "form", but its action: it is creating for a purpose:

> Further, where a series has a completion, all the preceding steps are for the sake of that. Now surely as in intelligent action, so in nature; and as in nature, so it is in each action, if nothing interferes. Now intelligent action is for the sake of an end; therefore the nature of things also is so. Thus if a house, e.g. had been a thing made by nature, it would have been made in the same way as it is now by art; and if things made by nature were made also by art, they would come to be in the same way as by nature. Each step then in the series is for the sake of the next; and generally art partly completes what nature cannot bring to a finish, and partly imitates her. If, therefore, artificial products are for the sake of an end, so clearly also are natural products.
>
> (Aristotle, trans. 1983, Physics, Book II, Part 8)

It is interesting to notice how in translation, technē is referred to as art. In fact the semantic area covered by the word technē in ancient Greek is overlapping substantially with the idea of craftsmanship and fine arts, which is arguably far away from the notion of technology we have today. To understand this semantic gap, when must address the fundamental change in the relationship between man and nature introduced by the Copernican Revolution.

By adopting the scientific method, humans transitioned from being mere observers of an unchangeable world, as the ancient Greeks were, to actively probing nature for answers through technology (e.g., the telescope). This transformation drastically altered the way humans perceived their surroundings. The ancient Greeks deemed themselves vastly inferior to nature, which resulted in a humble and submissive attitude towards the natural world (Galimberti, 1989). This modest perspective was also mirrored in Greek mythology, where the gods rarely concerned themselves with human affairs, and when they did intervene, it was driven by their own interests rather than concern for mortals.

Centuries later, the Judaic-Christian tradition altered this dynamic, portraying nature as God's gift to humankind so that they could prosper and fulfill God's plan(Galimberti, 1989). This anthropocentric view of the world encouraged a different attitude towards nature, positioning it in service of humans rather than something to be feared. Nevertheless, the divine mysteries restrained human knowledge, as God's plan might not be fully comprehended by humans who simply had to accept it (Galimberti, 1989).

Galileo and the Copernican revolution challenged this worldview, demonstrating that, with technology and rational thought, mankind could investigate the mysteries of nature without the mediation of gods. The Age of Enlightenment placed human rationality at the center, while nature was no longer a humbling source of awe and wonder but instead viewed as something to be understood and controlled (Galimberti, 1989; Ihde, 1990). The role of technology in this dynamic is crucial because it is through technology that humans exert this control.

These ideas are captured in Heidegger's thought as pointed out by Lovitt: "We ordinarily understand modern technology as having arisen subsequently to science and as subordinate to it. We consider it to be a phenomenon brought about through scientific advance. Heidegger points out that, on the contrary, modern science and machine technology are mutually dependent upon one another. More importantly, technology, in its essence, precedes and is more fundamental than science." (Lovitt, 1977, p. xxviii)

### 2.2.2 Phenomenology and post-phenomenology

Modern philosophical views of technology cannot avoid discussing the thought of Heidegger. Martin Heidegger was a German philosopher and a seminal thinker in the continental tradition of philosophy. He is best known for his contributions to phenomenology, hermeneutics, and existentialism. Heidegger's work has had a profound influence on 20th-century philosophy, particularly on the fields of existentialism, deconstruction, and post-modernism. He is also infamously known for being involved with Nazism, which makes him a controversial author to discuss. However, this thesis is mainly concerned with his critiques of technology and modernity, rather than his political views. If the reader has any issue with this philosopher, they are welcome to reach out to me with their observations.

Heidegger's view of technology is complex and multifaceted. In "The Question Concerning Technology" he deems the Greek definition (technē) unfit to describe modern technology: "In opposition to this definition of the essential domain of technology, one can object that it indeed holds for Greek thought and that at best it might apply to the techniques of the handcraftsman, but that it simply does not fit modern machine-powered technology" (Heidegger, 1977, p. 13). He then points out a mutual relationship:

> It is said that modern technology is something incomparably different from all earlier technologies because it is based on modern physics as an exact science. Meanwhile we have come to understand more clearly that the reverse holds true as well: modern physics, as experimental, is dependent upon technical apparatus and upon progress in the building of apparatus. (Heidegger, 1977, p. 14)

He suggested that technology is as a way of understanding the world that shapes the human experience. The *modus operandi* of scientific progress habituates us to look at the world in terms of how it can be used and transformed, according to Heidegger, it is never neutral (Heidegger, 1977). For example:

> [t]he forester who, in the wood, measures the felled timber and to all appearances walks the same forest path in the same way as did his grandfather is today commanded by profit-making in the lumber industry, whether he knows it or not. He is made subordinate to the orderability of cellulose, which for its part is challenged forth by the need for paper, which is then delivered to newspapers and illustrated magazines. The latter, in their turn, set public opinion to swallowing what is printed, so that a set configuration of opinion becomes available on demand. (Heidegger, 1977, p. 18)

He also points out how modern technology can change the essence of nature:

> In the context of the interlocking processes pertaining to the orderly disposition of electrical energy, even the Rhine itself appears as something at our command. The hydroelectric plant is not built into the Rhine River as was the old wooden bridge that joined bank with bank for hundreds of years. Rather the river is dammed up into the power plant. What the river is now, namely, a water power supplier, derives from out of the essence of the power station. (Heidegger, 1977, p. 16)

Heidegger's view of technology is a rather pessimistic one. Several philosophers have commented and continued the work of Heidegger attempting to reconcile the human and non-human in different forms. Among them, Don Ihde, an American philosopher, revisits Heidegger's thought on technology in his book "Technology and the Lifeworld" (1990) offering an new perspective on the control we have on technology. According to Ihde,

> [t]he reason technology cannot be *controlled* is because the question is wrongly framed. It either assumes that technologies are *merely* instrumental and thus implicitly neutral, or it assumes that technologies are fully determinative and thus uncontrollable. Both extremities are involved in the current debates, but

> both miss the point of the human-technology and the culture-technology relativities that would reconstitute the debate (Ihde, 1990, p. 140).

In other words, Ihde's reframes the question "Can technology controlled?" into "Can *cultures* can be controlled?" (Ihde, 1990, p. 140).

Ihde points to the interdependence that exists between culture and technology observing that "[t]echnologies, by providing a framework for action, do form intentionalities and inclinations within which use patterns take dominant shape" (Ihde, 1990, p. 141). As an example he discusses how different writing instruments affect the style of the composition. An ink pen, a typewriter, and the word processor each present varying speed of writing and ease of editing, which in turn affect the writing style (Ihde, 1990, p. 142). In the case of the word processor, Ihde points out that "[p]recisely because the editing process is made easy, composition now provides a focal temptation. The ease of rewriting becomes a way to see the whole project as more malleable and thus unfixed" (Ihde, 1990, p. 142). A similar point was made by Heim (1999) as he argued in his book "Electric Language" that writing with the aid of computers implies a completely different notion of the writing task itself. Along the lines of this post-phenomenological definition of technology, scholars such as Bruno Latour and Peter-Paul Verbeek bring their attention to the mediating role of technology.

Bruno Latour is a French philosopher, anthropologist and sociologist. He is best known for his work in the field of science and technology studies. Latour argues that technology is "society made durable" (Latour, 1990) in the sense that a purely *social* world can never exist. The assemblage of a heterogeneous network of humans and non-humans is what produces stability. The example he brings up is that of a door. If a door is removed, a significant amount of work would be required by the human to fulfill the same purpose. A new hole would need to be made and bricked back up to go indoors. With the door, one is able to walk in through the combined efforts of both the human and the non-human. To go through, the door must present itself in a way that it can be opened *AND* the human must interact with it in a specific way to open it. According to Latour, the symmetry of this encounter is what creates stability in society.

Peter-Paul Verbeek is a Dutch philosopher who also has written extensively on the role of technology in society, following Ihde's lead. In his book "Moralizing Technology: Understanding and Designing the Morality of Things", Verbeek proposes an ethical framework for designing technologies which takes into account both human values and technological capabilities (Verbeek, 2011). He argues that technologies should not only serve utilitarian purposes but also take into account moral considerations such as justice, fairness, autonomy and responsibility. By doing so, he believes we can create more meaningful relationships between humans and their environment through technology.

Both philosophers revisit the idea already present in Heidegger and Ihde, that technology is not just a means to an end, but it also shapes the way we live in society, thereby defining our reality. According to this post-phenomenological view of technology, there is no separation between the human and the non-human, rather a mediation. Under these assumptions, what can be said about computing technology? In what mediating relationship do humans and computers stand in the context of creative practices? In what way does the mediation afforded by a brush differ from that afforded by a computer? The next subsections address these question with particular focus on the impact that different approaches to computation affect the creative process.

### 2.2.3 Good old-fashioned AI

The debate about whether machines can produce anything new at all dates back to Charles Babbage's invention. According to a famous passage from Ada Lovelace's notes: "[t]he Analytical Engine has no pretensions whatever to originate anything. It can do whatever

**Figure 2.1:** Physical representation of Turing's machine, an iconic device that exemplifies the principles of early computing.

we know how to order it to perform" (Menabrea et al., 1843, Note G). On the other hand, Turing (1950) thought that ultimately, robots might be able to simulate human-*like* reasoning to the point that a human would not be able to tell the difference. Turing's prototypical computer, much like Babbage's analytical engine, is a symbol manipulator (2.1). It adheres to the operator's set of rules (i.e. program) when reading and writing characters to an infinitely long tape. Its working effectively embodies CT, in the sense that crisp rules govern the process to a unique conclusion or result that is deduced analytically from the premises.

Even today, it is not uncommon to hear the analogy between a computer and a human mind. In the early days of computer history, Turing was perhaps the first to assimilate human thinking as computation: a man provided with paper, pencil, and rubber, and subject to strict discipline, is in effect a universal machine (Turing, 1937). After him, many were fascinated by the possibility that through this new technology we could automate even a part of our cognitive processes. Turing's vision launched an entirely new area of study known as Artificial Intelligence (AI).

The origins of AI are widely attributed to the Dartmouth College Conference, which took place during the summer of 1956. This gathering brought together a group of notable scientists, including John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon, to explore the possibility of creating machines that could emulate human thought processes. The conference was grounded in the philosophical concepts put forth by Turing, which posited that computers could be programmed to solve problems using methods akin to those employed by humans. The ideas generated at this event laid the foundation for much of the subsequent research in the field of AI.

The era of AI research that followed the Dartmouth College Conference is commonly known as Good Old-Fashioned Artificial Intelligence (GOFAI). During this period, researchers focused on symbolic reasoning and problem-solving. The outcomes of their investigations laid the groundwork for modern AI technologies, which have found application in various domains. However, the GOFAI approach also encountered opposition and criticism. Some scholars raised objections, particularly in light of the emergence of PT and the fallacies of CT that it exposed. In fact, some of the issues associated with CT are directly linked to those presented by GOFAI.

1. According to Searle (1980) and his Chinese room thought experiment, symbolic systems just need knowledge of the proper rules of manipulation rather than necessarily requiring comprehension of the symbolic references.

2. Since it is unable to determine how each new piece of information connects to a particular idea, GOFAI is constrained in its ability to update its opinions about preexisting concepts. As the number of concepts grows, the combinatorial explosion makes the issue impossible to solve with logic alone. This is referred to as the *frame problem* (Dennett, 1984) and is an epistemological byproduct of Plato's problem, discussed

in 2.1.1 and the problem of ignorance and error.

3. Symbolic reasoning "sustains no creative inductions, no genuinely new knowledge, and no conceptual discoveries" (Gärdenfors, 2000, p. 211). This is because "the epistemological origins of the initial logical predicates is never addressed" (Gärdenfors, 2000, p. 211). This echoes Lovelace's intuition that the machine only can produce what we ask it to, it does not question if the instructions have a valid reference to the world.

4. The rigidity of the definitional approach adopted by GOFAI and its reliance on analytic processes makes it unfit to model conceptual fuzziness and graded categorization. The first prototype theorists (Oden, 1977; Rosch & Mervis, 1975) suggest fuzzy-set theory as a complementary theory to PT that could better model such phenomena.

Overall, the successes of GOFAI are still to be praised. Perhaps among the many achievements of the program, the most remarkable was to beat a human at chess. It is the case of IBM's Deep Blue, a supercomputer that was first defeated by Kasparov 4-2 in 1996, but then won in a rematch only a year later by 3½–2½.

GOFAI's attempts at generating original media are numerous and span multiple domains. For example, it is widely accepted that the 1957 string quartet composition Illiac Suite, later known as String Quartet No. 4, was the first musical score created using an electronic computer. Lejaren Hiller and Leonard Isaacson worked together to create the compositional material using the ILLIAC I computer at the University of Illinois at Urbana-Champaign, where both authors were professors (Hillier & Isaacson, 1959).

The piece is divided into four movements, each of which corresponds to one of four experiments: the first movement deals with the creation of *cantus firmi*, the second with the generation of four-voice segments using a variety of rules, the third with rhythm, dynamics, and playing instructions, and the fourth with a variety of models and probabilities for generative grammars and Markov chains (Hillier & Isaacson, 1959). The use of pseudo-randomness combined with rules is one of the ways in which GOFAI is able to explore novel combinations while still maintaining an overall structure. The same approach is also found in the visual arts, where some of the philosophical ideas attuned to GOFAI and CT set in motion a new movement.

### 2.2.4 The birth of generative art

In parallel to the AI advancements discussed in the previous section, the visual arts community also embraced forms of computation. The *Generative Art* movement began in the 1960s and has since become an important part of contemporary art. It is a form of art that uses computer algorithms to create unique works, often with unpredictable (but deterministic) results. Its origin is specifically linked with the appearance of the programmable plotter, a machine capable of drawing on paper following procedural instructions.

The main exponents of generative art are artists such as Manfred Mohr, Frieder Nake, Georg Nees, and A. Michael Noll. The movement started in the late 1960s when these artists began experimenting with plotters to create visual art, inspired by the philosophy of German philosopher and mathematician Max Bense. His work focused on the idea that art should be generated and evaluated with mathematical principles (Nake, 2012). Rather than relying on traditional artistic sensitivity, he advocates for *information aesthetics*, an interdisciplinary concept of developing exact, scientific measures for introducing objectivity into aesthetics (Klütsch, 2012; Nake, 2012). He argued that this approach would allow for greater creativity and expression in art, as well as providing an opportunity to explore new forms of aesthetic experience.

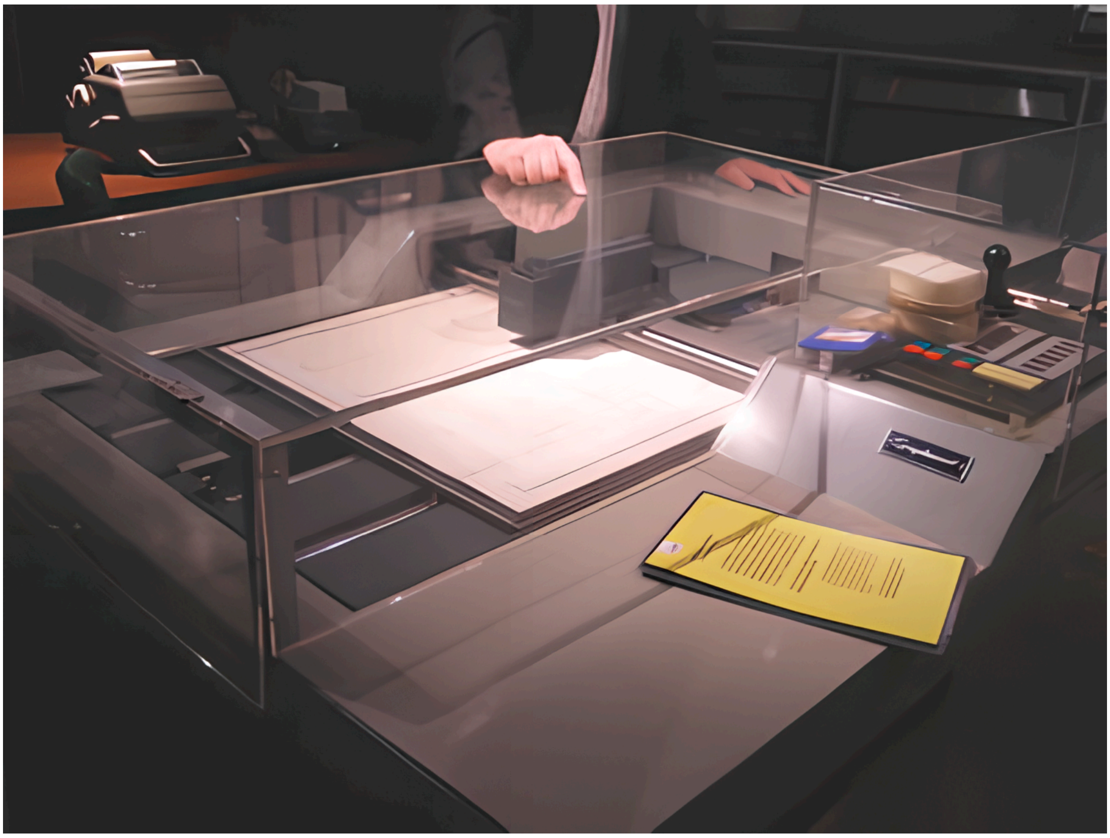

**Figure 2.2:** Image of the Zuse Graphomat Z64 plotter, a versatile device originally designed for technical drawing and other industrial applications. However, some artists and designers repurposed the plotter for creating generative art, a style of art that relies on algorithms and mathematical formulas to generate complex patterns and designs.

Early in 1965, Bense's lecture rooms hosted the first-ever display of computer-generated art, featuring a dozen or so abstract, black-and-white designs created algorithmically by Nees on the recently released Zuse Graphomat Z64 plotter (fig. 2.2) (Klütsch, 2012). Just as those lithic tools from 40 kyr ago (Rodríguez-Vidal et al., 2014), the plotter has been repurposed for non-utilitarian use, an analogy that holds in the form but perhaps not in the content. Ihde would argue that a lithic tool affords a substantially different mediation from that of a plotter. Indeed the rock would not carve a wall without being attached to a hand, yet for a plotter it is only necessary to input the code and press a button. In other words,

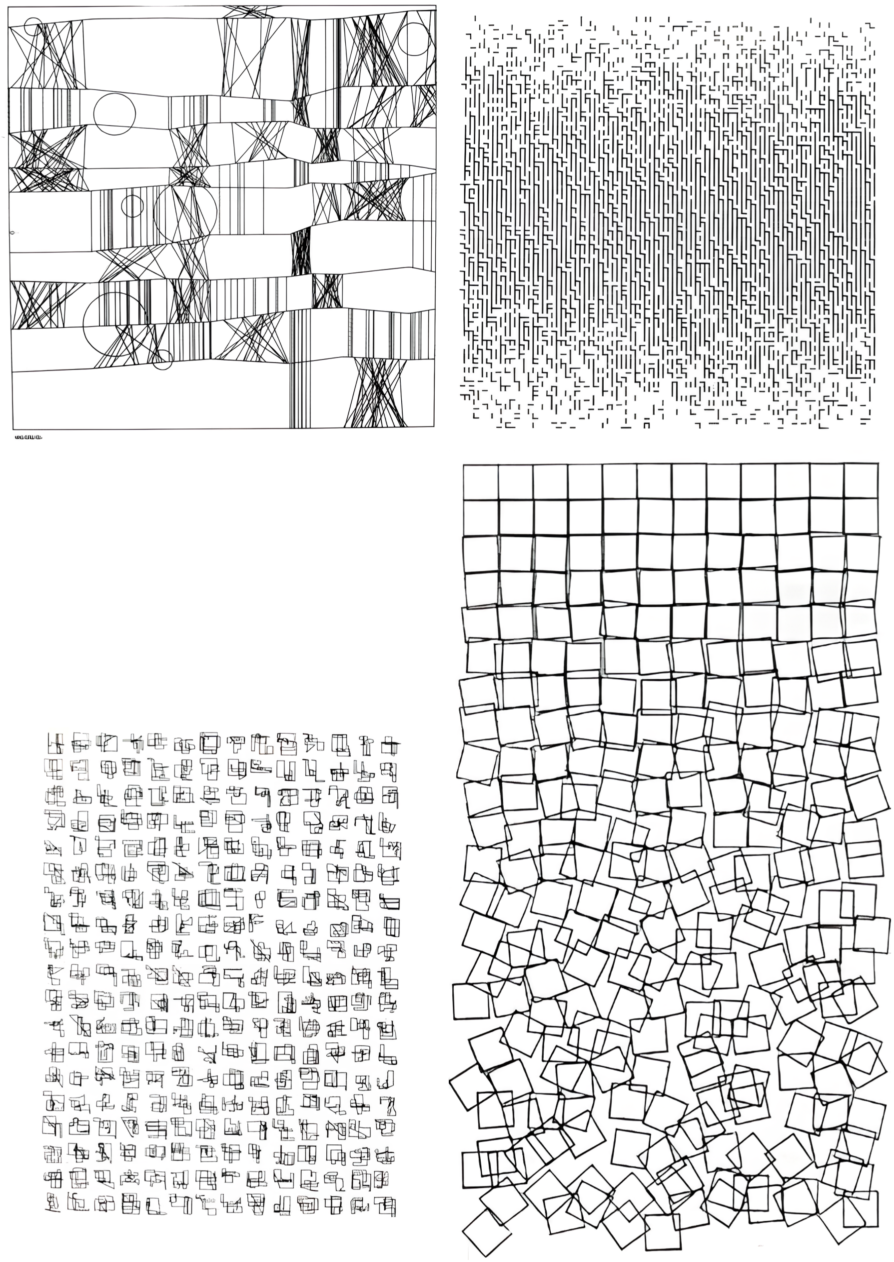

**Figure 2.3:** From left to right, top to bottom: “Hommage à Paul Klee” by Frieder Nake (1965), “Walk-Through-Raster” by Frieder Nake (1966), “23-Ecke” by Georg Nees (1965), “Shotter” by Georg Nees (1968). These four artworks are a selection of some of the landmarks in the history of computer-generated art, featuring intricate geometric patterns and designs generated by algorithms programmed by the artists. The artworks represent a significant milestone in the development of digital art and design, showcasing the potential of computers as a tool for artistic expression.

the actions of the author on the plotter affect the output only indirectly.

A post-phenomenological reflection is due here. GOFAI and the trend of symbolic reasoning offer a new way to see the world: as computation. Analytical thinking is in many ways married to CT as discussed earlier, so Bense's attempt to define aesthetics objectively (Klütsch, 2012; Nake, 2012) falls perfectly in line with the program. He is seeking a formal definition of Beauty, with necessary conditions for its application. The generative art movement is set out to expose the aesthetics of mathematics, logic and computation.

Conceptually, this was perceived both as strength and weakness in the early days of generative art. The innovative aspect in these artistic designs was the focus on the process, as Nake puts it: "The individual human subject simply did not exist anymore, once he or she had set the boundary conditions for the image to be computed" (Nake, 2010, p. 62). However, to the audience this may not be obvious by just looking at the output (some examples presented in Figure 2.3) and at Bense's exhibitions the reactions were intense (Klütsch, 2012). To appreciate this kind of artwork one must understand the relationship between the author, the program and the machine that produced it.

While the human-technology relation in generative art is not entirely dissimilar to the "human-pencil-paper" combination, the programmable plotter introduces a level of abstraction (i.e. the program) that is understood by the functioning of the plotter. Much like Latour's door (Latour, 1990), it is through the effort of both the human symbolic abstraction and the machine's embedded compatibility with it that these generative designs come into being. It is important to note that computer code is a necessary abstraction when using a plotter: it is perfectly possible to draw geometrical shapes using a pencil without knowing a formal procedure that generates them, but the same is not true for a plotter. This necessity shifts the attention towards the technology itself, leaving the output as an indirect byproduct of the interaction.

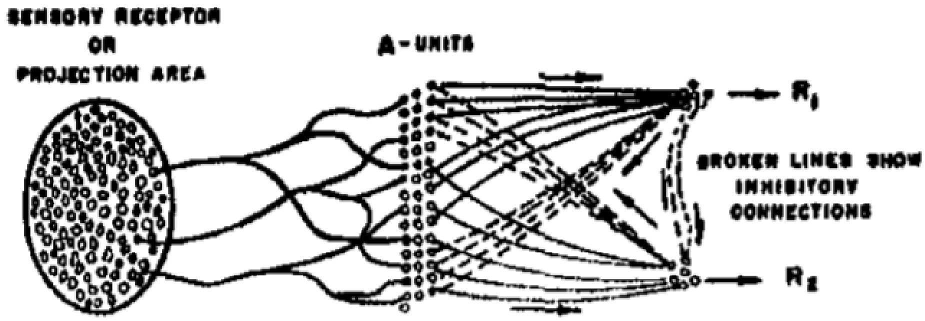


**Figure 2.4:** Image of Rosenblatt's first model of artificial neural network, a groundbreaking concept in the field of AI. The image showcases the basic architecture of the model, which comprises input nodes, output nodes, and hidden nodes that process and transmit information between them. The model was designed to simulate the behavior of biological neurons, and has since become a foundational concept in the development of machine learning and neural networks.

### 2.2.5 A bio-inspired mind

Alongside GOFAI another stream of research was concerned with a different agenda: modeling the biology of a human brain as a network of its primary components, the neurons. The *perceptron*, developed by Rosenblatt (1958), is considered to be the first formalization of a neural network, although at the time it did not receive much attention. Two decades later, in light of the criticisms directed towards CT and the GOFAI approach, a new stream of researchers set out to explore the use of neural networks as alternative implementation of human cognition. This scientific agenda is known as connectionism, a branch of cognitive science that emerged in the 1980s. Connectionism is based on the idea that mental processes can be implemented as networks of interconnected neurons. Some of the main exponents of connectionism are David Rumelhart, James McClelland, and Geoffrey Hinton.

Connectionism embraces a biologically-inspired framework of the mind, diverging from the logical model prevalent in GOFAI. The primary objective of this research endeavor is to comprehend human cognitive processes through simulation rather than the development of thinking machines. This perspective, commonly known as weak-AI, stands in contrast to strong-AI, which is dedicated to attaining Artificial General Intelligence (AGI).

Within a neural network, each neuron operates autonomously, resulting in decisions that are inherently local. Consequently, interpreting the values of individual neurons becomes challenging. This intrinsic locality attribute contributes to the limited dependability of neural networks in performing logical operations, a limitation that persists to this day. For example, Brown et al. (2020), state "GPT-3 achieves 21.3% accuracy at single digit combined operations (for example, 9*(7+5))". This should not be surprising because GPT is not intended to be used as a calculator, rather capture probabilistic features of language tokens and their relationships with other tokens.

Connectionism asserts that cognition can be understood as an emergent property of neural networks, which can learn to recognize patterns in data through backpropagation. This process involves adjusting the weights between neurons to minimize prediction errors. According to connectionists, this type of learning enables neural networks to generalize from examples, making them more powerful than traditional symbolic models that rely on explicit rules.

These claims have been supported by numerous studies conducted over the past few decades. For example, Rumelhart et al. (1987) demonstrated how backpropagation could be used to train a network to recognize handwritten digits with high accuracy. Similarly, Waibel et al. (1989) showed how neural networks could be used to recognize phonemes for speech recognition. Finally, McClelland et al (1987) demonstrated how a network could be trained to recognize the structure of sentences in natural language. These studies provide evidence that neural networks can learn complex patterns from data and generalize them to new situations, making them powerful tools for solving cognitive tasks.

Parallel distributed processing, as advocated by McClelland and Rumelhart, defines a different computing paradigm that contrasts the established GOFAI approach. Much like CT and PT, the two schools rely on different assumptions about how concepts should be represented. In practical terms, the GOFAI approach was at the time much more developed and established due to its commercialization, while parallel processing hardware was complex, expensive and relegated to research. The first wave of excitement about neural networks was primarily due their scientific explanatory power into the biology of the brain, but the high computational cost, their complexity and unpredictability kept the enthusiasm at bay, at least until the early '00s.

### 2.2.6 Hybrid theories

A compelling hybrid theory has been proposed by Gärdenfors (2000). The author posits that concept formation cannot be accurately depicted by either the GOFAI or connectionist approach, as neither adequately addresses the concept of *similarity*. Gärdenfors recommends incorporating a third intermediate layer that utilizes geometric structures to better model similarity relations. In this framework, concept formation is closely associated with identifying convex regions in space, as demonstrated in Figure 2.5. While convex regions arise from physical properties such as COLOR or SHAPE for raw perceptual dimensions, the same cannot be assured for abstract dimensions like JUSTICE or HONESTY. The fact that convex regions are found for a given conceptual space validates the conceptual-space and its dimensions, but this might lead to confirmation bias (see the swans example in Section 2.1.2).

In spite of its shortcomings, Gärdenfors (2000) theory of conceptual spaces still provides great insights into modern AI. A growing number of authors made attempts to integrate symbolic reasoning and deep learning, employing different strategies which, can be grouped into "*semantic characterizations* (i.e., define a logic whose formulas capture deep

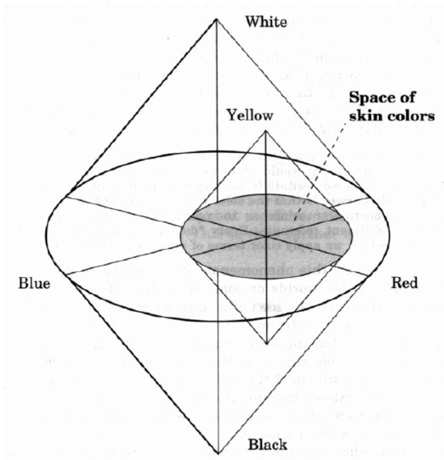


**Figure 2.5:** Image illustrating Gärdenfors's conceptual space of colors, a theoretical model that organizes colors based on their perceptual and conceptual properties. The image focuses on the convex subset region that represents the diverse range of skin tones, highlighting the potential of the model for understanding and analyzing color perception and representation.

learning), *constrained learning* (i.e., integrate symbolic constraints in deep learning), and *hybrid methods* (allow neural computations and symbolic reasoning to co-exist separately, to enjoy the strengths of both worlds)" (Fuxjaeger & Belle, 2019, p. 1). Some of these architectures, such as the neuro-symbolic concept learner by Mao et al. (2019), may be considered almost direct implementations of Gärdenfors' framework.

The intuition that concepts may be described by topological properties finds at least partial confirmation in the reliance of modern *deep learning* on the idea of *latent space*. *Latent spaces* are in fact topological entities that encode training data to a space with less dimensions, a process that may be considered a form of *compression* and *abstraction*. The last decade has seen remarkable development in the understanding of *latent spaces* and their application.

### 2.2.7 Deep learning

A new wave of research on neural networks came around in the early 2010s, also known as Deep Learning (DL). DL is a subset of Machine Learning (ML) that uses neural networks with numerous layers to learn from data and make decisions. Since then, it has become one of the most popular and influential fields in AI research, with an exponential growth in the number of papers published every year (Krenn et al., 2022). Geoffrey Hinton, Yann LeCun and Yoshua Bengio were credited as its founding fathers and received the ACM A.M. Turing Award in 2018 for their contribution.

Deep Learning algorithms are used for various tasks such as image recognition, natural language processing (NLP), speech recognition, autonomous driving and robotics. The major achievements of DL include breakthroughs in computer vision tasks such as object detection and segmentation; NLP tasks such as machine translation; speech recognition tasks such as automatic speech recognition; and reinforcement learning tasks such as game playing.

One of the most influential papers in the field was published by Hinton et al. in 2006 titled "A Fast Learning Algorithm for Deep Belief Nets". This paper introduced a new algorithm called *deep belief nets* which allowed for faster training times than traditional neural networks. This paper laid the foundation for many subsequent advances in deep learning research.

Krizhevsky et al. (2012) published a paper titled "ImageNet Classification with Deep Convolutional Neural Networks" which introduced the AlexNet architecture. This paper demonstrated that deep learning could be used to achieve state-of-the-art performance on image classification tasks. This breakthrough was followed by many other papers in the field, such as Szegedy et al.'s 2014 paper "Going Deeper with Convolutions" and He et al.'s

2015 paper “Deep Residual Learning for Image Recognition”, which further improved the performance of deep learning models on image recognition tasks.

In 2016, Google Brain researchers published a foundational paper titled “Attention Is All You Need” which proposed a new type of neural network called Transformer (Vaswani et al., 2017). This model was able to achieve state-of-the-art results on various NLP tasks such as machine translation and text summarization without using any recurrent layers or convolutions. One of the key aspects that this architecture introduces is the concept of self-attention. Self-attention layers learn how different elements in a sequence are related (or unrelated) to other elements. Compared to recursive neural networks (RNNs) the transformer-attention architecture processes sentences all at once, rather than word by word, which allows for better parallelization, leading to more efficient training and inference time.

In light of what has been discussed about compositionality in the previous section, it is worth noting that self-attention in seq2seq (transformers) models seems to capture at least some aspects of language compositionality. From a theoretical standpoint, token embedding dimensions as intended in DL could be assimilated to Jackendoff’s *semantic fields* that are described by probability distributions. Self-attention in neural networks is able to capture only salient relationships between tokens in a sequence therefore enabling specific inter-dependent predictions starting from any pair of tokens.

These achievements mark a trend that has been on rise for almost two decades. One of the factors that contributed to this upward trend is also the increased accessibility of hardware that can afford the kind of computation neural networks require. In the early 2000s a small number of academics began looking into the use of graphics processing units (GPUs) for DL algorithms (another *lithic tool* gets repurposed!). These devices are well suited for it despite being initially created for rendering graphics because they are engineered to run mathematical operations of linear algebra (matrix multiplications, vector dot products, etc...) leveraging parallel computation.

Another fundamental contributing factor is the availability of large amounts of data and metadata in digital form accessible through the internet. The Web 2.0 era allowed for users to become producers of their own media through mobile phones, flooding the internet with user generated content. This perhaps commercially driven push towards ML-based data analysis stimulated innovation aimed at dealing with the scale and magnitude of the accumulated information. In fact, a defining characteristic of popular DL tools of today is precisely that they have been learning at massive scale. Such is the case for GPT3, trained on 45TB of text data, or Stable Diffusion, trained on 5.85B images.

These factors, and possibly many others, have lead us to the tools we have today which are based on deep learning. Of particular interest for creativity studies is the subfield of Generative Deep Learning (GDL), which is particularly concerned with deep learning models that can generate output rather than simply classify data.

GDL offers a novel method for creating unique artifacts. The generated content retains a consistent set of characteristics, such as a face, while allowing for variations from the original training dataset, such as altering the hairstyle or transforming a frown into a smile (Pumarola et al., 2019). An additional advantage of GDL is that these large unsupervised models can extract features without requiring an annotated dataset. For instance, the authors of StyleGAN2 assert: "the new architecture leads to an automatically learned, unsupervised separation of high-level attributes" (Karras et al., 2018). In practice, unsupervised learning enables Deep Neural Networks (DNNs) to process extensive corpora of text, images, or sounds and acquire a set of *parameters* that can be modified and employed for synthesizing new outputs.

The precise nature of these *parameters* learned by DNNs remains somewhat elusive, particularly given the challenges that TOCs face in defining concepts and their properties. At the most basic level, it can be inferred that these parameters represent the properties and compositional relations (utilizing self-attention) of the input data they aim to replicate. However, questions arise regarding how DNNs extract these dimensions, the meaning of concept learning in this context, and whether this process can be deemed inherently creative (Hoorn, 2023).

### 2.2.8 Non-human creativity

The creative process requires a medium: when we create, we always create *something*. Creative artifacts exist in relation to their supporting technology and are therefore bound to the support's affordances. When a new technology emerges, the exploration of its novel affordances might bring disruption in what an audience believes it means to be creative. In many cases, the introduction of a new technology also produces a shift in the aesthetic sensitivity, and sometimes even assembles a new community of reference. As the progress of technology becomes part of the background of our society, the audience will naturally shift the selection criteria for what is considered novel and valuable.

Indeed, any attempt to create an *intensional* definition for the term *creativity* may necessitate an *audience of reference* responsible for evaluating specific attributes such as *novelty* or *value* (Boden, 2003; Newell et al., 1959; Rhodes, 1961). Similar to Meno (see Section 2.1.1), the audience may encounter difficulties in formulating such definitions. Consequently, differing perceptions of *novelty* and *value* may arise among various audiences. It appears nearly inescapable that, when discussing creativity, we must accept an *extensional* definition and acknowledge that evaluations of creative output will invariably reflect human bias (Hoorn, 2023).

Over the past two decades, advancements in computing technology have led to several modifications in how creativity is assessed. The Computational Creativity (CC) community has revisited the concept of creativity to accommodate non-human artifacts: "[t]he performance of tasks which, if performed by a human, would be deemed creative" (Wiggins, 2006a, p. 451). This criterion is agent-agnostic and focuses solely on the output. The underlying assumption is that different instances of the creative process will generate artifacts with varying creative value. Colton also asserts: "While it is not a necessary requirement, there is an implicit assumption that to produce the most pleasing artefacts, aspects of human creative behaviour will have to be simulated" (Colton, 2008, p. 1).

However, a certain ambiguity arises when attributing creative qualities to artificial agents based solely on their output. Notably, AlphaGo's (Silver et al., 2016) move 37 was met with astonishment by commentators, the audience, and Sedol himself. This move could arguably be considered creative, but it did not originate from a simulation of human Go-playing strategies[1]. AlphaGo Zero (Silver et al., 2017) surpassed AlphaGo by eliminating human bias and relying solely on self-play to develop its strategies. Nevertheless, an agent must incorporate a minimal set of assumptions about the world, which will determine the potential or desired outcomes. Even in the case of AlphaGo Zero, which was not exposed to human gameplay during training, the rules of Go constitute the minimal instructions required for task performance. Once this space is defined and explored by an agent without human intervention, non-human solutions that could be deemed creative may emerge.

As counter argument, it is then fair to ask whether AlphaGo's move 37 can be deemed as not creative. After all it is the algorithm's goal to explore game-winning strategies, the fact that it came up with its own distinctive ones is in line with the instructions it was given, how is this surprising? From this perspective, move 37 may be considered just a display of an *alien* way of thinking (Fazi, 2018) that perhaps just is to be deemed as *intelligent* rather

[1]This may be the reason for AlphaGo's success and why these moves were considered *novel*

than *creative*. Furthermore, AlphaGo's *alien* strategies may not be of any value to humans, since we do not have the computing power in our brains to come up with those moves ourselves.

We meet here at an ontological fork in the road. What does it mean for a technology to be described as creative? Is it something about the form of its internal processes? Or is it just about the judgment of its output by experts? How does the human interaction with technology come into play? The field of CC has been trying to address these theoretical questions and provide space for technical solutions to be shared and discussed. In order to better understand the discourse around computational creativity and its reception of the probabilistic turn, a systematic literature review has been conducted.

### 2.2.9 Computational creativity: a systematic literature review

The systematic literature review presented in this section has two main purposes. The first one is to identify the research questions and topics discussed in the field and the second is to expose the effects of current shift towards probabilistic models in CC.

The corpus of 386 papers under review has been retrieved from Scopus search using the query string `KEY("computational creativity")` as of March 2023. The full reference, abstract and number of citations of each paper has been retrieved via API using custom code provided in Appendix A. The papers in the corpus were categorized according to three dimensions:

- The medium addressed in the paper. The list of possible choices of medium has been defined as:
  - No medium
  - Visual, images and movies
  - Music and musical composition
  - Design, urban design, architecture

  - Writing, narrative and language
  - Game design
  - Concepts
  - Culinary recipes
  - Multi-modal

- The theoretical scope of the paper:
  - **Evaluation**: the paper discusses the evaluation of artifacts.
  - **Theory**: the paper discusses a particular theory or hypothesis about creativity.
  - **System**: the paper presents a technical implementation of a specific creative system.
  - **Other**: the scope of the paper does not fit any of the other categories.

- The computational approach adopted or discussed in the paper:
  - **Rule-based**: these are methods that follow deterministic rules and definitions, such as expert systems.
  - **Evolutionary algorithms**: genetic algorithms and other methods inspired by evolution.
  - **Data-driven**: this includes deep learning, machine learning and other probability based approaches.
  - **Other**: the paper does not belong to any of the above or there is no specific approach.

The categorization process utilized can be summarized as follows:

1. Skim all papers and identify the categories listed above.
2. Ask GPT3.5 to categorize each paper and provide an explanation for its choices.
3. Review categorization, assess accuracy, adjust if necessary.

The use of GPT3.5 to support the categorization task is to be considered mostly as an exploration of the capabilities of this tool, rather than a validation. In fact, a preliminary categorization exercise was already performed in early 2020 at the beginning of my doctoral journey. The addition of this intermediate step was inspired by the curiosity of how

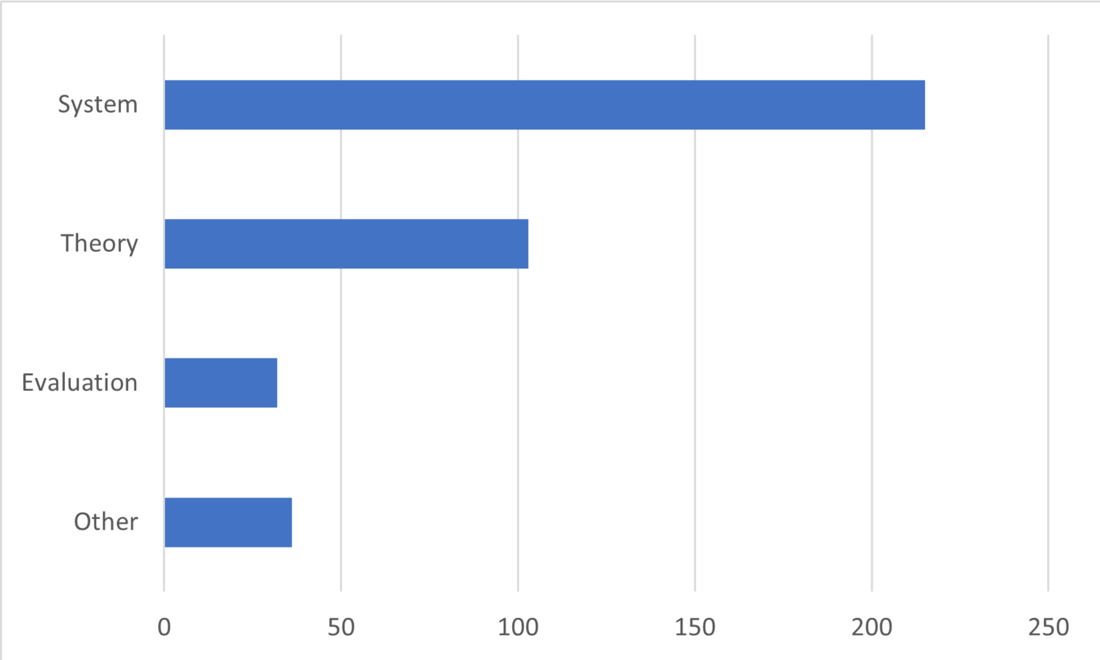


**Figure 2.6:** Horizontal bar chart depicting the distribution of *Scope* in the systematic literature review. The X-axis represents the frequency of occurrences. The chart highlights the variation in the distribution of *Scope* among the papers reviewed, providing insights into research focus of the literature in the field.

accurate GPT's categorization might be, given that traditional methods such as topic modeling did not yield satisfying results. GPT was tasked with assigning to each paper a value for each of the dimensions specified above (the prompt used can be found in appendix A). After reviewing the assigned, GPT's classification was given a score from 0 to 3 based on how many dimensions were considered correct. According to this measurement, GPT's accuracy over the whole corpus was $\approx 88\%$. Considering that GPT did not have visibility over the full text of each paper and that some abstracts did not have all the information required to make an accurate judgment, this method proved to be rather effective in assisting to categorize the corpus, especially for those cases that are relatively straightforward.

Figure 2.6 and 2.7 summarize the number of papers by scope and media respectively. Over half ($\approx$55%) of the papers examined discuss the implementation of a specific system,

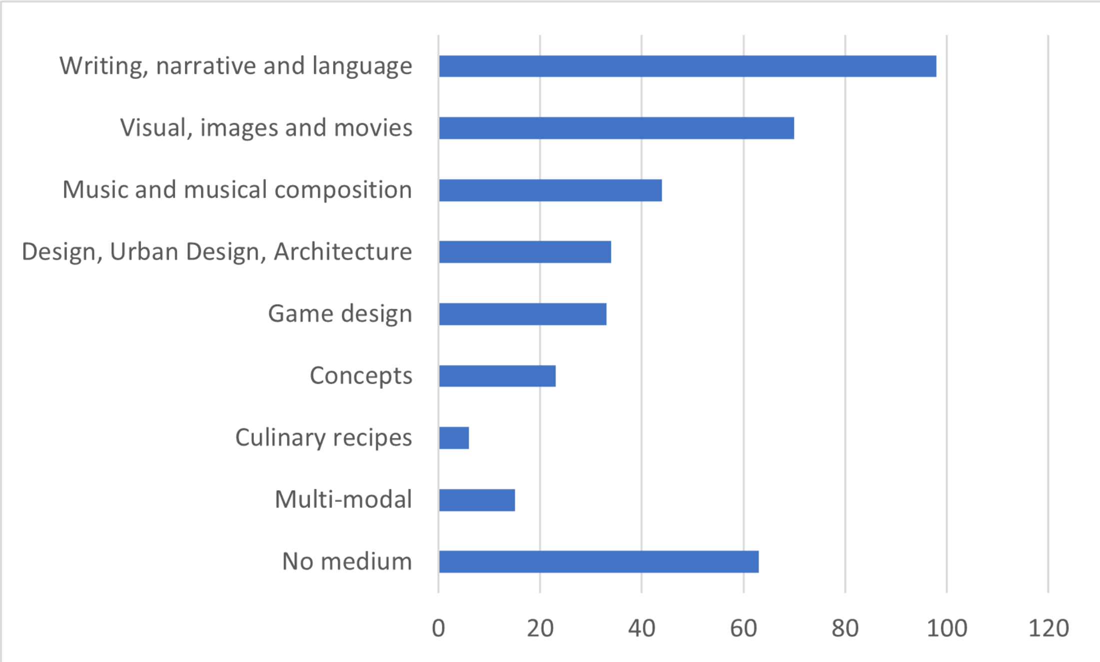


**Figure 2.7:** Horizontal bar chart depicting the distribution of *Medium* in the systematic literature review. The X-axis represents the frequency of occurrences. The chart highlights the variation in the distribution of *Medium* among the papers reviewed, providing insights into the domains and topics of the literature in the field.

indicating that the community provides space for practical experimentation. The rest of the papers address theoretical issues regarding either creativity itself (≈26%), the topic of evaluation of creativity (≈8%), and other topics such as reviews of the field or historical perspectives (≈10%). As for the media distribution, the top three are unsurprisingly language (≈25%), images (≈18%) and music (≈11%), followed closely by design (≈8%), game design (≈8%) and concepts (≈6%, mostly discussing conceptual blending). Not all papers discuss creativity within the context of a specific medium, 16% of them take a more abstract approach and focus on exploring creativity purely from a theoretical standpoint. Further analysis of the breakdown shown in Figure 2.8 summarizes the medium based on each of the identified scopes. As expected, most theory papers do not discuss a specific medium, rather tackle the topic of creativity as an abstract process, disentangled from a specific ap-

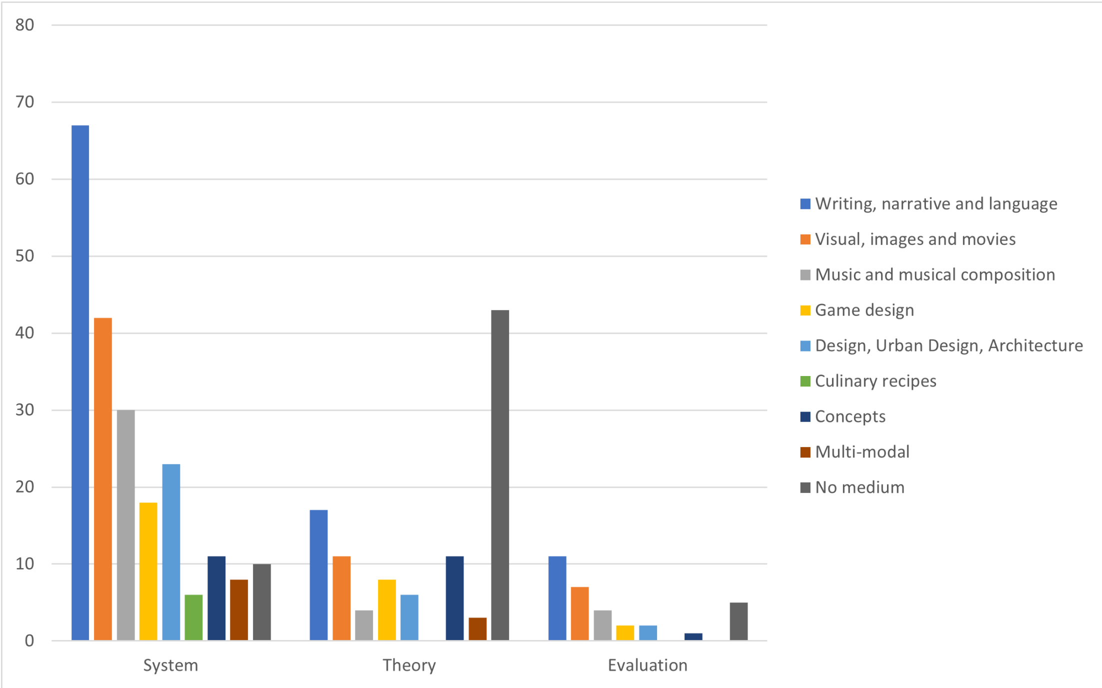


**Figure 2.8:** Vertical bar chart illustrating the breakdown of each *Scope* by *Medium* in the corpus of papers. The chart reveals trends and patterns in the literature based on *Scope*.

plication. The majority of papers discussing evaluation of creativity, however, seem to be primarily situated in a specific domain, with only 5 papers identified as addressing the topic without focusing on a medium.

Figure 2.9 displays the number of papers adopting a specific approach, broken down by year. This chart is particularly interesting in relation to the probabilistic trend emerging in recent years. The sharp decline of rule-based approach papers makes space to new discussions around data-driven methods. This trend also produces a shift in the scope of the papers, as shown in 2.10. It seems that not as many new system implementations have been produced in the last 4 years as in the previous years, but instead the discussion seems to have shifted towards theory and evaluation of systems that are not built by the community but are publicly available, such as GPT (Radford et al., 2019). The different ratio between the amount of systems and the amount of papers discussing theory or evaluation across

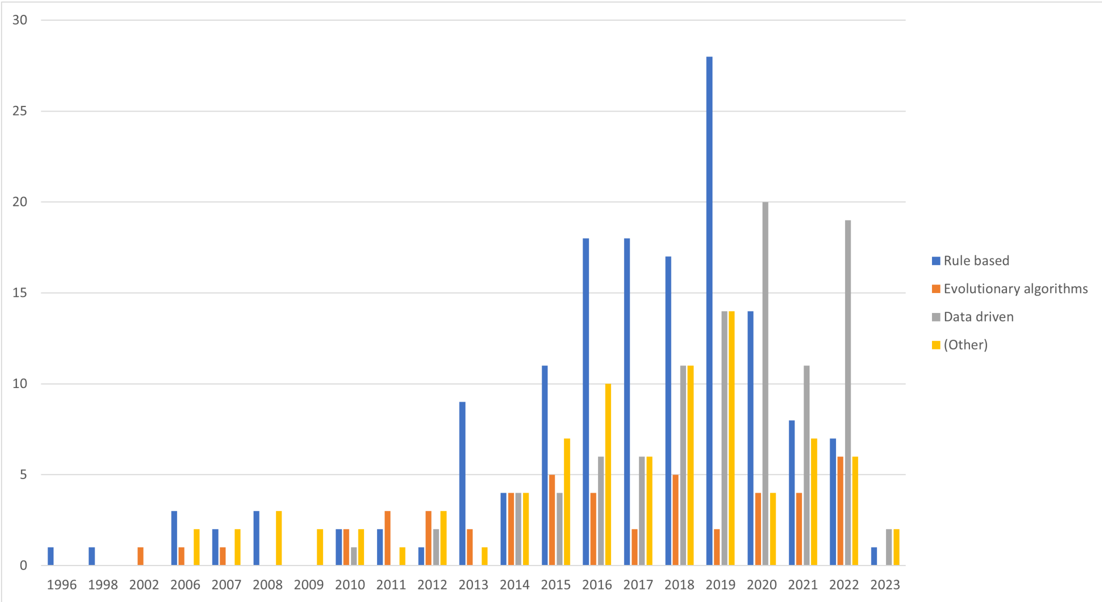


**Figure 2.9:** Vertical bar chart illustrating the breakdown of each computational *Approach* by *Year* in the corpus of papers. The chart reveals trends and patterns in the literature based on *Approach* over time, providing insights into the evolution and focus of computational methods in the field.

different approaches is also noticeable in the breakdown presented in 2.11, which seems to confirm this intuition.

What does this mean for the CC community and the future of research in this field? If this trend continues, it is plausible that the community will develop a deeper understanding of how DL systems fit into the creative process. It is also possible that the ecosystem of computational creativity will display less diversity in the range of tools used, as the big players converge towards a fully multi-modal experience, such as the one offered by GPT4. Perhaps this will allow the community to focus on questions regarding the implications of using these tools in our everyday life. In any case, it seems evident from this data that, in the last 4 years, the field has witnessed a sharp turn towards a new form of computational creativity, which is not implemented directly by the community itself. Academic research is perhaps now chasing the forefront of development rather than leading it.

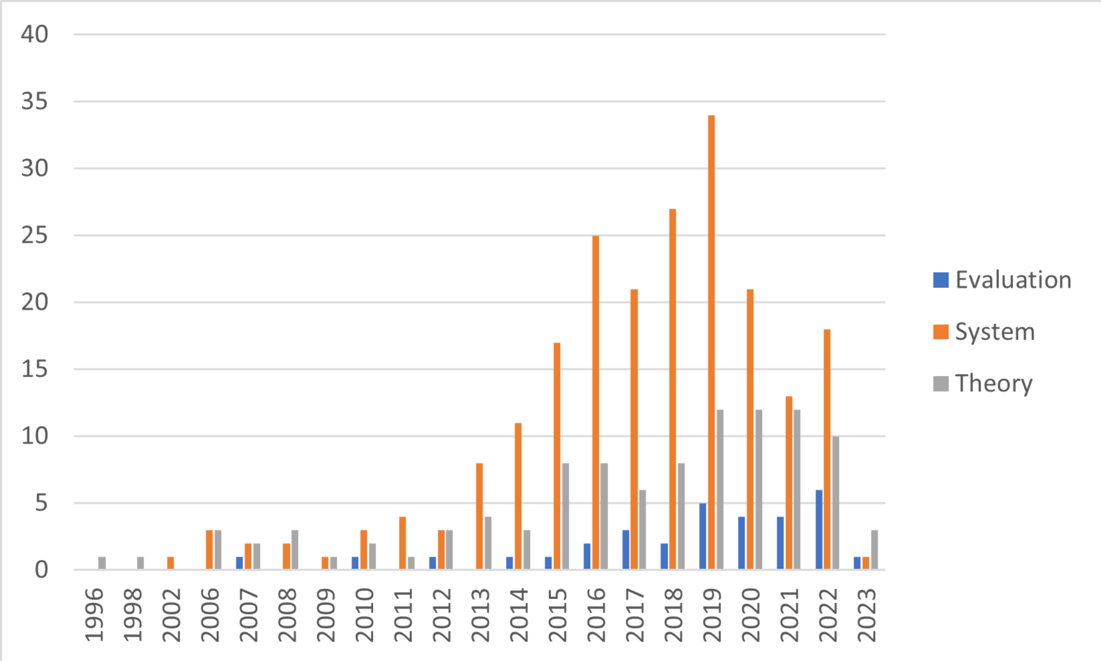


**Figure 2.10:** Vertical bar chart illustrating the breakdown of each *Scope* by *Year* in the corpus of papers. The chart reveals trends and patterns in the literature based on *Scope* over time, providing insights into the evolution and focus of the field.

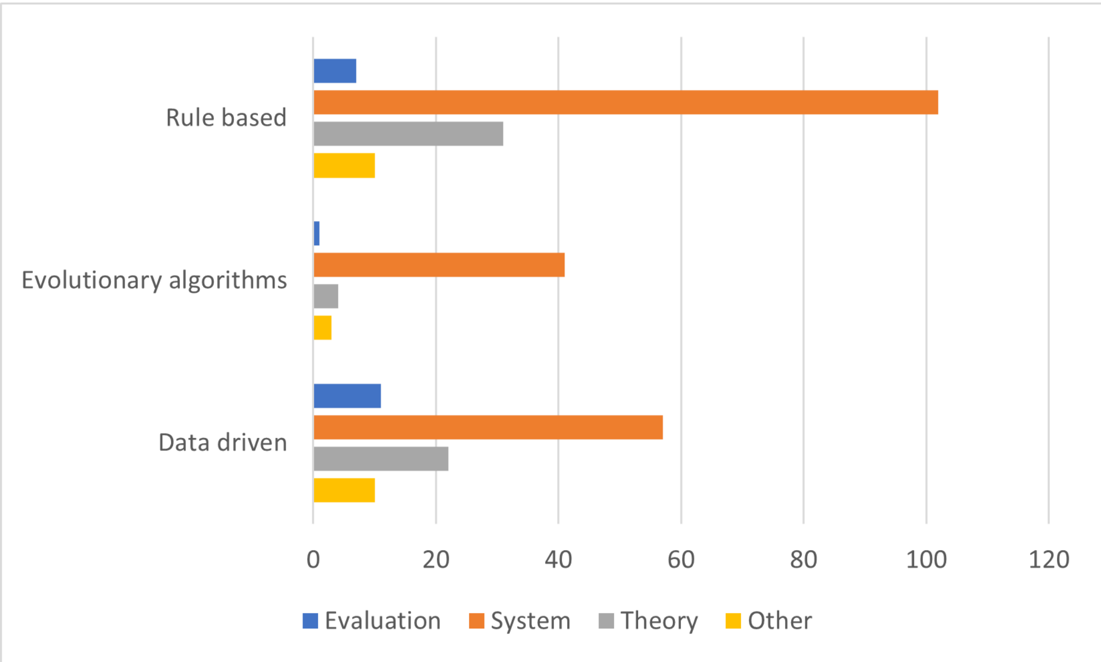


**Figure 2.11:** Vertical bar chart illustrating the breakdown of each computational *Approach* by *Scope* in the corpus of papers. The chart reveals trends and patterns of research focus based on computational *Approach*.

## 2.3 Summary

This literature review addressed several bodies of work in different fields with broad strokes, attempting to find trans-disciplinary links between them. It highlighted possible links between the study of conceptual categorization in humans, different approaches to AI and the post-phenomenological view of technology. It also exposed the DL trend that is unfolding in the field of CC, which arguably embeds a theory of concept of its own. This review also shows the possibility for a fruitful encounter between the post-phenomenological view of technology and non-human creativity theory, which so far has not been explored, to the best of my knowledge. In the next chapter I will attempt to frame the different approaches to AI and their underlying assumptions about what concepts are under the post-phenomenological interpretation, with specific focus on technologies that mediate human creativity.

*Definitions get you into that time trap, and I'm very much more process-focused. Take Lucy, for example. Lucy is famous largely because she has almost a total skeleton. The more sophisticated we get with instruments, the more we can find out. Through CT scans of her skeleton, they now think she died falling out of a tree because of the way her bones are broken. If nineteenth and twentieth century technologies can retroactively transform our bodiment, what then do the technologies we now use do?*

Don Ihde

# 3 Methodology

Two things should be evident at this point. First, that there is a fundamental distinction between the rule-based and the data-driven approach, and that this distinction is grounded in the different TOCs that they subscribe to. Second, that there is a growing trend shifting the attention towards data-driven technologies. This chapter presents the methodology (see Figure 3.1) and the theoretical framework used in this thesis to investigate the impact of this trend on the creative process.

To understand the impact of the current data-driven trend on the creative process, this research starts by establishing a theoretical framework capable of addressing the distinction, within the context of the creative process. The next step is to identify critical factors of concept representation in data-driven technologies. This is accomplished in the first and

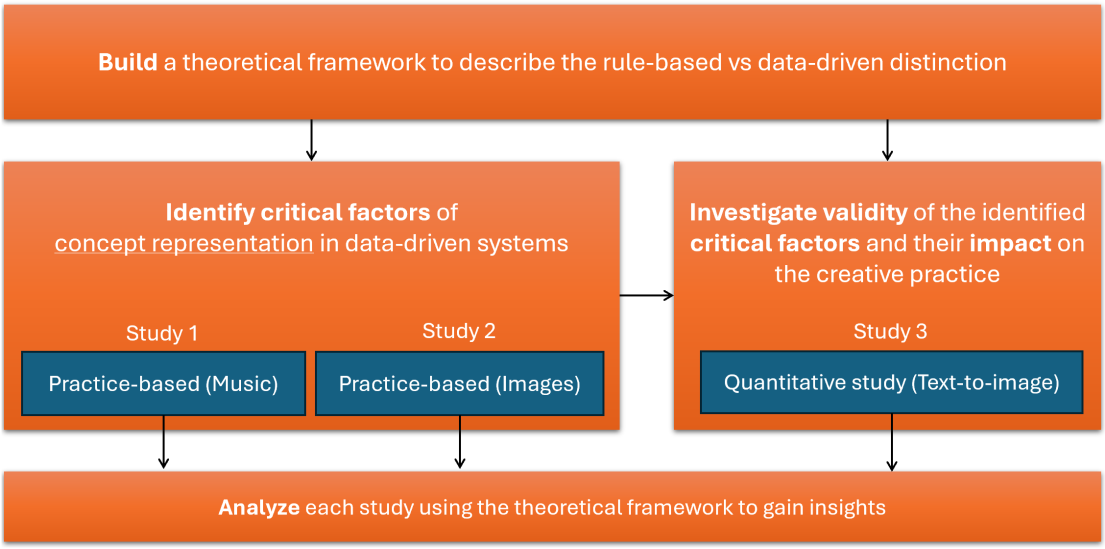


**Figure 3.1:** The diagram shows the different components of the methodology and their connections.

second study using qualitative research. The identified critical factors are then investigated in the third study through quantitative methods. The theoretical framework is also used to analyze each study and gain further insights.

## 3.1 Technological mediation

The methodology employed in this doctoral thesis is firmly grounded in post-phenomenology, a philosophy of technology that understands technologies in light of how they mediate human-world relations by co-constituting the subjectivity and objectivity of experience (Rosenberger & Verbeek, 2015). As discussed in the previous chapter, addressing computational creativity under this perspective proves valuable, as it offers a robust framework for analyzing and interpreting the intricate relationships between technology, human experience, and the social context. Moreover, it can accommodate a wide variety of possible relations between humans, technology, and the creative process.

In this section, the application of mediation theory is explored as a means to provide a comprehensive understanding of both the GOFAI-CT (Good Old-Fashioned Artificial Intelligence and Classical Theory) paradigm and the PT-DL (Prototype Theory and Deep Learning) approach. Mediation theory, as proposed by Ihde (1990), offers a systematic framework to analyze the interaction between human cognition, technology, and the world, making it a suitable lens through which to examine these two distinct paradigms in the field of artificial intelligence and cognitive science. Ihde's schematization of mediation can be summarized as follows:

- Unmediated perception: I—World
- Mediated perception: I—Technology—World

He provides an account of four fundamental types of relations in which humans, technology and the world stand in specific relation to one another, summarized in Table 3.1. In *embodiment* relations, technologies unite with a person and point their unity outward at the outside world. For example, we talk on the phone with other people rather than to the phone itself, and we view objects through a microscope rather than at it. As defined by Ihde, *hermeneutic* relations are those in which people interpret how technologies reflect the world, such as an MRI scan that depicts brain activity or a metal detector's beeping that denotes the presence of metal. In this situation, technology unites with the environment rather than the person using it. People are drawn to the ways that technology depicts the universe. In a third category of human-technology-world relations, which Ihde refers to as the *alterity* relation, people engage in technological contact with the outside world acting as a backdrop. Instances include interacting with robots, withdrawing cash from an ATM, and using machinery. In actuality, one of the major areas of interaction design can be seen as this relationship. Fourth, Ihde makes a distinction between the *background* connection and how technologies frame human experiences and behaviors. In each of these instances, technology is a context for human existence rather than being directly experienced by the user. The sounds of air conditioners and refrigerators, the cool air from heating systems, the warm air from heating installations, the notification sounds from cellphones during a conversation.

| Name | Form | Definition | Examples |
|---|---|---|---|
| Mediated Embodied | (I — T) → W | Broaden the area of sensitivity of our bodies to the world | Glasses, a dental probe, a paintbrush |
| Mediated Hermeneutic | I → (T — W) | Provide a representation of the world that we need to interpret | Thermometer, watch |
| Alterity | I → T (— W) | Humans are related to or with technology as a quasi-other | ATM, robots |
| Background | I (— T / W) | Shapes the context of our experience in a way that is not consciously experienced | Refrigerators, central heating system |

**Table 3.1:** A summary of the relations types proposed by (Ihde, 1990) and their formalization. In the examples in the second column, **I** represents the human, **T** stands for Technology, and **W** refers to World.

The theory also comes with notation system defined as follows:

| | |
|---|---|
| — | simple connections between entities |
| → | interpretation of one by the other |
| ( ) | being experienced together |
| / | being in the background of another entity |
| [ ] | being already contextualized in some way before being processed |

As a starting point, a possible formulation of the embodied relations examples introduced in the previous chapter could be:

- (I — Pencil) → Drawing
- (I — Door Handle) → Door

Such relations describe the tools used as being ready-at-hand, as if they are becoming a direct extension of our bodies. Through this symbiosis between the human and the non-human we can act on the world and realize the potential of the object of our attention (Drawing and Door).

When observing a human interaction with a computer or a plotter under the mediation theory lens, we could describe it as an *alterity* relation:

- I → Computer (— Result)
- I → Plotter (— Printed Design)

In *alterity* relations, human attention is directed towards the tool itself, as we need to interpret (→) its interface to obtain the desired output. These relationships could be expanded further, taking into consideration that *alterity* relations require the technology to exhibit some form of autonomy, or else there is no possibility of interaction with a *quasi-other*. In literature, this idea has been discussed in various forms by different authors. Eede (2010) provides an overview of the different notions of *transparency* and *opacity*, as discussed in various strands of post-phenomenology.

Discussing Ihde, the author suggests that "in the case of embodiment relations — Ihde partly builds upon Heidegger's tool analysis here — a technology must be transparent enough, that is, to the person embodying it, for one to be able to embody it" (Eede, 2010, p. 148). Eede continues: "with alterity relations, the transparency moves even further away—here the human interacts with the technology itself. The technology has thus become opaque" (Eede, 2010, p.149). This technological opacity is effectively overlapping with the idea of autonomy in the sense that if a technology is experienced as a *quasi-other* then the human does not need to see through it. Furthermore, "in background relations, Ihde (1990, p.109) says, [...] the role of background presence is not displaying either what I have termed a transparency or an opacity. Instead, here the technology is, phenomenologically speaking, present as an absence" (Eede, 2010, p. 149). By looking at what is acting in the *background* of *alterity* relations we can gain a more detailed understanding of how the technology interprets human behavior. Here is a potential expansion:

- I → Computer / (CPU → [Computation]) (— Result)
- I → Plotter / (PlotterSoftware → [Design]) (— Printed Design)

For example, when we ask a Computer to calculate $5 + 8$, the symbols have to first be translated into the corresponding operations that are needed for the CPU to produce the result. This involves transforming symbols representing decimal numbers into byte values, allocating memory to store the result and possibly many other tasks that are not evident to the human which could be considered as acting in the background (hence the / symbol). The operator [] (contextualization) in this case effectively coincides with the programming language interface used to express the computation. This hidden abstract layer between the computer and the human is what enables the entire system to produce the result.

In a similar way when we try to describe a DL-PT typical interaction this expansion may look like this:

- I → GPT / (Model → [Language]) (— Response)
- I → Stable Diffusion / (CLIP → [Image Description]) (— Generated Image)

In the case of a language model such as GPT, the interface is simply natural language, which is interpreted by the model. Similarly for large diffusion models in a text-to-image pipeline, CLIP first translates the description into a latent space vector which then is used to guide the diffusion. What does the operator [] stand for in this case? It seems that this step is where the difference in the two approaches to computation lies. Instead of interpreters and compilers that translate the programs into machine instructions, this step in neural networks is governed by statistical inference, which requires prior data (i.e. the model weights). Referring back to Kant's distinction discussed in Section 2.1.1, it could be argued that, under the GOFAI paradigm, a machine uses formal logic, rules, and symbols to return logical conclusions to our questions based on some initial definition: "Are bachelors married? False (Because you told me what a bachelor is, I have never actually seen one)". This contrasts with the deep learning data-driven approach, which relies on statistical models and data to make predictions, in a loose sense *learning from experience*[1]. Probabilistic models typically represent information extracted during training as vectors in a high dimensional space and operate through inference: "Are bachelors married? Probably not (Because I have not seen a married bachelor yet)".

This thesis explores the hypothesis that the two computational approaches play a role in the technological mediation. To formalize this distinction, I introduce a new notation to describe these two modalities of information processing:

- $R[...]$ for rule-based (GOFAI-CT analytic) context and
- $D[...]$ for data-driven (DL-PT synthetic) context

So we can now express all of the above without confusion:

- I → Computer / (CPU → **R[Computation]**) (— Result)
- I → Plotter / (PlotterSoftware → **R[Program]**) (— Printed Design)
- I → ChatGPT / (Model → **D[Question]**) (— Answer)
- I → Stable Diffusion / (CLIP → **D[Image Description]**) (— Generated Image)

[1]It would be controversial to claim that machines experience anything at all. Colton et al. (2020) provides an interesting perspective on this particular issue, yet this is beyond the scope of this section. The term *experience* is only used as an analogy to facilitate the reader in understanding the different approaches that GOFAI and DL are bound to.

In prior studies addressing this matter, Benjamin et al. (2021) put forward an extension of the post-phenomenological approach to characterizing machine learning, inspired by Heidegger, Ihde (1990), and Rosenberger and Verbeek (2015). They suggest that, compared to other forms of technological mediation, ML embeds an interpretation of the world within the context of data. For example, the original version of CLIP has been trained on "400 million (image, text) pairs collected from a variety of publicly available sources on the Internet" (Radford et al., 2021, p. 3). As a result, the alignment between image and text produced by that version is dependent on how the dataset is constructed and what it contains. If there are no examples in Chinese, then D[Chinese] would simply "fail" silently, in the sense that it would still generate an image, but the output will be poorly aligned with the input prompt, as the system cannot make a well-informed guess.

Building upon the previous examples, it is possible to identify two distinct archetypes of technological mediation that broadly correspond to the two computing paradigms under investigation:

- **GOFAI**: Human → Technology / (Program → R[Input]) (— Output)
- **DL**: Human → Technology / (Model → D[Input]) (— Output)

This distinction and its notation shall serve in the coming chapters to discuss how the two forms of contextualization affect the creative process. To understand how these operators influence the technological mediation, it is essential to see through an *alterity* relation's *opacity* and identify the nature of the contextualization (i.e., asking: R[] or D[]?). For example, naive users may struggle to differentiate or see through the technological opacity of DL tools due to their novelty and inherent non-logical nature.

Because this thesis is particularly concerned with the impact of these two types of technologies on the creative process, mediation theory is not sufficient to describe the entire picture. Addressing this challenge, the examination of various creativity theories and their integration with mediation theory can provide valuable insights into the roles of R[] and D[] within the creative process. The following section will present different theories of creativity and extend the theoretical framework just established, so that it can more accurately and fruitfully describe the technological mediation that is typical of the creative process.

## 3.2 Creativity theories

As highlighted in the literature review of CC, the community has explored and analyzed various ways to frame creativity. On one hand, some authors have discussed creativity as a non-anthropocentric idealized process, such as search or formal combination making (Besold, 2017; Hoorn, 2014; Wiggins, 2006a), while others have focused on the interaction between human and non-human elements, addressing topics such as co-creation and evaluation (Davis, 2016; Jordanous, 2012; Kantosalo, 2019; Saunders, 2012). Many of these authors refer to well-known creativity theories in their arguments, attempting to re-contextualize them for non-human creativity. For this doctoral thesis, it is crucial not only to mention these foundational theories but also to connect them with the post-phenomenological interpretation of technology. The following sections will discuss three theories to establish a foundational layer for the studies presented in the subsequent chapters. Although these three theories represent only a limited subset of the available theories, they offer a well-rounded overview of how creativity has been conceptualized by recognized experts in the field.

### 3.2.1 Rhodes' four Ps of creativity

According to Rhodes (1961), there are four perspectives on creativity: the person, the process, the product, and the press. When Rhodes talks about *perspective* in the context of creativity, he means a way of looking at or understanding creativity. The four perspectives that Rhodes identifies each emphasize different aspects of creativity and provide unique insights into how it works. By considering all four perspectives together, researchers and practitioners can develop a more nuanced understanding of what drives creative thinking and how it can be fostered in individuals, communities and organizations. I present below a non-exhaustive list of authors and theories addressing each perspective. It is rare for an author to discuss only one perspective, in fact most of the researchers mentioned below developed comprehensive theories which cover more than one. The purpose of this list is to identify each perspective and its scope, rather than give a full account of each author's theory.

### Person

The **person** perspective focuses on the individual characteristics, traits, and abilities that contribute to creativity. This perspective emphasizes the importance of personality traits, such as openness to experience, and cognitive abilities, such as divergent thinking, in fostering creativity.

Some examples of authors and their theories addressing this perspective:

1. Mihaly Csikszentmihalyi: Csikszentmihalyi is a psychology professor who has written extensively about creativity and its relationship to personality traits. He suggests that individuals with certain personality characteristics, such as openness to experience and non-conformity, are more likely to be creative than others (Csikszentmihalyi, 1990; 1996).

2. Teresa Amabile: Amabile is a professor at Harvard Business School who has conducted extensive research on creativity in the workplace. She argues that intrinsic motivation plays an important role in fostering creativity, and that individuals who

feel passionate about their work are more likely to generate creative ideas (Amabile, 1997).

3. Dean Keith Simonton: Simonton is a psychology professor who has studied the relationship between intelligence and creativity. He suggests that while high intelligence is necessary for creative thinking, it may not be sufficient on its own; other personal factors, such as openness to experience and perseverance, also play an important role (Simonton, 1985; Simonton, 1994; 1999).

### Process

The **process** perspective focuses on the steps and stages involved in creative thinking and problem-solving. This perspective emphasizes the importance of various cognitive processes, such as incubation and insight, in the creative process.

Some examples of authors and their theories addressing this perspective:

1. Graham Wallas: Wallas was an early theorist of the creative process and his book "The Art of Thought" (1926), is considered a classic in the field. He proposed that there are four stages to the creative process: preparation, incubation, illumination, and verification.

2. Robert Sternberg: Sternberg is a psychologist who has studied creativity extensively throughout his career. He has proposed a number of models of the creative process over time; one influential model suggests that there are six stages to creative thinking: problem-finding, problem-definition, idea generation (using divergent thinking), idea evaluation (using convergent thinking), implementation planning and taking action (Sternberg, 1988; 2003).

3. Mihaly Csikszentmihalyi: Csikszentmihalyi's work on flow also encompasses aspects of the creative process; he suggests that individuals often experience flow states during periods of intense focus on a task or problem-solving challenge (Csikszentmihalyi, 1990).

### Product

The **product** perspective focuses on the creative output or outcome of the creative process. This perspective emphasizes the importance of evaluating the quality and originality of the creative product. Some examples of authors that have discussed this perspective in depth within their theories.

1. Mark A Runco: Runco's research on creativity includes work on evaluating creative products across domains such as science, art and literature. Runco has developed several measures for assessing creativity in various domains, such as science, art, and literature. These measures are designed to evaluate products based on their originality, fluency (i.e., the number of ideas generated), flexibility (i.e., the range or diversity of ideas generated), and elaboration (i.e., how much detail or complexity is present). Runco's work also emphasizes the importance of using expert judges to evaluate creative works. He suggests that experts can provide valuable feedback on the quality and originality of creative products, and that their judgments are often more reliable than those made by non-experts (Runco, 2014).

2. James C Kaufman: Kaufman is another well-known researcher in the field of creativity who has studied the evaluation of creative products across domains including visual arts, music, writing and humor. One area of his research that is particularly noteworthy is his focus on identifying specific characteristics of creative products

that can be used to assess their quality and originality. For example, he has identified a number of features that are common across highly-rated paintings such as harmony and complexity. Kaufman has also developed several measures for assessing creativity in different domains. These measures are designed to evaluate creative products based on criteria such as novelty, appropriateness and value (Kaufman, 2016; Kaufman & Sternberg, 2019).

### Press

The **press** perspective focuses on the environmental factors that influence creativity. This perspective emphasizes the importance of social and cultural factors, such as organizational climate and societal norms, in fostering creativity.

1. Keith Sawyer: Sawyer is a psychologist who has studied the role of collaboration and group dynamics in fostering creativity. He argues that social factors, such as trust and communication, can play an important role in promoting creative thinking (Sawyer, 2017).

2. James C. Kaufman: Kaufman also studied cultural influences on creativity around the world. His research on this aspect of creativity suggests that different cultural values and norms can either encourage or discourage creative thinking, depending on how they prioritize individualism versus collectivism (Kaufman & Sternberg, 2019).

3. Teresa M. Amabile: Amabile has also written extensively about the social and organizational factors that can influence creativity. Her research emphasizes the importance of providing individuals with autonomy, resources, and support in order to foster creative thinking (Amabile, 1996). Furthermore, in their book "The Progress

Principle" (2014) Amabile and Kramer describe how organizational factors like positive feedback, social support, and meaningful work help people stay motivated and engaged in creative work.

4. Richard Florida: Richard Florida is a well-known urbanist and social scientist who has studied the role of creativity in economic development. His work focuses on how cities can attract and retain *creative class* professionals, such as artists, designers, and tech workers. Florida's research emphasizes the importance of creating environments that are conducive to creativity. He argues that cities that offer a high quality of life (e.g., good schools, affordable housing, access to cultural amenities) are more likely to attract creative professionals than those that do not. He also suggests that cities with strong social networks and diverse populations are more likely to foster innovation and creativity (Florida, 2014).

### 3.2.2 Boden's 3 types of creativity

Margaret Boden is a renowned cognitive scientist and philosopher who has extensively studied creativity. According to her theory of creativity, there are three types of creativity: combinatorial, exploratory, and transformational (Boden, 1996; 2003).

1. Combinatorial Creativity: This type of creativity involves taking existing elements, concepts, or ideas and combining them in a novel way to create something new. Combinatorial creativity is often seen in fields like art, music, and literature where artists or creators take inspiration from previous works but add their own unique twist to it by mixing elements from different styles or genres. Analogies and metaphors also included in this category, for example, Niel Bohr's model of the atom borrows the idea of the solar system to describe how electrons revolve around the nucleus.

2. Exploratory Creativity: This type of creativity involves exploring a conceptual space working within accepted rules of procedure looking for something that no one else has discovered before (A. I. Miller, 2019). Exploratory creativity is often seen in scientific research or technological innovation where researchers push the limits of what we know by experimenting with new ideas and theories. For example, when Marie Curie discovered radium she was exploring an area that had not yet been fully understood by science.

3. Transformational Creativity: This type of creativity involves fundamentally changing how we view or understand something. Transformational creativity often arises from challenging existing assumptions or paradigms and proposing a radically different perspective on a subject matter. For example, when Einstein developed his theory of relativity, he transformed our understanding of space and time, challenging the long-held assumptions of Newtonian physics.

It is important to note that these types of creativity can also overlap or intertwine with one another. For example, a scientist may use combinatorial creativity when combining existing theories to form a new hypothesis before exploring it further with exploratory creativity. Similarly, an artist may achieve transformational creativity by challenging established artistic norms and then use combinatorial creativity to create something novel within this new framework.

Boden's taxonomy is a recurring theme in creativity research, it is hard to find books or articles that do not mention it, for better or for worse. Among the critics, Hoorn (2023) challenges Boden's (and many before her) view that creativity is the "ability to come up with ideas or artefacts that are *new, surprising and valuable*" (Boden, 2003, p. 1). Hoorn argues with regards to combinatorial creativity: "Two things can be put together but the observer does not know this has already been done before. Two things can be put together and the outcome is new to the observer but unsurprising because it was predicted from the premises. A novelty may be regarded invaluable at invention. For instance, Hertz demonstrated Maxwell's electromagnetic vortices but deemed them useless" (Hoorn, 2023, p. 3). This echoes what has already been discussed in Section 2.2.8 about the notions such as *novelty* and *value*: these criteria are dependent on the audience of reference and, more in general, the context. To sidestep these and other issues, Hoorn promotes a model of creativity that is non-anthropocentric and modular.

### 3.2.3 Hoorn's ACASIA model

Hoorn's ACASIA model of the creativity described in his book "Creative Confluence" (2014) is a relatively new theory attempting to describe the creative process as a set of six different components. These modules may be described as follows:

1. **Association**: This module refers to the capacity to generate images, words, meanings, and other semantically related features in response to a stimulus or entity. It involves connecting different ideas or concepts that are related in some way.

2. **Combination**: This module involves establishing connections between associations of disparate entities by matching their (fuzzy) feature sets and measuring perceived similarity and dissimilarity. It is the process of combining different ideas or concepts into something new.

3. **Abstraction**: This module involves bringing certain phenomena onto a conceptual level where connections can actually take place. At higher abstraction levels the similarity between entities is increased, thereby making new links possible.

4. **Selection**: This module involves dismissing those features that cannot be used to make the combination acceptable in the eyes of the creator or the audience, hence affecting the measure of dissimilarity between disparate entities. It means choosing which aspects of different ideas or concepts will be included in the final product.

5. **Integration**: This module involves attaching the features of one entity to another entity literally, creating a cohesive whole from disparate parts.

6. **Adaptation**: This module involves changing individual features such that the transition from one entity to another would become acceptable by optimizing the blend between them. It entails modifying certain aspects of different ideas or concepts so they fit together better in the final product.

One defining aspect of ACASIA is its modularity, which allows for flexibility in implementing all or only some of the components in a creative system. For example, a non-human agent might only generate random combinations, while humans perform the evaluation steps (Selection, Integration, Adaptation). This feature also enables the model to describe natural phenomena, such as evolution or chemistry, as creative systems within the same framework. For this reason, the ACASIA model is very well suited to compare the two forms of technological mediation discussed in Section 3.1.

### 3.2.4 A post-phenomenological view of computational creativity

To better understand the differences between $R[]$ and $D[]$, a comparison of these approaches in the context of the ACASIA model is presented below. Table 3.2 demonstrates how each ACASIA module might be implemented based on the two paradigms. This comparison connects existing CC literature with Hoorn's model and the post-phenomenological interpretation, offering insight into the contrasting nature of the paradigms and their potential impact on the creative process.

| **Module** | Program → **R**[Input] | Model → **D**[Input] |
|---|---|---|
| Association | CT-based similarity models, such as fuzzy sets or semantic networks, typically employ search algorithms to produce meaningful associations. Weights can be adjusted according to goals and concerns, so *Press* elements may guide this step. | *Latent space* is constructed from datasets so the similarity space is predetermined during training. Associations turn out to be a reflection of the most probable associations found in the dataset, which may not be ideal for creative purposes. |
| Combination | Crisp compositional rules can produce a large number of combinations which preserve a desired structure. Rule-based systems might yield unexpected results because they are not normally influenced by typical instances. | As seen in Section 2.1.2, compositionality is problematic when using statistical methods. However, transformers and attention can solve this issue to some degree, for example, by capturing some elements of the compositionality of language and using this *latent space* to condition the generation (see Section 2.2.7). |
| Abstraction | Analytical methods are already *abstract*. Semantic networks are typically constructed by humans, so rule-based systems can only perform second-order abstractions as instructed by the user. | Unsupervised learning architectures such as VAEs excel at abstraction (see Section 2.2.7). *Latent spaces* constitute an viable ground for making connections between concepts using geometrical methods in a multi-dimensional space. |
| Selection | Selection criteria may be implemented formally (e.g. Max Bense's aesthetic principles, see Section 2.2.4) based on given properties that need to be defined objectively, which leads to issues about definitions (see Section 2.1.1). | The primary selection criteria used in DL is the optimization of a loss function, typically calculated as a distance from ground truth represented in the training dataset. This makes the selection process very efficient, but also not quite explainable, because *latent space* is generated during training and it is not intelligible by humans. |

| | | |
|---|---|---|
| Integration | In rule-based systems a set of instructions for integration must be provided. While it can be relatively easy and efficient to provide rules for integration in a settings where the entities that need to be glued together are relatively simple, it might become an issue as complexity increases, as the number of rules might suffer from combinatorial explosion. | Integration in DL systems is performed seamlessly in *latent space* and the complexity it can handle depends on the size of the model (number of dimensions of the latent space). Attention algorithms can capture integration strategies that are typical in the training data. |
| Adaptation | Rule-based systems might employ fuzzy sets or other forms of graded adaptation to gain optimal similarity. | Generative DL models are naturally adaptive as they rely on one or more loss function(s) that can guide the adaptation process during inference in a 0-shot learning scenario. For example, LDMs like Stable Diffusion can do this out-of-the-box. |

**Table 3.2:** A comparison of how ACASIA modules are implemented using $R[]$ and $D[]$.

These theoretical considerations will be expanded more in detail with examples in each of the three studies presented in the coming chapters. It is important to note that not all of these modules need to be automated in the creative process. In fact, none of the studies address a fully automated system that performs all of these functions. In each study there are shared responsibilities between the algorithms, the users and myself in the role of researcher and practitioner in support of the users. The exploratory scenarios presented next can be considered as multi-agent systems combining humans and non-humans, where each of the entities might be in charge of one or more of these modules. Because my role as technical expert in the studies does also occasionally overlap with the my role as researcher, it is important to frame the methodology of this doctoral thesis within the larger picture of practitioner research.

## 3.3 Practitioner research

As it should be evident from the literature review, computational creativity is a rapidly evolving field that encompasses various disciplines, including artificial intelligence, cognitive science, and the arts. Given its dynamic nature, traditional research methodologies may not be sufficient to capture the nuances and complexities of this field. As already noted by Doorst over a decade ago, "[t]he purely analytical models of science that we have been using will only get us so far: in the face of such an immensely complex area as design, only experimental methods can bring the clarity and understanding we are seeking. We need to re-engage with practitioners, and get involved in experiments within the rapidly changing design arena." (Dorst, 2008, p. 11) Practitioner Research (PR), a form of insider research, has emerged as a valid methodology of inquiry due to its ability to blend theory and practice, allowing for a more nuanced understanding of the field (Candy, 2011). This section argues that PR is a valuable approach to studying computational creativity, given the fast pace of development in this field.

PR is particularly relevant in computational creativity due to its emphasis on self-reflection. Through self-reflection in the form of rigorous doubt about one's own way of practicing, it is possible to gain new insights and experiment with new ideas in a short cycle. The role of analytical thinking in this dynamic is to maintain coherence and crystallize theories, while practice enables the exploration of new conceptual spaces. PR allows practitioners to engage in critical self-reflection, examining their assumptions and biases and seeking to improve their practice through continuous reflection and experimentation.

In "Theory construction in design research: criteria, approaches, and methods", Ken Friedman (2003) delves into the significance of practice-based research in design and its connection to theory construction. Friedman posits that design is an interdisciplinary field that intersects with various domains, including natural sciences, humanities, social and behavioral sciences, human professions and services, creative and applied arts, and technology and engineering. This interdisciplinary nature of design highlights the importance of practice-based research, as it enables designers to apply knowledge from different fields to solve specific design problems. Friedman (2003) explains that practice-based research involves solving problems, creating new things, or transforming less desirable situations into preferred ones. However, understanding how things work and why requires analysis and explanation, which is the purpose of theory. Theory construction is crucial in design research as it provides a framework for understanding and interpreting design phenomena.

While practice-based research is essential in design, Friedman (2003) argues that it is not enough on its own to develop theory. He explains that "instead of developing theory from practice through articulation and inductive inquiry, some designers simply argue that practice is research and practice-based research is, in itself, a form of theory construction" (Friedman, 2003, p. 519). However, the author contends that this approach is insufficient as it fails to account for the critical inquiry and reflective insight necessary for theory construction. He emphasizes that "even though design knowledge arises in part from practice, it is not practice but systematic and methodical inquiry into practice—and other issues—that constitute design research, as distinct from practice itself" (Friedman, 2003, p. 512). Therefore, to reach from doing to knowing requires the articulation and critical inquiry that leads a practitioner to reflective insight.

Reflective insight is the ability to critically examine one's own experiences, assumptions, and beliefs. It involves a deep understanding of one's own practice and the ability to articulate that understanding to others. Reflective insight is crucial in theory construction as it enables designers to identify patterns, make connections, and develop frameworks for understanding design phenomena. Friedman's view on the necessity of self-reflection in practice-based research is rooted in the idea that such research involves a deep engagement with one's own professional practice. According to Candy, "the concept of reflective practice (Schön, 1983) provides a link between action research and practice-based research" (Candy, 2006, p. 19). Reflective practice, as defined by Schön, involves an individual's reflection on his or her own professional practice, rather than broader situations.

In particular, practice-led research is a form of PR that focuses on the nature of practice and leads to new knowledge with operational significance for that practice (Candy, 2006, p. 1). This type of research includes practice as an integral part of its method and often falls within the general area of action research (Candy, 2006, p. 19). It is essential to distinguish practice-led research from practice-based research, which emphasizes the use of creative artifacts as the basis of contribution to knowledge. In contrast, practice-led research results may be fully described in text form without the inclusion of a creative work (Candy, 2006, p. 1).

The methodology of practice-led research involves using practice as an integral part of the research method (Candy, 2006, p. 19). Practice-based researchers should devise a clear set of methods and techniques for collecting and analyzing data (Candy, 2006, p. 19). The personal process is a crucial element of practice-led research, and data collected should include initial starting points or motivation for the project or work, prior models or theories about how to create, perform or realize a creative artifact, time frame for the work or works to be created, role of the creative artifact in the creative process, environments and tools required to achieve the output, information to be gathered about the thinking, methods, tools, resources, support, collaboration, methods for collecting and collating data gathered, methods for analyzing collated data, expected outcomes of the research process, and the relationship of the practice outcomes to the argument of the thesis (Candy, 2006, p. 19).

## 3.4 Studies overview

The studies presented in the next three chapters all focus on observing interactions between people and technology in the context of various creative endeavors. The first two studies (Chapters 4 and 5) were conducted within the PR framework and address exploratory practices in the field of DL as a context for self-reflection. Both studies closely examine the interactions that emerge from the encounter between an expert in the field and data-driven technology. My role as a practitioner in both these studies was to provide technical solutions in response to the expert's needs. The third study (Chapter 6), however, adopts a more traditional approach, addressing the interactions of a larger group of participants and their behavior through quantitative measures. My active role in the third study is minimal and essentially limited to managing the platform where interactions take place.

Due to the inherently exploratory nature of these studies, my role as a researcher has been in constant evolution. The field of ML/DL is experiencing rapid growth in the number of tools and solutions available, making it a rather challenging task to stay up to date with the forefront of research and development. During interactions with experts, my contribution (and bias) primarily consisted of assessing their needs and crafting a viable solution to achieve the desired output. One of the main difficulties in this process was establishing a common language to present and explain the technology. This challenge was not surprising, as the technological *opacity* of computational creativity tools, particularly those based on DL, conceals the inner functional elements, making it harder for people without a technical background to understand why the system behaves the way it does.

Moreover, all three studies come with the unavoidable drawback of being immediately outdated, somewhat ephemeral, context-dependent, and highly subjective in nature, given that the environment is in constant flux. There is simply not enough time to prepare a well-structured experiment with tested protocols because, within a month or two, everything can change quite radically. For example, in the summer of 2022, generating an image from a text prompt required 1-3 minutes. It was simply impossible to observe a group of individuals using text-to-image technology in a creative setting, considering budget constraints and time limitations. Back then, only evaluating pre-generated images seemed like a possible strategy to understand the impact of this type of tool. As Stable Diffusion was released in August 2022, from one day to the next, it became possible to generate an image in approximately 4 seconds. This improvement opened up the possibility of running workshops with 20-30 people generating images as an iterative process, producing hundreds of images within a couple of hours.

For this reason, the studies are not particularly concerned with the specific technology being used, but rather with the broader implications of data-driven technologies for creativity and human-technology interaction. These studies are designed to contribute to the construction of a theory that describes how data-driven technologies differ from the rule-based technologies that we are accustomed to interact with. By examining the interactions between people and technology in creative contexts, these studies aim to shed light on the unique characteristics of data-driven technologies and their potential impact on creative practices. The theory that emerges from these studies aims to provide a framework for understanding and interpreting the phenomena associated with data-driven technologies in creativity, taking into account the nuances and complexities of this rapidly evolving field.

*My interest in making music has been to create something that does not exist that I would like to listen to. I wanted to hear music that had not yet happened, by putting together things that suggested a new thing which did not yet exist.*

Brian Eno

# 4

# Study: Music

MUSIC SITS AT THE INTERSECTION BETWEEN COMPLETE FREEDOM OF EMOTIONAL EXPRESSION AND RIGOROUS COMPOSITIONAL RULES. The physical properties of sound expose the mathematical symmetries of matter and vibrations. Yet, music can feel organic and deeply tied to our inner emotional experience. Evidence suggests that specific ratios between frequencies are at least culturally recognized by human brains (Jacoby et al., 2019). These physical properties lay out multiple dimensions of expression accessible through combination making.

## 4.1 Scope

Prior to the invention of recordings, the transient nature of sound limited the exploration of the conceptual space of musical composition to those who had the ability to perform. The first musical instrument is considered to be a flute made of bone, dating back some 35,000 years ago (Conard et al., 2009). The instrument was capable of producing a set of tones to which humans presumably could sing or dance along. The earliest known system for musical notation dates back to the Greeks in the 5th century BC. This system allowed composers to notate rhythms and melodies, and eventually enabled others to perform works composed by someone else.

The invention of recording technology enabled us to explore this conceptual space even further by allowing us to capture and store sound indefinitely. Music has gone through many iterations since then, from jazz and blues in the early 20th century, through rock n' roll, punk, hip-hop, EDM, etc..., all exploring different combinations of sounds in an attempt to create something new and unique. Recording technology also enabled us to mix multiple tracks together and add effects such as reverb or delay which add more complexity to the combination space. This innovation has removed the need to perform with the musical instrument in hand by providing a technological interface to sound, much like the plotter that inspired the generative art movement in the in the 1960s.

Similarly to other artistic domains, technological advancements seem to have shaped the evolution of sound and music allowing for a diversity of audiences to appreciate many forms of technologically mediated creativity. We could think of some examples of *embodied* technological mediation, such as:

- (I — Bone Flute) → Musical Notes (Possibly random pitches)
- (I — Violin) → Musical Notes (Any pitch in range)

- (I — Piano) → Musical Notes (Well-tempered pitch system)

It is interesting to observe in these examples how each instrument possesses a distinct pitch range, timbre, and note interval system, yet maintains a degree of interoperability. This is because they ultimately share the final medium (air), which facilitates a common experience. In fact, while some musicians are appreciated for their solo performances, an ensemble of musicians arguably delivers a more intricate and rich experience, offering something extra to the audience.

There are also plenty of examples of *alterity* relations in the music domain:

- I → Synthesizer (— Synthesized Sounds)
- I → Sequencer (— Repeating Note Patterns)
- I → Digital Audio Workstation (— Music Track)

These examples are based on a fundamentally different method of representing sound and music, which relies on technology. To better understand this idea, we may examine the two examples below:

- (I — Classic Guitar) → Classic Guitar Sound
- (I — Electric Guitar) → (Amplifier — Electric Guitar Sounds) / Analog Signal

Arguably, the transition from the first to the second example introduces a whole new expressive range that the musician can now control through the knobs and pedals of the amplifier, which processes the signal from the guitar. The signal coming from the electric guitar is not audible per se, but contains information about the vibrations to be reproduced. A musician must interact with the amplifier to produce sound; however, I would not consider this an *alterity* relation because the interaction still directly affects the output. For this reason, I would categorize the (Amplifier — Electric Guitar Sounds) as an *hermeneutic* relation. A human still needs to hold the guitar and play it, regardless of whether the instrument is a traditional acoustic one or an amplified one. On the other hand, when using a synthesizer, the output signal is produced indirectly from minimal human action. The triggering of notes in a synthesizer might even be controlled by other components, such as arpeggiators or sequencers, making the whole process better understood as a complex system rather than a direct linear flow. Digital systems are also ambivalent in this sense, as they can be fully transparent to the performer, except for perhaps a few milliseconds of latency introduced by the Analog-to-Digital and Digital-to-Analog Conversion. However, they can also be fully opaque, as is the case for an entirely digitally produced track that reuses sampled sounds.

So I believe it is important to expand the *alterity* examples above to include these foundational technologies as acting in the background of analog/digital tools, coupled with rule-based (R[]) interpretation of inputs:

- I → Synthesizer / (Analog Signal — R[Controls]) (— Synthesized Sounds)
- I → Sequencer / (MIDI — R[Notes]) (— Repeating Note Patterns)
- I → Digital Audio Workstation / (Digital Signal — R[Samples]) (— Music Track)

As discussed in Chapter 3, GDL tools can be likened to these forms of *alterity* relations, bearing in mind that the D[] operator will appear somewhere. For example, MusicVAE from Magenta, a DL-based toolkit for music generation developed by Google (Roberts et al., 2018), can predict possible continuations of a sequence of notes. The mediation provided by this tool can be described as:

- I → MusicVAE / (Model → D[InitialSequence]) (— Sequence continuation)

This study is set out to explore how the introduction of the D[] operator affects the creative process within the context of music composition.

## 4.2 Collaboration in practice

The opportunity for this study emerged from a conversation with Vicky Fung, a Hong Kong composer who has written the melodies for over 40 Canto-pop hits performed by various artists over the last 20 years (see Figure 4.1). Her contribution to the local scene is well-recognized within the community, and her passion for the industry was palpable from our first meeting. Like most creative ideas, this project began with a simple conversation about how our expertise could complement each other to explore the new creative landscapes offered by emerging technologies. Her motivation was rooted in the desire to preserve and maintain the Canto-pop genre, as she had sensed a gradual decline in interest over recent years. The conversation also revealed her longing to relive the early days of her career when her composition style had nuances she felt she could no longer replicate. My proposition to her was to explore the possibility of training a machine learning model that could capture her style and its evolution throughout her career. In theory, this would enable a form of *time-travel*, as we could ask the model to generate musical scores according to different periods of her life. In hindsight, this was an extremely ambitious goal, given the technological constraints and my limited knowledge of the field at the time. However, it proved to be a motivating objective that set in motion a fruitful collaboration.

Our first meeting on May 27th, 2020, marked the beginning of this journey, which culminated over a year later with the presentation of our work, "Co-creating Musical Compositions with an Artificial Agent: Time-travel through Machine Learning", at Artmachines 2. The conference was held at the School of Creative Media at CityU University of Hong Kong from June 10th to 14th, 2021. Due to the COVID-19 pandemic and other health-related issues, meetings were interrupted during the summer of 2020. However, we resumed our discussions online in October and November and met up again in December

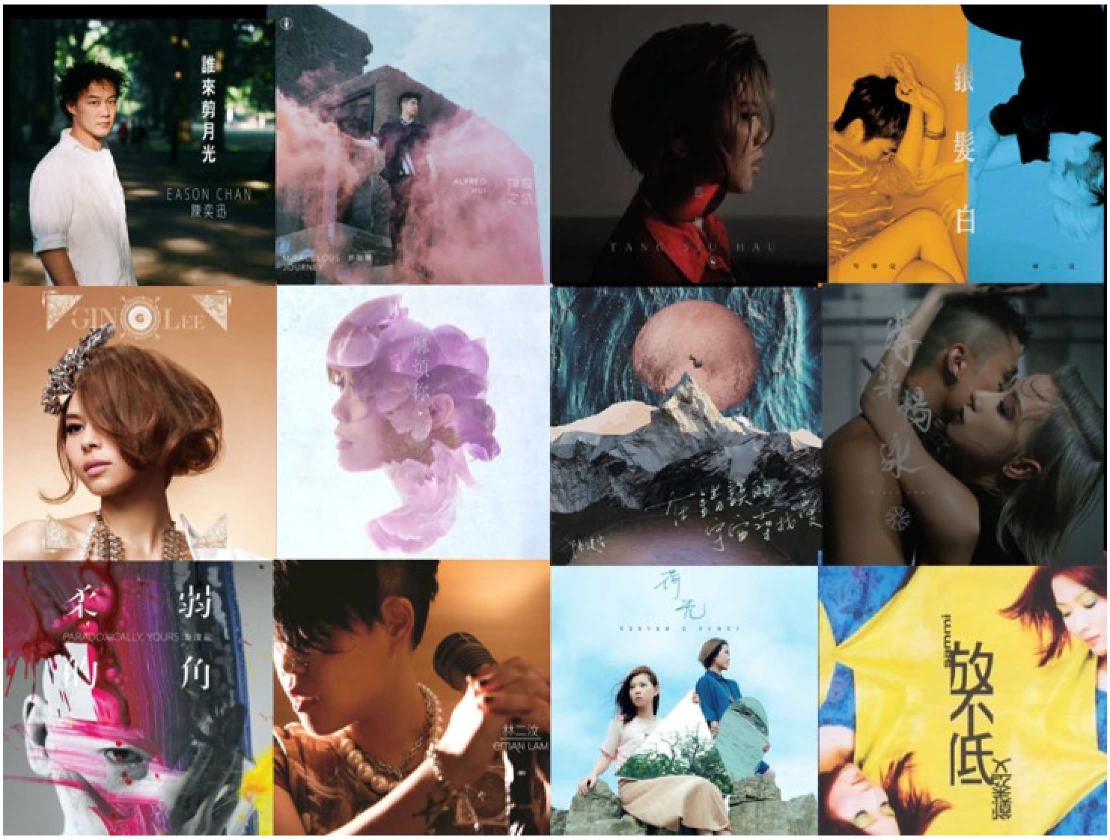


**Figure 4.1:** A compilation of album covers featuring Canto-pop melodies composed by Vicky. The images showcase the creativity and impact of Vicky's work within the Canto-pop genre. The figure highlights the success and influence of Vicky's contributions to the music industry.

and January to finalize our proposal for Artmachines 2. Upon acceptance of our submission, our collaboration intensified in 2021 during the months of April and May, when most of the development and output evaluation took place. During this period, we met regularly every two weeks and also communicated via online messaging about our progress. Due to access restrictions at PolyU campus, meetings had to be planned two days in advance, which was not always possible.

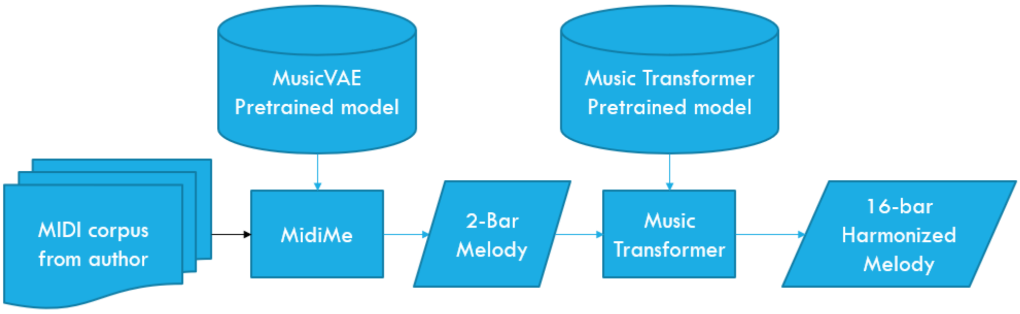


**Figure 4.2:** Schematic representation of the generation pipeline used in the first iteration. The diagram illustrates the various stages involved in the process, providing an overview of the workflow and the key components involved in the music generation process.

## 4.3 Method

Our collaboration can be divided into two phases, corresponding to two development iterations that experimented with different systems. Before we could train anything, our first task was to prepare a custom dataset with Vicky's melodies. This preliminary step proved to be rather time-consuming, as our attempts to automate MIDI melody extraction from audio files did not yield acceptable results. We ultimately created the MIDI files by hand after several failed attempts. Once we had a clean set, our first attempt utilized two models provided in the Magenta toolkit (C.-Z. A. Huang et al., 2018; Roberts et al., 2018), namely MusicVAE and Music Transformer. The MusicVAE 2-bar model was fine-tuned with 40 of Vicky's songs so that we could generate novel melodies in the same style. These 2-bar samples were then fed into Music Transformer for expansion and harmonization. The entire architecture can be visualized in Figure 4.2.

One significant limitation of this method was the restricted length of the VAE-generated melodies. This constraint was imposed by the VAE architecture in conjunction with the amount of GPU memory available to us at the time. MusicVAE is easy to train, but it is not very efficient at processing sequences. On the other hand, the transformer architecture, designed for language tasks, can be adapted for use with longer and more complex note sequences. After listening to several batches with Vicky, she was impressed by the quality of the generated samples but did not feel that the melodies produced reflected her style. She sensed a generic classical music feel characterizing the output, which was not representative of her work. Reflecting upon the causes of this behavior, it became clear that the influence of the training data used for Music Transformer was overshadowing the stylistic components extracted from the dataset we assembled. This realization prompted us to search for solutions that we could train entirely by ourselves, initiating our second development iteration.

As the transformer architecture gained popularity, several model implementations became available to the general public. We decided to try training our own model from scratch, testing out different code-bases. The most promising solution we identified at the time was X-transformers (a transformer implementation combining several techniques and optimizations documented at https://github.com/lucidrains/x-transformers[1]). However, during this iteration, we encountered another issue. Due to the limited number of songs in our dataset, the resulting output did not learn enough about general musical theory. For example, the generated output did not adhere to a single tonality, instead spanning all 12 notes rather freely. We surmised that this was a legitimate difficulty for the model since we provided only 40 melodies to work with. To mitigate this issue, we expanded the training set with an additional 200 melodies selected by Vicky. The choice of melodies to include was based on the artists and songs that were most influential in Vicky's musical upbringing. The resulting melodies were marginally better, yet still not exactly what we hoped to achieve.

[1]The training script we used can be found in this repository: https://github.com/asigalov61/Music-Transformers-Library

## 4.4 Results and discussion

In terms of ACASIA modules, this collaboration might be described as follows:

- **Association, combination and abstraction as D[Corpus] prediction**. In the first iteration, we trained our own MidiMe model (a VAE) to reproduce typical note patterns found in Vicky's corpus. Similarly, in the second iteration we have trained a transformer to make predictions about the next note in a sequence. In both cases training the model performs an abstraction, which enables the production of new associations and combinations during inference. In post-phenomenological notation our iterations would map to:
  - $MusicTransformer \to D_{pretrained}[(MidiMe \to D_{vicky}[Corpus])]$
  - $XTransformer \to D_{vicky}[Corpus]$
- **Selection**. Vicky, as the artistic soul of the project, was responsible for guiding the selection. However, we did not have control over the input selection of the pre-trained model.
- **Integration and adaptation**. These components were managed by humans, except for the integration between different steps of the pipeline, which was automated through a computer script ($R_1$):
  - $Vicky \to (I \to R_1[D_1, D_2, ..., D_n]) - AudioFiles$

For both of us, this exploration led to a deeper understanding of the potential benefits and limitations of using DL for music generation. Despite the difficulties we encountered and the limited quality of the output produced, our collaboration provided us with extremely valuable insights about the process of crafting our own $D_{vicky}[]$. These insights are summarized below.

- **Datasets are key**. When working with DL, most of the attention and time end up being directed towards matters related to the data we feed the algorithms during training. For pre-trained models with the scale required for adequate generalization, we cannot control this (as a popular DL mantra recites, “garbage in, garbage out”). The size of the dataset is also a significant factor in determining the quality of the output. Because our task is bound to a small dataset, we encountered difficulties in training a good model from scratch, as there seems to be no obvious way to augment the dataset without introducing unwanted elements. Recent models for audio generation provide text-based conditioning, which might constitute a solution to this issue (I will touch upon this again in Chapter 6 and 7).

- **Artistic identity may not be about information**. During our joint evaluation sessions, we discussed whether the generated output we were listening to could be considered *aligned* with her identity, as well as whether it could be deemed as *original*, since the output felt quite generic. Perhaps something is lost in the process of abstraction, which, for the sake of more efficient similarity, sacrifices details that matter for the selection process.

- **Evaluation of audio content is time-consuming**. Compared to the visual or language domain, evaluating music and more in general audio content is a much longer process. This is a trivial but important aspect of working with generative music. We can look at a picture and have an almost immediate reaction to it, but evaluating one minute of audio content takes at least one minute by definition. This generates a heavy load on the selection stage, which cannot be avoided.

- **Music theory not included**. When using $R[]$, rules of music can be defined. When using $D[]$ and the dataset is small, GDL may be able to capture some elements of the style, but incorporating musical theory and the geometrical symmetries of harmonic patterns in probabilistic models requires a much larger scale. A possible solution we did not explore is fine-tuning a pre-trained model.

Within the context of music, these observations emphasize the importance of rules and structure for efficient generation. Many of the issues we encountered working with DL are related to the inherent tonal structure of Western music, which is much more efficiently expressed through formal rules rather than inferred from data. The inherent symmetries of traditional 7-note scales used in the vast majority of pop songs could be easily described in terms of CT. In comparison, the Illiac Suite (see Section 2.2.3) created 55 years ago sounds much more pleasant than anything we could train from scratch using state-of-the-art transformers. While large-scale pre-trained models produce music consistent with music theory, their output tends to be merely *typical* and lacks novelty. Once again, DL, due to its inherent PT-inspired nature, reflects the limitations of the probabilistic approach by failing to capture the diversity within the large corpora they learn from.

In this study, we also confirmed the importance of compositionality in music. Despite the issues we encountered, the fact that transformers can generate long and complex musical sequences is testament to the architecture's ability to capture some level of compositionality that exists in music. The effectiveness of self-attention and scale might hint at the need to expand existing TOCs in this direction. While musical theory can be expressed by formal rules, it is not obvious how we develop these intuitions. This relates back to (Rey, 1993) (see Section 2.1.5), pointing out that humans seem to have the capability to at least grasp analyticity.

In conclusion, this collaboration led both of us to grapple with central questions about what it means to experience the creative process through an *alterity* mediation. It also exposed us to some of the practical limitations of working with the tools available at the time. Most importantly, it highlighted the difficulties that DL has when working with smaller datasets that practitioners create themselves.

*There can be no such thing as a naive, unconceived act of photographing. A photograph is an image of concepts.*

Vilém Flusser

# 5

# Study: Images

Image-making is arguably one of the primal forms of creative practice. The invention of cameras and the subsequent development of photography revolutionized the way we relate to visual content. Unlike human-drawn paintings, photographs could capture an instant of reality with remarkable speed and accuracy. This new form of image-making was a significant departure from traditional art forms that relied on subjective interpretation and representation. Through a camera, we can fix a moment in time, providing an illusion of objectivity (Batchen, 1997).

## 5.1 Scope

From a post-phenomenological perspective, the use of a camera introduces a certain degree of technological *opacity* (Eede, 2010) (see Section 3.1), which conceals elements of interaction that bring about a subjective bias. The intimate relationship that photographers develop with their favorite cameras, or the habits that they unconsciously form through daily use, are inevitably reflected in the output they produce. As Flusser put it, "Photographs are received as objects without value that everyone can produce and that everyone can do what they like with. In fact, however, we are manipulated by photographs and programmed to act in a ritual fashion in the service of a feedback mechanism for the benefit of cameras." (Flusser, 2000, p. 64) Furthermore, the social ecosystem around photography informs about the ways the technology can and should be used, subtly influencing the act of taking photos.

Vilém Flusser was a Czech-Brazilian philosopher and media theorist who is best known for his work on communication, technology, and culture. He wrote extensively about the impact of new media technologies on human perception and understanding of reality. According to Flusser (2000), a photograph is not simply an image captured by a camera, but rather the result of a complex set of programs that determine how the camera operates and how the resulting image is produced. These programs are embedded within what he calls the photographic *apparatus* - a system that includes not only cameras and film, but also printing processes, distribution networks, and cultural institutions that shape our understanding of photographic images.

*Relationality* is a key factor in Flusser's understanding of the photographic *apparatus*. He argues that the *apparatus* is not simply a technological system, but rather a social and cultural one that shapes our relationships with each other and with the world around us

(Flusser, 2000). Flusser suggests that the photographic *apparatus* creates a network of relationships between photographers, subjects, viewers, and cultural institutions. These relationships are shaped by the programs embedded within the apparatus - for example, by determining what kinds of images are considered valuable or meaningful. Once again we see elements of *Press* governing the selection module of the creative process.

Using the post-phenomenological framework, we can compare traditional image-making technologies to photography by formalizing them as different forms of mediation:

- (I — Brush) → Painting
- I → Camera / Apparatus (— Photo)

With the advent of digital cameras and mobile phones, the *apparatus* has grown to include social media and communication platforms, creating a vast ecosystem that could produce and distribute visual content at exponentially growing rates. The selection displayed through these channels is often the result of algorithms that suggest visual content based on what is most "relevant" to the user, based on previous data. Generative imaging tools, might expand the *apparatus* to a whole new level, redefining the conceptual boundaries of photographs and digital images.

Image generation with ML is a relatively recent field, which emerged from the progress made in computer vision (see Section 2.2.7). One of the first examples of DL tools for image generation is the Generative Adversarial Networks (GAN) architecture developed by Goodfellow et al. (2014). Similarly to the collaboration presented in Chapter 4, the purpose of this study is to understand how the process of image-making changes when we introduce $D[]$ in the mediating relation. We could describe the technological mediation offered by this type of technology in this form:

- **Generation:** $I \rightarrow GAN/(Model \rightarrow D_g[seed])(-Image)$

- **Training:** $I \rightarrow GAN_{training}/(R_{training}[D_d, D_g] \rightarrow DataSet)(-D_g)$

where $D_g$ is the generator component of a GAN that has been trained against a discriminator $D_d$, $seed$ is a variable intended as the random seed that guarantees a specific output image in a deterministic way, and $R_{training}$ is the training script. Building on the experience gained while working with DL tools in the previous study, our focus in this exploratory practice is shifted towards the act of building $D_g$ as context for self-reflection.

## 5.2 Collaboration in practice

The insight gained from the collaboration in the music domain had revealed the difficulties of working with *small datasets*. After sharing the insights from the first study with a fellow PhD researcher, we looked at what was the smallest amount of images that would constitute a viable dataset for image generation. We found that FastGAN (B. Liu et al., 2021) claimed it was able to produce coherent output using datasets of as little as 100 images. This discovery inspired us to lead investigation into dataset curation utilizing reflective practices.

Our intuition was that curating our own dataset could reveal our inherent biases in defining a coherent visual concept. In order to do so we identified a unique, locally-specific feature and compiled a dataset of 50 images that attempted to capture its essence. Our objective was to observe the output produced by FastGAN after training it with these photos and engage in a reflective exercise.

As theoretical framework of reference we adopted Schön's theory of reflection (Schön, 1983). According to Schön, reflection *on* action refers to the process of reflecting on past experiences after they have occurred. This type of reflection involves looking back at what happened, analyzing it, and drawing conclusions that can be applied to future situations.

Reflection *on* action is often used as a tool for learning from experience and improving performance over time. On the other hand, reflection *in* action refers to the process of reflecting while engaged in an activity or task. This type of reflection involves being aware of one's own thought processes and actions as they are happening, and making adjustments based on this awareness.

We performed two iterations of photo-taking, model-training and output evaluation. After the first iteration, we had a session dedicated to reflection *on* action. During this which set us in motion for the second iteration, with increased self-awareness about how and what we were choosing to include in the dataset.

## 5.3 Method

The first dataset was composed of images that my collaborator deemed to be representing of a specific type of storefront in Sham Shui Po, a local Hong Kong neighborhood (Figure 5.1). The images were then cropped to 1:1 ratio for the training process. We generated intermediate steps and checkpoints every 1000 learning iterations to check the progress. We sampled images from different checkpoints: 15,000, 25,000, 50,000 iterations (Figure 5.2). After joint evaluation of the output we identified that 25,000 iteration was the ideal value for our purposes. We noticed that checkpoints with more iterations were over-fitting the data and simply regenerated almost exactly a few training examples.

After this first iteration of training and evaluation we performed a reflection *on* action. We looked back at our process, paying attention to particular aspects of the photo-taking identified by our reflection *in* action and how they were linked the results. We noticed that FastGAN had picked up on the visual similarity across some the photos in the first set due to the specific type of framing that my colleague had used: the majority of the generated images of storefronts would have a darker element in the center, highlighting a pattern that


**Figure 5.1:** Photograph of a storefront in Sham Shui Po, chosen as the inspiration for the dataset used in the study. The image captures the unique aesthetic and atmosphere of the location. The choice of this particular storefront showcases the importance of selecting meaningful and contextually relevant stimuli in research. Originally published in M. Miller and Lion (2022).

was originally overlooked. Most shops photographed indeed portrayed a dark corridor at the center of the shot, which was a characteristic element in our set due to both the choice of shops and the type of framing that was habitual to the photographer.

By observing this recurrent characteristic of the generated images we were able to gain some insight about how our own biases were reflected in the selection of images for our dataset. Enriched by this reflective experience, we then proceeded to our second iteration. We first identified images that were not matching the typical emerging pattern and removed them from the set. For example, we took out a few images of closed shops which we identi-

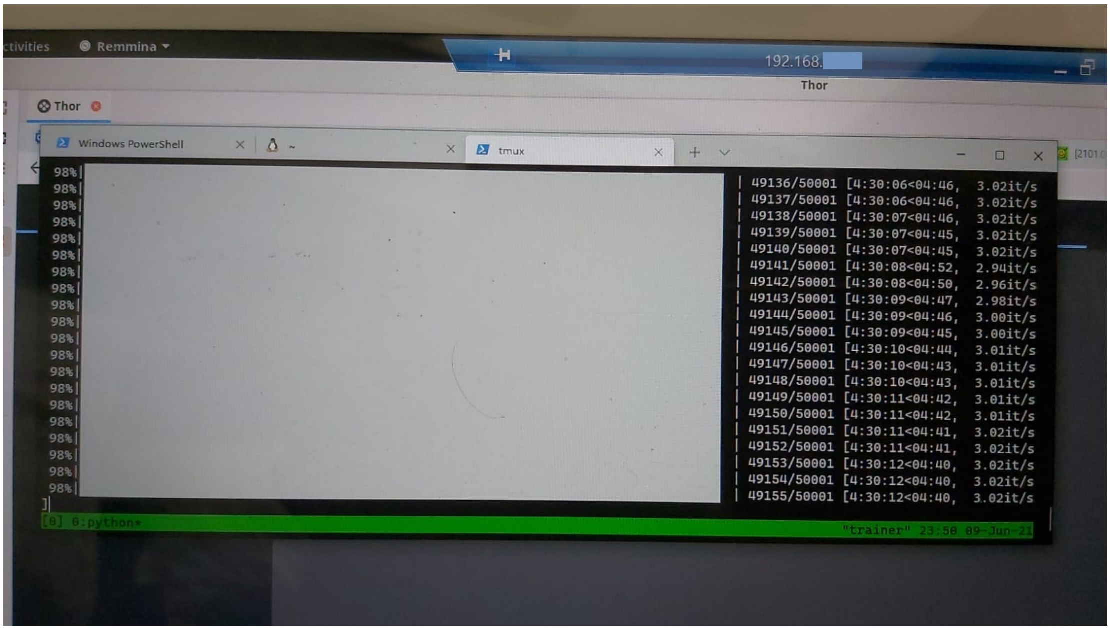


**Figure 5.2:** A photo of the training process for the model, showing the progress as it nears completion. The image documents our reflection *in* action as we are engaged in the act of training the model. Originally published in M. Miller and Lion (2022).

fied as not helpful for generation. Next, we integrated the original 50 images with another set that was taken after our reflective session. The set used for the second iteration was ultimately composed of 100 images selected among all the images taken during his two sessions.

After running the training script again, we evaluated the generated pictures after 25,000 iterations. We found that the images had more variety and they all contained visual elements typical of the shops we used as input. The central corridor, the many items hanging on display on both sides, cluttered shelves and white neon lights (see Figure 5.4).

In our final reflection *on* action, we focused on identifying which elements of our process affected the creation of the dataset and its coherence. In addition to what emerged from the first iteration in relation to the act of photo-taking, we recognized how our criteria of selecting the images that would go into the dataset had changed after we had experienced its output. What became clear to us is that the algorithm was primarily concerned with

**Figure 5.3:** A collection of sample images generated from the first iteration of the model, revealing the influence of photo-taking biases and framing habits on the training data. The majority of the generated images of storefronts exhibit a darker central element, reflecting a pattern that was initially overlooked in the dataset.

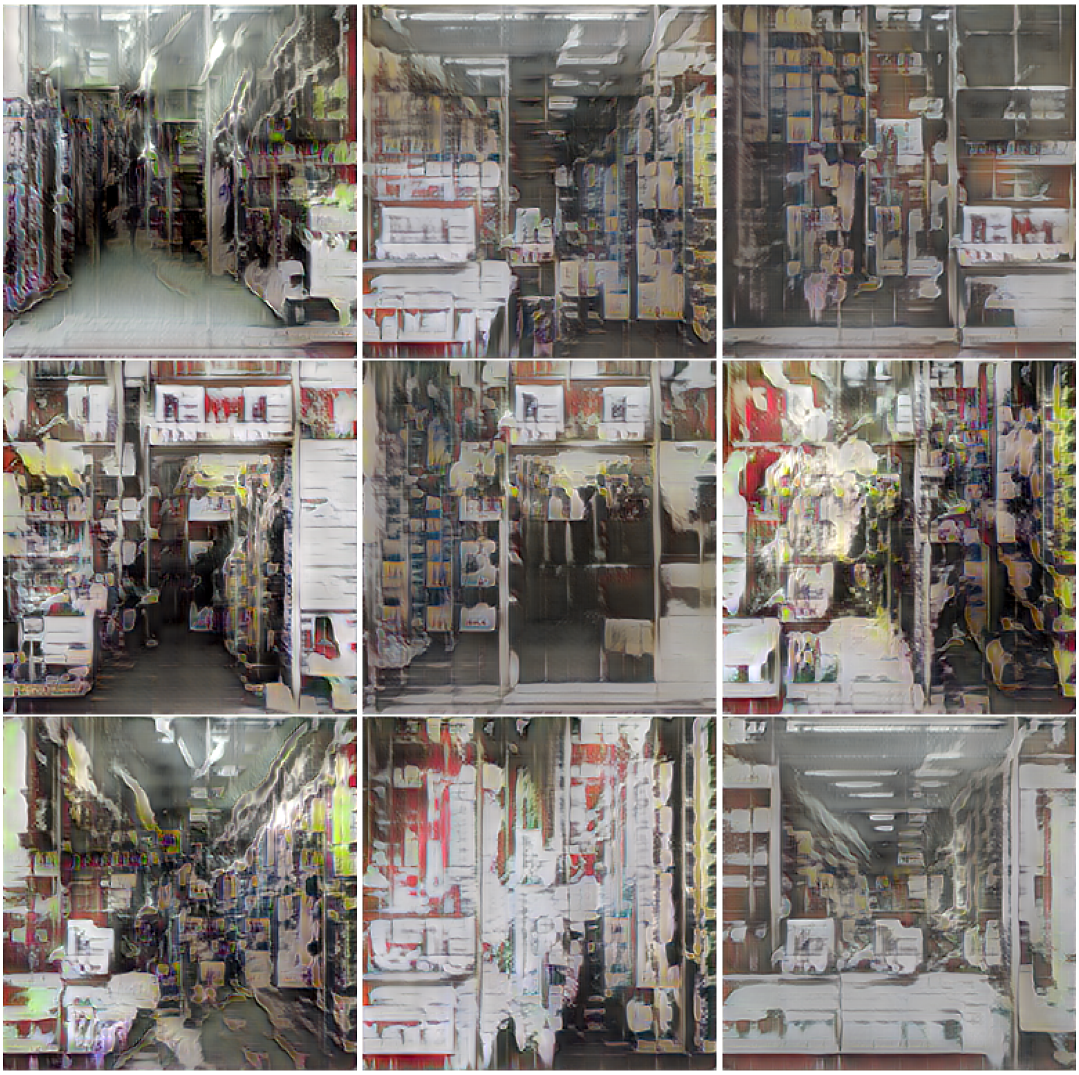

**Figure 5.4:** A collection of sample images generated from the second iteration of the model, after refining the dataset and addressing the identified biases. The images exhibit more variety while retaining visual elements typical of the input shops, such as the central corridor, items hanging on display, cluttered shelves, and white neon lights. The figure demonstrates the improvement in the model's performance in generating diverse and visually interesting content that reflects the characteristics of the input dataset.

pixel-based visual similarity rather than the more abstract conceptualization of similarity that we had in mind. This realization made us change the way we took and selected photos for our second iteration in order to figuratively "make the algorithm happy".

## 5.4 Results and discussion

In terms of ACASIA modules, this collaboration might be described as follows:

- **Association and selection as dataset curation**. In this study we took charge of directing the association making component of ACASIA by assembling a dataset according to our own selection of images, which was governed by our on our own definition of *similarity*.
- **Combination and abstraction through GAN training and generation**. The abstraction component is mediated through the GAN algorithm, which, once trained, performs the combination making step accordingly. Our reflection sessions uncovered the differences between our notion of *similarity* and the algorithm's one, which is based on quantifiable measures utilizing pixel information, rather than conceptual interpretation.
- **Integration and adaptation**. The integration and adaptation components of ACASIA did not come into play during this study. We observed the generated images as they came out from the generator without any additional step. Given the objective of this study was to understand more in depth the process of dataset curation, these components were not considered relevant to our goal.

As a result of our exploratory practice, we gained valuable insight regarding how our notion of *similarity* might include subjective elements which the algorithm did not pick up on. In fact, the loss function used by FastGAN deals with quantifiable measures purely based on perceptual qualities of pixels in the images, rather than on the conceptual elements which our mind habitually abstracts to. For this reason, the algorithm exposed our own bias towards what constitutes the *similarity* found in our dataset, a fact that became evident in our reflective sessions.

The insights gained through this study could be summarized as follows:

- **Dataset curation as reflective practice**. We have shown that the act of creating and curating a dataset can be conducive to fruitful self-reflection. By observing the generated output we were able to reflect on our own bias regarding the properties of a coherent set of photos representing a concept.

- **GAN similarity measures operate at pixel-level**. Our choice of tool, FastGAN, revealed that the primary distinction in the algorithm's image interpretation is its focus on pixel-level representation rather than a conceptual one. Although this has provided insights into our understanding of *similarity*, it also raises the question of whether this method aligns with how humans form associations based on *similarity*.

- **A balance between coherence and variations.** In our evaluation sessions we identified a very clear trade-off between the amount of training and the *novelty* of output produced. There is an optimal point during training that maximizes both coherence and originality of generated images. Before that point, the model tends to be incoherent, while further training beyond that point causes the generated images to look exactly like the ones in the dataset, due to over-fitting. It seems impossible to predict where this point is *a priori*, the only way that we could identify the ideal balance was through observation. It is possible that this tipping point is fully dependent on the subject chosen.

Overall, this collaboration proved productive, as it generated valuable insights into dataset curation within the context of small sample sets. Most GANs typically require a minimum of 10,000 images; however, assembling such datasets can be challenging for non-experts. FastGAN has enabled us to experiment with image generation on a smaller scale, which was previously unattainable. This opportunity has allowed us to delve deeply into understanding the concept of visual similarity from a human perspective and compare it with machine-based computations.

The results of this study also highlight the limitations of probabilistic approaches when dealing with concepts that are very local and specific. As pointed out for PT in Section 2.1.2, humans are indeed able to apply concepts even with few or no examples available. In this study, the similarity space is defined solely in terms of visual aspects within the concept and does not cover the extensive scope needed for comparisons with other concepts.

Furthermore, as in the previous study, we had no way to guide the generation towards a desired location in *latent space*. The technology simply does not provide a means to control the output, other than curating the training dataset. This mirrors the discussion in Section 2.1.2 regarding the challenges PT faces when dealing with compositionality. To understand an image and its representation, it may be necessary to break it down into smaller components and their relationships. The type of compositionality found in the visual domain is not entirely analytical (unless we use a plotter, for example), making it suitable for representation by a probabilistic model. GANs, however, only decompose images at pixel-level, rather than at a conceptual level, which makes it more difficult for humans to relate to. As the themes explored in the next study will demonstrate, language appears to be the ideal medium for connecting humans with visual concepts.

*The individual human subject simply did not exist anymore, once he or she had set the boundary conditions for the image to be computed.*

Frieder Nake

# 6 Study: Text-to-image

LARGE DIFFUSION MODELS ARE REDEFINING THE WAY WE CREATE, MODIFY AND UNDERSTAND IMAGES by providing an interface to visual content based on natural language. This study explores the usage patterns and creative interactions that field experts tend to have while using text-to-image technology. The study was conducted using a Discord server over the course of 4 months, gathering 76 participants ranging from different related fields such as graphic design, photography and digital illustration. The participants were presented with a text-to-image bot and given a private space to use the technology for their own purpose. Qualitative data, collected through interviews and meetings, was complemented by quantitative measures extracted from server-gathered interaction data, including usage, prompts used, and types of operations (text-to-image, image-to-image, prompt interpolation). The study results reveal that the distribution of images generated per user (usage) follows an exponential pattern, and there is a negative correlation between expectations of text-image alignment and the Tolerance for Ambiguity (TA) personality trait (Furnham & Marks, 2013; Herman et al., 2010; Norton, 1975; Zenasni et al., 2008). These findings constitute some preliminary evidence suggesting that the adoption pattern of this new technology is dependent on social factors. Additionally, the findings suggest that expectations regarding this technology could be linked to TA, potentially resulting from an individual's attunement to a specific TOC.

## 6.1 Scope

Text-to-image technology made a substantial impact in the field of computer vision since its inception. On Jan 6, 2021, OpenAI released CLIP, a model that was trained on 400 million image-text pairs, able to create a joint *latent space* combining visual and textual information (Radford et al., 2021). CLIP works in both directions, in the sense that it is not only capable of identifying objects in a picture, but also to guide the generation of images starting from text. Since the public release of CLIP, many open-source implementations of image generators based on its capabilities have been released, and their generated output has started to flood the internet.

CLIP and, more generally, the idea that we can control a generative model by conditioning its output on text (or, more precisely, token embeddings) constitutes a dramatic interface improvement. As it should be evident from all the TOCs reviewed in Section 2.1, language seems to play central role when dealing with concepts. CLIP effectively provides a *latent space* that unifies *lexical* and *visual* concepts. This multi-modality binds the compositional elements of language to the visual domain, allowing the generation of novel content obtained through the sophisticated combination making capabilities afforded by the pre-trained token embeddings.

Naturally, the limitations that neural networks generally encounter when dealing with analytical concepts are still in place. For example, study participants quickly find out that quantifiers have little effect on image generation: the prompt "two circles" yields a random number of circles (see Figure 6.1). This should not be surprising, given the nature of the $D[]$ operator, yet the technological *opacity* might conceal the reasons of this behavior. Our habitual way of thinking about intelligent systems perhaps creates the expectation that they must be analytical, yet in this case we are dealing with something completely different.

Referring back to the birth of *generative art* and the experiments with programmable plotters, it seems that $R[Program]$ and $D[Language]$ are very different, yet complementary, ways to manipulate visual concepts. Ultimately, both methods make use of randomness to generate variety, but in the former probabilities are defined by the programmer, whereas in the latter they are determined during training. For the sake of completeness, the post-phenomenological form of these two mediated interactions might be summarized as follows (from Section 3.1):

- I → Plotter / (PlotterSoftware → **R[Design]**) (— Printed Design)
- I → Stable Diffusion / (CLIP → **D[Image Description]**) (— Generated Image)

**Figure 6.1:** A collection of images generated using Stable Diffusion v1.5 and Automatic1111 UI, based on the prompt "two circles" (Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 3828270218, Size: 512x512, and Model hash: cc6cb27103). Despite the prompt, the images display a random number of circles, indicating that the image generation model does not accurately capture cardinality. The figure reveals the limitations of LDMs in generating content that strictly adheres to prompts that contain analytic concepts such as numbers, highlighting the need for further improvements in understanding and representing cardinality in image generation models.

After the release of CLIP, the ML community began exploring its capabilities in various directions. Ryan Murdock combined CLIP with BigGAN (Brock et al., 2018) to create BigSleep[1], which was one of the first publicly available tools in this space. Katherine Crowson experimented with CLIP guidance in a novel method for image synthesis known as diffusion, as proposed by Dhariwal and Nichol (2021). As 2021 drew to a close, numerous forums and user groups devoted to generative art began discussing image generation and their workflows as Python scripts that could be run on Google's Colab, a platform that provides free (or nearly free) GPU computing time. Many of these Colab scripts utilized Crowson's work, with Disco Diffusion[2] proving particularly popular among the pioneers. As the community continued to experiment with CLIP and its potential applications, its impact on the field of ML and beyond remained a topic of ongoing interest and discussion.

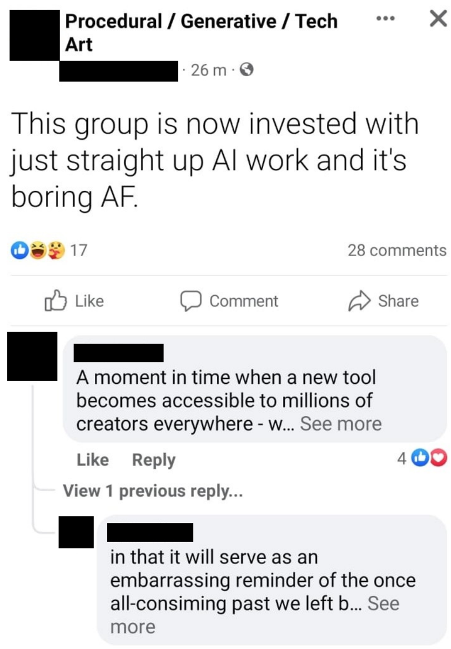


**Figure 6.2:** This is a screenshot of a post from March 19th, 2022, in which a user expresses their frustration about the amounts of AI content posted in the group (https://www.facebook.com/groups/procgenart). This post was eventually removed from the group by moderators.

[1]Source code available at: https://github.com/lucidrains/big-sleep
[2]Source code available at: https://github.com/alembics/disco-diffusion

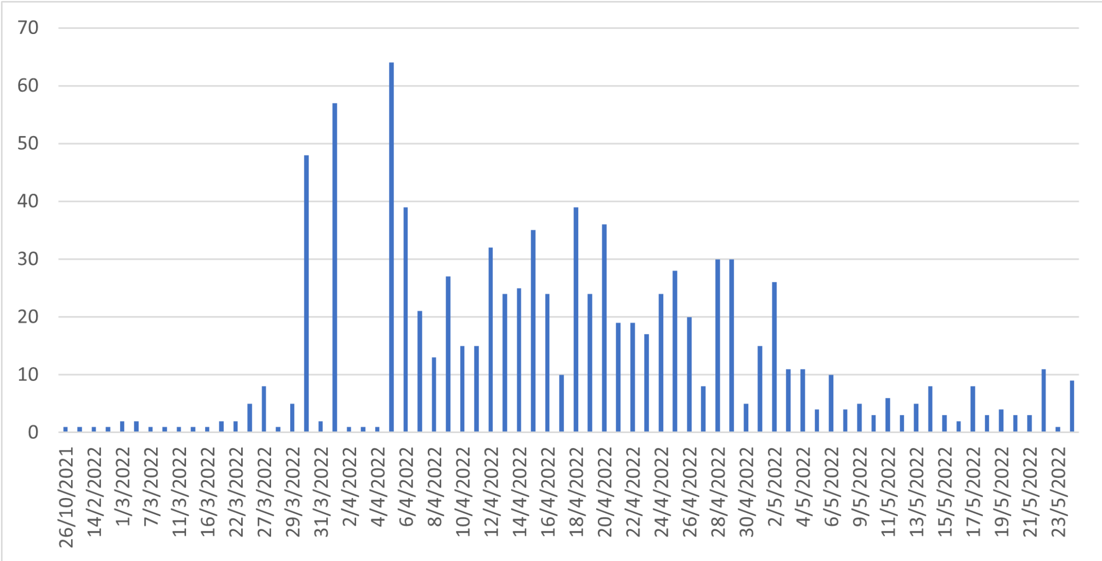


**Figure 6.3:** Bar chart illustrating the number of posts per day in the "Procedural / Generative / Tech Art" Facebook group (https://www.facebook.com/groups/procgenart), with a noticeable explosion of posts attributed to the trend of AI-generated images. The chart highlights the increase in group activity related to AI-generated images in April and May 2022, followed by an abrupt decline as moderators took countermeasures to direct this type of content to a dedicated group named "AI Generated Art". The figure demonstrates the impact of AI-generated images on the group's activity and the moderation challenges faced by the online community.

As a member of the Facebook group "Procedural / Generative / Tech Art", I witnessed first hand the wave of controversial *AI Art* that took over the group in the months of April and May 2022. Around that time, images generated with Disco Diffusion were pretty much the only content being posted in that group, to the point that some members started to complain about it vigorously (e.g. Figure 6.2). Official statistics for the group are not available, but I was able to scrape the public posts and the trend is evident from the collected data, presented in Figure 6.3. The moderators of the group implemented a new strategy to handle the overwhelming volume of posts related to AI-generated images. They created a new dedicated group called "AI Generated Art" and stopped publishing these types of images in the original group. As a result, there was a noticeable decrease in the number of posts around May, but this did not indicate a decline in interest. Rather, it was a deliberate effort to reduce the workload of the page admins who were struggling to keep up with the moderation demands.

### 6.3 Stable Diffusion

On August 22nd 2022, Stability.ai released to the public Stable Diffusion (SD) v1.4 (Rombach et al., 2021), a LDM that was trained on LAION-5B, a "dataset of 5,85 billion CLIP-filtered image-text pairs" (Schuhmann et al., 2022, p. 1). While OpenAI, Google, Meta, and Microsoft have all made public releases of their research and development efforts, it is rare to see them share pre-trained models with the public. These companies defend their position by citing the potential dangers of unmanaged use, and at the same time safeguard their investments from the competition. Training models at large scale is undoubtedly a big expense that is hardly justifiable if it does not produce returns. Stability.ai, however, made the bold move of releasing their model to the public allowing anyone to use it without any cost. Since its release, SD has taken the creative industry by storm.

SD outperforms the generation speed of other openly-available models by several orders of magnitude because it conducts the diffusion process in *latent space*, unlike previous solutions, such as Disco Diffusion, that operate at pixel-level. It is able to do so because it uses a VAE to encode and decode images and token embeddings to and from *latent space*[3]. This significant speed increase, coupled with the large-scale dataset it was trained on, make SD one of the most exciting technologies that we have seen in this field[4]. Indeed, it is remarkable that such a small file (less than 4GB) can encode the information of almost 6 billion images that humans have shared on the internet.

The opportunities that this technology has opened up for the general public are im-

[3]For an image of resolution 512x512, the pixel space is $3 \times 512 \times 512$, which is then encoded/decoded by the VAE to a $4 \times 64 \times 64$ *latent space* where the diffusion process takes place more efficiently.

[4]At the time of this writing.

mense. Many developers around the world have spontaneously developed user interfaces, extensions, performance improvements and also trained alternative models that can be used to generate specific subjects or in specific styles, all in less than a year since its public release.

In September 2022, it became possible to observe this technology in action in a study that could accommodate a large group of participants simultaneously. Unfortunately, the popular interfaces of SD (official webui[5], automatic1111 webui[6], invokeAI[7]) are not designed for group use. Projects in this space often use Discord as a platform to let users try the technology in a social context. For example, Midjourney, an image generation service based on SD, is a tool that existed only as Discord bot up until very recently (at the time of this writing, Midjourney's API access was just announced). Discord's interface enables programmatic interaction through *bots*, which can respond to predefined commands and execute arbitrary code. As SD was released, many open source solutions featuring image generation pipelines via Discord became available on GitHub[8] and with all these components ready, a study observing interactions with SD became possible at a larger scale.

## 6.4 Method

A Discord server named **MediaBots** was opened on September 2nd 2022 to host the study. Participants were selected among design visual arts practitioners, photographers and design students. They were invited to a 2-hour workshop, during which they had the opportunity to use text-to-image technology. During the study, 76 people from 5 different workshops accessed the server. The workshop and participant details are summarized in Table 6.1.

[5] Project page: https://beta.dreamstudio.ai/generate

[6] Project page: https://github.com/AUTOMATIC1111/stable-diffusion-webui

[7] Project page: https://github.com/invoke-ai/InvokeAI

[8] GitHub (https://github.com) is a web-based platform that provides version control and collaboration services using Git, a distributed version control system. It allows developers to create, manage, and collaborate on repositories (projects) that contain source code, documentation, and other files.

| Date | Participants | Expertise | Location |
|---|---|---|---|
| Sep 29th 2023 | 35 | Design (MSc) | HK PolyU, School of Design |
| Oct 28th 2023 | 10 | Photographers | Online |
| Nov 11th 2023 | 15 | Design (BA) | HK PolyU, School of Design |
| Nov 25th, 2023 | 7 | Informatics (MSc) | Masaryk University |
| Jan 8th, 2023 | 9 | Photographers | Current Plans Art Gallery (HK) |

**Table 6.1:** Overview of workshops during the study, including the date, number of participants, their expertise, and the location of each workshop. The workshops were conducted at various locations, including universities, art galleries, as well as online. Participants had diverse expertise, ranging from bachelor design students to professional photographers.

The server was setup so that participants would be required to join the study in order to see the content on the server. To join the study they were asked to confirm they have read the information sheet and then complete a questionnaire[9] measuring:

- General prior knowledge about text-to-image technology
- *Expectations of Human Compatibility* (EHC) when interpreting keywords
- *Expectation of Alignment* (EA) between images and text
- *Tolerance for Ambiguity* (TA) scale (Furnham & Marks, 2013; Herman et al., 2010; Norton, 1975)

The purpose of including these measures was to explore whether any of these factors was correlated with the usage patterns observed in the Discord server. In particular the TA scale is hypothesized to be a proxy for a participant's affinity towards a non-analytical approach. The intuition is that ambiguity is more characteristic of $D[]$ than $R[]$ so this personality trait might have some influence in the way the tools are used and perceived.

[9] See Appendix B for the full list of items in the questionnaire.

Once the participants completed the questionnaire, they were shown how to generate images through the bot's *slashcommands*, essentially special messages that begin with a / that are recognized by the bot as structured instructions.

The SD bot was named "Botticello", a wordplay in homage of Sandro Botticelli which translates from Italian to roughly "cute little bot" (much like Dall-E[10] is an homage to Salvador Dalí and the Pixar animation character Wall-E). Under the hood, Botticello is using `yasd-discord-bot`[11] and `dalle-flow`[12] to interpret the commands and handle the generation queue. The choice of this architecture was made because the `jina`[13] interface offered by the back-end supports queuing, making it possible for several users to use the same bot simultaneously, a crucial feature in a workshop setting.

The commands offered by Botticello are:

- `/image`: This is the first command presented to participants and, in its simplest form, it enables to generate an image based on a text prompt as shown in Figure 6.4. The user can also adjust other generation parameters, such as the image resolution, the strength of the text conditioning and number of iterations. Once the command is sent, the generated images are returned by the bot as shown in Figure 6.5.

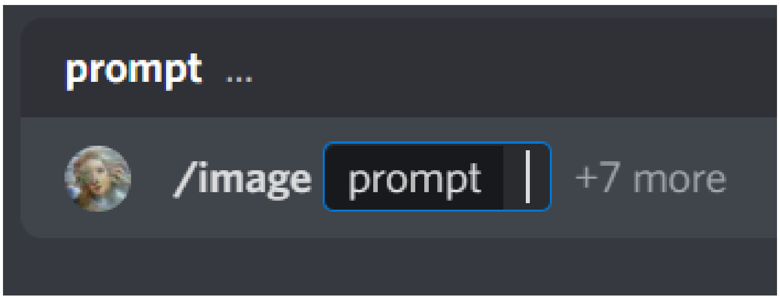


**Figure 6.4:** An example of the `/image` *slashcommand* used in the interface.

[10]DALL-E is a machine learning-based image generation system created by OpenAI.
[11]Source code available at: https://github.com/AmericanPresidentJimmyCarter/yasd-discord-bot
[12]Source code available at: https://github.com/jina-ai/dalle-flow
[13]Source code available at: https://github.com/jina-ai/jina

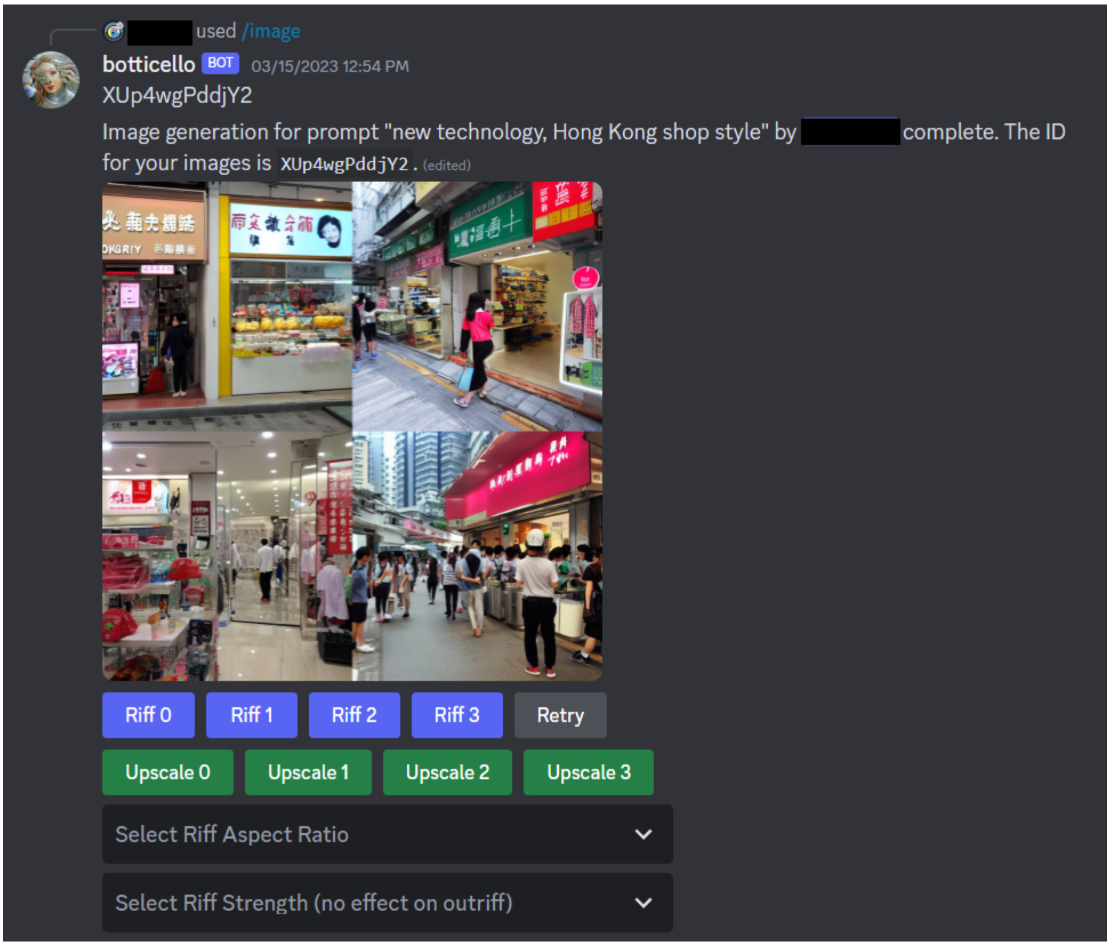


**Figure 6.5:** An example of the output generated by the `/image` command.

- `/riff`: This command is dedicated to image-to-image operations, that is, image generations that take an image as starting point. It can be used with or without a text prompt to generate variations from an image. For example, Figure 6.6 is used as starting point for `/riff` in Figure 6.7. By adjusting the *strength* parameter it is possible to control how much of the original image will be included in the output. While the `/riff` command is available as a button below every output, it is also possible to run `/riff` on any image uploaded into the channel by referring to its assigned identifier.


**Figure 6.6:** An example of the initial image uploaded by the user to be used with `/riff`.


**Figure 6.7:** An example of the output of a `/riff` *slashcommand* referencing the initial image in Figure 6.6.

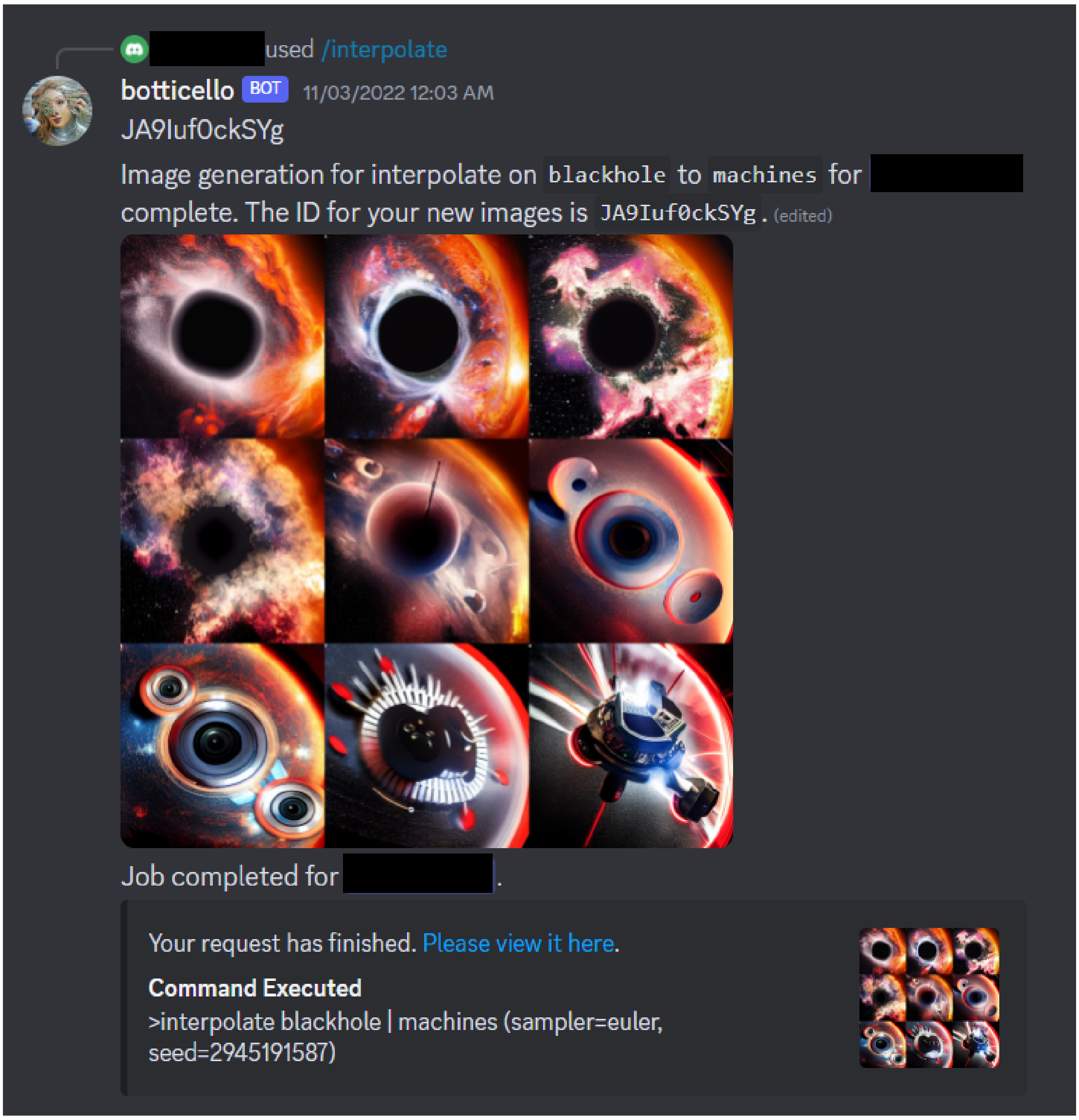


**Figure 6.8:** An example of the `/interpolate` *slashcommand* interpolating between *blackhole* and *machine*.

- `/interpolate`: This command can generate images that progressively shift between one prompt and another as demonstrated in Figure 6.8. This method can be used to experiment with visual conceptual blending as it provides combinations that span the whole range of ratios between two concepts.

## 6.5 Results and discussion

The analysis relative to the data collected through the questionnaires during workshops and tool usage data logged by the bot is presented as follows:

1. For general knowledge, intuition and expectation items, distribution bar charts are presented in Figure 6.9. As both *understanding* and *intuition* are skewed towards higher values, it appears that most participants have an idea of how this technology works or at least they can intuitively grasp it.

2. TA scale shows good reliability when all items were included ($N = 10$, Cronbach's alpha = 0.797). Constructed scale follows normal distribution (see Q-Q plot in Figure 6.10).

3. Number of images generated per user (or *tool usage*) is not normally distributed. Q-Q plot and Kolmogorof-Smirnoff confirm exponential distribution (see 6.11). Two outlier cases visible in the charts have been removed with a filter on number of generated images with condition $<= 400$.

Compared to earlier studies (Furnham & Marks, 2013; Herman et al., 2010), the TA scale was found to be slightly less reliable, which may be partly attributed to the impact of the COVID-19 pandemic. For instance, responses to certain items such as "I would like to live in a foreign country for a while" and "I like parties where the attendees are strangers" may have deviated from the rest of the scale due to travel and gathering restrictions that have been in place in recent years.

The two outliers identified also reveal an interesting story. After an informal conversation with one of them, they explained the situation. They are a couple who run an Insta-

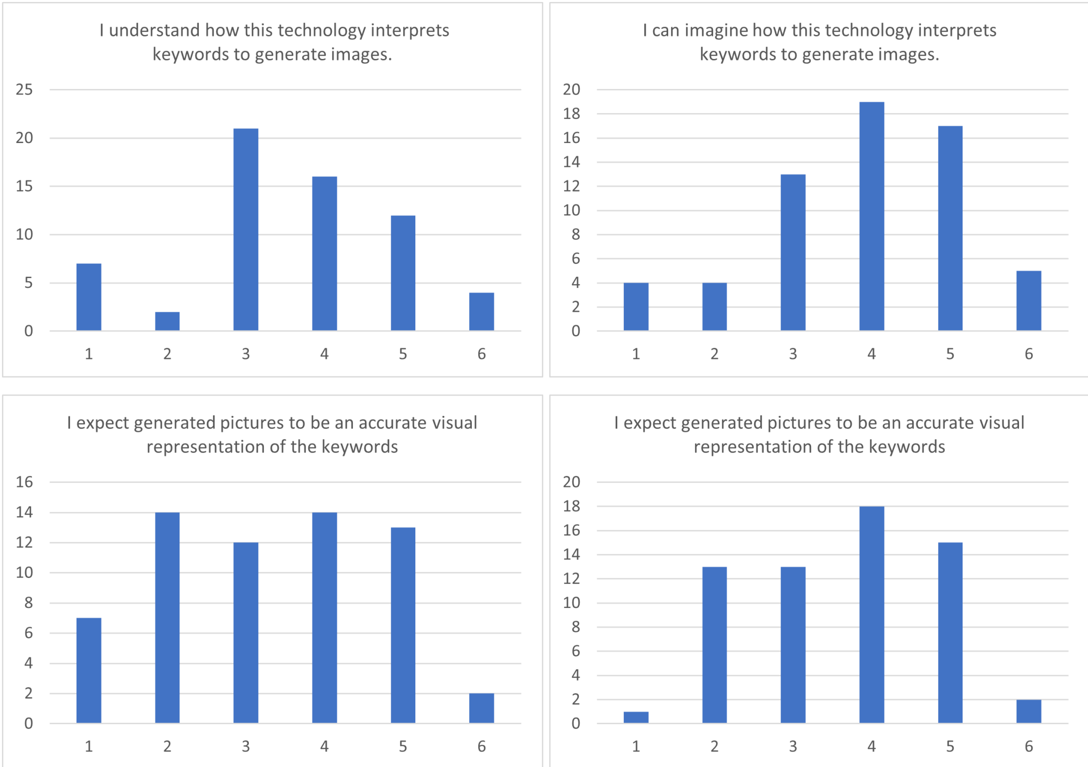


**Figure 6.9:** In order from top to bottom, left to right: distribution of understanding, intuition, expectation of human compatibility, and expectation of alignment.

gram page and the reason they had generated so many images was because they were posting content on social media and generating images on behalf of their friends.

The discovery of the exponential distribution in user usage suggests that a small number of individuals are responsible for the majority of the interaction. Despite the diverse range of participants, it is noteworthy that the Q-Q plot appears as a near-perfect straight line, as depicted on the right side of Figure 6.11. The exact cause of this phenomenon is not entirely clear. However, one possible explanation, based on the observation of two outliers, is that network effects play a crucial role in predicting tool usage. The relationship between the *Person* and *Press* elements could provide added incentive for individuals who are well-connected in a social network. It is worth noting that there is no significant correlation between usage and any other measured variables in the data collected, indicating that external factors may be influencing participants' use of the bot.

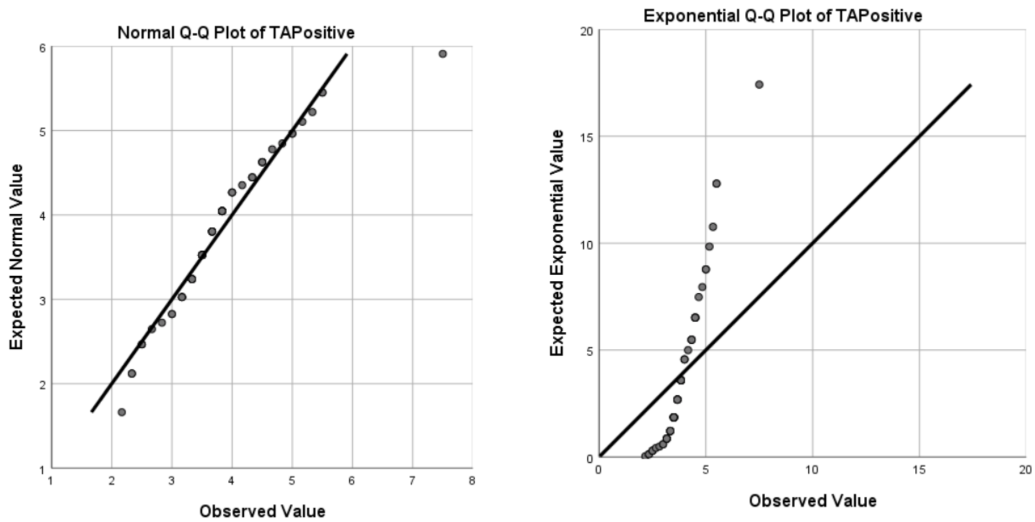


**Figure 6.10:** Q-Q plots for tolerance for ambiguity. Left: normal distribution. Right (values adjusted to be in a positive range after recoding): exponential distribution.

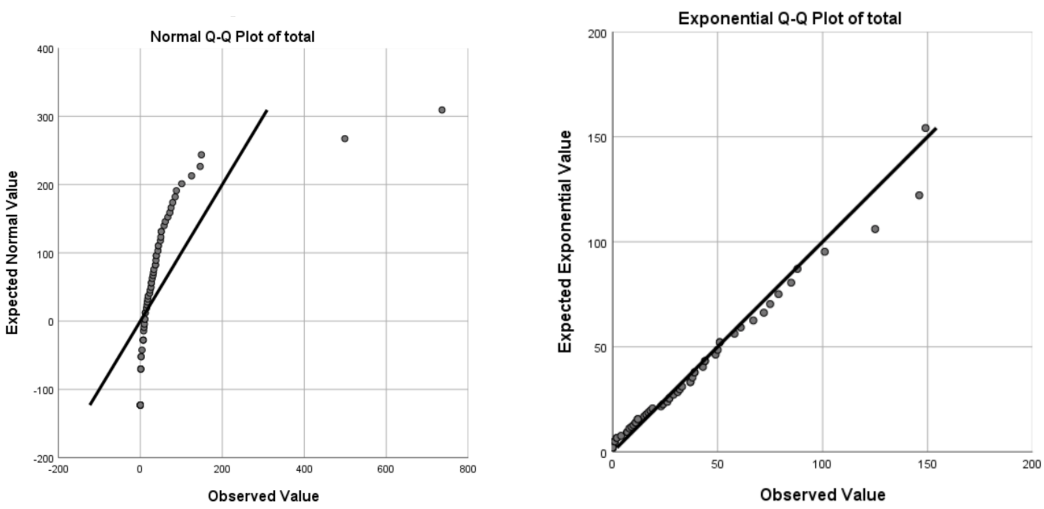


**Figure 6.11:** Q-Q plots for total number of generated images per user. Left: normal distribution. Right: exponential distribution.

The data reveals a significant correlation between the *Expectation of Alignment* item, "I expect generated pictures to be an accurate visual representation of the keywords", and the TA scale variable. Additionally, the *Expectation of Human Compatibility* item, "I expect this technology to interpret the keywords in a similar way as humans do", exhibits a significant correlation with EA, but not with TA. This indicates that TA and EHC may exert independent effects on EA, a relationship that warrants further investigation. To evaluate this interaction, a linear regression analysis was conducted using SPSS, with the findings illustrated in Figure 6.12.

**Coefficients**[a]

| Model | | Unstandardized Coefficients | | Standardized Coefficients | | |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | t | Sig. |
| 1 | (Constant) | 2.479 | .343 | | 7.228 | .000 |
| | TA | -.368 | .149 | -.280 | -2.468 | .017 |
| | EHC | .328 | .097 | .384 | 3.382 | .001 |

a. Dependent Variable: EA

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .513[a] | .263 | .238 | 1.051 |

a. Predictors: (Constant), EHC, TA

**Figure 6.12:** Results for linear regression on *Expectation of Alignment* with *Tolerance of Ambiguity* and *Expectation of Human Compatibility* as independent variables.

The negative coefficient attributed to TA is particularly noteworthy concerning the potential impact of this personality trait on the perception of text-to-image technology. A plausible interpretation of this result is that individuals with high TA may exhibit lower expectations regarding image alignment, as they already anticipate and accept the inherent ambiguity present in both language and images. Conversely, those who display intolerance to uncertainty are more likely to expect a clear alignment between language and images, projecting this expectation onto the technology. Nonetheless, none of the three variables exhibit a significant correlation with $log(usage)$, implying that these expectations do not ultimately influence the inclination to use the technology.

In conclusion, we may summarize the findings according to the ACASIA modules as follows:

- **Association and combination**. Both humans and the model play important roles in these components. As discussed in Chapter 4, pre-trained models build a *latent space* during training, while humans interpret the generated output, which can lead to unexpected associations. Conditional generation based on token embeddings enables the integration of language compositionality with the visual domain. This integration allows the user to visualize new combinations that can potentially stimulate further associations. When this process is iterated, it leads to a virtuous creative cycle that is incredibly fast, thanks to the efficiency of *latent space* diffusion, creating a feedback loop that leads to constant improvement and increasingly complex outputs.

- **Abstraction**. As highlighted before, SD is an outstanding example of compression of information, which arguably implies abstraction. Furthermore, a language interface provides fertile ground for human abstraction. Once again, the interplay between human and non-human joint efforts in this component leads to a virtuous cycle. Yet, the abstraction afforded by $D[]$ is not of the analytical kind, which may mislead users that have expectations of analyticity.

- **Selection**. This component is primarily human-led. Ultimately, the users make the decisions of what images to keep or share with others. Of course, the model is making some form of selection beforehand, on behalf of humans, and arguably SD is more efficient than the previous models at selecting pixels combinations that are visually relevant to the token embeddings.

- **Integration and adaptation**. Some aspects of these components are handled by the rule-based interaction with the Discord bot. Botticello affords a set of parameters that allow for creating small variations and minor adjustments. Ultimately, the final steps of integration and adaptation into finish product still need to be made by humans using traditional tools for graphic design and photo editing, such as Adobe Photoshop.

## 6.6 Limitations and further research

Overall, this study's limitations are related primarily to the lack of a structured methodology to address this novel type of interaction. In fact, the concept of image *alignment* with prompts is unique to this new form of generation and calls for further investigation. *Alignment* in this context has both objective and subjective components and its operationalization is dependent on the TOC that a person might be attuned with. In a creative context, *Alignment* perhaps is not critical or even desirable. A user could be looking for novel *misaligned* associations and combinations. In this sense, the non-analytic nature of $D[]$ can be considered an asset, echoing the hypothesis that $D[]$ is an inherently creative operator, as Hoorn (2023) suggests.

Moreover, the initial evidence discovered in this study suggesting the hypothesis of a socially driven adoption of this tool, remains inconclusive. The absence of correlation between number of images generated per user (tool usage) and any of the measured variables appears to rule out factors related to personal aspects, such as technology acceptance (TA) and expectations (EHC and EA), as drivers of use. However, the qualitative evidence emerged from this study in support of *Press*-driven adoption is not definitive and can only indicate a direction for further research.

In conclusion, the relationship between EA, TA and EHC constitutes an unexpected but relevant finding of this research. It points to a possible link between personality traits such as TA and expectations about the behavior of text-to-image technology. Further research is needed to investigate whether this connection is generalizable to all technologies adopting $D[]$. This finding also suggest that there might be individual differences in how we expect concepts to behave, implying that humans develop an unconscious preference for a TOC, which they then tend to ascribe to the technology they use.

*[Design] is never a process that begins from scratch: to design is always to redesign. There is always something that exists first as a given, as an issue, as a problem.*

Bruno Latour

# 7

# Discussion

Human—Technology relations are ubiquitous in creative endeavors. The studies presented offer evidence as to why we should adopt a post-phenomenological lens when examining human and non-human creativity in the context of art and design. The post-phenomenological perspective provides a way of looking at computational creativity that does not separate humans and society from the technology that shapes them, allowing us to form a more comprehensive and trans-disciplinary understanding of the phenomena we observe in this context. This is particularly important in today's world where the boundaries between the human and the non-human are becoming increasingly blurred. In particular, it was highlighted how the trending data-driven approach might bring about new opportunities and challenges for creative practices as it affects the way output is consumed and evaluated by the audience. This chapter further discusses the key insights that emerge from the studies in relation to TOCs and creativity theories, with the overarching goal of providing an outlook over what the future of computational creativity might evolve into, if this trend continues.

## 7.1 Datasets curation

Datasets are the central aspect of $D[]$ contextualizations. Computational systems that *learn*, must inevitably learn from something. In small-scale scenarios, such as those described in Chapters 4 and 5, the need for large amounts of training data in most implementations makes it extremely difficult for practitioners to effectively use data-driven tools without relying on external data sources. While on one hand we can blame the inherent limitations of DL for this, on the other we must take into account that creative practices do not exist in a vacuum. Artists and designers often have a degree of visibility into other methods and products, previous and contemporary, which might consciously or unconsciously affect them. Boden's (2003) account of *combinational* and *exploratory* creativity aptly describe this process of repurposing and representing old elements in a new way. Bruno Latour also points out the inherent *remedial* nature of design: "To design is never to create *ex nihilo*. [...] The most intelligent designers never start from a *tabula rasa*. [...] Designing is the antidote to founding, colonizing, establishing, or breaking with the past. It is an antidote to hubris and to the search for absolute certainty, absolute beginnings, and radical departures" (Latour, 2008, p. 5). What artists and designers do in their practice is perhaps better understood not as *creation*, but as *selection* of the old, *abstraction* and *combination*, to put in ACASIA terms. The emphasis here is on the *selection* of something as input which is then decomposed and reconstructed anew.

Putting together a large-scale dataset can be a daunting task, let alone having the resources to train a model at the scale we are witnessing in the recent years. It is remarkable that, as the scale increases, the prediction accuracy does not *plateau*, making tools like GPT, Dall-E and SD possible, but this race for scale in the long run might also have detrimental effects in terms of market concentration and accessibility. The impact of this is already evident from the literature review of CC, showing a decreasing amount of novel systems being developed. First hand experience also confirms this difficulty, as presented in the first two studies.

The response of the open-source community to this problem has perhaps shown a possible solution to this issue, both in practical and theoretical terms. For example, textual-inversion (Gal et al., 2022) provides a way to add token-embeddings and hyper-networks based on as little as 3-4 images. These custom-trained concepts can then be combined with other tokens to generate virtually infinite visual "possibilities". Dreambooth (Ruiz et al., 2022) provides a way to also fine-tune SD with custom visual content, yet this results in an entirely new model, which is inconvenient for sharing and makes merging with other models difficult. Low-rank adaptation of LLMs, or LoRA (Hu et al., 2021), was applied to SD and released in February 2023. LoRA has solved the issue of having to re-release the whole model when fine-tuning, while also providing a way for users to extend SD concept library with reduced memory requirements allowing training on consumer hardware. Alongside fine-tuned models, LoRAs are then easily published as relatively small files on online sharing platforms like Civitai (https://civitai.com).

LoRAs can also be combined with one another, thus enabling interoperability across user-defined visual concepts. This opens up the possibility for community powered compositionality of user-trained models. In a sense, such configuration is not unlike what Jackendoff (1989) and his neo-classical theory would predict: a set of core concepts (SD base model) and a series of semantic-fields (user trained models) which adjust and complete the base model incorporating examples that are more personal and specific (see Section 2.1.4).

In this context, I bring forward a consideration about what the shift of attention towards $D[]$ might bring for artists and designers. It may very well be that, in the future, practitioners will choose to release their particular graphic style, interior design, or product aesthetic as a generative model, rather than as a set of fixed outputs such as images, soundtracks, text or 3D models. Much like the generative art movement in the 1960s was not about the content, this wave of user-generated text-to-image models might take the same path and shift the attention towards the *Process* perspective and away from a *Product*-centric view, because content is becoming way too abundant to be of any value.

In a not so distant future, content producers of today might be required to evolve their skills to better understand the foundations of DL and, more specifically, learn how to build their own datasets and train their own models. Extending large pre-trained models with unique subjects or styles might be one of the ways individual artists can provide creative value to their audiences. Similarly, fine-tuning weights with further examples to integrate models into the flow of applications and interfaces might become the order of the day for creative practitioners and designers. Customization of the generative capabilities of these tools requires both knowledge and skill. The selection of "good" examples representing a coherent visual concept is perhaps an art that can be learned and mastered, a process that can reveal our biases and challenge our imagination. The combination-making capabilities of language might then expand the reach of our creations and let us explore conceptual spaces through computable compositionality.

By understanding the use of computational tools in the creative process as an act of engaging in human-technology-world relations, we can further refine and give meaning[1] to

[1] In the sense of *hermeneutic intentionality* described by Ihde (1990).

the interactions we have with any tool or instrument. In particular, as suggested by the differentiation between $R[]$ and $D[]$ considered in this thesis, the contextualization acting in the *background* of a specific interaction is an important aspect to consider when engaging with computational tools for creative tasks. Furthermore, the fitness of the two contextualizations varies based on the ACASIA component that they are applied to.

As highlighted in all studies, training a model can be thought of as the act of creating a conceptual space by assessing the similarities across all examples provided. By doing so, the model is performing its own *alien*[2] abstraction (as data compression), so then associations and combinations can happen in *latent space*. The modularity of the ACASIA (Hoorn, 2014) meta-model of creativity is particularly useful here because it allows us to compare the two approaches by looking how individual modules are implemented. The studies suggested that, while in $R[]$ contextualizations the abstraction has already been performed by humans defining analytical rules, in $D[]$, high-level human-compatible abstraction is achieved through large-scale datasets and training. Because of this, $D[]$ only affords the illusion of analyticity (GPT is an unreliable calculator, SD struggles drawing exactly two circles, etc...) and its abstraction is much more attuned to the complexities of human *perception*, as opposed to *cognition*.

In conclusion, attention to *Process* is required in order to select examples, form new combinations, iterate and learn. While the evaluation of output (*Product*) is necessary to improve, typically, the adjustments are to be made in the *Process*. Without a perspective over the technological relations that exist between the human and non-human, the intersection of technology and creativity is hard to navigate. As shown in Chapter 5, reflective practices act as a compass, making us become aware of how we *expect* these tools to behave and learn to distinguish it from how they *actually* work. This developed self-awareness can increase

[2]The word *alien* is used here figuratively, referencing the metaphor used by Fazi (2018).

the effectiveness and range of expression of using technology for creative endeavors. With intimate knowledge of the creative *Process* of a specific technology is possible experiment with its boundaries, in search for unexpected behaviors (Hoorn, 2023).

### 7.2 Language as interface for concepts

As emerged from the text-to-image study (Chapter 6), language can be a powerful interface for visual concepts. Similar types of text conditioning have been successfully applied to audio (Borsos et al., 2022; Dong et al., 2023; Ghosal et al., 2023; J. Huang et al., 2023; R. Huang et al., 2023; H. Liu et al., 2023) and music (Agostinelli et al., 2023; K. Chen et al., 2023; Melechovsky et al., 2023; Schneider et al., 2023), which suggests that language can be used to interface in a multi-modal way to multiple forms of perception. However, as the literature review has thoroughly exposed, it would be wrong to state that we can describe the nature of concepts only in linguistic terms. Doing so would require to either rely on formal definitions, hence falling into all the problems of CT, or limiting the study of concepts to lexical concepts, one of the limitations of NT. Reflecting on the studies conducted, it became clear that having language as interface for conditioning DL models was a tremendous asset, which unlocked the immense creative potential existing in *latent space*. A hypothesis that could follow from these ideas is to think of language as a technology in itself, so we could write:

- $(I - Language) \rightarrow Speech$

Ihde (1993) is cautious about addressing language as a technology and argues that because language involves human intentionality and meaning-making processes - something he calls “hermeneutic intentionality” - it cannot simply be treated as an objectified tool like other forms of technology. Instead, understanding how people use language requires an appreciation for the cultural contexts in which they operate, including shared assumptions about meanings and values. Ihde notes that unlike most technologies which have specific functions and purposes (e.g., a hammer is designed for pounding nails), language has an open-endedness and flexibility in terms of what it can express. This means that while we use language to achieve particular ends (such as conveying information or persuading someone), there is no limit to what we might say using this medium. Ihde also warns that addressing language purely as a technology risks missing out on its deeper dimensions such as creativity, playfulness, and poetic expression.

Researchers in other fields such as biolinguistics (Koster, 2009) and evolutionary linguistics (Mufwene, 2013) also brought forward the idea that language can be seen as technology. These views typically follow from Chomsky (1957) and his theory that language is an innate ability that humans are born with - it is part of our genetic makeup or universal grammar. Assessing language as a technology heavily depends on what we intend by LANGUAGE and TECHNOLOGY, which are both quite expansive and controversial concepts themselves, but as it should be evident since Section 2.1.1, looking for exact formal definitions seems to be a dead end.

I will explore here the perspective that a LANGUAGE can be broadly understood as a compositional system for concepts. I will also adopt the perspective that a TECHNOLOGY is anything that mediates our experience of the world. These two perspectives might then allow for LANGUAGE to be a TECHNOLOGY that mediates our experience of concepts. According to the post-phenomenological notation, we might formalize how language is acting as an interface for concepts as follows:

- $I \rightarrow Mathematics/(L_M \rightarrow R[C_1, C_2, ..., C_n])(-Result)$
- $I \rightarrow Storytelling/(L_S \rightarrow D[C_1, C_2, ..., C_n])(-Story)$

Where $L_x$ is a form of language and $C_n$ can be any concept, such as NUMBER, ADDITION, CAUSE, CHARACTER, DIE and so on. While in the first example the use of $R[]$ is reflecting the *analytic* nature of mathematics, the use of $D[]$ in the second example highlights the *synthetic* nature of storytelling, a process influenced by the author's experience and cultural context. In both examples the *hermeneutic intentionality* (Ihde, 1993) sets the concept contextualization required by the specific form of language.

It could be argued that for the storytelling example there are some elements of $R[]$ as well. A more obvious case is that of a poem that follows a specific meter, such as a sonnet. It seems that $R[]$ and $D[]$ can happily coexist in the same mediation. It is perfectly possible to have:

- $I \rightarrow Poetry/(L_P \rightarrow (R[Meter] - D[Theme]))(-Sonnet)$
- $I \rightarrow Music/(L_{Mu} \rightarrow (R[Tempo] - D[Theme])(-Song)$
- $I \rightarrow Architecture/(L_A \rightarrow (R[Materials] - D[Aesthetics]))(-Building)$

Once again, language is intended here very broadly as a compositional system, grounded in concept tokens and a representation of their relationships. It seems that the ability to deal with both $R[]$ and $D[]$ contextualizations of compositional systems (i.e. languages) is perhaps its most remarkable feature as it allows to bridge perception and cognition. As the experience of the first two studies suggests, similarity spaces constructed without language-like elements in support of compositionality are relatively difficult to control. In the third study, which leverages the compositional properties of word tokens and large-scale pre-trained models, the relationships between certain words and the visual representations of concepts they refer to, becomes immediately available to the users as an interface to visual concepts. Arguably, this is why in the third study most participants picked up relatively easily on how to generate content using Botticello.

As final remark, it should be noted that as LLMs evolve to be more powerful and capable, they will also become multi-modal. GPT4 already incorporates the ability to describe, generate and modify images based on natural language. It is only a matter of time before the compositional capabilities of these tools unify other domains under the same roof. However, it is still unclear how future DL solutions will integrate or achieve the ability to deal with concepts *analytically* and whether this is possible at all using machine learning.

### 7.3 Blurring human and non-human creativity

The concepts of ownership and authorship in art and design are constantly challenged by technological advancements and the current DL trend is perhaps a challenge that will leave a long trace. For example, there are many concerns regarding whether the output of image generators such as Dall-E or SD is to be considered derivative work because it relies on existing visual content that has been taken from the internet without explicit consent from the authors (Brittain, 2023). There are also many cases in which using these tools can be considered as fair use, especially when there is substantial tweaking of the generation parameters, custom datasets or manual post-processing efforts. In all studies presented, the boundaries of human and non-human efforts are being twisted and blurred in different ways. This section addresses the implications of this ambiguity and how existing creativity theories might adjust to accommodate for it.

In the attempt to frame CC systems under the four Ps of creativity, Jordanous (2016) suggests to change Person into Producer, in order to accommodate non-human agents as potential actors or co-actors. While this extension is welcome, it perhaps also leaves out other aspects that are characteristic of CC, for example how a specific technology (*Process*) might affect the criteria that the audience adopts to deem an artifact as creative (*Press*). Furthermore, communities of practice (*Press*) might foster individual learning (*Person*), which in turn will affect the way technology is used or even its core functioning, as seen in numerous open-source projects spawned after the release of SD (*Process*). Separating the different perspectives might have its benefits, but also promotes a reductionist view of creativity, which gives an incomplete picture.

The ACASIA meta-model by Hoorn (2014) does not, by definition, take into consideration the nature of the *agent* performing each module. ACASIA's agnostic take on the nature of the agent performing the tasks in each module makes the model better suited to describe these interactions. I believe that it might be valuable to express more specifically how the interaction between the human and the non-human shapes the creative process, within ACASIA modules. For example, in the third study, the interaction loop happened much faster compared to the first two studies in which the technology only allowed two iterations in total. As discussed in Section 6.5, the shorter generation time of SD affords both human and non-human association and combination in extremely short cycles. This fast paced interaction dramatically increases the speed at which a technology's *background* relations can become more *transparent* to the user.

Under the post-phenomenological view, the use of a technology for creative endeavors can be both personal and impersonal. On one hand, the non-human can be perceived as external to us, as is the case for *alterity* relations where technology is to be considered as a *quasi-other*. On the other hand, a technology can become more transparent to its user by gaining insights into its *background* relations, to the point where *hermeneutic intentionality* becomes possible through the interaction with it. With regards to the studies presented in this thesis, the affordances of specific DL tools seem to have had significant impact in defining these boundaries. In addition to the increased speed of the iteration cycle discussed earlier, the introduction of text-based conditioning provides a more *transparent* interface, given that the compositionality language is so familiar to us.

In the last two years, it is not uncommon to hear the term "prompt-engineer" referring to someone who specializes in the craft of writing *prompts* for LLMs or LDMs. Arguably, this skill requires a deeper understanding of a $D[]$ contextualization, which is acquired primarily by experimentation. Prompt-engineering is characterized by the ability to see through the *opacity* of *alterity* relations and gain intimate knowledge of the technologies acting in the *background* of DL models. By stepping into the technological mediation and understanding its components, one is then able to see the bigger picture and have full control over the technology. Seeing through *alterity* relations enables us to become aware of a technology's inner functioning and this awareness affects the mediation itself, making us perceive the output as more *ours*.

However, there is no shared understanding of authorship and ownership in the community of practice because most users approaching GDL use algorithms and models *off the shelf*, often with little knowledge about their internal functioning. If one comes up with the idea of an avocado shaped chair and uses a LDM to visualize how it may look like, is it still possible to argue that the final image was "made" by the human just because they typed the right sequence of words as prompt? Is the model simply solving a *production* problem? Can a "prompt-engineer" be considered the *author* of a prompt? These are extremely difficult questions to answer, but one thing is for certain: the value of merely executing existing ideas is diminishing rapidly in the face of these technological advancements. As DL tools become more accessible and widespread, the importance of fostering unique and innovative approaches to utilizing these technologies in creative practices becomes increasingly crucial. Consequently, the focus should shift towards developing a deeper understanding of the underlying mechanisms that drive these systems. In doing so, we can strive to ensure that the resulting creative outputs maintain a sense of authenticity and originality, even as the lines between human and non-human contributions continue to blur.

*I think we should think of AI as the intellectual equivalent of a backhoe. It will be much better than us at a lot of things.*

Geoffrey Hinton

# 8

# Conclusions

The purpose of this study was to investigate how different approaches to computation and their embedded TOCs affect the technological mediation in a creative process. Throughout this investigation, the technology associated with the topics addressed in this thesis has been in a constant state of evolution. There are no clear signs that the improvements in this area will stop or slow down, making it very challenging to come up with conclusions that can stay relevant for an extended period of time. Nevertheless, in this final chapter I will attempt to summarize some theoretical considerations emerging from this research that should endure, while also highlighting the many limitations that characterized this doctoral journey. The hope is that these insights can foster further research and collaborations to better understand and anticipate the future challenges that creative practices may encounter in the face of rapidly advancing technology.

## 8.1 Research summary

When this research began in 2019, it was set out to explore how different computational approaches might affect the creative process, with the assumption that they implicitly embed diverging concept ontologies (TOCs). Compositionality had emerged early on from the literature review as one of the factors that are fundamental to concept formation. While traditional rule-based approaches to computation implement crisp analytical rules of composition which are pre-defined by humans, DL originally struggled in achieving compositionality due to the inherent locality of neuron states in a network. During the course of this doctoral journey, the innovation of self-attention and its ability to scale over massively large datasets proved to be an effective way to address compositionality using data-driven technologies.

In order to describe different computational approaches and their respective nuances, an extension to mediation theory (Ihde, 1993) that addressed two different contextualizations of input ($R[]$ and $D[]$) has been proposed. This distinction has been discussed in relation to existing TOCs as well as models of creativity (Boden, 1996; 2003; Hoorn, 2014; Rhodes, 1961) and used in the studies to analyze specific forms of mediation afforded by DL technologies such as VAEs, Transformers, GANs and the most recent LDMs. The insights that emerged from each study highlight several themes that can be related back to the TOCs discussed in the literature review as well as provide further research directions for existing creativity theories.

As part of this doctoral thesis, three studies have been conducted. The first and second study embraced a practice-led research approach (Candy, 2011), in the form of close collaborations with a musician and a photographer. These studies addressed existing DL tools that could be trained on datasets we created ourselves. In the second study, the act of putting together a dataset was used as a *reflective practice* (Schön, 1983) to investigate how $D[]$ contextualizations form concepts from a series of examples. The third study focused on the usage patterns of a Discord bot powered by Stable Diffusion (Rombach et al., 2021), with particular attention to the construct of Tolerance for Ambiguity (TA) (Norton, 1975) and its relationship with user expectations about text-to-image technology. The data collected from 76 participants shows an exponential distribution of amount of images generated per user and a weak correlation between TA and expectations of image alignment. No significant correlation between TA and usage was found in the collected data. These results suggest the hypothesis that social factors, such as network effects, might be influencing usage, while personal traits (TA) only affect expectations about text-to-image technology capabilities.

## 8.2 Post-phenomenology of creativity

One of the main gaps exposed in the literature review was the lack of a non-dual way to represent human and non-human roles in creative endeavors. The adoption of the post-phenomenological lens allows for a description of the creative process that does not separate them, but instead puts them in a mediating relationship. Existing models of creativity taken under consideration either formalize the distinction, for example Rhodes (1961), or are agnostic about the nature of the actor performing the step as proposed by Hoorn (2014).

Adopting a post-phenomenological approach might provide an explanation to why certain aspects of a technology might be overlooked or not fully understood by its users, as discussed by Eede (2010) with regard to *technological opacity*. Indeed the difference between *alterity* relations that use $R[]$ or $D[]$ might not be immediately obvious. For example, users might be inclined to think that GPT is capable of analytical reasoning and consider its output as bound to some kind of *truth*, while the reality is that GPT is most definitely not fact-checking its output in any way and will likely make up a probable answer if it has not seen any data related to a topic during training. Understanding this is roughly equivalent to being able to discern $R[]$ from $D[]$ technologies.

The post-phenomenological framework discussed in this thesis is able to explain some of the unexpected results that emerged from the third study. As already highlighted by Flusser (1999, 2000) the *Apparatus* is to be considered part of the technological mediation. This perspective provides some insights into the unexpected finding that network effects might be a factor that affects tool use (see Section 6.5). This explanation also extends to the discussion about the increasingly important role datasets curation presented in Section 7.1, as the increasing popularity of model sharing sites like Civitai seems to confirm.

Adopting a perspective that allows to see language as a technology, the framework might also provide some insight into its ability to bridge between $R[]$ and $D[]$, thus mediating our experience of concepts. By enabling both *analytic* and *synthetic* contextualizations, language has the unique potential to bridge perception and cognition. This also highlights its key role in the development of multi-modal deep learning models that can effectively interface with various forms of human experience and understanding.

Finally, as highlighted throughout Chapter 7, possessing intimate knowledge of how technology affects creative work is crucial for future generations of artists and designers, as the attention of communities of practice is shifting from Product to Process (Rhodes, 1961). The extension of mediation theory (Ihde, 1990) proposed in this thesis should be useful to educators to promote a more conscious and effective use of both types of computational approaches. I believe it is also particularly important to teach how the two can integrate with one another. $R[]$ and $D[]$ have strength and weaknesses that are complementary and, as seen in Table 3.2, they might be more or less suited for a task, based on which ACASIA module they are embedded in. For the educators in this field, a teaching objective that this research suggests is to provide students with an understanding of technological *opacity* and *background* relations, so that they can master their creative expression through any technology.

## 8.3 Limitations

The studies address new technologies as they develop. This is not ideal, as discussed in Section 3.4, because the methods used to conduct a research must continuously change and adapt to new advancements in order to remain relevant. The methodology addressing this topic is still in its infancy. Constructs such as alignment between prompt and image are yet to be formalized and operationalized properly and by the time they will be, there might be new solutions which renders them obsolete. It is simply not possible to keep up to date with the latest release of every tool and conduct rigorous research at the same time. Fast research is not good research, but slow research becomes irrelevant fast in this field.

Regarding the first two studies, the practice-led approach has proven to be limiting in terms of potential generalization of results. While the experience gained through practice turned out to be valuable, $N = 1$ studies are very specific by nature and the results cannot be discussed in objective terms. In hindsight, the first study focused too much on the output rather than the process, which might be hindering its generalizability even further. Practice-led research conducted in this form is effectively limited to self-reflective exercises, which can only be valuable at a later stage.

Another limitation that was felt for all the studies was the lack of adequate parallel computing power and memory (i.e. GPUs) to run experiments. Unfortunately, the ability to run certain DL models depends hardware at hand, which, in turn, is contingent on budget constraints and accessibility. For example, in the first study, we could use a VAE capable of producing only two bars of melody because of the limited amount of memory of our GPU. Coincidentally, at the time there was also a global shortage of GPUs due to the COVID-19 pandemic. Again, in the second study, the resolution of the images we could generate was limited for the same reason. When we were finally able to buy a better GPU for the third study, as we put online the Discord bot, we committed our only available unit to this task, which prevented us from running other experiments without creating disruptions to users. Having only one GPU also imposed a limit to the amount of users we could host in a workshop.

A limiting factor for the third study was also the use of Discord as interface for generation. While on one hand the platform provides a relatively easy interface to operate the bot, on the other hand it does require some explanation and the user experience might not be very intuitive for new users. This introduces another layer of complexity which is unrelated to the study topic, potentially introducing a confounding variable. Unfortunately, there is no easy way around this if the goal is to expose a group of participants to these technologies. Most UIs for SD are designed for one user only and therefore do not provide a queue system which is necessary in the case of simultaneous requests.

Another limitation of this research is the lack of a comparative study between $R[]$ and $D[]$ technologies. A pilot workshop aiming to formalize a protocol for this kind of experimental research was conducted in early 2022, but the gathering restrictions imposed by the COVID-19 pandemic made it difficult to put it in place. Fortunately, it is very likely that this study will eventually take place within 2023.

## 8.4 Further research

This doctoral thesis opens many questions but leaves them only partially answered. A topic that demands further inquiry is the study of technologies that integrate $R[]$ and $D[]$. While this can be formulated as a practical problem, it does indeed hold many philosophical implications. Some examples of this type of hybrid approach are already surfacing after the release of GPT4, such as AutoGPT[1] which combines $D[]$ elements of GPT with traditional $R[]$ elements by giving it access to internet search, a memory system, a way to execute code and other scripted functionalities. It is possible that this type of integration will evolve into something more sophisticated in the future which could be able to encompass both analytic and synthetic realms.

An additional avenue for future research stemming from this doctoral study involves exploring community sharing and compositionality in user-trained models. As highlighted in Section 7.1, the active involvement of community members in developing personalized model extensions (such as custom concept embeddings, hypernetworks, LoRA, etc.) is transforming the way in which artists and creators present content to their audiences. If this trend persists, it is conceivable that these sharing platforms may evolve into a central hub for both content creators and consumers.

Furthermore, there are implications arising from the popularization of tools such as SD and GPT with regards the understanding of *truth*. For example, while in the early days of photography the artifacts produced by cameras could be considered as somehow bound to reality, as of today, it is almost impossible to discern a generated image from a real one. A similar problem is evident today in the education sector, as students make use of GPT for their assignments without necessarily questioning the origin of the information that is re-

[1] https://autogpt.net/

turned by the model. While to some extent these issues already existed before DL platforms became available, the problem is now becoming extremely obvious because these tools are so easy to use and millions of people are adopting them. How will these tools change our understanding of truth?

Another direction for further research that naturally follows from the journey presented in this thesis is to investigate multi-modality in DL. As discussed in Section 7.2, language could become the unifying interface which allows us to generate all types of media using just a single tool. For example, a future version of the popular streaming site Netflix could offer the possibility to describe a setting, upload a few photos of ourselves or our friends and generate a whole TV series where we are the protagonists. While this may sound like wild speculation, it is not so hard to believe that this will be possible one day. Crude yet functional text-to-video is already a reality (Esser et al., 2023) and so are language models for audio (Borsos et al., 2022) and music (Agostinelli et al., 2023) generation.

Finally, there is one last important question to be asked, which is what, if anything at all, can slow down or stop this trend. On March 22nd, 2023, an open letter was published on the internet asking to *Pause Giant AI Experiments*[2] which warns:

> AI systems with human-competitive intelligence can pose profound risks to society and humanity, as shown by extensive research and acknowledged by top AI labs. As stated in the widely-endorsed Asilomar AI Principles, Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources. Unfortunately, this level of planning and management is not happening, even though recent months have seen AI labs locked in an out-of-control race to develop and deploy ever more powerful digital minds that no one – not even their creators – can understand, predict, or reliably control.

These are legitimate concerns that the creative industry can fully relate to. It seems we

[2]https://futureoflife.org/open-letter/pause-giant-ai-experiments/

are witnessing the beginning of a new era, where, once again, the impact of technology on society is about to redefine what it means to be creative.

In conclusion, I would like to share a personal reflection that may leave readers on an optimistic note. I experienced an enlightening moment while observing a stunning sunset on a tropical beach. Many individuals around me were viewing this breathtaking scene through their mobile phone screens, attempting to preserve the moment in digital form. However, given my experience working on this thesis, I felt no such compulsion, confident in my ability to generate a virtually limitless number of sunset images, even featuring myself. Thus, I was content to merely immerse myself in the experience, allowing the beauty of the moment to spark the creative inspiration.

# A

# Supplement to the literature review

This appendix contains the code used to fetch the literature review data using scopus API as well as the prompt used to categorize the papers.

## A.1 Scopus search retrieval

```
require 'typhoeus'
require 'json'
require 'bibtex'
```

```
require 'active_support'
require 'active_support/core_ext'


API_KEY = "xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx"
API_URL = "https://api.elsevier.com/"
DOI_BASE = "https://doi.org/"

DEFAULT_OPTIONS = {
  method: :get,
  headers: {
    'Accept': 'application/json',
    'X-ELS-APIKey': API_KEY
  }
}

BIBTEX_TYPES = {
  "Conference Paper" => :inproceedings,
  "Article" => :article,
  "Editorial" => :article,
  "Note" => :article,
  "Review" => :article,
  "Chapter" => :inbook,
  "Article in Press" => :article,
  "Short Survey" => :article,
  "Book" => :book
}

@search_string = 'KEY ("computational creativity")'
@bib = BibTeX::Bibliography.new
@urls = []

def make_keyword_request(sstring, cursor = 0)
  options = DEFAULT_OPTIONS.dup
  options[:params] = {
    query: sstring,
    field: "prism:url",
    count: 25,
    start: cursor
  }
  request = Typhoeus::Request.new(api("content/search/scopus"), options)
  request.on_complete do |response|
    if response.success?
```

```
    parsed_response = JSON.parse(response.body)["search-results"]
    if parsed_response["entry"]
      parsed_response["entry"].each do |result|
        puts result["prism:url"]
        @urls.push result["prism:url"]
      end
      make_keyword_request(sstring, cursor + 25).run
    end
  elsif response.timed_out?
    # aw hell no
    puts "got a time out"
  elsif response.code == 0
    # Could not get an http response, something's wrong.
    puts response.return_message
  else
    # Received a non-successful http response.
    puts "HTTP request failed: " + response.code.to_s
    puts response.body
  end
 end
 request
end

def api(path)
 API_URL + path
end


def make_abstract_request(url)
 puts url
 options = DEFAULT_OPTIONS.dup
 options[:params] = {
   view: "META_ABS",
   apiKey: API_KEY
 }
 request = Typhoeus::Request.new(url, options)
 request.on_complete do |response|
   if response.success?
    parsed_response = JSON.parse(response.body)["abstracts-retrieval-
        response"]
    coredata = parsed_response["coredata"]
    puts "Found #{coredata["subtypeDescription"]}"
    puts coredata
```

```ruby
if is_valid_record(coredata)
  puts "Valid record"
  entry = {}
  entry[:bibtex_type] = BIBTEX_TYPES[coredata["subtypeDescription"]]
  entry[:issn] = coredata["prism:issn"]
  entry[:journal] = coredata["prism:publicationName"]
  entry[:author] = format_authors(coredata["dc:creator"]["author"])
  entry[:abstract] = coredata["dc:description"]
  date = Date.parse(coredata["prism:coverDate"])
  entry[:year] = date.year
  entry[:month] = date.month
  entry[:title] = coredata["dc:title"]
  entry[:volume] = coredata["prism:volume"]
  entry[:number] = coredata["issueIdentifier"]
  entry[:pages] = "#{coredata["prism:startingPage"]}--#{coredata["prism:
     endingPage"]}"
  entry[:citedby] = coredata["citedby-count"]
  if coredata["prism:doi"]
    entry[:doi] = coredata["prism:doi"]
    entry[:url] = DOI_BASE + coredata["prism:doi"]
  end
  entry[:eid] = coredata["eid"]

  entry[:publisher] = coredata["dc:publisher"]

  #entry[:file] = build_file_url(coredata["dc:identifier"])
  if parsed_response["authkeywords"]
    if parsed_response["authkeywords"]["author-keyword"].instance_of?
       Array
      entry[:keywords] = parsed_response["authkeywords"]["author-keyword
         "].map {|key| key["$"].titlecase.strip }.join(", ")
    else
      entry[:keywords] = parsed_response["authkeywords"]["author-keyword
         "]["$"].titlecase.strip
    end
  end
  @bib << BibTeX::Entry.new(entry)
else
  puts "Invalid record!"
  puts url
  puts "Continuing"
end
```

```
    elsif response.timed_out?
      # aw hell no
      puts "got a time out"
    elsif response.code == 0
      # Could not get an http response, something's wrong.
      puts response.return_message
    else
      # Received a non-successful http response.
      puts "major fail"
      puts response.body
      puts "HTTP request failed: " + response.code.to_s
    end
  end
  request
end

def format_authors(authors)
  authors.map do |author|
    "#{author['ce:surname']}, #{author['ce:given-name']}"
  end.join(" and ")
end

def is_valid_record(data)
  data["dc:creator"] && BIBTEX_TYPES[data["subtypeDescription"]]
end

make_keyword_request(@search_string).run

#urls = File.open("artificial_creativity.txt").read
@urls.each_with_index do | url , i |
  puts "Querying #{i+1} of #{@urls.length}"
  make_abstract_request(url.strip).run
end

#make_abstract_request("https://api.elsevier.com/content/abstract/scopus_id
    /84900523833").run
#print @bib.to_s
File.write("#{@search_string.remove('"').squish}.bib", @bib.to_s)
```

## A.2 GPT categorization prompt

```
A media is intended as the field of application addressed by the paper in
    question.
The theoretical scope of a paper is intended as the focus of the paper's
    research question.
The computational approach is intended as the broad category of algorithms
    and implementations discussed in the paper.
I need your help identifying the domain of this paper:

Title: {title}

Abstract: {abstract}

Journal: {journal}

Please assign it to one of the following media:

  - No medium
  - Visual, images and movies
  - Culinary recipes
  - Design, Urban Design, Architecture
  - Writing, narrative and language
  - Music and musical composition
  - Game design
  - Concepts
  - Multi-modal

Please also help me assign this paper to one of the following theoretical
    scopes based on the
research question that the paper is addressing:
  - Evaluation: the paper discusses the evaluation of creative artifacts
  - Theory: the paper discusses a particular theory or hypothesis about
     creativity
  - System: the paper presents a technical implementation of a specific
     system or algorithm developed by the authors
  - Other: the scope of the paper does not fit any of the other categories

Please also assign one of these computational approaches:
  - Rule based: these are non-data driven methods that follow deterministic
     rules and definitions, such as expert systems
  - Evolutionary algorithms: genetic algorithms and other evolutionary
     methods
  - Data driven: this includes Deep learning, machine learing and other
     methods that are based on datasets
```

```
  - Other: does not belong to any of the above or the approach is not
      clearly specified


Return the domain and the explanation of your choice in this format:
[ the media of the paper] | [ the theoretical scope ] | [the computational
    approach] | [ your explanation]

for example:

Multi-modal | Evaluation of creativity | Data driven | This paper discusses
     [...] so it belongs to [...]

Follow this format strictly and do not add any other words or prefix such
    as "Scope:" or "Medium:" before the domain and category you pick.
```

## A.3 Top cited paper and their research question

The following table displays the 25 papers with most citations in the corpus. The number of citations was not considered in the systematic literature review, because of the possible bias it may introduce. All topics have been considered as a equally important regardless of whether the paper was cited or not. Citations also are depended on publication date, which may yield further bias in the analysis.

| Author (Year) | Paper Title | Citations | Research Question |
|---|---|---|---|
| Ritchie (2007) | Some empirical criteria for attributing creativity to a computer program | 193 | Is it possible to empirically assess whether a computer program is capable of creative activity? |
| Jordanous (2012) | A Standardised Procedure for Evaluating Creative Systems: Computational Creativity Evaluation Based on What it is to be Creative | 129 | How can we evaluate the creativity of computational systems? |

| | | | |
|---|---|---|---|
| Wiggins (2006a) | Searching for computational creativity | 110 | How can traditional AI search methods be used to explore the relationship between Boden's account of creativity and Wiggins' formalisation of it? |
| Machado (2002) | All the truth about NEvAr | 109 | How can Evolutionary Art Tools, such as NEvAr, be used to generate images and what are the benefits of using such tools? |
| Boden (2009) | What is generative art? | 91 | What are the major categories of generative art, and what are the appropriate aesthetic criteria and locus of creativity for each category? |
| Neves et al. (2007) | The halt condition in genetic programming | 79 | What is the role of divergence and convergence in creative processes, and how can they be implemented within creativity programs in the Genetic or Evolutionary Programming paradigm to address the Halt Condition in Genetic Programming? |
| Jennings (2010) | Developing creativity: Artificial barriers in artificial intelligence | 55 | How can developers of creative artificial intelligence systems convincingly argue that their software is more than just an extension of their own creativity? |
| Davis (2016) | Empirically studying participatory sense-making in abstract drawing with a co-creative cognitive agent | 54 | How can participatory sense-making be used to model and understand open-ended collaboration between humans and computers in the context of abstract drawing? |
| Kowaliw (2012) | Promoting creative design in interactive evolutionary computation | 52 | How can creative design be promoted in interactive evolutionary computation? |

| | | | |
|---|---|---|---|
| Eppe (2018) | A computational framework for conceptual blending | 51 | How can modern answer set programming methods and optimality principles be used to develop a computational framework for conceptual blending? |
| L. Chen (2019) | An artificial intelligence based data-driven approach for design ideation | 49 | How can artificial intelligence and data mining techniques be used to enhance design ideation? |
| Peinado (2006) | Evaluation of automatic generation of basic stories | 47 | How can the utility of an automatic story generation system be measured in terms of quality and originality of the generated artifact? |
| Olteţeanu (2015) | ComRAT-C: A computational compound Remote Associates Test solver based on language data and its comparison to human performance | 44 | Can a computational compound Remote Associates Test solver based on language data be used to solve RAT queries and how does it compare to human performance? |
| Yang (2020) | On the evaluation of generative models in music | 43 | How can generative music systems be evaluated and compared in a reliable, valid, and reproducible way? |
| Sturm (2019) | Machine learning research that matters for music creation: A case study | 41 | How can machine learning be applied to music creation in a way that is useful and impactful for real-world practitioners? |
| Olteţeanu (2016) | Object replacement and object composition in a creative cognitive system. Towards a computational solver of the Alternative Uses Test | 41 | How can object replacement and object composition be used to create a creative cognitive system that can solve the Alternative Uses Test? |

| | | | |
|---|---|---|---|
| Liapis (2019) | Orchestrating game generation | 39 | How can a computational process orchestrate the various computational creators of different creative domains in order to generate a digital game with desired functional and aesthetic characteristics? |
| Colton (2008) | Emotionally aware automated portrait painting | 38 | How can a machine vision system and a non-photorealistic rendering (NPR) system be combined to automatically produce portraits which heighten the emotion of the sitter? |
| Cook (2017) | The ANGELINA videogame design system-part i | 37 | How can cooperative coevolution be used to automate game design and produce content that complements each other? |
| Grace (2015) | Data-intensive evaluation of design creativity using novelty, value, and surprise | 37 | How can data-intensive methods be used to evaluate the creativity of a new design in terms of novelty, value, and surprise? |
| al-Rifaie (2012) | Creativity and Autonomy in Swarm Intelligence Systems | 37 | How can swarm intelligence algorithms be used to create novel drawings of an input image, and what implications does this have for creativity and autonomy? |
| Jordanous (2016) | Four PPPPerspectives on computational creativity in theory and in practice | 34 | How can the Four Ps of creativity (Person/Producer, Product, Process and Press/Environment) be used to take a broader perspective on computational creativity in theory and in practice? |
| Köbis (2021) | Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry | 33 | Can people distinguish and prefer algorithm-generated versus human-written text? |

| | | | |
|---|---|---|---|
| Saunders (2012) | Towards Autonomous Creative Systems: A Computational Approach | 32 | How can autonomous computational creativity be developed to model personal motivations, social interactions, and the evolution of domains? |
| De Nicola (2019) | Creative design of emergency management scenarios driven by semantics: An application to smart cities | 31 | How can semantics-based techniques be used to support the creative design of emergency management scenarios in smart cities? |
| J. R. Smith (2017) | Harnessing A.I. for augmenting creativity: Application to movie trailer creation | 30 | How can Artificial Intelligence (AI) be used to augment creativity in the context of movie trailer creation? |
| Pasquier (2016) | An introduction to musical metacreation | 30 | What are the challenges and opportunities for the field of musical metacreation? |
| Carnovalini (2020) | Computational Creativity and Music Generation Systems: An Introduction to the State of the Art | 29 | What are the current state-of-the-art systems for Computational Creativity and Music Generation, and what open challenges remain to be addressed? |
| Lamb (2018) | Evaluating computational creativity: An interdisciplinary tutorial | 29 | How can computational creativity be evaluated using interdisciplinary theories of creativity? |
| Deterding (2017) | Mixed-initiative creative interfaces | 29 | How can mixed-initiative creative interfaces be designed to broaden and amplify creative capacity for all? |

**Table A.1:** A list of the 25 most cited paper found in the corpus of used for the literature review and their research questions extracted by GPT3.5

# B
# Questionnaire items and details

These are the items of the questionnaire required to access MediaBots. With the exception of the first item, which is in the form of an open-ended question, all other items are in the form of a 6-point Likert-scale question with options:

1. Strongly Disagree
2. Disagree
3. Slightly Disagree
4. Slightly Agree
5. Agree
6. Strongly Agree

The first 5 items are aimed at measuring the participant's knowledge and intuition about how this technology works, as well as their expectations about image *alignment* with the text prompts.

1. Based on what you currently know, how will text-to-image technology affect your creative process? [Open question]
2. I understand how this technology interprets keywords to generate images.
3. I can imagine how this technology interprets keywords to generate images.
4. I expect this technology to interpret the keywords in a similar way as humans do
5. I expect generated pictures to be an accurate visual representation of the keywords

The next 10 questions are attempting to measure the psychological construct of tolerance for ambiguity, adapted from Herman et al. (2010):

6. I like to surround myself with things that are familiar to me.
7. The sooner we all share similar ideals the better.
8. A desirable job is one where what is to be done is always clear.
9. I prefer a life with few surprises.
10. What we are used to is preferable to what is unfamiliar.
11. I prefer settings where people share my values.
12. I like parties where the attendees are strangers.
13. I can enjoy being with people whose values are different from mine.
14. I would like to live in a foreign country for a while.
15. I can be comfortable with all kinds of people.

# References

Agostinelli, A., Denk, T. I., Borsos, Z., Engel, J., Verzetti, M., Caillon, A., Huang, Q., Jansen, A., Roberts, A., Tagliasacchi, M., Sharifi, M., Zeghidour, N., & Frank, C. (2023, January). MusicLM: Generating music from text. https://doi.org/10.48550/arXiv.2301.11325

Agrawal. (2019). What is deemed computationally creative? *IBM Journal of Research and Development*, *63*. https://doi.org/10.1147/JRD.2019.2900640

Algarni, A. (2020). Neuroscience of creativity in human computer interaction. In K. Arai, R. Bhatia, & S. Kapoor (Eds.), *Proceedings of the Future Technologies Conference (FTC) 2019* (pp. 248–262). Springer International Publishing.

al-Rifaie, M. M. (2012). Creativity and autonomy in swarm intelligence systems. *Cognitive Computation*, *4*, 320–331. https://doi.org/10.1007/s12559-012-9130-y

Amabile, T. M. (1996). *Creativity in context*. Westview Press.

Amabile, T. M. (1997). Motivating creativity in organizations: On doing what you love and loving what you do. *California Management Review*, *40*(1), 39–58. https://doi.org/10.2307/41165921

Amabile, T. M., Kramer, S., & Williams, S. (2014). *The progress principle*. Brilliance Audio.

Aristotle. (1983, October). *Physics* (J. L. Ackrill & L. Judson, Eds.; W. Charlton, Trans.). OUP Oxford.

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, *13*(3), 263–308. https://doi.org/10.1016/0010-0277(83)90012-4

Atwill, J. M. (2009, September). *Rhetoric reclaimed*. Cornell University Press.

Batchen, G. (1997). *Burning with desire: The conception of photography*. The MIT Press.

Benjamin, J. J., Berger, A., Merrill, N., & Pierce, J. (2021). Machine learning uncertainty as a design material: A post-phenomenological inquiry. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3411764.3445481

Besold, T. R. (2017). Towards a computational- and algorithmic-level account of concept blending using analogies and amalgams. *Connection Science*, *29*, 387–413. https://doi.org/10.1080/09540091.2017.1326463

Boden, M. A. (1996). *Dimensions of creativity*. The MIT Press.

Boden, M. A. (2003, September). *The creative mind*. Taylor & Francis Ltd.

Boden, M. A. (2009). What is generative art? *Digital Creativity*, *20*(1-2), 21–46. https://doi.org/10.1080/14626260902867915

Boden, M. A. (2010). The Turing test and artistic creativity. *Kybernetes*, *39*, 409–413. https://doi.org/10.1108/03684921011036132

Borsos, Z., Marinier, R., Vincent, D., Kharitonov, E., Pietquin, O., Sharifi, M., Teboul, O., Grangier, D., Tagliasacchi, M., & Zeghidour, N. (2022, September). AudioLM: A language modeling approach to audio generation. https://doi.org/10.48550/arXiv.2209.03143

Brittain, B. (2023, December 1). *Artists take new shot at Stability, Midjourney in updated copyright lawsuit*. https://www.reuters.com/legal/litigation/artists-take-new-shot-stability-midjourney-updated-copyright-lawsuit-2023-11-30/

Brock, A., Donahue, J., & Simonyan, K. (2018, September). Large scale GAN training for high fidelity natural image synthesis. https://doi.org/10.48550/arXiv.1809.11096

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020, May). Language models are few-shot learners. https://doi.org/10.48550/arXiv.2005.14165

Candy, L. (2006). Practice Based Research: A Guide. *Creativity and Cognition Studios Report*, *1*.

Candy, L. (2011, November). Research and Creative Practice. In L. Candy & E. Edmonds (Eds.), *Interacting: Art, Research and the Creative Practitioner* (pp. 33–59). Libri Publishing UK.

Carey, S. (1991). Knowledge acquisition: Enrichment or conceptual change? In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition* (pp. 257–291). Lawrence Erlbaum Associates, Inc.

Carey, S. (2009). *The origin of concepts*. Oxford University Press.

Carnovalini, F. (2020). Computational creativity and music generation systems: An introduction to the state of the art. *Frontiers in Artificial Intelligence*, *3*. https://doi.org/10.3389/frai.2020.00014

Chen, K., Wu, Y., Liu, H., Nezhurina, M., Berg-Kirkpatrick, T., & Dubnov, S. (2023). MusicLDM: Enhancing Novelty in Text-to-Music Generation Using Beat-Synchronous Mixup Strategies. https://doi.org/10.48550/ARXIV.2308.01546

Chen, L. (2019). An artificial intelligence based data-driven approach for design ideation. *Journal of Visual Communication and Image Representation*, *61*, 10–22. https://doi.org/10.1016/j.jvcir.2019.02.009

Chomsky, N. (1957). *Syntactic structures*. Mouton; Co.

Colton, S. (2008). Creativity versus the perception of creativity in computational systems. *AAAI Spring Symposium: Creative Intelligent Systems*, 14–20.

Colton, S. (2012). Computational creativity: The final frontier? *Frontiers in Artificial Intelligence and Applications*, *242*, 21–26. https://doi.org/10.3233/978-1-61499-098-7-21

Colton, S., Pease, A., Guckelsberger, C., Mccormack, J., & Llano, M. T. (2020). On the machine condition and its creative expression. *ICCC2020: Proceedings of the 11th International Conference on Computational Creativity*.

Conard, N. J., Malina, M., & Münzel, S. C. (2009). New flutes document the earliest musical tradition in southwestern Germany. *Nature*, *460*(7256), 737–740. https://doi.org/10.1038/nature08169

Cook, M. (2017). The ANGELINA videogame design system - Part I. *IEEE Transactions on Computational Intelligence and AI in Games*, *9*, 192–203. https://doi.org/10.1109/TCIAIG.2016.2520256

Csikszentmihalyi, M. (1990). *Flow: the psychology of optimal experience*. Harper & Row.

Csikszentmihalyi, M. (1996). *Creativity: flow and the psychology of discovery and invention*. HarperCollins Publishers.

Davis, N. (2016). Empirically studying participatory sense-making in abstract drawing with a co-creative cognitive agent. *International Conference on Intelligent User Interfaces, Proceedings IUI*, 196–207. https://doi.org/10.1145/2856767.2856795

De Nicola, A. (2019). Creative design of emergency management scenarios driven by semantics: An application to smart cities. *Information Systems*, *81*, 21–48. https://doi.org/10.1016/j.is.2018.10.005

Dennett, D. (1984). Cognitive wheels: The frame problem of AI. In C. Hookway (Ed.), *Minds, Machines and Evolution*. Cambridge University Press.

Deterding, S. (2017). Mixed-initiative creative interfaces. *Conference on Human Factors in Computing Systems - Proceedings*, *Part F127655*, 628–635. https://doi.org/10.1145/3027063.3027072

Dhariwal, P., & Nichol, A. (2021, May). Diffusion models beat gans on image synthesis. https://doi.org/10.48550/arXiv.2105.05233

Dinculescu, M., Engel, J., & Roberts, A. (Eds.). (2019). *MidiMe: Personalizing a Music-VAE model with user data*.

Dong, H.-W., Liu, X., Pons, J., Bhattacharya, G., Pascual, S., Serrà, J., Berg-Kirkpatrick, T., & McAuley, J. (2023). CLIPSonic: Text-to-Audio Synthesis with Unlabeled Videos and Pretrained Language-Vision Models. https://doi.org/10.48550/ARXIV.2306.09635

Dorst, K. (2008). Design research: a revolution-waiting-to-happen. *Design Studies*, *29*(1), 4–11. https://doi.org/10.1016/j.destud.2007.12.001

Eede, Y. V. D. (2010). In between us: On the transparency and opacity of technological mediation. *Foundations of Science*, *16*(2-3), 139–159. https://doi.org/10.1007/s10699-010-9190-y

Eppe, M. (2018). A computational framework for conceptual blending. *Artificial Intelligence*, *256*, 105–129. https://doi.org/10.1016/j.artint.2017.11.005

Esser, P., Chiu, J., Atighehchian, P., Granskog, J., & Germanidis, A. (2023, February). Structure and content-guided video synthesis with diffusion models. https://doi.org/10.48550/arXiv.2302.03011

Fazi, M. B. (2018). Can a machine think (anything new)? Automation beyond simulation. *AI & SOCIETY*, *34*(4), 813–824. https://doi.org/10.1007/s00146-018-0821-0

Feenberg, A. (1991). *Critical Theory of Technology* (J. K. B. O. Friis, S. A. Pedersen, & V. F. Hendricks, Eds.). Oxford University Press.

Florida, R. L. (2014). *The rise of the creative class: Revisited and expanded*. Basic Books.

Flusser, V. (1999, October). *The shape of things*. Reaktion Books.

Flusser, V. (2000, October). *Towards a philosophy of photography*. Reaktion Books.

Fodor, J. A. (1975). *The language of thought*. Harvard University Press.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford University Press, USA.

Fodor, J. A. (2008, August). *LOT 2*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199548774.001.0001

Foss, D. J. (1969). Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. *Journal of Verbal Learning and Verbal Behavior*, *8*(4), 457–462. https://doi.org/10.1016/s0022-5371(69)80089-7

Frege, G. (1892). Über sinn und bedeutung [On sense and reference]. *Zeitschrift für Philosophie und Philosophische Kritik*, *100*, 25–50.

Friedman, K. (2003). Theory construction in design research: criteria: approaches, and methods. *Design Studies*, *24*(6), 507–522. https://doi.org/10.1016/s0142-694x(03)00039-5

Furnham, A., & Marks, J. (2013). Tolerance of ambiguity: A review of the recent literature. *Psychology*, *04*(09), 717–728. https://doi.org/10.4236/psych.2013.49102

Fuxjaeger, A., & Belle, V. (2019, November). Logical interpretations of autoencoders. https://doi.org/10.48550/arXiv.1911.11629

Gal, R., Alaluf, Y., Atzmon, Y., Patashnik, O., Bermano, A. H., Chechik, G., & Cohen-Or, D. (2022). An image is worth one word: Personalizing text-to-image generation using textual inversion. https://doi.org/10.48550/arXiv.2208.01618

Galimberti, U. (1989). *Invito al pensiero di Martin Heidegger [Introduction to Martin Heidegger]*. Ugo Mursia Editore.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. MIT Press.

Ghosal, D., Majumder, N., Mehrish, A., & Poria, S. (2023). Text-to-Audio Generation using Instruction-Tuned LLM and Latent Diffusion Model. https://doi.org/10.48550/ARXIV.2304.13731

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014, June). Generative adversarial networks. https://doi.org/10.48550/arXiv.1406.2661

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories.* The MIT Press.

Grace, K. (2015). Data-intensive evaluation of design creativity using novelty, value, and surprise. *International Journal of Design Creativity and Innovation*, *3*, 125–147. https://doi.org/10.1080/21650349.2014.943295

Hallo, W. W. (1971). *The ancient Near East: a history*. Harcourt Brace Jovanovich.

He, K., Zhang, X., Ren, S., & Sun, J. (2015, December). Deep residual learning for image recognition. https://doi.org/10.48550/arXiv.1512.03385

Heidegger, M. (1977). *The question concerning technology, and other essays*. Harper & Row.

Heim, M. (1999). *Electric language: a philosophical study of word processing*. Yale University Press.

Herman, J. L., Stevens, M. J., Bird, A., Mendenhall, M., & Oddou, G. (2010). The tolerance for ambiguity scale: Towards a more refined measure for international management research. *International Journal of Intercultural Relations*, *34*(1), 58–65. https://doi.org/10.1016/j.ijintrel.2009.09.004

Hillier, L. A., & Isaacson, L. M. (1959). *Experimental music composition with an electronic computer*.

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*(7), 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527

Hoorn, J. F. (2014). *Creative confluence*. John Benjamins Publishing Company.

Hoorn, J. F. (2023). Computer-vision classification-algorithms are inherently creative when error-prone. *Proceedings of the 18th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*. https://doi.org/10.1145/3574131.3574444

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). Lora: Low-rank adaptation of large language models. https://doi.org/10.48550/arXiv.2106.09685

Huang, C.-Z. A., Vaswani, A., Uszkoreit, J., Shazeer, N., Simon, I., Hawthorne, C., Dai, A. M., Hoffman, M. D., Dinculescu, M., & Eck, D. (2018, September). Music transformer. https://doi.org/10.48550/arXiv.1809.04281

Huang, J., Ren, Y., Huang, R., Yang, D., Ye, Z., Zhang, C., Liu, J., Yin, X., Ma, Z., & Zhao, Z. (2023). Make-An-Audio 2: Temporal-Enhanced Text-to-Audio Generation. https://doi.org/10.48550/ARXIV.2305.18474

Huang, Q., Jansen, A., Lee, J., Ganti, R., Li, J. Y., & Ellis, D. P. W. (2022, August). MuLan: A joint embedding of music audio and natural language. https://doi.org/10.48550/arXiv.2208.12415

Huang, R., Huang, J., Yang, D., Ren, Y., Liu, L., Li, M., Ye, Z., Liu, J., Yin, X., & Zhao, Z. (2023). Make-An-Audio: Text-To-Audio Generation with Prompt-Enhanced Diffusion Models. https://doi.org/10.48550/ARXIV.2301.12661

Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Machine Learning*, *110*(3), 457–506. https://doi.org/10.1007/s10994-021-05946-3

Ihde, D. (1990, May). *Technology and the lifeworld*. Indiana University Press.

Ihde, D. (1993). *Postphenomenology: essays in the postmodern context*. Northwestern University Press.

Jackendoff, R. S. (1989). What is a concept, that a person may grasp it? *Mind & Language*, *4*(1-2), 68–102. https://doi.org/10.1111/j.1468-0017.1989.tb00243.x

Jackendoff, R. S. (1995). *Languages of the mind: Essays on mental representation*. The MIT Press.

Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdés, J., Ossandón, T., & McDermott, J. H. (2019). Universal and non-universal features of musical pitch perception revealed by singing. *Current Biology*, *29*(19), 3229–3243.e12. https://doi.org/10.1016/j.cub.2019.08.020

Jennings, K. E. (2010). Developing creativity: Artificial barriers in artificial intelligence. *Minds and Machines*, *20*, 489–501. https://doi.org/10.1007/s11023-010-9206-y

Jordanous, A. (2009). Evaluating machine creativity. *C and C 2009 - Proceedings of the 2009 ACM SIGCHI Conference on Creativity and Cognition*. https://doi.org/10.1145/1640233.1640285

Jordanous, A. (2012). A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative. *Cognitive Computation*, *4*, 246–279. https://doi.org/10.1007/s12559-012-9156-1

Jordanous, A. (2016). Four PPPPerspectives on computational creativity in theory and in practice. *Connection Science*, *28*, 194–216. https://doi.org/10.1080/09540091.2016.1151860

Kant, I. (1998, January). *Critique of pure reason* (P. Guyer & A. W. Wood, Eds.). Cambridge University Press. https://doi.org/10.1017/cbo9780511804649

Kantosalo, A. (2019). Experience evaluations for human–computer co-creative processes–planning and conducting an evaluation in practice. *Connection Science*, *31*, 60–81. https://doi.org/10.1080/09540091.2018.1432566

Karras, T., Laine, S., & Aila, T. (2018, December). A style-based generator architecture for generative adversarial networks. https://doi.org/10.48550/arXiv.1812.04948

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2019, December). Analyzing and improving the image quality of StyleGAN. https://doi.org/10.48550/arXiv.1912.04958

Kaufman, J. C. (2016, March). *Creativity 101*. Springer Publishing Company.

Kaufman, J. C., & Sternberg, R. J. (2019). *Cambridge handbook of creativity*. Cambridge University Press.

Klütsch, C. (2012, September). Information Aesthetics and the Stuttgart School. In *Mainframe Experimentalism* (pp. 65–89). University of California Press. https://doi.org/10.2307/j.ctv114c7b3.9

Köbis, N. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior*, *114*. https://doi.org/10.1016/j.chb.2020.106553

Koster, J. (2009). Ceaseless, unpredictable creativity: Language as technology. *Biolinguistics*, *3*(1), 061–092. https://doi.org/10.5964/bioling.8683

Kowaliw, T. (2012). Promoting creative design in interactive evolutionary computation. *IEEE Transactions on Evolutionary Computation*, *16*, 523–536. https://doi.org/10.1109/TEVC.2011.2166764

Krenn, M., Buffoni, L., Coutinho, B., Eppel, S., Foster, J. G., Gritsevskiy, A., Lee, H., Lu, Y., Moutinho, J. P., Sanjabi, N., Sonthalia, R., Tran, N. M., Valente, F., Xie, Y., Yu, R., & Kopp, M. (2022, September). Predicting the future of AI with AI: High-quality link prediction in an exponentially growing knowledge network. https://doi.org/10.48550/arXiv.2210.00881

Kripke, S. (1980). *Naming and necessity*. Harvard University Press.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, & K. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (Vol. 25). Curran Associates, Inc. https://doi.org/10.1145/3065386

Lamb, C. (2018). Evaluating computational creativity: An interdisciplinary tutorial. *ACM Computing Surveys*, *51*. https://doi.org/10.1145/3167476

Latour, B. (2008). A cautious Prometheus?: A few steps toward a philosophy of design (with special attention to Peter Sloterdijk). *Annual International Conference of the Design History Society*. http://www.bruno-latour.fr/node/69

Latour, B. (1990). Technology is society made durable. *The Sociological Review*, *38*(1_suppl), 103–131. https://doi.org/10.1111/j.1467-954x.1990.tb03350.x

Latour, B. (1993). *We Have Never Been Modern*. Harvard University Press.

Laurence, S., & Margolis, E. (1999). Concepts and cognitive science. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 3–81). MIT Press.

Liapis, A. (2019). Orchestrating game generation. *IEEE Transactions on Games*, *11*, 48–68. https://doi.org/10.1109/TG.2018.2870876

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A review of machine learning interpretability methods. *Entropy*, *23*(1), 18. https://doi.org/10.3390/e23010018

Liu, B., Zhu, Y., Song, K., & Elgammal, A. (2021, January). Towards faster and stabilized GAN training for high-fidelity few-shot image synthesis. https://doi.org/10.48550/arXiv.2101.04775

Liu, H., Chen, Z., Yuan, Y., Mei, X., Liu, X., Mandic, D., Wang, W., & Plumbley, M. D. (2023). AudioLDM: Text-to-Audio Generation with Latent Diffusion Models. https://doi.org/10.48550/ARXIV.2301.12503

Lovitt, W. (1977). Introduction. In *The question concerning technology, and other essays*. Harper & Row.

Machado, P. (2002). All the truth about NEvAr. *Applied Intelligence*, *16*, 101–118. https://doi.org/10.1023/A:1013662402341

Mao, J., Gan, C., Kohli, P., Tenenbaum, J. B., & Wu, J. (2019, April). The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision. https://doi.org/10.48550/arXiv.1904.12584

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist*, *44*(12), 1469–1481. https://doi.org/10.1037/0003-066x.44.12.1469

Melechovsky, J., Guo, Z., Ghosal, D., Majumder, N., Herremans, D., & Poria, S. (2023). Mustango: Toward Controllable Text-to-Music Generation. https://doi.org/10.48550/ARXIV.2311.08355

Menabrea, L., Babbage, C., Lovelace, A., & L, A. (1843). *Sketch of the Analytical Engine invented by Charles Babbage with notes by the translator*. R. & J. E. Taylor.

Miller, A. I. (2019). Margaret Boden's three types of creativity. In *The Artist in the Machine*. The MIT Press. https://doi.org/10.7551/mitpress/11585.003.0007

Miller, M., & Lion, G. (2022). Reflecting through the photo's un-making. In *[ ] With Design: Reinventing Design Modes* (pp. 1990–2000). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-4472-7_130

Mufwene, S. (2013, January). In search of universal grammar: From Norse to Zoque. In T. Lohndal (Ed.). John Benjamins Publishing Company.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*(3), 289–316. https://doi.org/10.1037/0033-295x.92.3.289

Nake, F. (2010). Paragraphs on Computer Art, Past and Present. *Electronic Workshops in Computing*. https://doi.org/10.14236/ewic/cat2010.7

Nake, F. (2012). Information Aesthetics: An heroic experiment. *Journal of Mathematics and the Arts*, *6*(2–3), 65–75. https://doi.org/10.1080/17513472.2012.679458

Neves, J., Machado, J., Analide, C., Abelha, A., & Brito, L. (2007, December). The halt condition in genetic programming. In *Progress in Artificial Intelligence* (pp. 160–169, Vol. 4874). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-77002-2_14

Newell, A., Shaw, J. C., & Simon, H. A. (1959). *The processes of creative thinking*. RAND Corporation.

Norton, R. W. (1975). Measurement of ambiguity tolerance. *Journal of Personality Assessment*, *39*(6), 607–619. https://doi.org/10.1207/s15327752jpa3906_11

Oden, G. C. (1977). Integration of fuzzy logical information. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(4), 565–575. https://doi.org/10.1037/0096-1523.3.4.565

Olteţeanu, A.-M. (2015). ComRAT-C: A computational compound Remote Associates Test solver based on language data and its comparison to human performance. *Pattern Recognition Letters*, *67*, 81–90. https://doi.org/10.1016/j.patrec.2015.05.015

Olteţeanu, A.-M. (2016). Object replacement and object composition in a creative cognitive system. Towards a computational solver of the Alternative Uses Test. *Cognitive Systems Research*, *39*, 15–32. https://doi.org/10.1016/j.cogsys.2015.12.011

Osherson, D. N., & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, *9*(1), 35–58. https://doi.org/10.1016/0010-0277(81)90013-5

Pasquier, P. (2016). An introduction to Musical Metacreation. *Computers in Entertainment*, *14*(2), 1–14. https://doi.org/10.1145/2930672

Peinado, F. (2006). Evaluation of automatic generation of basic stories. *New Generation Computing*, *24*, 289–302. https://doi.org/10.1007/BF03037336

Plato. (1998, March). *Meno* (G. Anastaplo & L. Berns, Trans.). Focus Publishing Pullins & Co.

Pumarola, A., Agudo, A., Martinez, A. M., Sanfeliu, A., & Moreno-Noguer, F. (2019). GANimation: One-shot anatomically consistent facial animation. *International Journal of Computer Vision*, *128*(3), 698–713. https://doi.org/10.1007/s11263-019-01210-3

Putnam, H. (1970). Is semantics possible? *Metaphilosophy*, *1*(3), 187–201. http://www.jstor.org/stable/24434824

Putnam, H. (1975). The meaning of "meaning". *Language, mind, and knowledge*, *7*, 131–193. https://hdl.handle.net/11299/185225

Quine, W. V. (1951). Two dogmas of empiricism. *The Philosophical Review*, *60*(1), 20. https://doi.org/10.2307/2181906

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021, February). Learning transferable visual models from natural language supervision. https://doi.org/10.48550/arXiv.2103.00020

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022, December). Robust speech recognition via large-scale weak supervision. https://doi.org/10.48550/arXiv.2212.04356

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. https://d4mucfpksywv.cloudfront.net/better-language-models/language-models.pdf

Ragot, M. (2020). Ai-generated vs. human artworks. A perception bias towards artificial intelligence? *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3334480.3382892

Rey, G. (1993). The unavailability of what we mean: A reply to Quine, Fodor and Lepore. In *Grazer Philosophische Studien* (pp. 61–101). Amsterdam: Rodopi.

Rhodes, M. (1961). An analysis of creativity. *The Phi Delta Kappan*, *42*(7), 305–310. http://www.jstor.org/stable/20342603

Ritchie, G. (2007). Some empirical criteria for attributing creativity to a computer program. *Minds and Machines*, *17*, 67–99. https://doi.org/10.1007/s11023-007-9066-2

Roberts, A., Engel, J., Raffel, C., Hawthorne, C., & Eck, D. (2018). A hierarchical latent vector model for learning long-term structure in music. *ICML 2018*. https://doi.org/10.48550/arXiv.1803.05428

Rodríguez-Vidal, J., d'Errico, F., Pacheco, F. G., Blasco, R., Rosell, J., Jennings, R. P., Queffelec, A., Finlayson, G., Fa, D. A., López, J. M. G., Carrión, J. S., Negro, J. J., Finlayson, S., Cáceres, L. M., Bernal, M. A., Jiménez, S. F., & Finlayson, C. (2014). A rock engraving made by Neanderthals in Gibraltar. *Proceedings of the National Academy of Sciences*, *111*(37), 13301–13306. https://doi.org/10.1073/pnas.1411529111

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2021). High-resolution image synthesis with latent diffusion models. https://doi.org/10.48550/arXiv.2112.10752

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization* (pp. 27–48). Erlbaum.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Rosenberger, R. J., & Verbeek, P.-P. (2015). *Postphenomenological investigations: Essays on human—technology relations*. Lexington Books.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408. https://doi.org/10.1037/h0042519

Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., & Aberman, K. (2022). DreamBooth: Fine tuning text-to-image diffusion models for subject-driven generation. https://doi.org/10.48550/arXiv.2208.12242

Rumelhart, D. E., McClelland, J. L., & the PDPResearchGroup. (1987). *Parallel distributed processing: Foundations* (Vols. 1). The MIT Press.

Runco, M. A. (2014). *Creativity: Theories and themes: Research, development, and practice*. Elsevier Science & Technology Books.

Saunders, R. (2012). Towards autonomous creative systems: A computational approach. *Cognitive Computation*, *4*, 216–225. https://doi.org/10.1007/s12559-012-9131-x

Sawyer, R. K. (2017). *Group genius the creative power of collaboration: the creative power of collaboration*. Basic Books.

Schneider, F., Kamal, O., Jin, Z., & Schölkopf, B. (2023). Moûsai: Text-to-Music Generation with Long-Context Latent Diffusion. https://doi.org/10.48550/ARXIV.2301.11757

Schön, D. A. (1983). *The reflective practitioner*. Basic Books, New York.

Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C., Wightman, R., Cherti, M., Coombes, T., Katta, A., Mullis, C., Wortsman, M., Schramowski, P., Kundurthy, S., Crowson, K., Schmidt, L., Kaczmarczyk, R., & Jitsev, J. (2022, October). LAION-5B: An open large-scale dataset for training next generation image-text models. https://doi.org/10.48550/arXiv.2210.08402

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489. https://doi.org/10.1038/nature16961

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, *550*(7676), 354–359. https://doi.org/10.1038/nature24270

Simonton, D. K. (1985). Intelligence and personal influence in groups: Four nonlinear models. *Psychological Review*, *92*(4), 532–547. https://doi.org/10.1037/0033-295x.92.4.532

Simonton, D. K. (1994). *Greatness: Who makes history and why*. Guilford.

Simonton, D. K. (1999). *Origins of genius: Darwinian perspectives on creativity*. Oxford University Press.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Harvard university press.

Smith, J. R. (2017). Harnessing A.I. for augmenting creativity: application to movie trailer creation. *Proceedings of the 2017 ACM Multimedia Conference*, 1799–1808. https://doi.org/10.1145/3123266.3127906

Smith, M. R., & Marx, L. (1994). *Does Technology Drive History? The Dilemma of Technological Determinism*. MIT Press.

Sternberg, R. J. (1988). *The nature of creativity: Contemporary psychological perspectives*. Cambridge University Press.

Sternberg, R. J. (2003, October). *Wisdom, intelligence, and creativity synthesized*. Cambridge University Press. https://doi.org/10.1017/cbo9780511509612

Sturm, B. L. (2019). Machine learning research that matters for music creation: A case study. *Journal of New Music Research*, *48*, 36–55. https://doi.org/10.1080/09298215.2018.1515233

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2014, September). Going deeper with convolutions. https://doi.org/10.48550/arXiv.1409.4842

Turing, A. M. (1937). On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *s2-42*(1), 230–265. https://doi.org/10.1112/plms/s2-42.1.230

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433–460. https://doi.org/10.1093/mind/lix.236.433

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352. https://doi.org/10.1037/0033-295x.84.4.327

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017, June). Attention is all you need. https://doi.org/10.48550/arXiv.1706.03762

Verbeek, P.-P. (2005). *What Things Do: Philosophical Reflections on Technology, Agency, And Design*. Pennsylvania State University Press.

Verbeek, P.-P. (2011). *Moralizing technology: Understanding and designing the morality of things*. The University of Chicago Press.

Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., & Lang, K. (1989). Phoneme recognition using time-delay neural networks. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *37*(3), 328–339. https://doi.org/10.1109/29.21701

Wallas, G. (1926). *The art of thought*. Solis Press.

Wiggins, G. A. (2006a). A preliminary framework for description, analysis and comparison of creative systems. *Knowledge-Based Systems*, *19*(7), 449–458. https://doi.org/10.1016/j.knosys.2006.04.009

Wiggins, G. A. (2006b). Searching for computational creativity. *New Generation Computing*, *24*, 209–222. https://doi.org/10.1007/BF03037332

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, *109*(1), 121–136. Retrieved December 22, 2023, from http://www.jstor.org/stable/20024652

Wittgenstein, L. (1953). *Philosophical Investigations*. Wiley-Blackwell.

Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., Liu, X., Wu, Y., Dong, F., Qiu, C.-W., Qiu, J., Hua, K., Su, W., Wu, J., Xu, H., Han, Y., Fu, C., Yin, Z., Liu, M., ... Zhang, J. (2021). Artificial intelligence: A powerful paradigm for scientific research. *The Innovation*, *2*(4), 100179. https://doi.org/10.1016/j.xinn.2021.100179

Yang, L.-C. (2020). On the evaluation of generative models in music. *Neural Computing and Applications*, *32*, 4773–4784. https://doi.org/10.1007/s00521-018-3849-7

Zenasni, F., Besançon, M., & Lubart, T. (2008). Creativity and Tolerance of Ambiguity: An Empirical Study. *The Journal of Creative Behavior*, *42*(1), 61–73. https://doi.org/10.1002/j.2162-6057.2008.tb01080.x

Zhang, L., & Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. https://doi.org/10.48550/arXiv.2302.05543