CREATING MONSTERS

CRAFTING GENDER AMBIGUOUS CHILD TOYS THROUGH REFLEXIVE DESIGNER-AI INTERACTION

Masterthesis Anne Arzberger,TU Delft

Creating Monsters-
crafting gender ambiguous child toys through reflexive designer-AI interactions

Design for Interaction Masterthesis
Faculty of Industrial Design Engineering
Delft University of Technology

Author
Anne Arzberger
Student number: 5382084

Graduation committee
Chair - Prof. dr. Elisa Giaccardi
Faculty of Industrial Design Engineering
Department of Human-Centered Design
Mentor - dr. Maria Luce Lupetti
Faculty of Industrial Design Engineering
Department of Human-Centered Design
Company - oio.studio
Company mentor - Simone Rebaudengo

project as part of the DCODE Network

December 2023

PREFACE

Dear reader,

An exciting journey is coming to an end. Not having met many people due to the pandemic, I am very grateful to have met just the right ones. Before I would like to introduce you to my graduation project, I would thus like to thank a couple of people.

A big thanks to my project supervisors Elisa Giaccardi, Maria Luce Lupetti and Simone Rebaudengo for all the great feedback and inspiring conversations. You very much encouraged me to take on new challenges and go beyond what I thought would be possible. Special thanks to my super-supervisor Maria Luce Lupetti, without you my study life would have been only half as interesting, challenging and fun. I am very grateful for all the great opportunities you gave me.

I also want to take this chance to thank my honours supervisors Peter Lloyd, Senthil Chandrasegaran and Vera van der Burg, who introduced me to the world of research, AI and conferences. Your motivational, positive and constructive feedback was invaluable and made the honours project one of the key parts of my studies.

Of course studying abroad during a pandemic would have not been possible without my fantastic family and friends. Thank you Mama and Papa for always encouraging and supporting my crazy ideas. Without you I would have never had the courage to move to a new country to study at a big technical university.

A special thanks to my partner and best friend. Thank you Jesse for always sharing your excitement, curiosity and endless support with me. Your encouragement was the most important for me.

Last but not least, special thanks to my friends who made this (online) study as much fun as possible. I would have not believed that I would find such deep friendships in such a short time.

All the best,

Anne

EXECUTIVE SUMMARY

Growing up, social constructs like roles, norms and values are being internalised and naturalised. Despite offering a sense of stability, such constructs also prohibit equality, justice and diversity, by pushing people into categories, roles and norms they do not represent. However, once internalised, social constructs fall under the surface of awareness, making their mitigation and re-framing a complex task. This also poses a great challenge for designers, who often aim to create fair and inclusive futures for those marginalised and discriminated against.

Attempts are made to mitigate bias by introducing artificial intelligence (AI). Technology however, often acts as a double-edged sword, having the abilities to both identify and mitigate bias, or amplify inequality, reinforce existing stereotypes and increase injustice.

Recognising both the potential but also the limitations of AI, this thesis explores the idea of reflexive designer-AI interactions, as a new form of human-machine interaction towards more reflective design practitioners who are able to surface, dismantle and re-think personal and collective imaginings. Seeing a key role in reflection, often criticised behaviour of AI, like inconsistency, unpredictability and confrontation, are being explored as potentially meaningful for triggering critical and self-reflective thinking and decision making in design.

Following a speculative and introspective research through design approach, this thesis explores such reflexive interactions in the context of gender representation in child toys. The hypothesis is that the introduction of reflection and a change in mindset when engaging with AI, can be productive in terms of mitigating gender bias in child toys. In order to envision the situated designer-AI interactions, a speculative vision of the first gender fluid child toy company is introduced. This vision serves as a tool for presenting queer future AI-design practices, but also as a critique of current gender stereotypes in the design of children's toys.

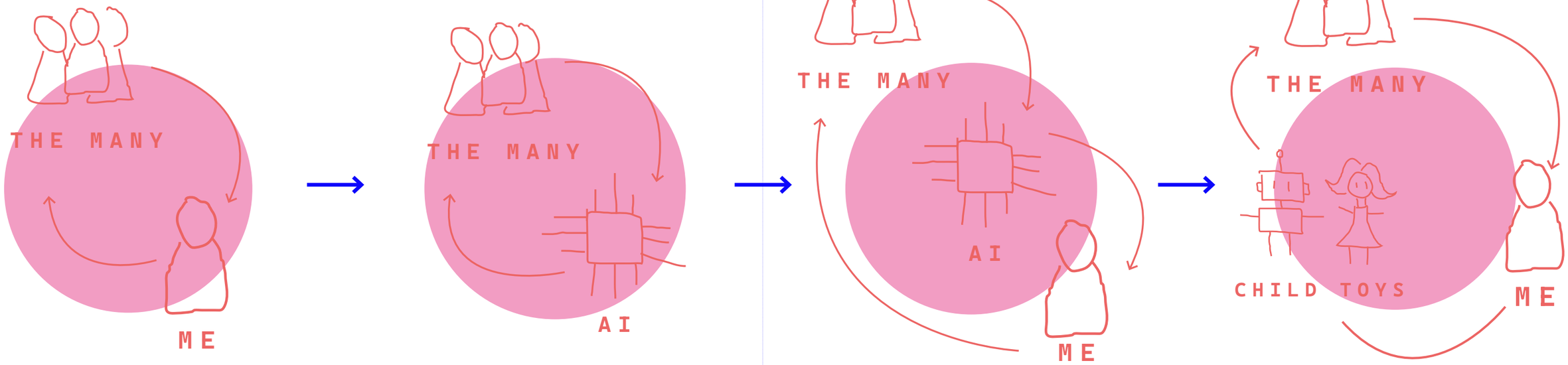
Oio.studio serves as an inspiration source for designer-AI collaborations, furthermore providing industry perspective in the adaptation of AI.

Technological exploration insights are translated into four design tactics, resulting in designer reflection and bias awareness. Those tactics are applied and explored in practice, by designing three gender-ambiguous child toys. Each toy represents a new reflexive design-AI workflow.

Each workflow differently illustrates human and non-human collaboration that surfaces, defamiliarizes and dismantles personal and collective imaginings of gender in toys. Additionally, these speculative practices also challenge the status quo in design, raising awareness about design and AI issues that need to be addressed further in the future.

Taking into account the insights from the experiments, as well as prototype testing with children and evaluating expert interviews, this project concludes that reflexive interactions – as proposed alternative to traditional human-AI interactions and in addition to the current design practice – are potentially productive to surface, dismantle and re-familiarize personal bias and collective imaginings. Furthermore, does this thesis suggest that AI's often negatively described behaviour like confusion and inconsistency, also carry the power to trigger reflective practices that help surfacing and challenge bias. However potentially limiting factors like ecological, economical and social cost as well as ethical concerns are discussed as well.

Figure 1 - project overview



problem:
perpetuation of collective imaginings

exacerbation:
normative human-AI interaction

research gap:
reflexive human-AI interaction

context:
stereotypical gender representation in child toys

INDEX

INTRODUCTION

01.1. INTRODUCTION

This thesis is about surfacing and dismantling social bias, as carried for instance in gender stereotypes, through reflexive designer-AI interactions.

Describing reality in mental categories is necessary and fundamental to the existence of all living beings (Gregor & Hackett, 2017). Ranging from binary categories like safe vs. dangerous to more complex forms of '*categorical discirmination*', they form the way in which we perceive reality (Hackett, 2019).

Arguing that cognitive functions originate in social interactions, Vygotsky (1987) disagreed with the idea that learning could be dissociated from its social environment, as held by cognitivists like Piaget and Perry. As a result, social categories and norms like gender roles and stereotypes, are being internalised naturally when growing up in society (Mareis, 2012).

Categorical thinking can be of use for maintaining the illusion of stability, on which social structures are based (Douglas, 2002). They can furthermore be seen as a product of human cognition, reducing the complexity of the world in order to allow humans to interact with it efficiently (Maqsood et al., 2004; Pfister et al., 2016). Afraid of ambiguity and uncertainty, individual members of a society usually obey to its categorical distinctions (Douglas, 2002).

However useful for keeping the cognitive load in everyday life low, internalised categories also create inequality and injustice, often for those already marginalised and discriminated against (Mareis, 2012; Canlı, 2018), by pushing people into categories that do not represent them. Trying to free oneself from those collective imaginings, designers often stumble over our own biases and internalised categories, failing in their ambitions to create diverse, inclusive and equitable futures (Moore Pervall, 2022).
Machine learning is being used more frequently in both the public and private sectors to make decisions, which has exposed new intersectional categories of protected identities as well as new categories of algorithmic discrimination (Mann & Matzner, 2019). Technological tools like artificial intelligence (AI), which are already widely used for inspirational purposes in design (Koch et al., 2019; Designs.ai - Creative Work Done Effortlessly, n.d.-b), also become more and more of interest as means to counteract our unconscious tendencies to categorise and discriminate (Arzberger et al., 2022; Turtle, 2022; AI Toys, n.d.-b).

Despite its many promising and interesting abilities, recent concern was raised about AI, further exacerbating social norms, values and categories by baking in the structures of power and prejudices, thus increasing injustice, rather than combating it (Mehrabi et al., 2021; McQuillan, 2018). Further issues arise due to inconsistent and unpredictable behaviour of AI (Amershi et al., 2019), calling for interaction guidelines that ensure meaningful human control over human-AI interactions (Cavalcante et al., 2022). Technology can thus be seen as a double-edged sword, having the power to both amplify inequality and further push people into categories that do not represent them, or potentially mitigate against such bias (De Cremer & Kasparov, 2022).
Traditionally being used in a more normative way, technology is often more agency and autonomy, giving it a more active stance in human-machine interaction. However, such interactions leave no room for reflection, which in turn is found to be a powerful weapon against unconscious and biassed decision making (Westberg & Jason, 1994). Despite AI's ability to exacerbate thinking in categories, e.g. gender identities and the fact that AI can misclassify or fail to detect gender identities that go beyond the binary distinction between male and female (Keyes, 2018), this work argues that binary and categorical thinking are human tendencies, not machine ones. As a result of the dominant normative approach, we lack discourse and alternatives for equity and inclusivity and how they are reflected in the tools we create and the interactions with these (Hagendorff, 2019). Technologies like AI or ML are not inherently bad, nor is the individual trying to utilise them. Hence, alternative ways are needed to complement current design and human-machine practices.

Contributing to filling this research gap, this thesis explores how the flaws of artificial intelligence – like bias exacerbation, inconsistency and unpredictability – can also be seen as an opportunity for human reflection and thinking beyond the categories we know. A more reflexive interaction between designer and AI is explored through several design experiments that aim to create a more reflective design practitioner who is able to surface personal and collective bias and dismantle the binary categories.

Whilst AI systems can be discriminatory towards race, abilities and other minorities or marginalized groups, the focus of this project lies on gender, as gender binary classifications are perpetuated through materialisation in a multitude od everyday products like clothing, tools or fast cars (Whisner, 1982;). Especially problematic are toys that, acting as 'cultural signifiers', introduce following generations early to narrowly defined gender identities (Almeida, 2017).

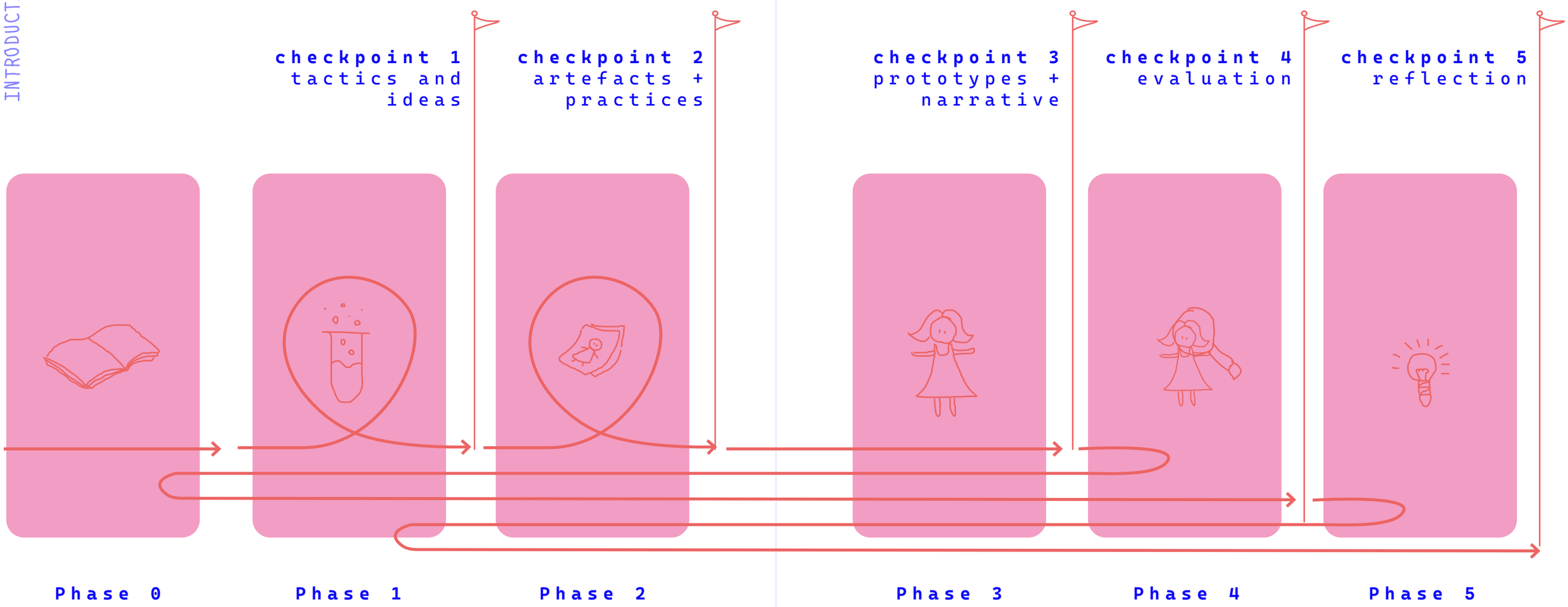
Using AI to dismantle our collective gender imaginings, this thesis introduces future design practices are formed through the creation of *'Monsters'* – queer and ambiguous toys that combine contradicting categories, thus challenge the collective stereotypes and de-famliarize the idea of the '*binary*' (Douglas, 2002) -, in order to recognize the identities beyond the binary categories of male and female. Children hereby form an especially vulnerable target group, who are often presented with heavily gender stereotypical toys, exacerbating gender categories and impeding the development of an individual identity (Wang & Degol, 2016) and early on training the next generation on common social norms, roles and values.
A definition of reflexive AI-design collaborations is seeked throughout the explorations and toy creations and further discussed and evaluated in interviews with experts from different domains and backgrounds, like toy design, traditional human computer interaction and queer futures.

Seeking inspiration, knowledge and discussions about alternative relations between designer and AI, this project is carried out in collaboration with oio.studio (Öiǫ / Studio, n.d.). Their knowledge and experience on innovative integrations of AI as a day-to-day design colleague and assistant will accompany me throughout my project.

As will be described in further detail in the Chapter 'Framework', this project follows a 'research through design' approach. Set up in five stages, alternative designer-AI collaborations will be first explored, and then tested in an introspective manner, to finally be illustrated and communicated through a speculative vision.

The setup of this project (as can be seen in Figure 2) was made in order to ensure the flexibility needed for exploring and curating future design practices. Phase 0 represents a really fundamental up-front literature study, meant to give a quick overview over current practices in HCI and design. However, the majority of knowledge and ideas is gathered during iterative explorations and experiments (Phase 1 + 2 in Figure 2). Phase 1 (as will be described in further detail in the chapter '*early explorations*') concerns all activities that help inspire, identify and understand designer-AI relations. Phase 2 builds on the knowledge and ideas generated in phase 1, by applying them to generate new practices and design ambiguous childtoys. In phase 2 the speculative vision will be created and further detailed, in order to contextualise the designer-AI experiments. In phase 3, the generated artefacts, practices and speculative vision will be further detailed and illustrated. In phase 4 the experiments will be critically evaluated. In phase 5–the last stage of the project – I will then discuss and reflect upon the outcomes and the process of this project.

Figure 2- project map



01.2. THE INITIAL ASSIGNMENT

The following describes the initial project idea as stated in the graduation proposal at the beginning of this project. As will be illustrated in the following chapters, the project has further evolved from this idea. This initial assigment closes with the personal motivation and ambition for this project.

01.2.1. INTRODUCTION

Designing is often referred to as a creative way of solving problems. However, the process of designing as well as the decision-making towards solutions is typically based on 'intuition', or as Michael Polanyi calls it 'tacit knowing' of the designer (Wong & Radcliffe, 2000). This challenges the ambitions of the design research community committed to generating knowledge and to making it accessible for other design practitioners. With a lack of awareness about one's own decision making furthermore comes the danger of introducing potential biases and fixations that can impede fair and inclusive designs.

With AI raising increasing concern about the perpetuation of biases in data by its tendency to bake in the structures of power and prejudices (McQuillan, 2018), it does not seem like a suitable technology to address the problem of tacit knowledge at first. However, the fact that AI can represent the biases of specific populations means that it potentially has a useful role to play in designing, exposing designers to viewpoints beyond the normal range of their experience and knowledge. The question therefore is: What can we learn about ourselves from machines?

The aim of this graduation project is therefore to address the challenges of tacit knowledge in design by reflecting on parts of the design process where tacit knowledge plays a decisive role. Focusing on the 'designers' language' the project will explore the design space evolving around the creation and communication of artefacts (e.g. sketches, prototypes, mood boards, etc.) and their inherent values, biases, fixations, etc.. As an active contributor to this reflective process, the AI will help to expose, confront and eventually understand and share one's tacit knowledge. The AI-confrontation is hereby not necessarily meant to ease the design process, on the contrary it is meant to irritate, surprise and provoke and thus create a more fair and inclusive way of designing.

However, these explorations and their effectiveness are bound to the accessibility of high quality off-the shelf AI tools. I am furthermore aware of the 'cost' that comes with the usage of AI and thus want to discuss the implications of my project on a system level at the end of my thesis, reflecting on the use of resources and other ethical consequences.

01.2.2. PROBLEM DEFINITION

Tacit knowledge is a 'key behind exercising judgment in human decision-making and employing intuition or 'gut-feeling" (Maqsood, & Armstrong, 2004). Tacit knowledge and experience can lead to effective intuitive decisions. However, there is a probability that 'tacit' decisions favour wrong judgment. The more complex, uncertain, and pressured a design process is, the higher the potential for bias, fixation and other cognitive heuristics, as an automatic reflex meant to lower the cognitive workload (Maqsood, & Armstrong, 2004).

Concluding that the ability to create fair and inclusive designs through effective decision making is closely bound to an understanding of tacit knowledge, the need for learning and understanding one's own subjective design world is being raised.

Knowledge though, is a very vague and 'messy' concept. Therefore, capturing it is a task fraught with difficulties. As Donald Schön describes it (Schön, 1992): 'designers know more than they can say, tend to give inaccurate descriptions of what they know, and can best (or only) gain access to their knowing-in-action by putting themselves into the mode of doing.' Not only are designers incapable of putting their knowledge into words, but one also typically remains unaware of their own experiences, interests and values that go into the formation of tacit knowledge and thus decision making. These underlying elements of decision making furthermore remain hidden for the designer's colleagues, collaborators, and next generations of design practitioners which can impede the share of knowledge.

01.2.3. ASSIGNMENT

To address the challenge of accessing and communicating tacit knowledge (and its inherent values, biases, fixations, etc.), I want to create an AI reflection and communication tool that exposes underlying patterns in the designer's language of artifacts. The project will unfold through practice-based explorations, where different ways of grasping tacit knowledge with an algorithm will be explored. Called upon to decipher (subjective) designerly worlds and its tacit knowledge, the AI activities will confront designers with traces of their own doing and evoke/provoke reflections on their patterns of action.

By changing perspective and looking at one's design practice through the lens of an AI, the designer can gain new insights into their own thought and decision process, as

well as the implications of the design's outcome. The tool should furthermore assist in passing on the gained awareness and understanding to others to help information flow and explained decision making.

The AI tool is expected to be explored in a conceptual way, mostly relying on of-the-shelf-AI tools, resulting in an experienceable prototype that can be used for testing and concept evaluation. The focus, however, should be on the outcomes of the designer's interaction with the AI. Therefore, the designer's creations, evolved through the joint design-AI confrontation, should be made tangible in a small physical exhibition.

01.2.4. PERSONAL MOTIVATION AND AMBITION

In previous projects, I explored the intersection of AI/ robotics/technology and design from an interactive perspective. Courses in Data Science and Machine Learning helped me to build up a theoretical and practical base to be further improved, challenged and applied in a project combined with design. In an attempt to look beyond the tools and methods designers use these days, main inspiration for my thesis project was taken from an AI-scholarship project about 'semantic design with AI' and the latest findings from my Honours Project 'using human-AI dialogue to stimulate problem framing in collaborative design' and my research elective focusing on 'non-human embodiments beyond anthropomorphism'.

Within the area of design, I hope to deepen my understanding of designerly ways of knowing, especially tacit knowing, both from a theoretical and a practice point of view. Outside the design discipline, I would like to learn about psychology, system and control theory (cybernetics) and AI (from a computer science background). Ideally, my knowledge and skills in data science, ML and programming can be improved throughout this project as well. I furthermore would like to challenge my usually very structured and thought through-way of working by following a very experimental - hands -on approach.

PROBLEM SPACE

Why do we fail do mitigate bias? This chapter discusses the nuances of bias perpetuation, from natrual internalization of collective imaginings, over automatic descision making, the limits of one's own perspective and AI bias.

PROBLEM SPACE

02.1. THE PROBLEM OF COLLECTIVE IMAGININGS

From a European perspective, today like never before, we live in a society that strives to enable all its members to live a life that is both fair and comfortable. Current headlines bear witness to the failure and complexity of this ambitious goal - from the restriction of self-determination for women in America (Sargeant, 2022) to the killing of people with different skin colors (Bunn, 2022) or sexual orientation (Iraq: Impunity for Violence Against LGBT People, 2022). The most frequently affected are also the least privileged ones, '*the ones already discriminated, disenfranchised and marginalised by the hegemonic order due to their gender, sex, sexuality, race, ethnicity, class, nationality, religion, ability, mobility, age and other social status and identity attributions*' (Canlı, 2018).

As designers, we often claim to be the advocates of those disadvantaged and marginalised. However, certain measurable barriers in conceptual design processes can impede those designerly ambitions of creating fair, equitable, inclusive futures. Bias –the '*inclination or prejudice for or against one person or group, especially in a way considered to be unfair*' (Oxford Languages and Google - English | Oxford Languages, 2022b) –is reflected in many forms in the design process. Task distribution, attribution, credit, and even the strategy of design teams can be impacted by biases resulting from discrimination towards people based on their gender, colour, or sexual orientation (Pfister et al., 2016). Design fixation for instance is known to cause a '*blind adherence to a set of ideas or concepts limiting the output of conceptual design*' (Jansson & Smith, 1991). Initial problem construal (frame formation) happens from projecting previous heuristics, allowing the creation of bias (Knoblich et al., 1999). Others show proof of bias for instance in the way we perceive colour (Olkkonen et al., 2014), use gender in conversational agents (Feine et al., 2020), or prefer information that confirms our existing world views (Confirmation Bias in UX, n.d.)[1]. However, in specific contexts, the human cognition uses prior knowledge and assumptions to interpret the world, especially when presented with complex and ambiguous input (Adams et al., 2004), or in emotionally charged situations (Epstein, 1999).

[1] *many more biases are known and illustrated for instance in Desjardins infographic (2021)*

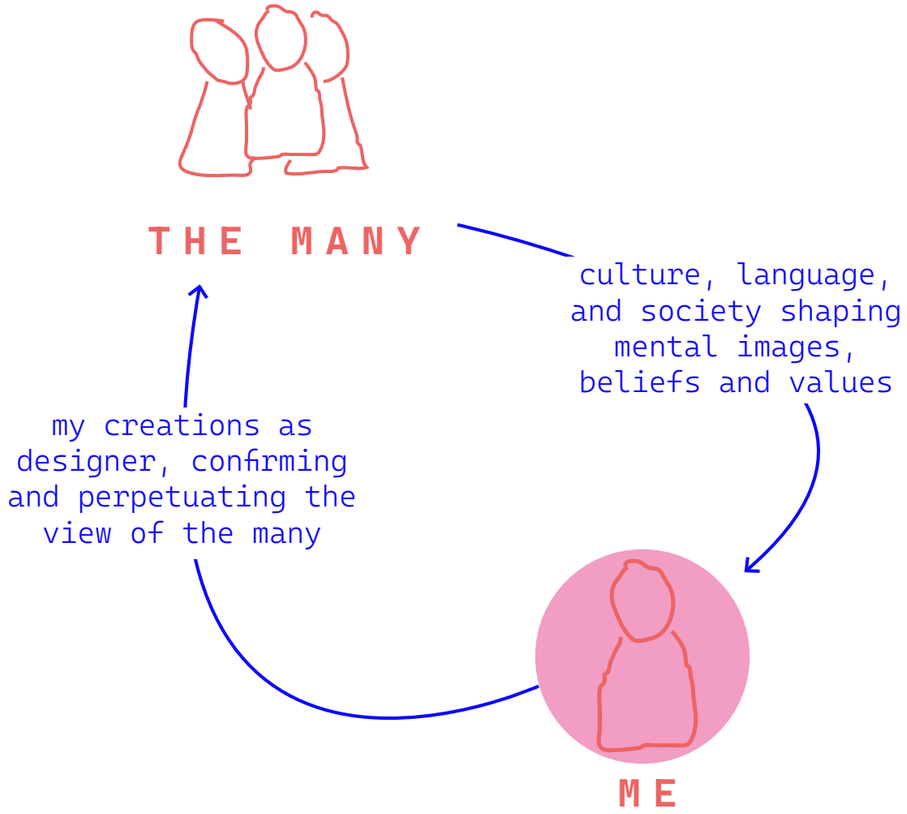


Figure 3 - problem of collective imaginings

Merely being aware of bias is insufficient to combat it. According to scientists, 95% of our decisions are made automatically and without conscious thinking (Zaltman, 2003). Because the tremendous amount of information that enters the human brain every second cannot all be processed consciously, a significant fraction is processed automatically and is thus concealed from our consciousness. As a result, we often choose passive, emotionally driven, and thoughtless decisions (Thaler & Sunstein, 2011; Pfister et al., 2016). Part of this susceptibility of human thinking to biases thus lies in the limitation of human cognition and its confined ability to perceive and interpret the surrounding, as well as to identify and challenge habituated and naturalised information (Maqsood et al., 2004; Mareis, 2012; Jansson & Smith, 1991).

There are tools for confronting bias, like the Harvard University bias test (Take a Test, n.d.), that measures bias through implicit associations. The assumption tested here is again, that bias is rooted mostly in unconscious thinking, thus most present in quick and automated decision making (Thaler & Sunstein, 2011; Pfister et al., 2016). The mission of Project Implicit is to educate the public about bias, by having people quickly link words to images. In a little self-experiment, I challenged my own gender perception through the Harvard University bias tests (Take a Test, n.d.). I discovered that I myself have a tendency to connect household more to the female and career more to the male (Figure 4).

Despite the the limits of the designer's individual knowledge, perception and experience[1] , it is the designers collective and tacit knowledge that impedes their good intentions (Jansson & Smith, 1991). Collective imaginings refer to a common, sometimes implicit, picture of the world, its dominant narratives, and its social, cultural, and political ramifications (Søndergaard & Hansen, 2018). From early childhood on, the limits of our individual behaviour, perception, and thought are predetermined by our habitus[2] (Mareis, 2012). Learning is the process by which learners

[1] The designer essentially considers information, knowledge and experience that is already familiar (Jansson & Smith, 1991)
[2] According to Bourdieu's definition of the habitus, all conventions, physical prowess, aesthetic and cultural preferences, and other non-discursive components of knowledge that are taken for granted by a particular social group are included (Mareis, 2012).

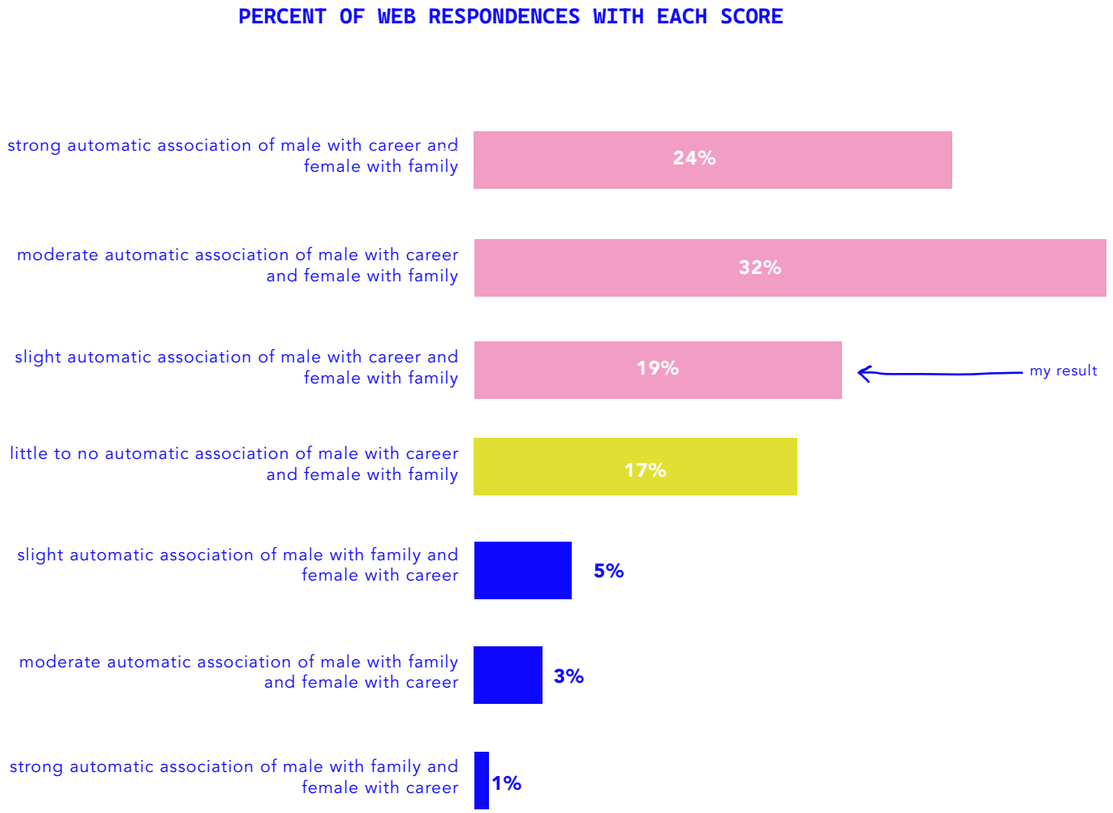


Figure 4- my gender bias (Take a Test, n.d.)

are incorporated into a knowledge community and introduced to collective imaginings and social constructs like gender; it does not just consist of assimilation and accommodation of new knowledge by learners. All cognitive processes are derived from social interactions (Vygotsky, 1978). Unlike knowledge that is taught or acquired formally, societal norms, values and stereotypes are learned unconsciously through observation and practice (Epstein, 1999). Design educators have historically been working designers who impart their knowledge, abilities, and values to students through an apprenticeship process. In small projects, design students '*play out*' the job of the designer while receiving instruction from more seasoned designers (Cross, 2006).

The values and norms thus not only remain hidden, their internalisation furthermore leads to their naturalisation. Once accepted as natural they become forgotten (Mareis, 2012). The acquired knowledge about societal constructs like norms, values, etc. is not openly stored, concrete rather than abstract and intuitive in nature (Wong & Radcliffe, 2000; Schön, 1984; Epstein, 1999). Also described as tacit knowledge, it '*could be available to conscious awareness and yet typically remains unarticulated*', making values and norms primarily experienced through someone else's actions (Wong & Radcliffe, 2000). Such learning through observing and doing allows bias and collective imaginings to remain unquestioned. As a consequence of this habituation and naturalisation, individual members of society, and especially designers, unconsciously perpetuate the view of the many.

A society furthermore has mechanisms to maintain its value system. For example, '*Taboos*', as Mary Douglas (2002) calls them in her book '*Purity and Danger*' , represent a spontaneous device for protecting distinctive categories[1]. Categories thus represent a necessity to maintain the illusion of stability, on which social structures are based. They furthermore allow the human brain to make fast and automated decisions, thus keeping the cognitive

[1] *Categories can be seen as a necessary classification of reality. A way to simplify the world, in order to make it more accessible and understandable for us humans. Categories however, always remain artificial and can never truly capture the reality in its complexity.*

load low (Thaler & Sunstein, 2021; Pfister et al., 2016). They reduce intellectual and social disorder (Douglas, 2002). '*Taboos*' protect the local consensus and collective imaginings. If something does not correspond to the predetermined categories, this triggers ambiguity and cognitive discomfort in the individual. Taboo forces the ambiguous into the sacred category. Threats and promises that are blatantly ludicrous are used to induce compliance, particularly in childcare. Breaking taboos can be dangerous and credible to reasonable people, if a taboo upholds morality or propriety (Douglas, 2002). As a result, an individual that is part of a value system, always strives to classify elements in one's environment into suitable, fitting and norm-compliant categories. Our prejudices reveal that we only have a fragmented understanding of the world.

Traditional inclusive design strategies, such as rehabilitation[1] design and design by storytelling[2], have come under fire for stigmatising populations with disabilities and creating specialised products that fail even when used by the target market (Keates et al., 2000). The goal of models for inclusive design is to routinely include data regarding end users who might be excluded by the developed artefacts (Keates & Clarkson, 2003).

Concluding the idea of the design advocate, we can say that however well intended the designer, design remains the site of prejudice, bias, misunderstanding, and fixation. Certain norms, values and categories are first internalised through habituation, and then, as the regulation mechanisms of a discourse remain silent, executed and perpetuated unknowingly (Mareis, 2012).

[1] Creating particular artefacts to address certain impairments.
[2] Before design and assessment, there is a process of observation and „understanding".

02.2. THE NORMATIVE HUMAN–AI INTERACTION

In an attempt to circumvent and overcome the biases of the many, and to rationalise our decision-making processes, the normative approach advertises the idea of technologies that can lead us beyond the frontier of human cognition.

Joint problem solving with machines has a long history. First mentioned by Licklider in 1960 (Licklider, 1960), it has been increasingly taken up in both theory and practice (Brynjolfsson & McAfee, 2014; Gerber et al., 2020). '*Man-Computer Symbiosis*' as it was first called, was thereby defined as '*close cooperative relationship between humans and machines that would be capable of thinking in ways no human brain had ever done and process data that machines of the time could not handle*' (Licklider, 1960; Gerber et al., 2020).

Today, innovations like artificial intelligence (AI) – the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages ('*Artificial Intelligence*' n.d.) – have increased the opportunities for such collaborations. Whereas computers for instance started beating humans in chess, it was proven not shortly after the firstly defended Kasparov, that humans started forming alliances with the '*enemy*'. This symbiosis proved to be even more powerful than a standalone performance by just a computer or just a human (Huang, 2022).

In design, recent research is for instance investigating the '*Co-performance*' and role of artificial agency in the design of everyday life (Kuijer & Giaccardi, 2018), or the emerging practices in designer-AI collaboration (van der Burg, 2022).

The question of how designers are influenced and affected by this shift in the design practice is also investigated in recent studies (Stembert & Harbers, 2019). AI has been applied in creative work in a wide range of scenarios, ranging from analysing creative work (Maher & Fisher, 2012), exploring the form language of a given product (Burnap et al., 2016), or producing ideas in generative design (Kazi et al., 2017).

Both the public and commercial sectors are increasingly using machine learning for decision-making, which has exposed new intersectional kinds of protected identities as well as new types of algorithmic discrimination (Mann & Matzner, 2019). Meant to augment our human limitations, those algorithms are found to '*bake in*' the social structures of power and prejudices, thus amplifying our biases

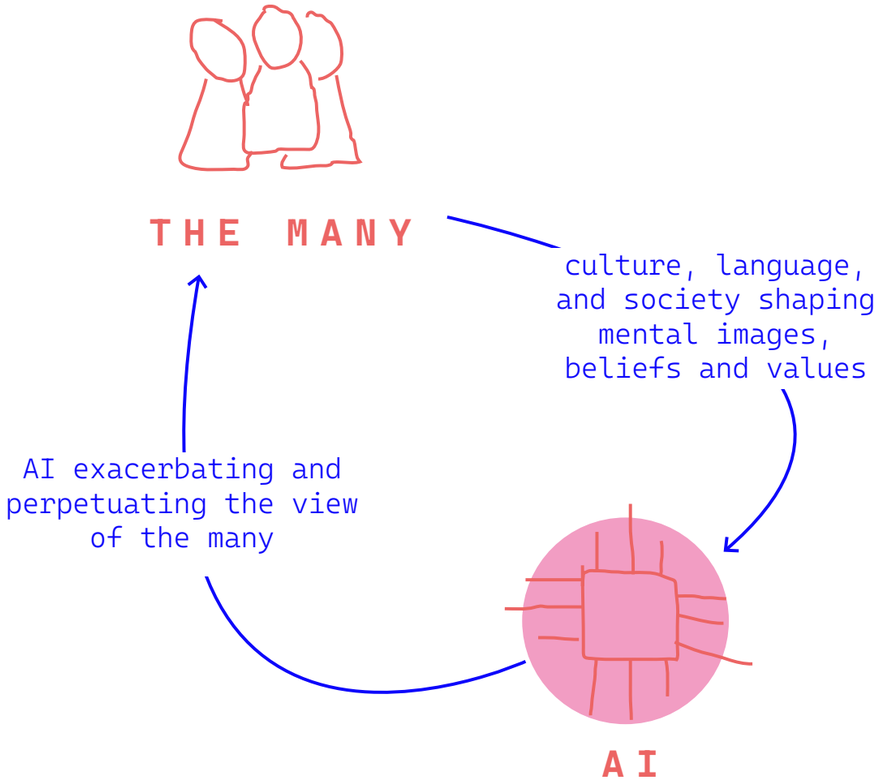


Figure 4 - the paternalistic human machine interaction

(Mehrabi et al., 2021; McQuillan, 2018). AI, even dedicated its own term, '*AI Bias*'[1], was, for instance, responsible for the perpetuation of race inequality in US healthcare. An American hospital used an algorithm to predict which patients would need additional medical care. Although race was not included as a demographic variable in the system, the algorithm showed a clear preference for white over black patients. The problem: cost/price of insurance policy was taken as the crucial variable for prediction treating white patients preferentially, as black people invest less in healthcare for a variety of reasons (Shin, 2021b).

Nevertheless, it is not the algorithm that is inherently bad or dangerous (Mehrabi et al., 2021). Failures of implementations have suggested that the algorithms pick up on all patterns in the data that they are trained on, necessarily including any of all human biases that guided and influenced the process of data collection and structuring. Whilst trying to create a desired future –in the case of the US hospitals, a fair system predicting additional care for patients– we provide the algorithm with data from the past we are trying to overcome (see Figure 7 -Timeline of approaches, for further illustration). Thus the algorithms are doomed to reinforce past biases of society (Enninga, 2022). In this sense, these technologies exhibit '*function creep*' (Dahl & Saetnan, 2009): they promise one thing, but as a result of the design, something else – often less ethical – sneaks in. These things, whether deliberately so or not, are political entities that raise certain ethical and philosophical issues that we must also explore via design. (Søndergaard & Hansen, 2018)

Limiting in creating useful data is our human tendency to think in categories that represent internalised norms, roles and values. Whilst we as humans experience a discomfort with non-categorical thinking, or an aversion to chaos itself, the technology does not, and is as such not inherently prawn to categorisation. For example, a generative algorithm (e.g. generative adversarial network (GAN)) is able to generate an almost infinite number of images that fall between two categories (Pieters & Wiering, 2018). '*Latent space*', a 2D representation of a multidimensional space, allows its users to explore this infinite space of elements that are all related to each other.

[1] *'AI bias is the underlying prejudice in data that's used to create AI algorithms, which can ultimately result in discrimination and other social consequences' (Shin, 2021b).*

Nearby things are hereby more closely related than distant ones. As Philipp Schmitt (Schmitt, 2019) describes his experience with this open space: '*Right here, I am king. Subtract the vector for man, add woman and I become queen. Somewhere else, I am Madrid. Subtract Spain, add France, and I become Paris*'. Even classification algorithms do not really show clear decisions for one or the other category, but rather probabilities that always express a certain ambiguity (Braak, 2021). Similar are generative language algorithms that are based on probability distribution, i.e. words and sentences are sampled according to their probability arising in the training data set, thus never expressing a categorical certainty (undefined [Computerphile], 2019).

Concluding that categories and biases are human, not technological tendencies, even the best algorithm perpetuates bias, if not trained on non-human data. The solution proposed by engineers is simple in idea: better technologies. As a result, we lack discourse on '*AI ethics such as care, equity welfare, or ecological networks*' (Hagendorff, 2019), and consideration of '*alternative alignments in design between humans and nonhumans*' (Coulton & Lindley, 2019; Giaccardi, 2020) as reviewed by (Nicenboim et al., 2020).

PROJECT APPROACH

How can we translate concerning AI behaviour into a useful design material? This chapter introduces the idea of reflexive interaction as an alternative human-AI collaboration that can help surface and dismantle current norms and roles. The context of designing for child toys is introduced, based on the understanding that the normative gender representation in toys has to be addressed.

PROJECT APPROACH

03.1. PROPOSING REFLEXIVE INTERACTIONS: A NEW APPROACH

Positioned within the emerging space of addressing equity, inclusion and diversity in and through AI, this thesis proposes a reflexive approach to human-AI interaction. The key component in bias mitigation is hereby the process of reflection. AI is seen in a assisting and reactive, rather than proactive and normative position. The fact that AI can present collective ideas and stereotypes has a potentially valuable role to play in designing, exposing designers to viewpoints beyond the typical range of their experiences and knowledge as well as surfacing assumptions and collective imaginings that otherwise remain hidden. Additionally, understanding AI as a probability based tool, uncertainties can help exploring spaces beyond and between existing categories. By surprising and confronting, the AI becomes an assistant for the designer, helping in surfacing and dismantling internalised and hidden constructs of society. The work in progress definition for reflexive AI interactions– as it will be re-defined throughout this work– is: '*a form of human-machine collaboration, where the AI is responsible for triggering and assisting the designer's process of identifying and challenging bias and collective imaginings, rather than actively proposing the ideal solution itself.*'

By introducing AI as a trigger for reflection, the goal of this project is to train more reflective design practitioners, who can in turn better identify and challenge bias and collective imaginings. The reflexive proposal uses technological limitations like exacerbation of the structures of power, mental images and stereotypes, as design material, and combines them with the AI's ability in representing the ambiguous spaces between categories –the *'in-between'*– as well as uncertainty, thus proposing an alternative perspective in the discussion about biases and the processes counteracting them.

Reflection hereby takes a key role in the proposed '*reflexive interactions*'. Designers work in specific contexts, with certain materials, and in a particular medium and language (Schön, 1984). In the process of making physically embodied thoughts, ideas, values, etc., consequences other than those intended can occur. Reflection is thereby a critical element, already familiar to designers. By reflecting, the designer steps into conversation with these unintended elements, and reshapes the situation, gaining new appreciation and understanding. '*In a good process of design, this conversation with the situation is reflective*' (Schön, 1984).

The designer considers not only the current choice but also a tree of subsequent choices that it leads to, each of which has a different meaning in relation to the sys-

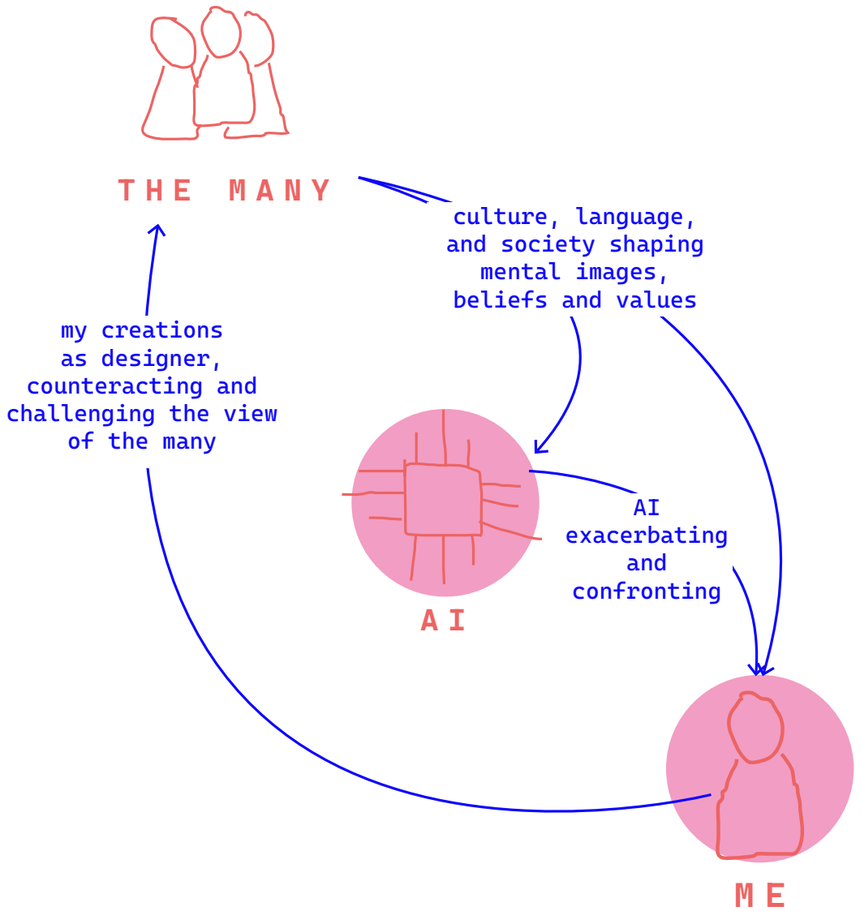


Figure 5 - the reflexive human-machine interaction

tem of implications set up by earlier moves. This is known as reflecting-in-action on the situation created by the designer's own previous design moves. As a result, the design practitioner also conducts research (Schön, 1984). Only then situations can be fully understood, generalised and adequately applied under different circumstances. As the designer frames a challenging scenario, the designer '*does not keep means and ends separate*', but defines them simultaneously (Schön, 1984). Thus, knowledge can only exist in connection to the one perceiving and employing it, rather than independently (Epstein, 1999). Studies comparing designerly ways of working with more science-like approaches came to similar conclusions. Whereas the scientist adopts a problem-focused strategy, the designer learns about the nature of the problem in a reflexive '*trying out*' kind of process (Cross, 2006). While such an iterative and reflexive problem-solving approach partially stems from a '*learning through observation and doing*' approach, the ill-defined, nonlinear and unstructured nature of design problems (Rittel & Webber, 1973) simply requires a more iterative approach.

Studies with medical professionals and physicians furthermore show that only with a certain level of self-awareness and self-knowledge, as it is the result of self-reflection and mindfulness, a practitioner is able to understand, communicate and express their core values (Epstein, 1999). Practitioners with a high capacity for such critical self-reflection in all aspects of their practice are shown to be more present with the user, better at solving problems, eliciting and transmitting information, making more evidence-based decisions, and defining their own values (Westberg & Jason, 1994).

The growth of self-identity, self-awareness, and personal agency are all significantly impacted by the changing relationship between oneself and one's work (Billett, 2010). The process of knowing how to learn and accepting one's key role in their own learning are thereby both facilitated by reflection (Brockbank & McGrill, 1998). A worker's ability to contribute to their role at work and the development of their long-term career goals are improved by working toward being a reflective practitioner (Schön, 1984), (Heyler, 2015).

This project argues that interactions with machines are essentially reflexive in nature. They represent a mirror held in front of us. Lacan (Jacques Lacan (Stanford Encyclopedia of Philosophy), 2018) describes mirroring— the looking at yourself through a reflexive shiny surface– as not just a visible physical phenomenon alone. Moreso, he sees in a mirror an '*image*' of oneself, a '*conveyed sense*

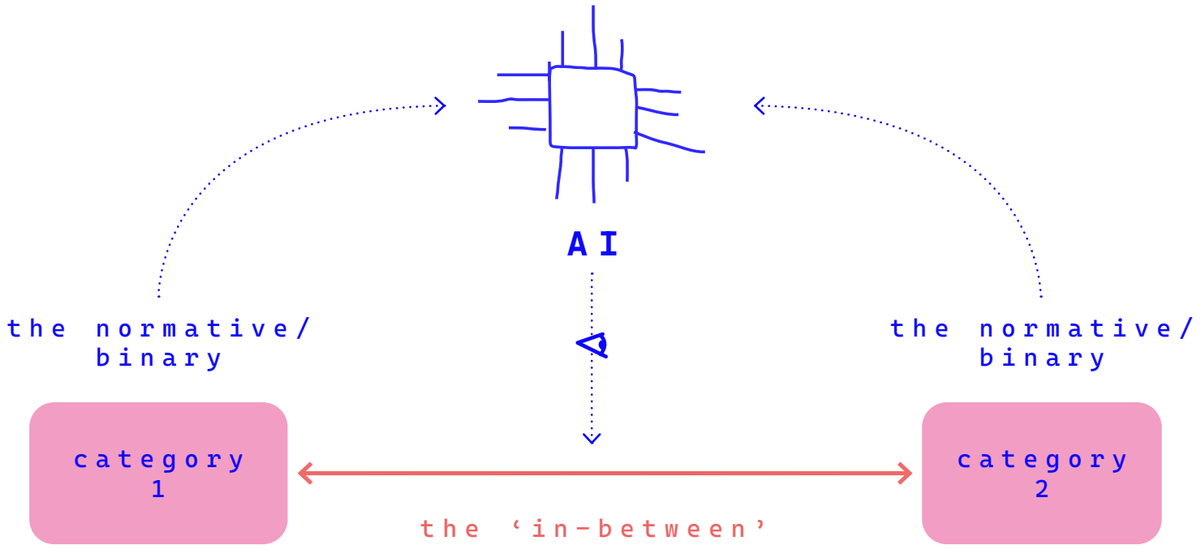


Figure 6 - AI assisting in exploring the 'in-between' social roles and categories

of how one "appears" from other perspectives' (Jacques Lacan (Stanford Encyclopedia of Philosophy), 2018). Recent work in design research, '*the mutant in the mirror*' (Turtle, 2022), has proven that an AI is capable of such mirroring. Others described their interaction with an AI like the following: '*Where I was trying to find an 'objective' speaker in [the AI], I ended up with the most subjective speaker of all. The circular interpretation between [the AI's] output and my interpretations resulted in nihilism (Why am Idoing this and when does it end?) only because the communication ended in a return-route towards myself*' (Van der Burg, n.d.). Utilising and expanding this reflexive nature, this work argues that we can use AI in new ways, by focusing on the AI's abilities that visualise spaces between categories, rather than using it to exacerbate our social constructs.

Stepping into conversation with machines, furthermore points to a new essential role of reflection. '*Behaviour by people and/or machines that is not guided or scripted in advance by designers and analysts, but emerges through discussion, experimentation, adaptations, and workarounds in groups or communities of practice. Emergent behaviour in relation to business processes and activities may occur even when the details of user interfaces are highly engineered* (Alter, 2010). Emergence occurs from the dynamic interaction between different organisms in an environment and can't be inferred from observing the isolated behaviour of an individual organism in that environment. Emergence, as a natural result of human-machine interaction thus calls for 'reflection-in-action'. Newly emerged properties have to be evaluated and taken into account, in order to allow human-machine interaction to live up to its full potential (Ghajargar et al., 2018).

However, emergence remains challenging. New processes are required in order to allow better judgement of what emergent properties to integrate and which to eliminate. Methods like research through design, with its flexibility in adjusting to surprising events, can provide a way of approaching the design for human-machine symbiosis (Stappers & Giaccardi, 2017).

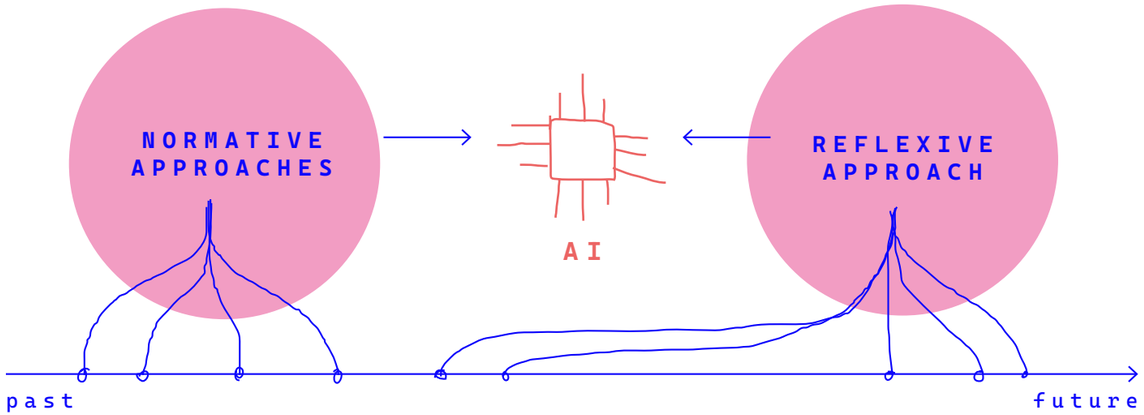


Figure 7 - timeline differentiation of approaches

Despite the critical role of reflection in the design process, one has to be aware that '*reflecting*' is usually associated with '*looking back*' and examining the past (Heyler, 2015). Whilst this can lead to an understanding about past mistakes and how to avoid them, it also carries the chances of repeating and perpetuating those flaws and mistakes that have already been naturalised and forgotten. Reflexive collaborations with machines are furthermore limited. Although often of great value, interaction properties can emerge, that can't be accounted for upfront:

The reflexive or generative approach furthermore requires a data-rethinking, modelling data that can represent or trigger[1] desired futures, rather than perpetuating outdated pasts. Figure 7 visualises this shift. Curating and generating those desired data also poses new challenges that have yet to be discovered.

[1] *Triggering the desired future can also require the conscious incorporation of past data. Unlike in the normative approach, this is done only intentionally.*

03.2. PROVIDING CONTEXT: QUEER FUTURES THROUGH AI

The problem and solution with regard to categories can be well illustrated and clarified using the example of gender. Unlike biological distinctions (*'sex'*), differences in gender (how someone internally identifies with social gender constructs, regardless their biological sex) are shaped by society (Butler, 1986). In almost all societies the idea that gender can be divided into two separate categories that are opposed to each other dominates: the masculine and the feminine. This claim may seem absurd to someone who is conflicted about their own gender; to such a person, it may seem preposterous that one's gender is established by the interpretation of others (Green, 2004). Not only can such binary distinction increase inner conflict for those who do not feel norm conform, they also form the basis for structural and social inequalities (Ravanera, 2020; How Dismantling the Gender Binary Can Help Eradicate Inequality, 2021). The gender binary can also drive analysts to perceive things via binary lenses when they would be better understood in other ways since it is a core ideological construct (Eckert, 2014).

Despite these primarily binary thought structures, it is also true in the area of gender that reality is clearly more plural and diverse. In her famous work '*The second Sex*' (Beauvoir et al., 2011), Beauviour also poses the question of what the ‚female' actually represents. Of course the answer to this question is as unclear as it is immutable.

Since Simone de Beauvoir described her idea of the woman as the second sex (Beauvoir et al., 2011), socially created vulnerability of women and non-binary people moved forward into the digital age, founding themselves perpetuated in modern data science (Beauvoir et al., 2011; Perez, 2021; D'Ignazio & Klein, 2020). The depiction of gender relations in advertising subtly and often unintentionally reflects and colours our understanding of cultural values, belief systems and social norms (Cortese, 2015; undefined [Filmanalyse], 2022). Modern media such as Instagram further disseminate these gender ideals (Sofia P. Caldeira et al., 2018), thus increasing the pressure on individuals to conform to the generally accepted categories of '*male*' - assertiveness, short hair and patriarchal roles - and '*female*' -e.g. grace, long hair and matriarchal roles.

Looking at the formation of individual identity, the fact that identity reflects the power relations that are part of the social practices of inclusion and exclusion, makes the formation of gender identity a great subject to the perpetuation of norms and values in a society (Paasi, 2000). Toys have evolved into '*cultural signifiers*' (Almeida, 2017) that symbolise contemporary society norms and values because of their positioning within prevalent narratives and stereotypes of interaction and how they get contextually configured in usage. These early socio-cultural experiences, that accumulate over time, perpetuate existing

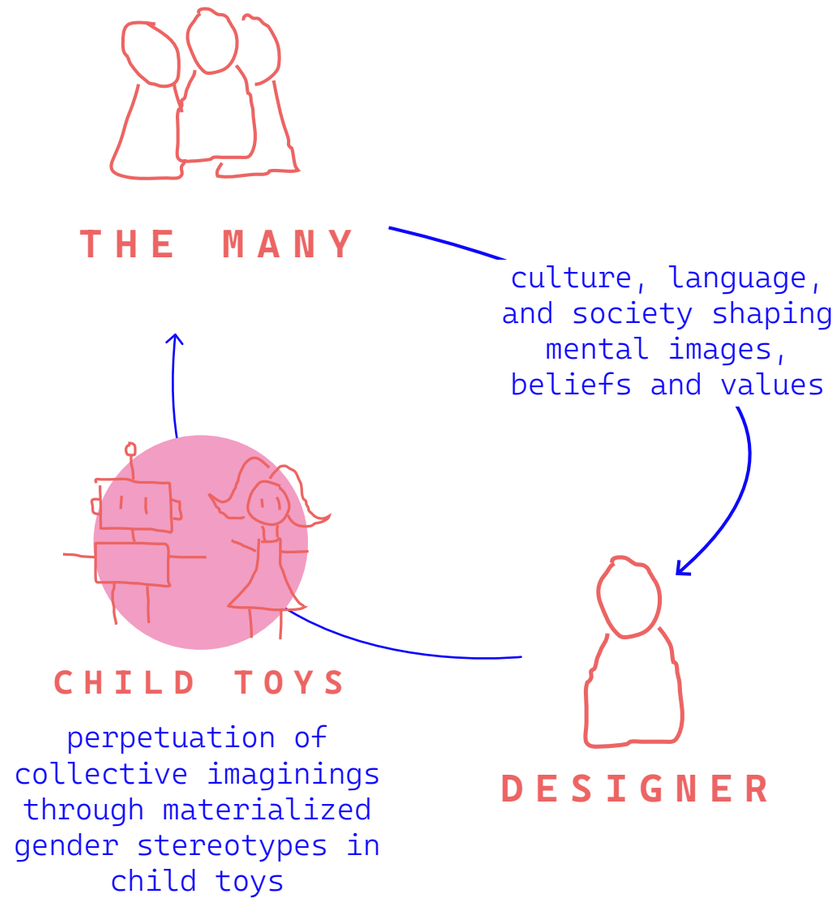


Figure 8 - toys as materialization of collective imaginings

gender norms and influence the formation of cognitive abilities and motivation of children (Wang & Degol, 2016).

Not until very recently, toy companies started searching for new narratives that somehow reflect new social realities. Attempts witnessed reach from new strong female disney heroes to Barbie advertisements presenting new female futures (Almeida, 2017; undefined [Barbie], 2015). More artsy approaches try to exaggerate current social norms in toys, already by using AI (AI Toys, n.d.-b). However, envisioning a new generation of child toys remains a challenge, since toys tend to reveal a great deal about the conceptions and representations of a culture where up to now, binary stereotypical gender ideas remain dominant (Almeida, 2017). A speculative approach seems needed, that can help envisioning toy futures beyond the traditional binary of masculine and feminine narratives.

Queer futures propose such a speculative tool. Opposing the idea of binary gender, '*queerness*' refers to the necessarily '*indeterminate, ambiguous and always in relation, denoting flexible spaces for the expression of all aspects of non, anti, contra, straight, cultural production and reception'* (Turtle, 2022). Since queer is the non-normative, it is per definition always in the future, a '*basic desire to live otherwise*' (Muñoz, 2019). '*Queer futurity, is less about expanding a range of choices (liberal freedom) than it is about transforming the kinds of beings we desire to be while embracing the [..] multitudes and entangled life-worlds that make up the pluriverse [...]*'(Turtle, 2022).

In pairing up with the AI, the idea is to get intentionally lost in ambiguity, in order to re-define what the '*masculine'* and the '*feminine*' or the in-between and beyond means to me as an individual. Re-framing those categories through those new and surprising worlds AI can offer, I want to integrate uncertainty and ambiguity in the artefacts I am designing, keeping the social implications and affordances deliberately low, in order to open up the categories not just for myself, but for the children and parents engaging with those artefacts as well. Uncertainty thus becomes a powerful tool to redefine our collective imaginings, rather than a weakness (Preciado, 2021). As Turtle states (Turtle, 2022) '*In this way, queerness is just as much about playing with the kinds of beings we desire to be as much as it is about offering a critique of the cultural, sociotechnical systems that shape our becoming with AI*. This makes a queering AI a tool to challenge our collective imaginings, and re-imagine what things can and have to be.

RESEARCH FRAMEWORK

A combination of research approaches and design research methodologies is chosen according to the projects layout. This chapter introduces the research through design approach, combine with speculative design elements and introspective design research. Limitations of this research framework are discussed and additional methods presented.

4

RESEARCH FRAMEWORK

04.1. RESEARCH THROUGH DESIGN APPROACH

To form my own interaction design research approach within the HCI research, I combine the research through design approach with an intersection of creative AI tools, introspection and speculative design methods (see Figure 9). Addressing the limitations of introspective research through design, additional generative research and user testing was applied when needed.

Research and Design or science and practice, were long seen as two distinct categories. However, as traditional scientific research is found to be unable to capture and address the projective, imaginary, and uncertain spaces design practice and research deal with (Prochner & Godin, 2022), design activities and artefacts are now becoming more and more accepted bodies of generating and communicating knowledge (Stappers & Giaccardi, 2017; Edelson, 2002). Research through design (RtD) is defined by Gaver (Gaver, 2012) as design practise applied to contexts selected for their theoretical and topical potential, with the resulting designs seen as embodying designer's judgments about appropriate approaches to address the possibilities and issues implicit in such contexts. Reflection on these results enables the articulation of a variety of topical, procedural, pragmatic, and conceptual insights.

Design Research, when it occurs through the practice of design itself, is a way to ask larger questions beyond the limited scope of a particular design problem (Zimmerman, 2003). Thus I decided to use research through design as my overarching approach for this project. As a parallel and retrospective process of reflection upon my design and its outcomes, research through design allows me to explore the emergent reflexive interactions between the designer and the AI (Edelson, 2002). This is also a common technique in Human Computer Interaction (HCI), which deals with the abstract issues brought on by the challenges of giving shape to the novel possibilities and complexity made possible by information technology (Stappers & Giaccardi, 2017).

Donald Schön describes the internal processes of acquiring design knowledge as a conversation with the materials of a situation (Schön, 1984). Running a set of local experiments, in which the designer builds representations of their ideas, they can shape the situation in accordance to unintended changes (Schön, 1984). As each local experiment explores implications of ideas, it contributes to the global experiment of understanding the problem and situation at hand (Schön, 1984). Integrating those principles in my research, I will set up a series of small exploratory experiments that, confronting me with an often unexpec-

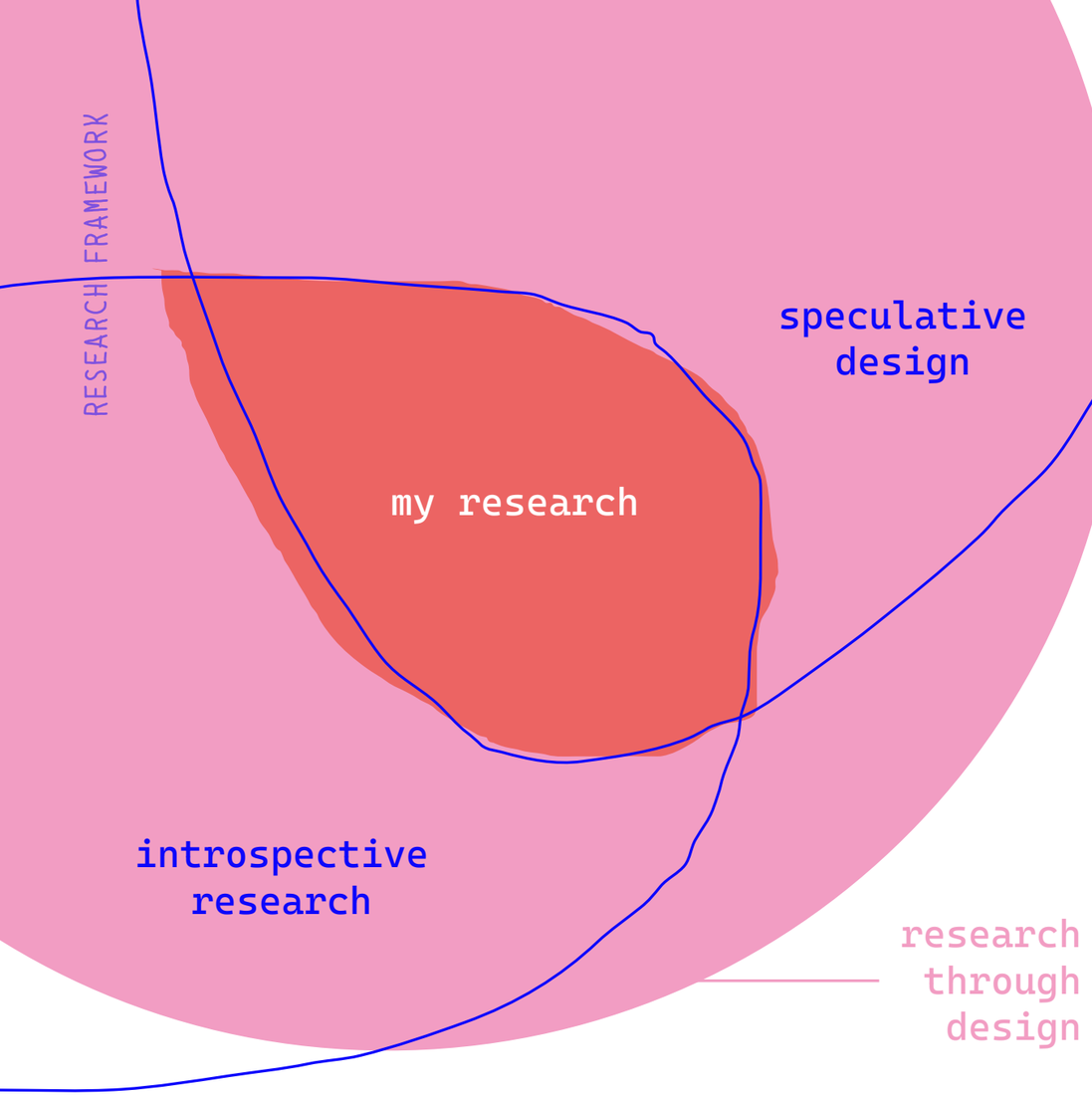


Figure 9 - research framework schema

ted reality, will help me to form new understanding and ideas in a co-evolutionary hands-on approach.

By creating situated artefacts, in my case the queer toys, the research process opens up to new opportunities, leaving room for unexpected events to be considered in the design (Schön, 1984; Stappers & Giaccardi, 2017): '*In evaluating the performance and effect of the artefact situated in the world, design researchers can both discover unanticipated effects and provide a template for bridging the general aspects of the theory to a specific problem space, context of use, and set of target users. [...] the designer(s) will have struggled with opportunities and constraints, with implications of theoretical goals/constructs, and the confrontation between these and the empirical realities in the world*' (Rosson, 2017).

While the artefact itself, and the testing of it, also plays an important role in this knowledge development, my focus will be on the designer-AI collaboration that brings the toy to life. The future design practices of designers and AI will thus be shaped in an evolutionary manner, by testing and trying out.

Through allowing unintended consequences to be utilised as design material, rather than having to deal with the unintended as an afterthought, gaining knowledge and understanding through creating artefacts can also help to account for the emergence[1] in human-machine practices (Stappers & Giaccardi, 2017; Van Alstyne & Logan, 2007). Additionally, creating research artefacts enables academics to tackle challenging or '*wicked*' issues and assess the potential human impact of emerging technology (Zimmerman & Forlizzi, 2008).

Despite its benefits, research through design also poses some challenges. Recent work explored such challenges in discussion with both research through design researchers and practitioners (Boon et al., 2020). As research through design is still being further developed, a current

[1] *The process of a higher level of structure emerging from the combination and interaction of lower level components, exposing novel behaviours or traits unrelated to the lower level components, is known as 'emergence' (Van Alstyne & Logan, 2007).*

challenge remains the documentation of research. As Kenneth Agnew said (Kenneth, 1993) research through design is '*hindered by the lack of any fundamental documentation of the design process which produced them. Too often, at best, the only evidence is the object itself, and even that evidence is surprisingly ephemeral. Where a good sample of the original product can still be found, it often proves to be enigmatic.*' To avoid this lack of documentation, my project is setup in an introspective way (as will be described in detail under '*Introspection*'), where every step along the way is stored and annotated on MIRO[1].

Furthermore, large parts of design study is based on a learning-by-doing and learning-through-observation approach (Schön, 1984; Wong & Radcliffe, 2000; Mareis, 2012). Thus large parts of the knowledge on which design is based remains tacit (Wong & Radcliffe, 2000; Mareis, 2012). Tacit knowledge however, cannot be communicated via words (Wong & Radcliffe, 2000; Mareis, 2012). While tacit knowledge could be made conscious, it usually remains hidden (Wong & Radcliffe, 2000; Mareis, 2012). This first-order information, or the subjective insights and understandings relevant to specific circumstances, is the experience knowledge acquired from engaging in design activities. This information may be conveyed and represented in certain ways, as through written accounts. However, the nature of these representations, which might be regarded as second-order knowledge, is fundamentally different (Höök et al., 2015). Research through design thus struggles with the challenge of how to articulate the gained designerly knowledge so it can be shared and scrutinised (Höök et al., 2015). This also entails the challenge of abstracting such tacit and situated knowledge in order to generalise it. A key skill and problem in any research is connecting the general and the specific, the abstract and the real (Stappers & Giaccardi, 2017).

Some argue that the artefacts generated in design and design research are itself the carrier of knowledge (Stappers & Giaccardi, 2017). In order to explain, authenticate, and constitute the design information obtained through research through design, some people suggest intermediate kinds of knowledge - knowledge between the tangible and the abstract (Stappers & Giaccardi, 2017; Höök et al., 2015). Such intermediate forms of knowledge can be annotated portfolios (Löwgren, 2013), conceptual constructs, experimental qualities and bridging concepts (Höök & Löwgren, 2012). Rather than being an instance of the abstract notion conveyed by the language, the physical objects shown in these texts might be understood as annotations to them (Gaver & Bowers 2012; Stappers & Giaccardi, 2017). Despite this, the field still has difficulty communicating information and ideas related to aesthetics, design expertise, designerly knowledge, politics, morals, and other intangible important elements in design practice (Höök et al., 2015).

While such annotated objects might generally serve as a sufficient means to communicate the acquired knowledge, the artefact in my projects are only representations of different forms of practice. As the creation of the artefact is thus of much more relevance, the artefact itself cannot present the entirety of knowledge that will be gained. As proposed by Höök (Höök et al., 2015), neighbouring disciplines like architecture can assist us in finding new ways of communicating such research through design knowledge. In architecture, the starting point when teaching is often an individual artefact: look at this particular building; it is in a typical Bauhaus style. This is then followed by a discussion on qualities, aesthetics, and functions—in other words, design criticism (Höök et al., 2015). Similarly, my artefacts (child toys) will thus only serve as discussion starters, being carried further by a speculative narrative that helps criticise current practices and illustrate alternative ones (more details under '*speculative design*').

The following will present the productive overlap of speculative design, introspection and creative AI and how they relate to my research through design approach, as well as to my project in the broader sense. Despite being described here as separate methods, they already intersect with one another.

[1] *Miro is an online collaboration platform, also allowing large amounts of image, text and web data to be visually stored and structured.*

04.2. INTROSPECTION

Introspection is defined as *'an ongoing process of tracking, experiencing, and reflecting on one's own thoughts, mental images, feelings, sensations, and behaviours'* (Xue & Desmet, 2019), used to describe and interpret artefacts, situations, experiences, beliefs and practices (Adams et al., 2017). Introspection builds on the understanding that 'many of the facts we come to know through ordinary sense perception are facts about objects we do not perceive' (Dretske, 1994). Meaning, that we are always able to relate and combine previously learned with the new materials of a situation: '*one [...] listents for the sound of a timer to learn when the cake in the oven is done*' (Dretske, 1994). Personal experience is furthermore infused by collective imaginates like expectations and cultural norms (Adams et al., 2017). As such Introspection, an instance of displaced perception, leads to better self-understanding in terms of values, experiences, as well as personal weaknesses, biases and fixations (Xue & Desmet, 2019; Xue, 2022; Dretske, 1994) and helps identify and interrogate with the intersections of the self and the many (Adams et al., 2017; Xue, 2022). Such critical reflexive practices are furthermore found to help to recognise the researcher's role in the production of knowledge, which can in turn be a necessary means to conduct responsible research (Bettany & Woodruffe-Burton, 2009).

Introspection often happens in an automatic way, and was thus often doubted criticised as '*lacking objectivity and therefore being unscientific'* (Xue & Desmet, 2019). However, it became apparent that describing and researching humans as merely rational and predictable, is not sufficient for experience design research (Xue & Desmet, 2019; Fulton, 2003).

Not only does this method provide tools for reflection and self-understanding, thus assisting in both explicit-making of internal knowledge and generalisation of such (Xue & Desmet, 2019). It also helps in properly documenting my project, by enforcing a structured reflection (Xue & Desmet, 2019). Introspection can thus serve as a great support structure for my umbrella research through design approach.

This research is interested in knowledge hidden under the surface of consciousness and expressibility, making standard user-research very difficult. Introspection can help, by providing guidance in self-researching my experiences with the AI (Xue, 2022). It poses an interesting method for designing with AI in a reflexive manner, as it surfaces internalised values, norms and the relations between oneself and the many. I am myself part of the target group, thus the discrepancy between my own thoughts and emotions and the experiences of the hypothetical user can be considered small (Xue, 2022).
However, introspective research comes with limitations such as the difficulty in generalising knowledge. As Blythe (2014) describes: '*Some people like this or that prototype but others do not like it at all. Such findings are inconclusive because the researchers do not seek to generalise.'*

Introspection could furthermore prohibit research through design attempts in generating mid-level knowledge, such as annotated portfolios. As such portfolios are *'typically being annotated in several different ways'* and by several different researchers, to '*reflect different purposes and interests*' (Bouwers, 2012). On the contrary, researcher introspection, in which the researcher acts as the only introspector, uses only their own relevant feelings, sensations, memories, ideas, or imaginations as the basis for analysis (Wallendorf & Brucks, 1993).

04.3. SPECULATIVE DESIGN

Speculative design is interested in possible futures. Reshaping our values, beliefs, attitudes, and behaviour in speculative and desired futures, can help understanding and transforming the present. Futures form a tool to aid imaginative thought and start a conversation (Dunne & Raby, 213). '*Dreams [...] can also inspire us to imagine that things could be radically different than they are today, and then believe we can progress toward that imaginary world*' (Duncombe, 2007). Design fiction can take the shape of narratives, short stories and films, but it can also come in the form of objects and prototypes (Blythe, 2014). Similarly to science fiction, it is about the ideas and thought experiments, as Stanislav Lem illustrates in his work Imaginary Magnitudes (Lem, 1973) that uses introductions and prefaces, instead of fully written books, as a way of playing with ideas. Haraway draws heavily from science fiction and uses storytelling—as well as feminist theories—to describe other realities (Haraway, 2016).

As reviewed by (Søndergaard & Hansen, 2018) designing the world or narrative around a future technology is also a technique used in design fiction to explore possible conflicts of future technologies (Bleeke, 2009; Lindley & Coulton, 2016; Tanenbaum, 2014). Design fiction explicitly condemns the seeking of '*solutions*' and problem solving technologies (Blythe et al., 2016). Instead, speculative de-

sign enables a designer and its audience to see the potential implications of adopting future technologies like AI, and critically reconsider them (Lindley et al., 2017; Nicenboim et al. 2020; Dourish & Bell, 2009). Fictional scenarios are also used to '*act out*' non-human agents, in order to investigate the impact of AI in everyday life, surfacing the technologies implicit interrelations with its surroundings (Nicenboim et al. 2020). Fiction can serve as a lens through which controversy, criticism and conflict can be explored, pointing out what is typically unsaid in the implicit social and political context of a design and forcing us to deal with questions of ethics and consequences arising in a world with a certain technology (Søndergaard & Hansen, 2018; Blythe, 2014). As reviewed by Blythe (Blythe, 2014) such, speculative design is of great interest for HCI (Sterling, 2009; Bleecker, 2009; Markussen, & Knutz, 2013). The future becomes a lens through which we may view ourselves and our society as it acts as a projection of the problems and conflicts of the present (Søndergaard & Hansen, 2018).

Whilst in my case the AI is asked to provide the ambiguous vision of toys, speculative design can create the narrative that gives voice to those alternative imaginings. Speculative design further allows me to explore alternative practices, without seeking solutions, rather critically consider the present paternalistic approaches in HCI. Speculative design accomplishes two unique goals: it allows me to see the future design practices and it critiques current practices like designing for child toys.

Processes concerning reflexive interactions with AI are furthermore abstract and intangible. A speculative narrative, provocative yet relatable, can assist me in communicating my ideas and most importantly, to start discussing, disagreeing, and forming common ground, understanding and ideas with others.

Not only will this fictional narrative help illustrate, annotate the practices and artefacts, it can furthermore be used to communicate my research knowledge to other disciplines. In combination with an academic paper for instance, it will encourage action and reflection, and inspire further design work (Höök et al., 2015).

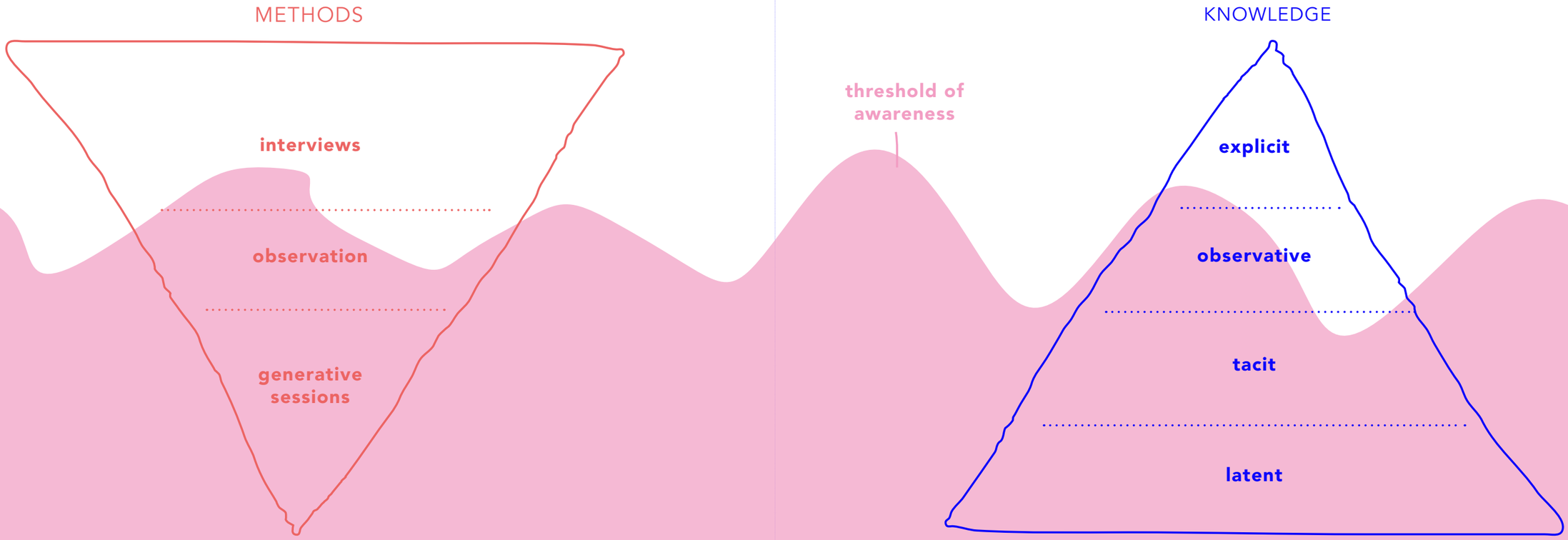
04.4. ADDRESSING THE FRAMEWORK'S LIMITATIONS

While the intersection of research through design, speculative design and introspection shows to be very suitable for this project, it does not come without limitations like problems in generalising knowledge due to researcher introspection and contextualised research. Addressing those limitations, I am integrating snippets from other known design and research practices. To name them:

user testing:
Integrating users, in this case other designers, in the design process, can help generalise knowledge. Observing the reactions, emotions and discussing the opinions and experiences of others, can open new perspectives and help relate the own experiences as an introspector, to the bigger picture. The integration of others in the process can happen in multiple ways. For this project, interviews, workshops and observations are planned.

generative research:
Studying biases can be a challenge. Rooted in our beliefs, values, interests, etc., they are abstract qualities that people can't or are not used to talking about (Sanders & Stappers, 2013). Due to their tacit nature, biases are thus often easier to observe in peoples' doing, then they are expressible by the person itself (Wong & Radcliffe, 2000). Based on this understanding I decided to expand my introspective research with generative design research methods, tools and techniques, helping me to surface biases in designerly doing. The generative design research approach brings '*people we serve through design directly into the design process in order to ensure that we can meet their needs and dreams for the future*' (Sanders & Stappers, 2013). I furthermore tend to use the creations of these sessions as data for my later experiments. The generative research approach hereby offers a set of methods and tools to help access knowledge below the surface of awareness and explicitness (see Figure 9). '*First, it prepares people for the generative session by getting them to remember and reflect upon their day. Second, it provides the foundation for layering through the levels of knowledge'* (Sanders & Stappers, 2013).

Figure 10 - methods for accessing tacit knowledge (Sanders & Stappers, 2013)



EARLY EXPLORATIONS

In addition to the literature research, as well as the following introspective research through design, an early stage analysis through practitioner and generative research was conducted. Expanding on the understanding gained in early readings, this short analysis phase helped in gaining a more holistic overview over the current practices and state of bias in the general design problem. Current technological tools and their impact on reflection were explored and tested.

5

EARLY EXPLORATIONS

05.1. INTERVIEW WITH OIO.STUDIO

In order to better understand the symbiosis of designer and AI, this graduation project is carried out in collaboration with oio.studio (Öiø / Studio, n.d.). Oio describes themselves as a *'creative studio made of designers, technologists and bots working on future products and interactions*' (Öiø / Studio, n.d.). The following examples of Designer-AI collaborations are discussed and analysed in order to better understand the challenges and limitations of human-machine symbiosis in design.

The descriptions of human-robot or designer-'*Robi*' (as oio's AI is called) interactions are abstracted and analysed in terms of their relationship and interaction cycles, as well as their incorporation in the design process.

I chose two ways of depicting the practice we discussed, in order to depict all the nuance expressed in the interview. Both graphics refer to the same process, yet reflecting and highlighting different aspects of it. In order to visualise the relationship between human and non-human entities I chose an entity-relationship notation (Bozzon, 2021) (Figure 11). Interaction cycle and integration in the design process on the other hand are represented in a simplified graphic (Figure12).

Figure 11 - entity relationship representation of designer-AI relations at oio.studio

An entity-relationship (E-R) model is the most common conceptual data model. It provides a series of constructs capable of describing the data, entities and relationships in a mini world (like the studio practices at oio). It is using graphical formalisms and is thus easy to understand. An entity is hereby defined as a class of objects (e.g. things, faces, designers). Each entity can have different instances that represent the entity. Multiple instances of the entity designer could be for instance, interaction-designer, graphic designer, etc.. A relationship is the logical link between two or more entities. Those relationships can have structural constraints. The cardinality for instance describes the maximum and minimum number of relationship instances in which an entity instance can participate. A participation furthermore be optional (single line) or mandatory (double line). All those elements can be found in Figure 11 (E-R model as introduced to by Bozzon, 2021)

As illustrated in both Figures 11 + 12, neither human nor robot performs standalone, but form a co-dependency, where the designer still remains in charge of both, grasping the problem, and forming the solution. Rebaudengo sees most productivity by combining the abilities of hu-

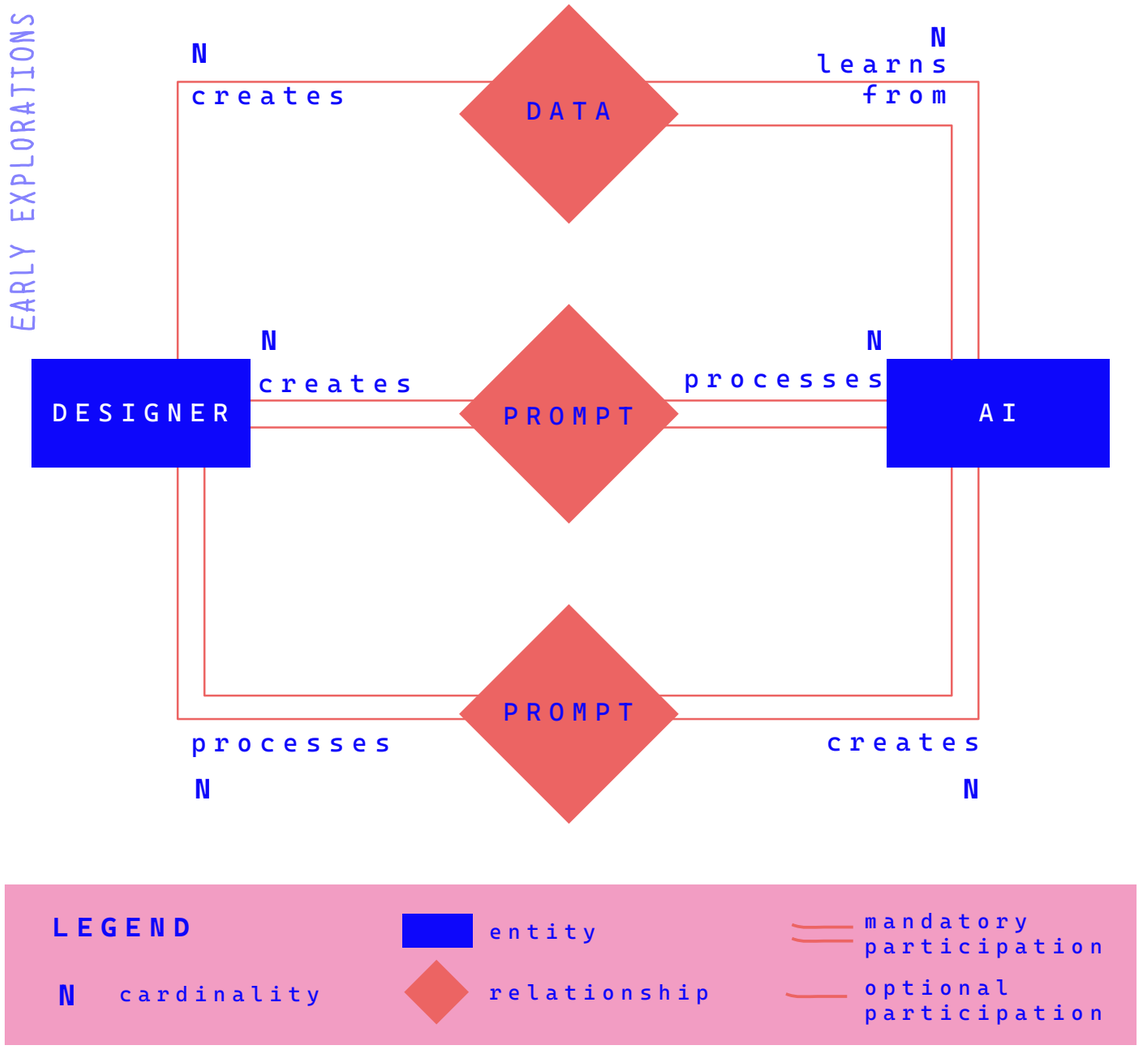


Figure 11 - research framework schema

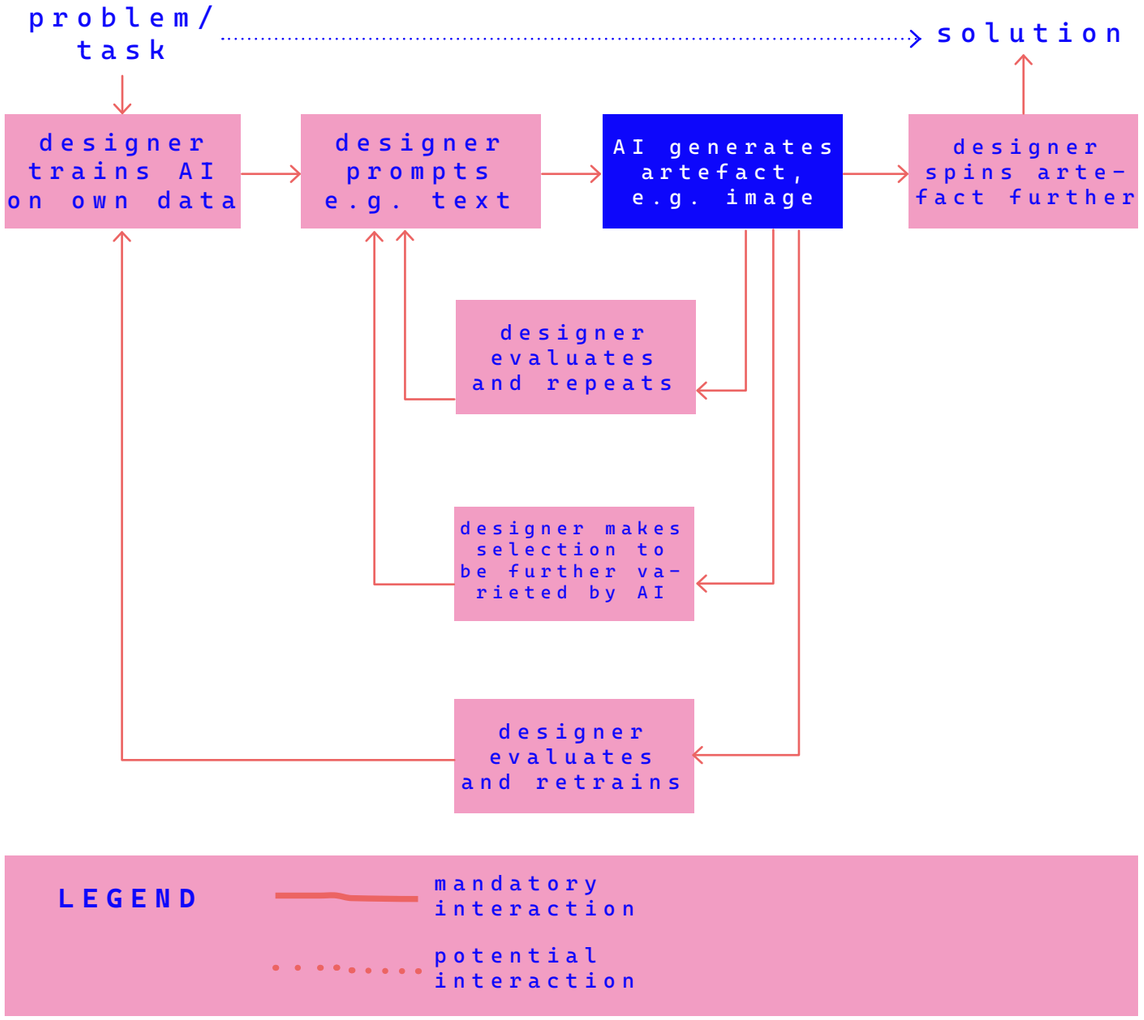


Figure 12 - research framework schema

man and machine: *'[computer algorithms] are fast and powerful, but if [the human] plays together with a computer, basically using the computer for prediction and processing, and the human to make the decision, then actually it is way more unbeatable then the human or the computer itself.'* The traditional ideation process, as being performed only by the designer, is no longer existing. Human and non-human entities rather form a new joint practice, where artefacts in form of data, text and image prompts are being passed from one to the other in an iterative and turn-taking manner (Figure 12). Rebaudengo (2022) emphasises the importance of human information processing: '*You can generate images with Midjourney, you can generate products, but the question is what did you [as the human designer] choose, how did you choose, how did you make it'*. The AI, as trained on a vast amount of data, is thereby able to inspire '*what you did not previously imagine' and can sometimes offer 'another point of few on your idea'* (Fioravanti, 2022). Fioranti (2022) even describes that emotional relationships between human and non-human can evolve: '*When someone asks Robi to write a poem, this generates some kind of emotional relationship with it [...]'*. Over time, the AI is being more and more integrated in the design processes and '*now part of our life*'.

Such a practice is hereby not simply planned upfront, but evolves through iterations of experiments and trial and error (Fioravanti & Rebaudengo, 2022). This exploratory approach is necessary due to aforementioned emergent properties which are difficult to design for, and that arise from the interplay of different elements. They also highlight the differences between the scientist and the practitioner. While the scientist asks '*to have a perfect model, and if there is any bias, you have to fight it, because you try to do something replicable and scientific'* but as a practitioner, starting to design with AI '*you need to create biases to progress with your project, [...], like some form of alchemy where you need to put data and try to figure out what happens*'. Only after many iterations of '*trial and error will you find a way [of establishing a productive workflow]'*. Marta Fioravanti also indicates the challenge of bias in algorithms. Instead of avoiding them, she rather aims to incorporate them in the process.

05.2. GENERATIVE DESIGN SESSIONS

Aiming to have a look at how biassed designers tend to think in the context of child toy design, I designed a toolkit for a generative design research session. As such this activity falls outside of the general research framework. However, I decided that generative design poses a great method to gain a deeper understanding of the context specific difficulties.

The session consisted of collage making, trying to learn about underlying associations designers have with certain elements of their design process (idk. colours, stereotypes in research,...). I created a set of words and images for participants to choose from, when creating their collages. In order for the participant to begin by selecting familiar stimuli that immediately connect to the region of experience, some triggers are chosen to be unambiguous. People may feel more at ease throughout the creative process as a result (Sanders & Stappers, 2013). Examples of less ambiguous triggers are shown in Figure 13.

Other triggers in a toolkit are intended to be ambiguous, allowing them to invoke a variety of associations and being picked for a variety of reasons. This offers two benefits: First, because the trigger is vague, the person is allowed to interpret it in light of his or her own experiences. Second, the ambiguity of the trigger allows them to explain their interpretation of it and considerations for picking it when they exhibit their collage or map to others (Sanders & Stappers, 2013). Examples of ambiguous triggers are shown in Figure 14.

The participants were given 2 out of 3 different scenarios. One of the three scenarios was obviously gendered with the potential for strong associations and biases to be exposed. Participants were hereby asked to put themselves into the mindset of a fashion label designer, creating the new summer dress for young girls. The second scenario was deliberately phrased less openly gendered, though still evoking high potential for strong gendered associations. Participants were asked to design the next Lego mind-storm robot. In both of these two scenarios the target group was not assumed. The third scenario was explicitly asking for a gender neutral adventure game design. The goal was to think of the main character that would master the adventure.

At this stage of the project, the generative session was used to explore biases in the front-end of the design process, more specifically the framing of target group and design criteria. Participants were hereby first asked to create a mood board through images or text, representing their target group/user. The images and text used for the

Figure 13 - sample of less ambiguous image and word trigger

Figure 14 - sample of ambiguous image and word trigger

representation of the target group are deliberately less ambiguous (see Figure 12). Secondly they had to create another mood board representing qualities of their design by choosing from either ambiguous text or ambiguous images (see Figure 13).

They first created mood boards for the gendered scenarios, afterwards for the gender-neutral task. After the selection of images or text, representing the target groups and design qualities of their imagination, participants were asked to classify the images or words into '*feminine'*, '*masculine*' or '*neutral'*. Participants were encouraged to make fast decisions and classify intuitively. Although told that these classifications will not be judged, some participants stayed mostly neutral, avoiding the binary decision making. Thus in a second round all the images and words that were classified as '*neutral*' had to be put in the category of either '*feminine*' or '*masculine*'.

In total 4 students participated in a session. All sessions were done separately, and only participants were chosen that were not aware of the goal of my thesis. Two participants were asked to use images for their collages and two were given words. Two of the participants were female, two male. One man and one woman each worked on scenario 1 and 2, all worked on scenario 3. The time was kept short to encourage the automatic system to make the decision. For the less ambiguous trigger 1min was given, for the ambiguous trigger 3min. The amount of triggers was appropriately adjusted given the time, so that participants had enough time to scan the images and words.

In Figure 15 one example of the entire work flow is provided. All the results can be found in Appendix A. In the following a few key insights will be presented.

The mood boards for task one showed strong stereotypically associations. Participant A for instance, described the target group for the next lego mind storm robot through images of science, tools and technology. In the description of the choice of images the participant even explicitly said that the target group was intended to be male. While expressing this, the participant became aware that although a male target group was not asked for, it was never a conscious decision for a male target group or against a female one. The participant described the choice as made intuitively.

Although specifically asked to design for girls, the mood board done by a participant imagining a new summer dress, also showed strong signs of gender stereotypical associations. Even more interesting were the results of

the gender neutral task each participant had to perform. One participant especially, showed strong difficulties in describing and imagining the target group and design qualities in a non-gender related way. Each decision was argued for in terms of its gender neutralness, instead of describing gender qualities like '*curiosity*', '*high energy*', etc..

When comparing the images and words people used to describe their gender neutral target group it was shown that all three participants had a tendency to use rather male words and terms to describe their target group and design. None showed an equal amount of male and female objects in their mood boards.

When comparing the gender classifications of the two participants working with images, they showed a huge overlap in the less ambiguous trigger set. Only four out of 26 images were classified differently. The more ambiguous images were found to be only half categorised similarly. The reasoning behind the choice of images and words also differed. In the classification of ambiguous triggers, art and architecture were classified most differently by participants. One was thereby basing the decision more on colour, the other more on shape. Some described personal experiences that influenced their decision, others seem to be based more on cultural stereotypes. Participants of all scenarios showed a similar ease in classifying their triggers.

Due to the small sample size no quantitative results could be gathered. Comparisons were made carefully and not generalised. Although participants were assigned to tasks in a mixed gender way, the amount of participants does not allow for comparisons between the performance of participants of different genders.The sessions showed that designers have unconscious associations in terms of gender. In the description of their choices, a male participant mentioned that he perceived the task as difficult, because he was never a girl.
Similar can probably be said for the other participants as well. Although of the same gender, childhood is for most designers distant and difficult to re-imaging without following gender stereotypes.

In the classification exercise participants showed discomfort in sorting elements in binary categories. This discomfort is interesting and is potentially interesting for sparking reflective processes. It was interesting to see that conceptualising child-products without any notion of gender at all was very difficult for the participants. These tests were focusing on the early stage of the design pro-

Figure 15 - example workflow generative session from participant A

cess. It was shown that participants at this stage did indeed show strong signs of stereotypical thinking, especially in relation to children as the target group.

However, to identify the ideal point of intervention, similar sessions need to be performed, challenging associations at later stages of the design process. A follow-up session focusing on the ideation part of the design process is thus planned.

05.3. TECHNOLOGY EXPLORATIONS

In a series of small experiments, I explored different off-the-shelf AI tools and ways to curate their input data and interpret their output data. The goal was to gain a better understanding of ways in which AI can confront and trigger reflection, while expanding my knowledge and skill in collaborating with AI. No design context was provided yet in these experiments.

The algorithms varied in terms of input-output data, and data processing. The aim was to play around with a large variety of algorithms.

05.3.1. Exploration 1: Visualising bias through text-to-image algorithms

While language models are able to look at biases beyond their visual representations, their outcome usually remains in written format. Designers' solutions have been seen to be influenced by the type of stimulus, the relationship the designer draws between the stimulus and the problem, and even the stimulus's recentness (Gonçalves et al., 2012). Designers favour visual stimuli because they find it simpler to draw connections between a source and goal notion (Gonçalves et al., 2014). These stimuli include shapes, textures, and gestalts that designers can recognise, memorise, abstract, and recall (Goldschmidt, 2015).

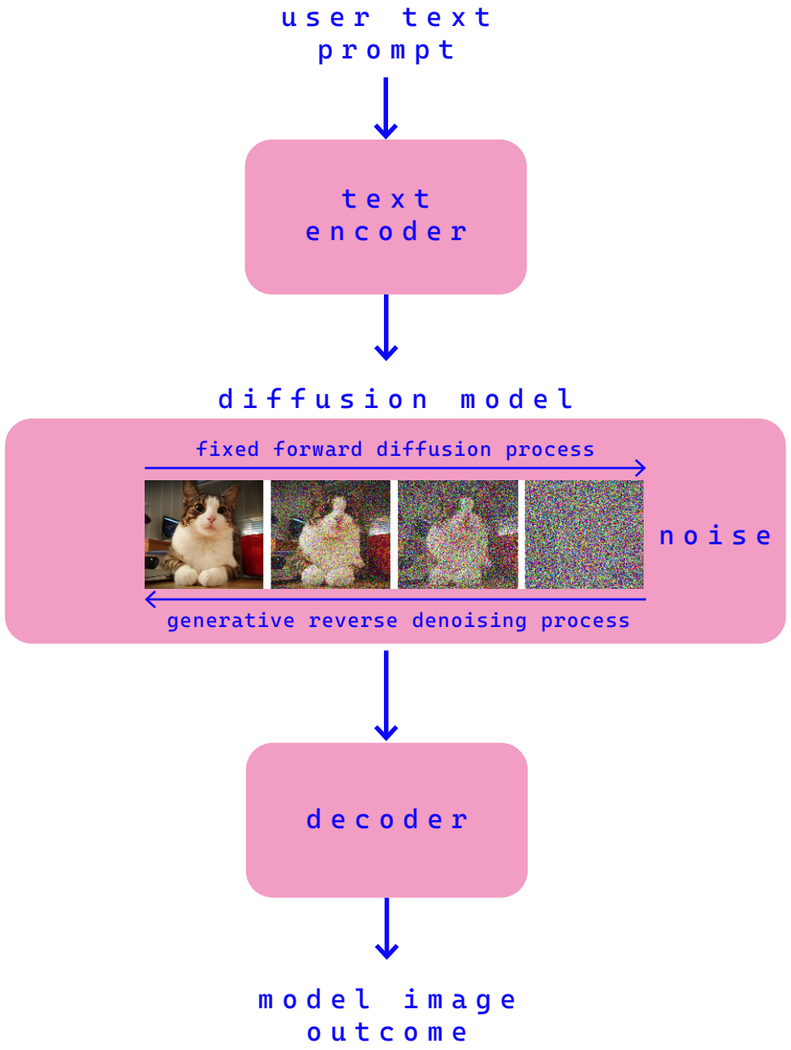


Figure 16 - Diffusion Models Schema (O'Connor, 2022)

However, generative adversarial networks which are able to generate eerily realistic images, can usually only learn from visual information as well, narrowing the spectrum of bias to aesthetics.

Algorithms like Dall-e (DALL·E 2, 2022), Midjourney (Midjourney, n.d.) or Stable Diffusion (Mostaque, 2022b) provide a perfect combination of text generators and GAN's, as they translate text prompts into images (Figure 16). Pre-trained on millions of stock images, those algorithms called diffusion models, use language models like GPT3 that are trained to generate images from text prompts by learning from text-image data pairs. Diffusion models carry out two separate tasks in succession. They try to reconstruct photographs after they have deconstructed them. Programmers provide the model with real visuals that have human-assigned interpretations, such as a dog, an oil painting, a banana, the sky, a 1960s couch, etc. They are diffused, or moved, by the model through a long series of sequential actions. Each stage in the diffusion process adds random noise in the form of scattershot, meaningless pixels to the picture that was given to it by the step before, then passes it on to the one after. When this is done repeatedly, the original image eventually turns into static and its original meaning disappears. (Fedor Indutny , n.d.)

Using Dall-E mini (DALL·E Mini by craiyon.com on Hugging Face, n.d.) for some of those explorations, the website itself informed me about the problem of bias: '*While the capabilities of image generation models are impressive, they may also reinforce or exacerbate societal biases. While the extent and nature of the biases of the DALL·E mini model have yet to be fully documented, given the fact that the model was trained on unfiltered data from the Internet, it may generate images that contain stereotypes against minority groups*'.

Utilising this idea of bias, I used the algorithm to challenge my own bias by comparing my mental images associated with words with the AI's generations. Despite my initial idea of words as rather neutral constructs, the results of my dall-e explorations revealed that words heavily affect my way of internally visualising and conceptualising gender (Figure 17 + 18). The examples in Figure 17 + 18 show how closely our mental images regarding gender are linked to words like *'strong'*, *'sexy'*. My explorations revealed a bias linking more active and technology related words like *'strong'* or '*science*' to male figures, while more passive related words to feminine figures. As shown, image-generating AI very easily picks up on the prejudices and toxicities ingrained in the millions of web-sourced images used to train them.

Figure 17, AI image generation based on input 'A photo of a strong person', generated with Dalle 2 (DALL·E 2, 2022)

Figure 18, AI image generation based on input 'A photo of a sexy person', generated with Stable Diffusion (Mostaque, 2022b)

'Trustworthy individual'

Figure 19, AI image generation based on participants design descriptions, generated with Dalle mini

However such generated images are often ambiguous, leading to the projection of my own gender bias in the interpretation of the algorithm's output. Despite this being potentially problematic for an easy to use bias visualisation tool, it poses great potential when my own gender expectations don't meet the generated ones. Integrated in a more reflexive approach, such clashes of expectation might become useful means to raise personal bias awareness.

In a small workshop (stepping outside of introspective research to compare my own experience with observations) I built on this idea of visualising bias, by having other designers use textual descriptions of their designs as input text prompts for Dall-e mini (Figure 19). I asked them to compare their internal ideas of what those words represent with the images they received from the algorithm to explore potential discrepancies between the designers intentions and reality. All results can be found in Appendix B.

Despite a lot of perceived similarity between the internal and the generated imagery related to words used to describe the designer's creations, a few observations were made. The interaction seemed to generally trigger some internal reflection. One designer pointed out confusion, when receiving AI imagery that depicted the opposite then the expected gender when asking for a '*trustworthy individual*'. This gap between the expected and the received was taken as main learning from this exploration.

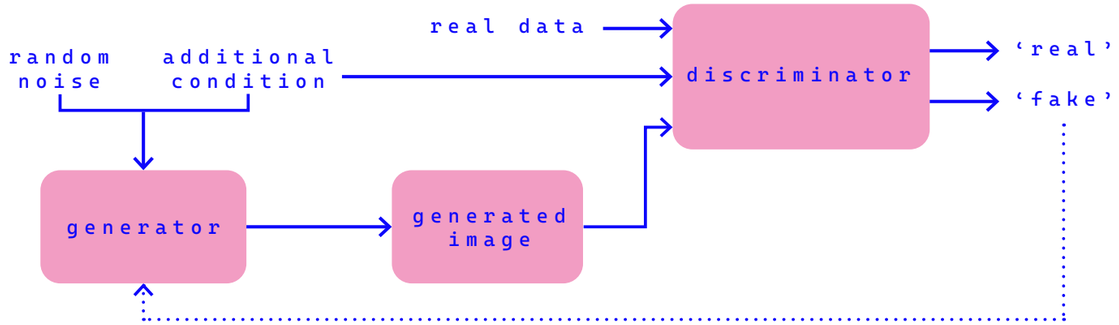


Figure 20 - cGAN Schema

05.3.2. Exploration 2: Using cGAN's to break with gender stereotypes

Starting my exploration with image-to-image algorithms, I wanted to use a cGAN to create the similar irritation observed in the previous exploration, when the designer received an AI image that was depicting the opposite of the expected.

'*GANs rely on a generator that learns to generate new images, and a discriminator that learns to distinguish synthetic images from real images. In cGANs, a conditional setting is applied, meaning that both the generator and discriminator are conditioned on some sort of auxiliary information (such as class labels or data) from other modalities. As a result, the ideal model can learn multi-modal mapping from inputs to outputs by being fed with different contextual information*' (Using Figment With PIX2PIX | Figment, n.d.) (Figure 20). Trained on those image-pairs like illustrated in Figure 21, the algorithm is for instance used to colour old black and white movies and images.

Based on this understanding of cGANs, I prepared a set of data that could potentially break with our preconceived ideas about gender roles and stereotypes. As illustrated in Figure 22, the idea is to replace stereotypical character or gender characteristic features like colour, with opposing ideas and images. As such, our own gender biases and internalised categories like male and female can be made obvious, confronting us with the alternatives we were not expecting. However, due to the complexity and need of powerful computing, I was not able to test this idea.

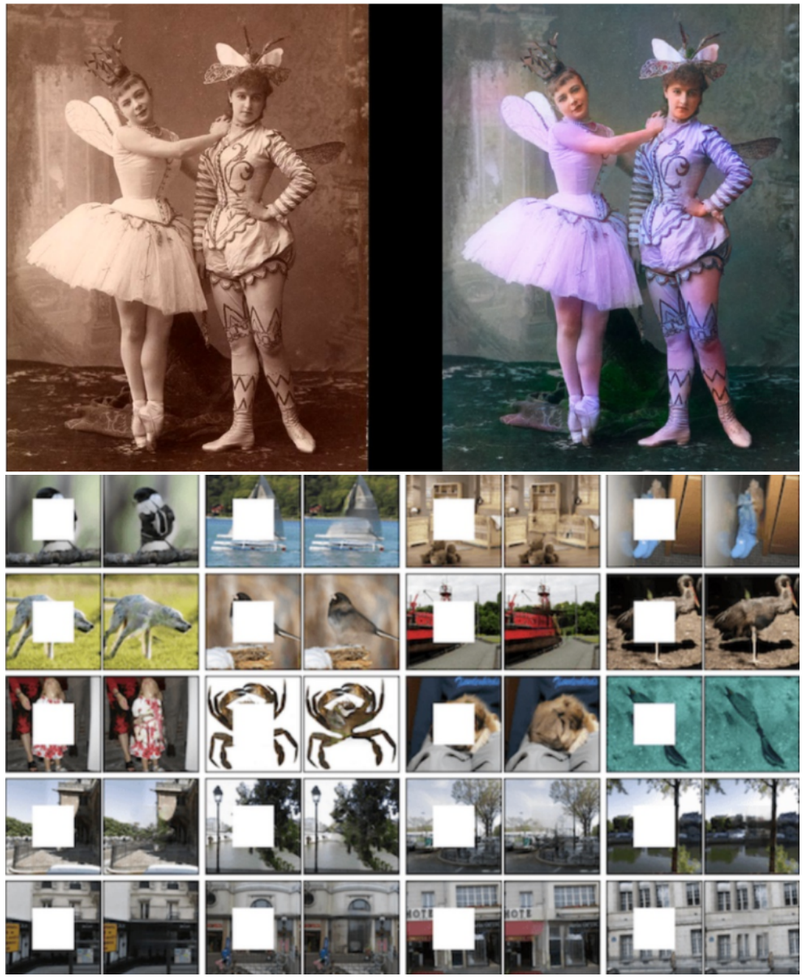


Figure 21 - examples of training data for cGANs (Using Figment With PIX2PIX | Figment, n.d.)

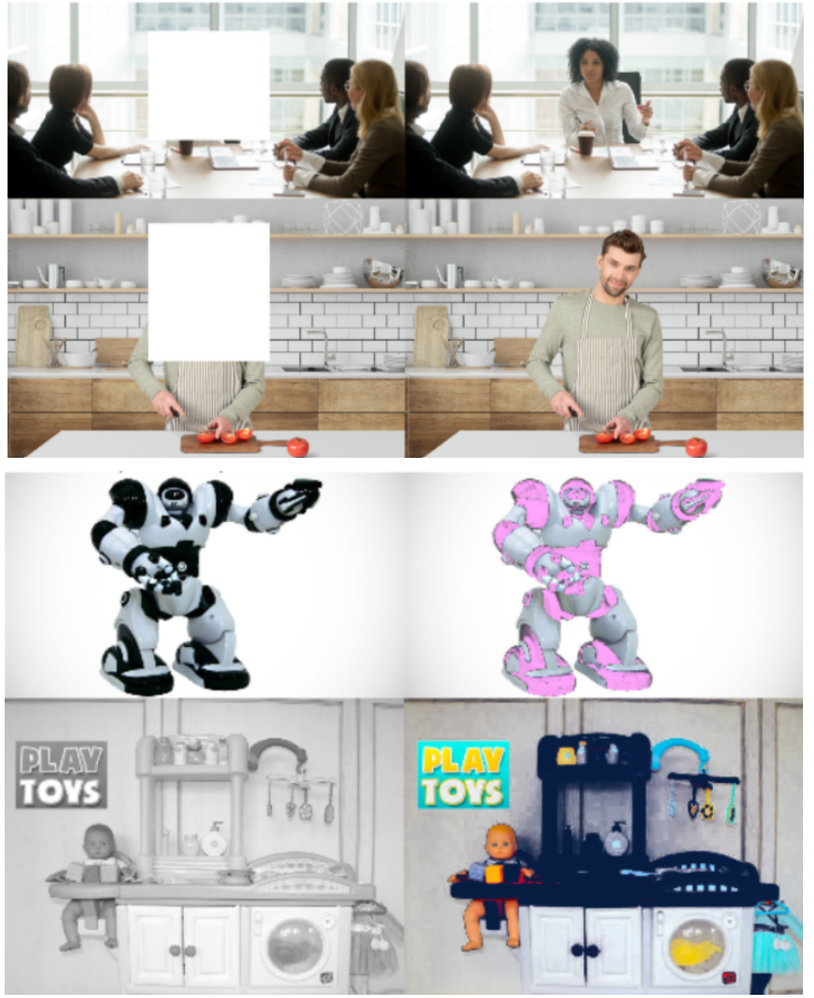


Figure 22- data explorations creating a clash of expectations

05.3.3. Exploration 3: the latent space

In my third exploration, I used a SyleGAN to imagine spaces between the masculine and the feminine. A StyleGAN, like the one of RunwayML (Runway, n.d.) I used, is a type of generative adversarial network. In addition to producing stunningly photorealistic, high-quality images of faces, the resultant model also provides control over the style of the created picture at various degrees of detail by adjusting the style vectors and noise (Brownlee, 2019) (Figure 23).

A styleGAN poses two interesting abilities I wanted to test in my exploration: 1) the latent space, 2) the latent walk. While latent space usually described the hidden and non-observable space of a neural network, RunwayML[1] found a way of visualising the multi-hundred dimensions of this space in a 2D plane (unfortunately I couldn't find any information on how that works exactly). It is a space filled with points. The beauty of this space is that the generative model learns to map these points to output images. All images are related, but the closer they sit next to each other in the latent space, the more similarity can be found in the images. A latent walk is simply a series of images that show a transition between two or more generated images, that you can choose upfront.

For my idea of visualising the in-between spaces, I trained a StyleGan, that was pre-trained on thousands of images of cars, on a dataset I created out of images of gender stereotypical toys. After the first training batch of 4 hours, I was then able to explore the images the algorithm generates in the latent space. Scanning through this infinite space, I was searching for images that I did not perceive as particularly masculine or feminine (example Figure 24).
Those could potentially serve as study material or inspiration for developing own designs that challenge current stereotypes. It was furthermore the experience of exploring this infinite space between images that made me realise how unrealistic the concept of categorising is. The training was not yet perfect, however, I was able to still define elements in the images as either masculine or feminine or neither (example Figure 25).

In the next step I searched the generated images for especially good representations of my ideas of femininity and masculinity (example Figure 26). Those I used as the ends of the spectrum I created through interpolating between them (through the latent walk). The resulting video served again as inspiration to imagine spaces between the binary, and as great visualisation of the absurdity of categories.

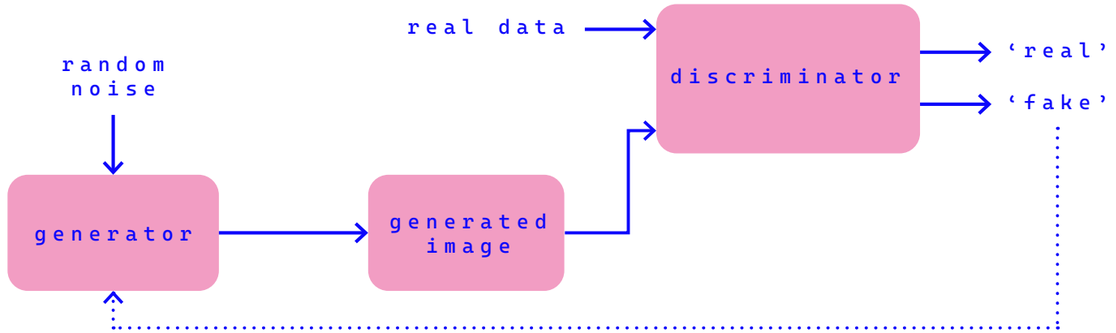


Figure 23 - StyleGAN Schema

[1] *Runway ML is a free software making Machine Learning accessible to artists and creatives (Runway, n.d.).*

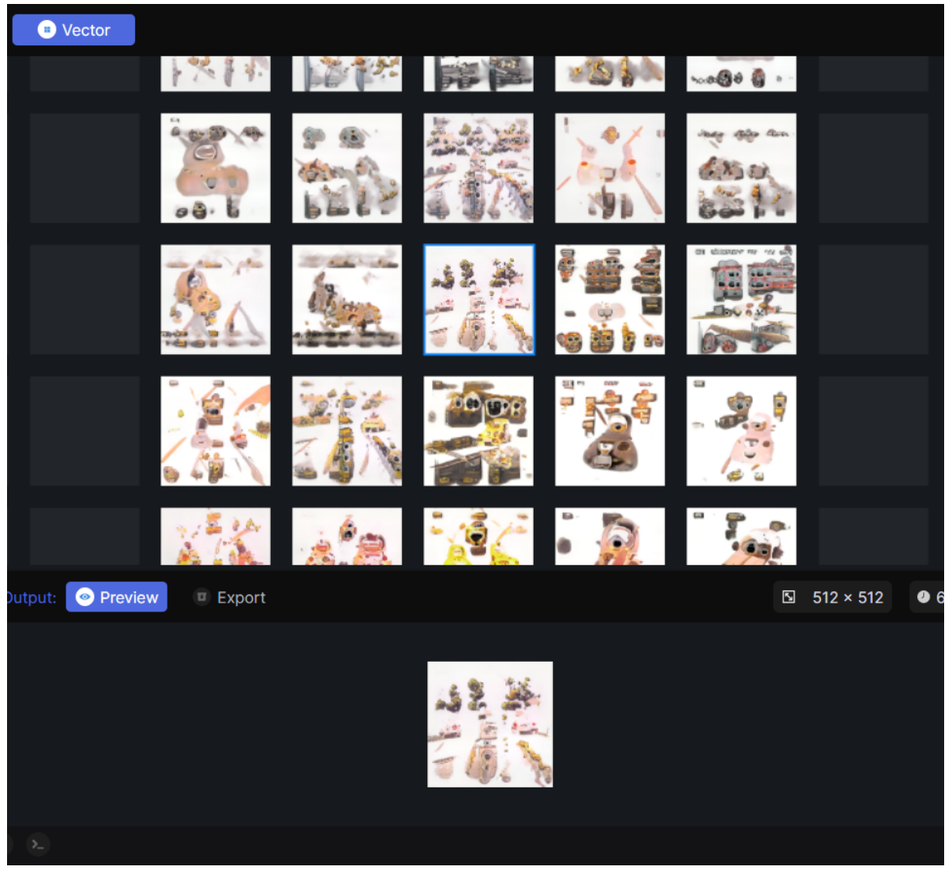


Figure 24 - searching the latent space for ambiguous images

Figure 25 - from left to right, two gender ambiguous images, two feminine perceived images, tow masculine perceived images, all from latent space

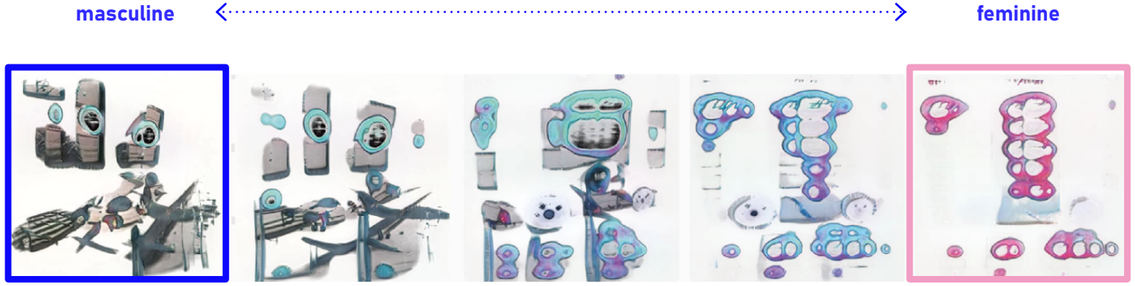


Figure 26 - series of images, morphing from masculine to feminine through latent walk

05.3.4. Exploration 4: the De-calssifier

The idea of my last exploration was to capture unconscious associations in my perception by encoding them into data, that I feed a classification model. This experiment was done with the open source tool teachable machine (Teachable Machine, n.d.).

Classification algorithms use a method called 'transfer learning'. Teachable Machine uses a pre-trained neural network, and the categories added by the user are effectively the final layer or step training of that network (Teachable Machine, n.d.). A software learns from the dataset or observations provided and then classifies additional observations into various classes or groupings in classification (Classification Algorithm in Machine Learning - Javatpoint, n.d.)(Figure 27).

I started this small exploration by classifying random objects in my environment into the binary categories of either feminine or masculine (Figure 28). I was surprised by the ease of classifying images of all levels of abstraction into the binary categories of maculine and feminine. Nonetheless, I am not able to give an explanation for perceiving some images as masculine and others as feminine. I also experienced the binary categories as challenging. Although being able to classify anything in '*male*' or '*female*', I noticed how this way of looking at things made me blind to many nuances in-between those two categories. I experienced the first step of classifying my images and creating my dataset as an important sensitising activity that already started to make me aware of the existence of bias in my thinking.

In the next step I used a teachable machine to train a classifier on my dataset. I had to make sure to include images of my camera background in both categories, and to keep my fingers out of the images to not confuse the results. I furthermore made sure to have an almost equal amount of images per category. With the new individual gender bias lens I created, I started testing a bunch of different objects from my environment to see what category they would fall into. I also started testing my own designs to see which category they would be assigned to.

Some of the classifications were surprising (Figure 29). Curious to find out why the classifier was thinking '*knitting is male*', I started to investigate the images in my male category. I found at least 3 images that illustrated tools of some kind, that if taken out of their context of use, did actually show quite some similarity with the knitting sticks.

The testing was also very helpful to gain a deeper understanding of my unconscious associations. I learned that my gender bias manifests itself a lot in colour associations. Whereas a lot of the black objects were classified as masculine, most of the more richly coloured objects were seen as feminine. The same object could thereby be identified as two different categories depending on which side of the object was facing the camera.

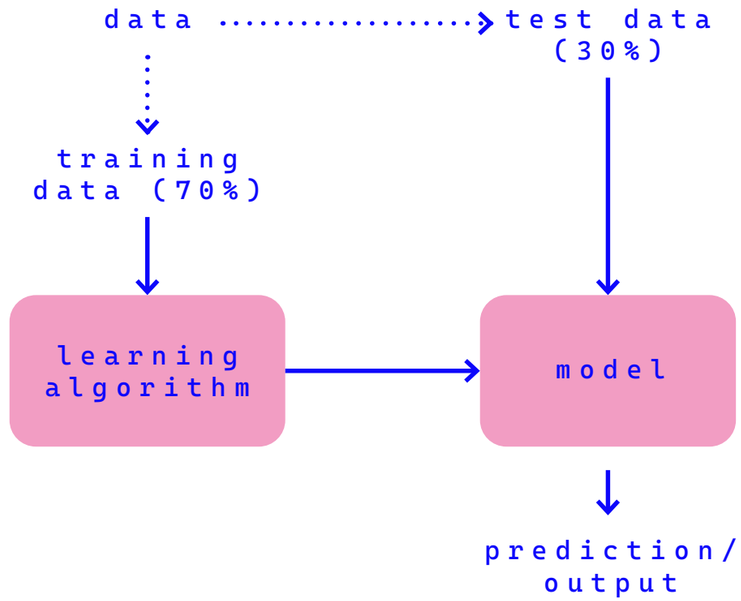
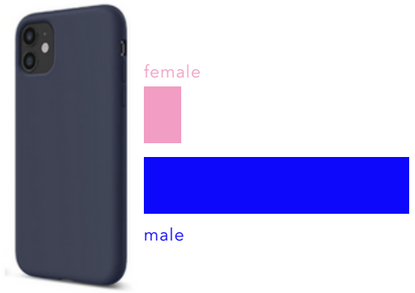


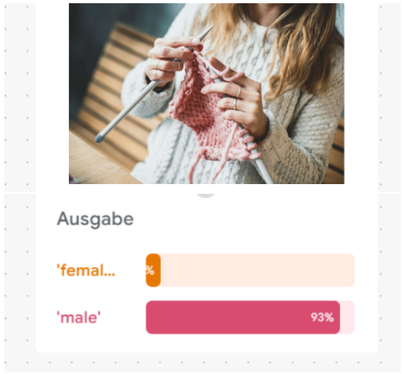
Figure 27 - Classification Model Schema

EXAMPLES OF „FEMALE“ DATA

EXAMPLES OF „MALE“ DATA

Figure 28 - encoding my gender bias in the training data





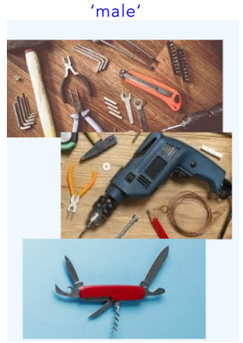


Figure 29 - surprising findings

05.4. WORKSHOP AT THINGSCON

Based on the gender-classifier experiment, I designed a workshop that was held as part of the '*Punk Bots, Radical Designers!*' workshop at Thingscon 2022 (ThingsCon 2022 – ThingsCon, n.d.). As such it forms one of the few activities that were considerably performed outside of the introspective research frame. This workshop therefore allowed me to expand on my introspective experiences with the classification algorithm, and observe how other designers, also in groups, react to those potentially confronting interactions.

The aim of Thingscon 2022 was raising questions and discussions about the fluid state of IoT, interaction in an increasingly networked world, and inspire new frontiers and experiences with technologies like Web3, the metaverse, and digital twins (ThingsCon 2022 – ThingsCon, n.d.). The workshop more generally dealt with problems of diversity, equity and inclusion in the design for embodied AI. The aim was to raise awareness and implement critical reflection in the design process. The workshop was given together with dr. Cristina Zaga, dr. Maria Luce Lupetti & DEI4AI collective (Diversity, Equity and Inclusion for Embodied AI, n.d.).

My part of the workshop was about 45 min long and divided into 5 steps. We had 13 participants, some of which were already familiar with AI-tools, others completely new to the field. First participants were asked to use images in magazines to create a set of 5 cards that represents their identity, interests and/or personality (Figure 30).

In the next step participants had to look for google images, and classify them into the binary category of either masculine or feminine. Participants were given the choice to pair up or work alone. The classified selections of images can be found in Appendix C.

Next teachable machine was trained on the earlier generated data. In the following participants were then exploring the model by presenting different images to the classifier (Figure 31). After the first round of explorations participants were then encouraged to test their identity cards on their model. In a next step they were then asked to walk around the other groups in order to see if other participants' classifiers and '*biases*' differed from their own.

Participants showed great engagement in the activities. Everyone paired up with other people, without any encouragement necessary. Participants were discussing biases together without any difficulties. The classification algorithm seemed to provide a safe space for everyone to openly, and potentially detached from oneself, discuss

Figure 30, the making of identity cards at thingscon

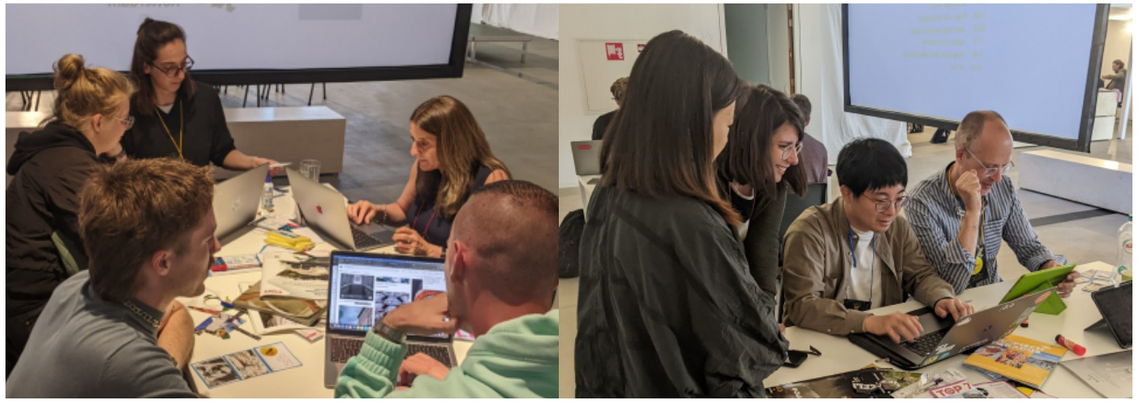


Figure 31, participants classifying trainings data and testing/discussing the outcome

Figure 32 - garden classes and classification results

biases. One group, investigating gardens, was especially surprised about the existence of gender associations in something as neutrally perceived as gardens (Figure 32).

All participants used the classification to investigate what elements in their data were defining the gender in their associations. For the ‘*garden*’ group, colour played less of a role than the ‘*messiness*’ of the objects in a frame. Other groups were really surprised when testing images that were classified opposite to the expected. This clash of expectation also sparked an investigation in the data, as well as discussions among other groups at the table. Some groups were able to find explanations for their classifications others did not.

In the follow up discussion with the whole group, a participant expressed some kind of discomfort with the ‘narrowness’ of the binary classification. Further discussions about biases arose when participants started testing their identity cards at other people‘s classifiers. The results of the different classifiers differed sometimes. Groups with less number of people seemed to find it easier to classify images in masculine or feminine, in groups with more than 2 participants the decision making took longer and more objects were identified as not fitting any given category. The question arose on what to do after these tasks were completed and how to integrate the findings into the individual design process. My answer to this question can be found in the chapter ‘*Design Experiments*’.

EMERGING TACTICS

Through what means can we trigger reflection? This chapter synthesizes the findings from literature and early explorations and translates them into four tactics.

EMERGING TACTICS

'How do we make sure that we can make room for designing and developing while being mindful of the biases, stereotypes and values we have about gender? How do we integrate these reflections in our processes rather than confining them as an afterthought? What practical actions can we take in our daily practices to integrate diversity, equity, and inclusion into the design process itself?' (Diversity, Equity and Inclusion for Embodied AI, n.d.)

Asking myself and experts those questions, I identified 4 potential design directions that can create ambiguity and help the designer to become aware of the existence of bias in their own '*object worlds*' (Bucciarelli, 1994) by reflecting. The design goals are furthermore illustrated through metaphors (Zijlstra & Daalhuizen, 2020). Integrated in the design process they can become potentially productive for critiquing and changing the existing common and controversial norms and provide tangible resources for practising diversity, equity, and inclusion in design.

Instead of using AI as a direct and explicit intervention - precisely identifying and communicating the flaws in ones subjectivity and associations-I intend to use the non-human counterpart as an instance of confrontation, irritation, provocation and surprise, visualising common and controversial norms, to be challenged and reflected upon by the designer itself.

The here described tactics can act as starting points and interaction guidlines for the design experiments. They don't exculde each other, neither does an interaction require all four to spark ambiguity and reflection.

06.1. AUTO-CONFRONTATION

'*Professionals can learn to articulate their personal knowledge by observing their own actions*' (Epstein, 1999). Auto-confrontation hereby refers to this process of confronting someone with the traces of their own doing.

Designing auto-confrontational experiences is already a common and important practice in wearable technology. It is argued that this methodology and approach could also become beneficial in other design and research processes (Martelaro & Ju, 2018). Such an auto-confrontation is not only a trigger for reflective processes, but can also enable users to '*recall and reflect on more than is recorded in the data, providing critical insight for further design and development.*'(Martelaro & Ju, 2018)

This mental process is also referred to as '*hyper-recall*', a moment of '*remembering*' in more details than just the initially present ones (Expertinterview Foster, 2022). Such auto-confrontation is furthermore practised as a means to trigger internal reflection. Clinicians who become aware of the blind spots of tacit knowledge for instance, gain insight into the influence of the observer, for example, when they review their own videotaped patient visits (Epstein, 1999).

Such a confrontation and its effect can be described by the metaphor of an astronaut looking down to earth. By looking at the entirety of humanity's traces on earth, he is suddenly able to experience the fragility and vulnerability of our world. This moment of confrontation might then lead to a change in behaviour or at least a reflection on such.

"I also see a lot that I don't like. A burning jungle, melting glaciers, lakes that used to be much larger. Fragile like a soap bubble. Shortly after arriving on the ISS, every astronaut becomes the „most intensive environmental ambassador one could wish for."

Figure 33- auto-confrontation metaphor

06.2. SHIFT IN PERSPECTIVE

A change in perspective in order to free the mind of its own biases and assumptions about the world is not a new idea. In her essay '*A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century*' Donna Haraway creates a cyborg lens to challenge essentialist/ biological concepts of gender and our traditional concepts of identity (Haraway, 1991).

By introducing the cyborg metaphor she breaks down distinctions between machine vs. organic, human vs. animal, natural vs. technological and social reality vs. fiction (Haraway, 1991). It is the shift in perspective through the means of a metaphor that can hereby allow critical reflection without constraints of common norms and expectations. Seeing yourself and the world around you through the eyes of someone or something else can free the mind from its blind spots.

For my metaphor I chose the concept of '*thought-experiment*'. Illustrated on the right you will find the '*brain in a tank*' experiment. Philosophers use this metaphor to free their mind from existing norms in order to freely evaluate questions like '*is life an illusion?* (Schweizer Radio und Fernsehen (SRF), n.d.). Creating a different reality thereby helps to shift perspectives.

Figure 34 - Shift in perspective metaphor "the brain in a tank" (Schweizer Radio und Fernsehen (SRF), n.d.)

06.3. CLASH OF EXPECTATIONS

Often reflection is prompted by a critical incident involving an error, a difficult situation, or an unexpected result of one's actions (Epstein, 1999). One form of creating error can hereby be a staged clash in expectation. This can be achieved, either by breaking with already existing conventions and expectations or by creating new expectations first, that are then destroyed on purpose.

The moment this expectation collapses creates a discomfort that will force the designer to think and reflect consciously. It can be seen as interruption of habits and well-known practices. Such moments of error are experienced regularly, and are furthermore not a new instrument in the field of human-machine interaction.

Martijntje Smits for instance describes the usefulness of the '*uncanny valley*' as a way of stimulating human behaviour and skill by sparking a moment of reflection (Figure 31). It is often those moments she argues, where new ideas are discovered. According to Martijntje (2020) those moments where you expect one thing, and get another, almost always evolve around new technologies. This idea also relates to Schön's reflective conversation with the materials of a design situation (Schön, 1984).

> "One might say that the prosthetic hand has achieved a degree of resem-blance to the human form, perhapson par with false teeth. However, oncewe realize that the hand that lookedreal at first sight is actually artificial,we experience an eerie sensation. Forexample, we could be startled during ahandshake by its limp boneless griptogether with its texture and coldness.When this happens, we lose our senseof affinity, and the hand becomes uncanny. [49]"

Figure 35 - clash of expectation metaphor "humanoid prosthetics (Mori et al., 2012)

06.4. CREATING MONSTERS

As dedicated the title of this project, another form of error that can spark reflection is the creation of '*Monsters*'. Mary Douglas shapes the word '*Monster*' as a creation of two excluding categories that can not be put back in place (Douglas, 2002). Just naturally we structure the world around us in categories, as a way of simplifying its complexity. Those categories thereby always remain a model of reality, and can never truly capture reality as such. It is therefore not surprising that we are now and then presented with situations that puzzle us. Moments where we are not able to assign a known category to things or people.

Not all of these clashes of categories have to be '*Monsters*'. Sand in the living room for instance, is such a situation of irritation, because neither sand belongs in the category of living room nor the other way around. It can nonetheless be easily resolved by, for instance, putting the sand back in the garden. However, moments of irritation and clashes of categories can also occur, without the possibility to resolve other than re-framing the set of categories.

It is such a moment of reflection, sparked by the wish to escape the unpleasant moment of irritation, that can break up a too narrow view on the world. My metaphor picks up on the examples given by Mary Douglas (Douglas, 2002) and illustrates the '*ready made*' as conceptualised by Marcel Duchamp (Just a Moment. . ., n.d.). By combining exclusionary elements he created '*Monsters*'.

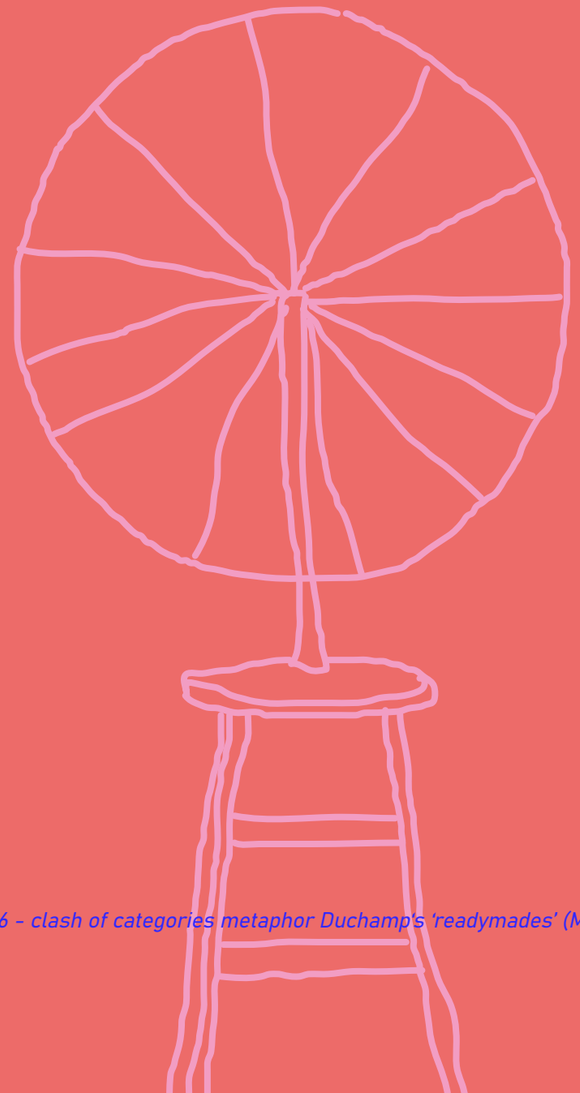


Figure 36 - clash of categories metaphor Duchamp's 'readymades' (MoMa, n.d.)

SPECULATIVE VISION

Answering the question of how the speculative body of this project is developed, this chapter introduces the first gender fluid child toy company as response to the problematic representation of gender stereotypes in toys.

7

SPECULATIVE VISION: THE FIRST GENDER FLUID CHILD TOY COMPANY

07.1. SPECULATIVE DESIGN AS RESPONSE TO TOYS AS CULTURAL SIGNIFIER

As stated at the beginning of this report, toys, books, and educational resources for kids frequently carry stereotype-based gender associations (Raj & Ekstrand, 2022). According to research (Pomerleau et al., 1990), males and girls encounter significantly different surroundings very early on in their development. The results showed that boys were more likely to receive tools, big and small cars, and athletic equipment. Girls had access to more dolls, fictional characters, kids' furniture, and interactive toys. They wore pink and colourful apparel more frequently, and they had more pink jewellery and pacifiers. Boys wore more blue, red, and white clothing than girls did. There were more blue pacifiers available. It is anticipated that these diverse surroundings would affect how children acquire certain skills and preferred activities. (Pomerleau et al., 1990)

Because of their location within prevailing interaction narratives and stereotypes as well as how they get contextually configured in use, toys have grown into '*cultural signifiers*' (Almeida, 2017) that represent norms and values of contemporary society. These early sociocultural experiences compound over time, maintaining gender stereotypes and affecting children's motivation and cognitive development (Wang & Degol, 2016).

Understanding the responsibility for design that comes with the understanding of toys as cultural signifiers and answering the question of how the fictional character of my work is conceptualised, I created a speculative vision: '*The first genderfluid child toy company*'.

This vision serves as a tool for presenting queer future AI-design practices, but also as a critique of current gender stereotypes in the design of children's toys. The aim is to integrate others into the discussion about alternative ways of human-machine interaction for a fairer society. Despite the fact that the company itself might never exist, nor the toys it creates, it can still serve as mental support in envisioning the not yet existing - the queer, non-normative future.

In order to build such a productive narrative, a targeted bridge to the perception and experience of the audience must be created, a speculative vision that informs the use of technology, aesthetics, behaviour, interaction and function of the design artefacts (Auger, 2013). The challenge is to find a connection to the audience's perception that presents itself as plausible, while being slightly controversial and provocative in order to not assimilate into the normative and thus remain unnoticed (Auger, 2013). The audience will not relate to the plan if the vision veers too

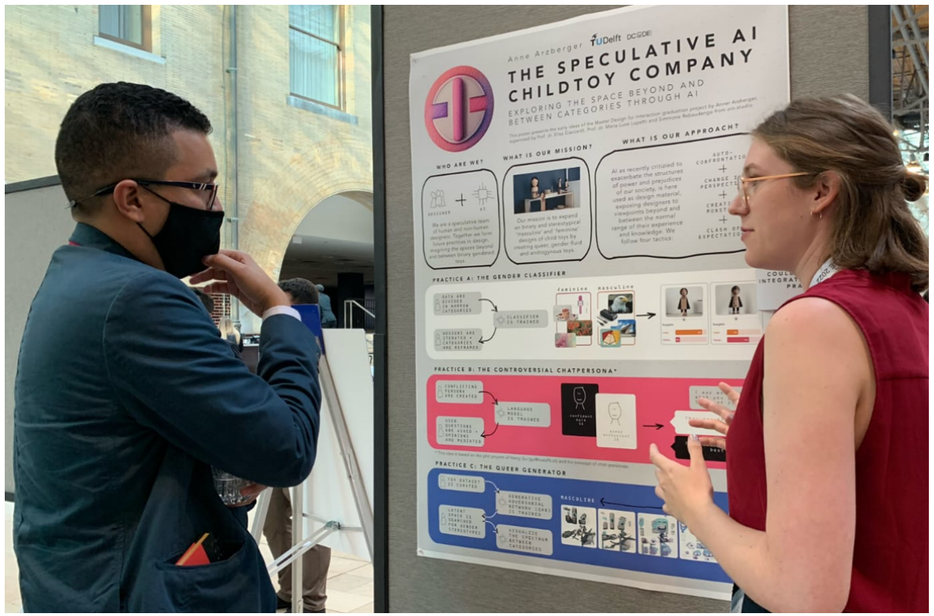


Figure 37 - testing of speculative vision in engineering context

far into the future to portray unrealistic notions or alien technological environments, which will lead to a lack of engagement or connection (Auger, 2013). The design solution is complex and paradoxical; it is familiar while yet being provocative. This contradictory human response that evokes a sense of familiarity while still being strange was dubbed *'uncanny'* by Sigmund Freud (1990). Whilst keeping this balance poses a great challenge, when achieved responses to the design concept tend to be both meaningful and strong (Auger, 2013).

My fictional company can serve such a controversial and norm-critiquing purpose by bringing in the notion of the '*first genderfluid toys'* – implying that current toys are rather stereotypical. Furthermore does the concept of queerness pose some controversy itself, as many are still believing in the idea of binary gender. Toys present a familiar concept, yet queer or genderfluid toys are alien enough to raise attention.

This speculative vision was furthermore tested in an engineering context (IDETC conference poster sessions), where the concept of queer AI futures stood out in the context of otherwise rather traditional engineering work (Figure 37). The engineering audience was intrigued by the idea of reflexive AI for child toys. Aiming to target such current human-AI practitioners, the speculative vision thus tested to be successful in raising critice at current practices, while leading the attention to an alternative future practice proposal.

07.2. INTRODUCING THE COMPANY

The following is meant as an introduction to the visual narrative of my speculative child toy company. Whereas the strategy of which was introduced in earlier chapters, this part is meant to highlight the designerly work that went into making this fictional narrative as immersive and relatable as possible. Such a strong immersive narrative is serving both communicational and creative purposes.

Visually the company primarily presents itself via a website, as common in the industry (Figure 38). The website provides information about the team, the mission, the design tactics and of course the product portfolio. Each product represents one of the experiments that will be described in the following part of this chapter. As such it provides an easy to grasp overview over the project in a compelling narrative, that helps visualising the purpose and procedure.

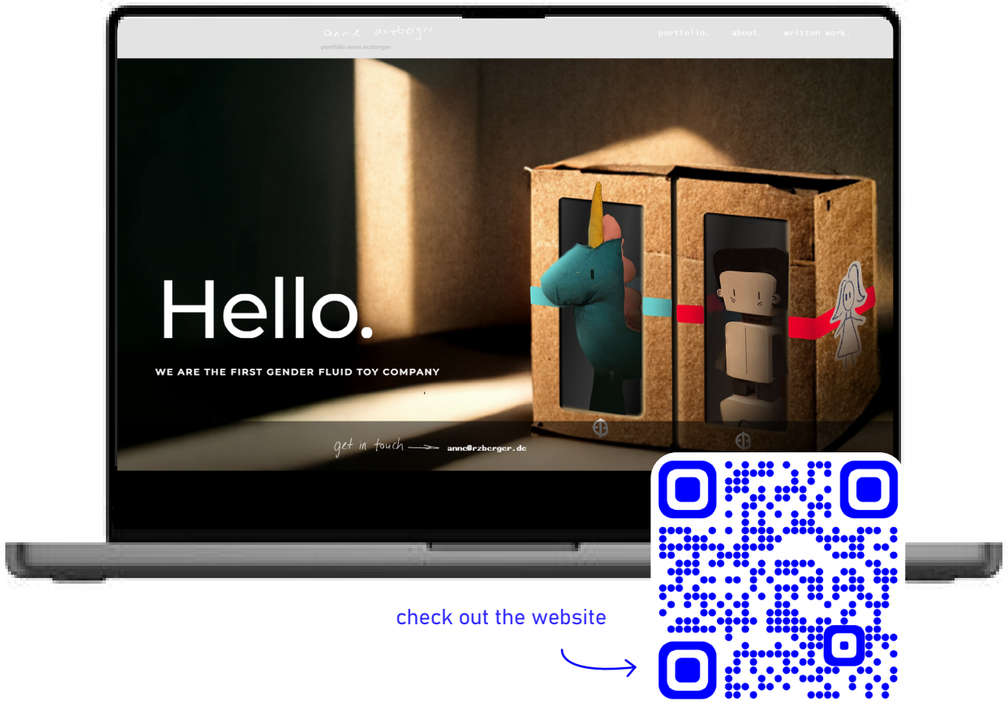


Figure 38 - Welcome page of company website

Figure 39 - The company's mission as presented on the website

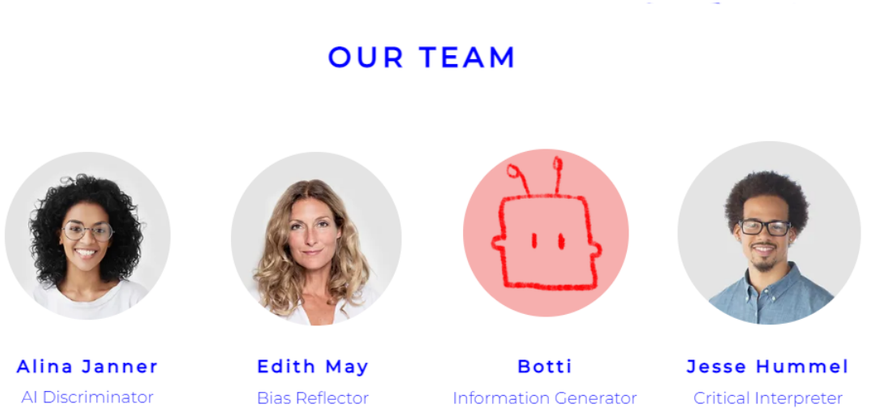


Figure 40 - the design team as presented on the website

DESIGN EXPERIMENTS

As core part of this project, this chapter illustrates the design experiments that explore new reflexive interactions between designer and AI in the context of gender fluid child toy design. Three experiments are shown, that each explore different nuances of reflexive interactions in context.

DESIGN EXPERIMENTS

08.1. PRACTICE A: THE ROBO-DOLL EXPERIMENT

As discussed throughout this report, classification algorithms represent the idea of AI as a categorical measure like no other algorithms. That made this kind of model especially interesting for presenting an alternative interaction with AI that helps designers re-frame their categories and create ambiguous objects.

The goal of this experiment was thus to design a queer toy, by using a classification algorithm. This experiment is based on one of the early explorations as described in Chapter 5. A more detailed description of the algorithm's functionality can also be found there. The idea is to use the procedure of data labelling to encode the designers gender related bias in the data. This individual perception of gender can then in turn be visualised in numbers of uncertainty by the classification algorithm, leaving the designer a trainable measure to mitigate against their bias.

As such this experiment is combining the tactics *'auto-confrontation'*, *'shift in perspective'* and *'clash of expecations'*.

This experiment furthermore integrates text-to-image algorithms like MidJourney in order to create and analyse visual representations of the designers mental ideas. This part of the procedure is meant to create up-front inspiration and early learnings on the individual gender perception of the designer.

I kicked-of the robo-doll experiment with a set of MidJourney generated images on my toy idea for a robot-doll like toy. Input text prompts reached from '*queer robot*', *'genderfluid toys'* over '*wooden robot-doll'* (Figure 41).

Figure 41 - inspirational input from MidJourney as starting point

Figure 42 - outcome of the first experiment

I then selected the images that were the closest to my imagination of the toy. Spinning this AI input further, I analysed those images in more detail (Figure 43). I paid most attention to my own gender perception, which I then tried to understand. Visualising the images I had in my mind helped me identify better which design elements affect the gender as I perceive it.

Taking the just gained understanding about what elements affect gender perception into practice, I then started sketching out some first ideas for the robo-doll (Figure 44). I paid attention to the proportions of the different body elements, as I found them to most effectively influence my perception of the toys gender.

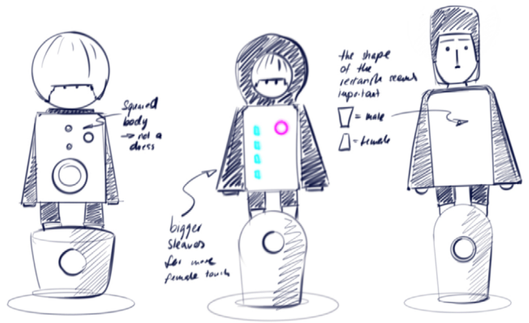


Figure 44 - First robo-doll sketched ideas

Figure 43 - analysis of Midjourney inspiration

As a consequence of the sketching exercise, I was able to identify the head of the toys as a very crucial element for my personal perception of the toys gender. I thus went back in a second iteration to Midjourney (Midjourney, n.d.) to generate more '*gender neutral*' heads (Figure 45). This again allowed me to better test what visuals work well for my personal perception of gender, since I was capable of judging existing images in terms of gender rather than coming up with neutral ones myself.

Figure 46 - Ideation of heads for robo doll

I was then able to translate the gained knowledge about heads and gender into own design ideas as sketched in Figure 46.

After this first part of the process, where I gathered understanding and inspiration both through traditional sketching and AI image generation, I pinpointed a few of the key design elements for my final design.

I aimed for a good balance between black and wood textures, the naturalness of the wood and the artificiality of the electronics, and of course a balance between the feminine and the masculine. Based on this decision I created a batch of three design ideas that I wanted to test with the classification algorithm.

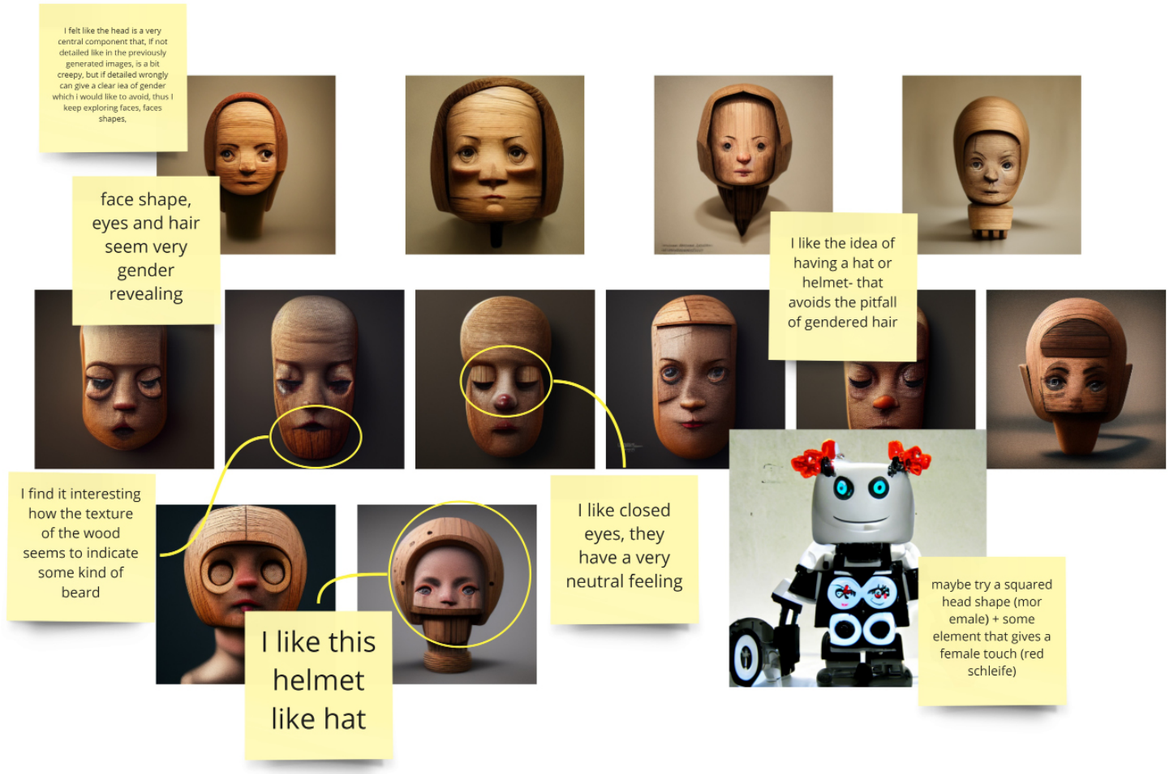


Figure 45 - Second iteration of Midjourney image generation studying gender perception of heads

However, to test those ideas I first needed to train my algorithm. I carefully selected a collection of images ranging from toys, to more ambiguous objects like art, food and tools (Figure 46). Once the selection of images was made, I started with the labelling process, in which I decided for every individual image whether I feel a more feminine or more masculine tendency. I ended up with a binary dataset of male and female images.

Figure 46 - Dataset of male and female objects for classification algorithm

After training those images through teachable machine[1] I was then able to test my first batch of three ideas as sketched before (Figure 47). Online algorithms like teachable-machine or RunwayML are usually pre-trained, meaning they already run several hours of training on millions of image data. Which data were used to train an AI can be seen for instance with the webpage '*Have I been trained?*' (Have I Been Trained?, n.d.).

This first testing (Figure 47) showed that a combination of a more masculine body with a more feminine head leads to the most ambiguous gender in comparison to the other two design options. However, a ratio of 34% masculinity and 66% femininity was not perfect yet. Aiming for a 50%-50% ratio, to be as much in between the two genders, I decided to extend this sketch-testing and went for a second iteration round. The changes made in each design variation were kept low in order to allow a guided design evolution. Some iterations showed a lower ambiguity than the previous design. In these cases I went back a step and tried a different design variation.

[1] *Online algorithms like teachable-machine or RunwayML are usually pre-trained, meaning they already run several hours of training on millions of image data. Which data were used to train an AI can be seen for instance with the webpage 'Have I been trained?' (Have I Been Trained?, n.d.).*

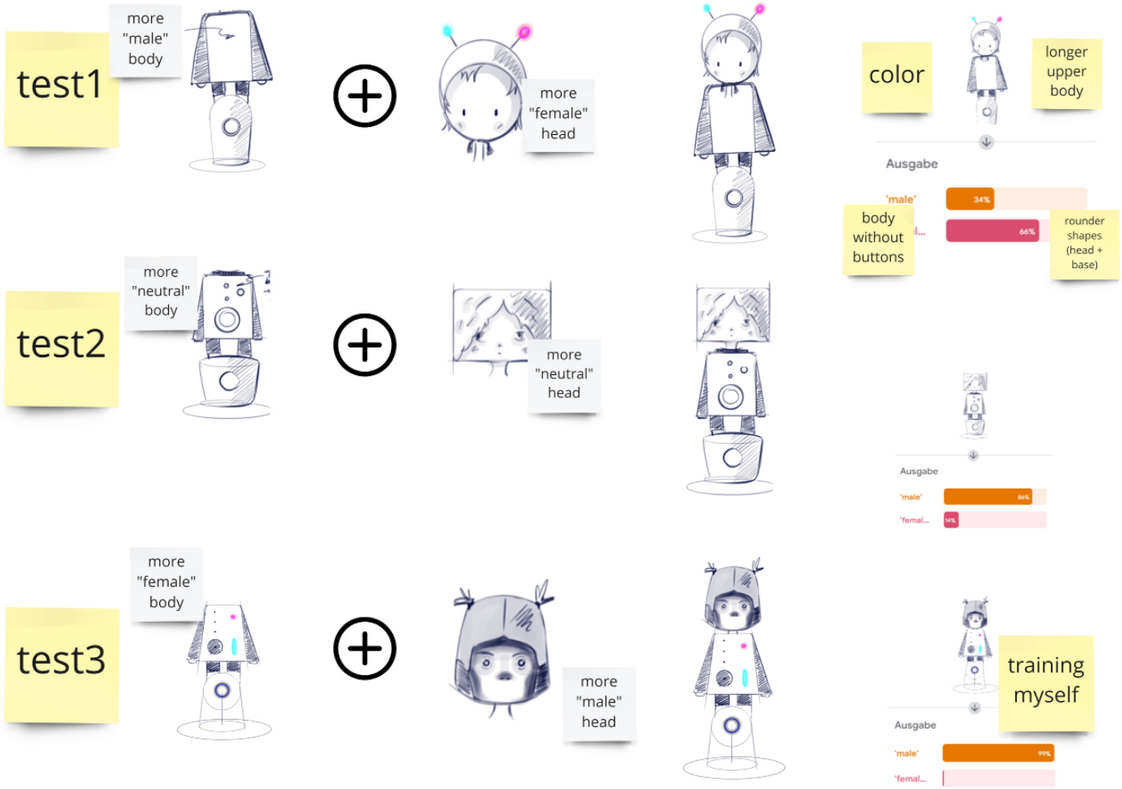


Figure 47 - First classification test batch

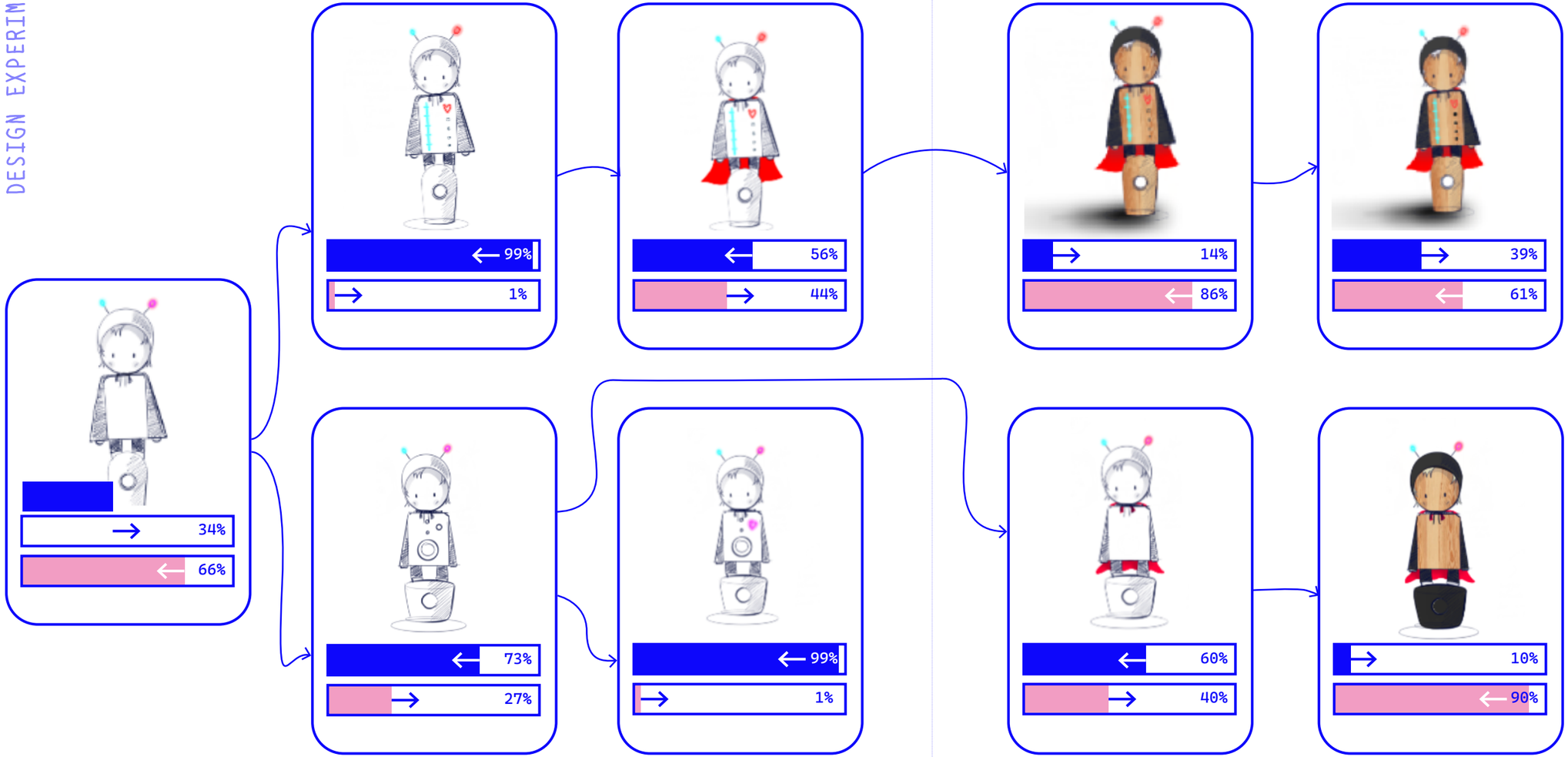
Having selected two sketches with promising ambiguity, I started with physical prototyping. To capture both three dimensionality and texture of the designs, I made a clickable prototype that allowed quick changing of differently scaled and sized body parts like arms, legs and bodies. For that, I laser cutted several design variations, sanded and painted them (Figure 48 + 49). While some pieces could just be clicked in place, others were held together by magnets, also allowing rotation of arms and head.

48 - clickable prototype through laser cutting

The different variations of the body parts were then systematically tested in front of the camera (Figure 49 + 50). A modified test setup was needed, due to the sensibility of the classification algorithm to slight changes in angle and scale. While the order of testing body parts is expected to have an influence as well, I focused on testing one sequence of variations.

Figure 49 - prototyping different design variations

Figure 50 - challenging physical design variations with initially trained classification algorithm



After the first sequence of tests (Figure 50), additional data were needed. First tests showed that the original collection of images was not sufficient to cover the variance in design I was trying to challenge. I thus selected additional image data in the categories of shape, texture and colour which I again classified in masculine and feminine (Figure 51).

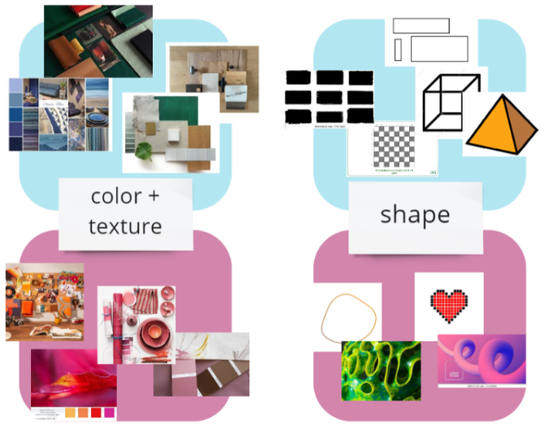


Figure 51 - additional training data being labelled

With my re-trained classifier, I ran a second test on my prototype variations (Figure 52, following page). This time I was able to indeed capture the nuances in design and make conscious decisions towards more ambiguity in the combination of different design elements.

While the final results were not yet showing the level of ambiguity sought, the set up of the experiment allowed me to well identify the elements in my design that need improvement. As can be derived from Figure 52, the cape showed to be a less ambiguous feature than initially suggested when testing the first sketches. This inconsistency can be grounded on either the re-training on additional data, or simply the additional dimensionality and texture of the prototype compared to the sketch. However, it points to the fact that the integration of a cape in the design should be more carefully considered.

As the testing furthermore suggested, more design ideas on the head are needed. Current ideas were not able to increase or simply keep the level of ambiguity. More iterations of ideas were developed and tested as can be seen in Figure 53.

Finally the best combination of body parts as tested with the classification algorithm was prototyped and electronic parts were added.

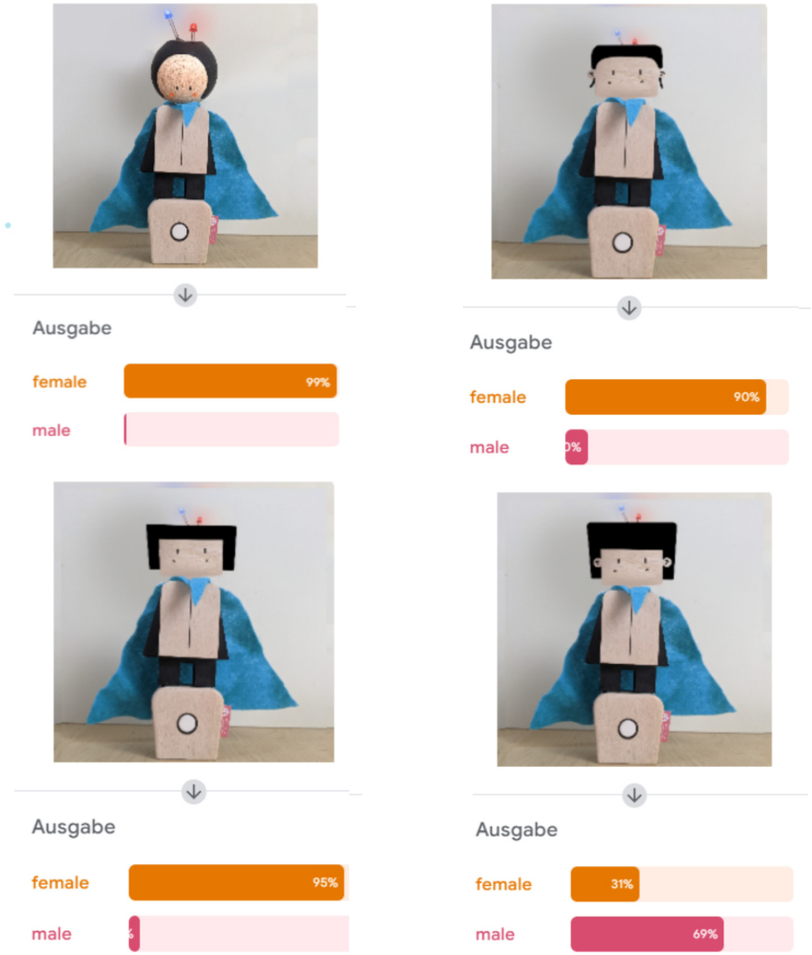
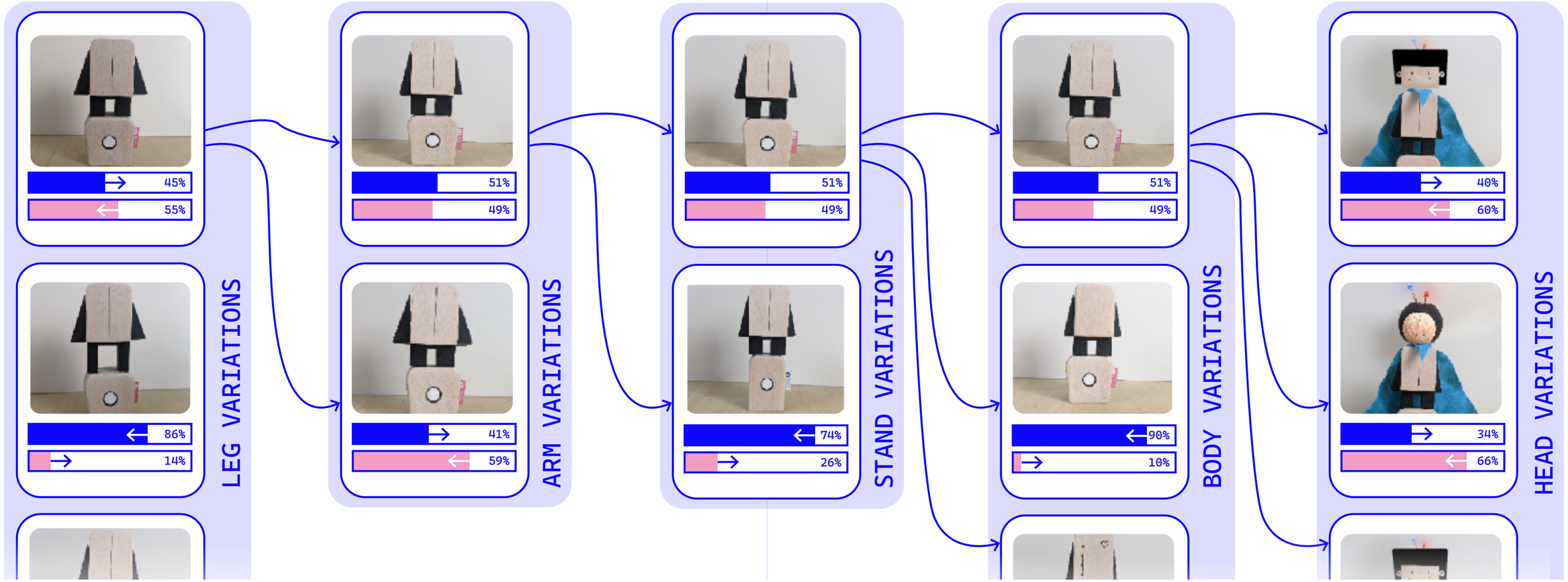


Figure 53 - more head design testings

Figure 52 - challenging physical design variations with re-trained classification algorithm



08.1.1. Reflection Practice A:

While the overall results generated by the AI could be used as measures for more conscious decision making, it was more the overall procedure and reflection in and on actions that seemed productive. Categorising the data for instance, revealed a level of biases in my own thoughts that I was previously unaware of. I was furthermore able to identify patterns in the categories of masculine and feminine myself, like for instance the link between colour, shape and gender. Furthermore the classifier acted almost like a mirror, reflecting my own bias in concrete and tangible numbers, making decision making much easier.

However, the level of sensitivity of the algorithm made consistent testing difficult. While I believe that no general consistency of the algorithm is needed, it has to stay persistent throughout the individual experiments. Slight changes in camera angle and position of the prototype however seemed to strongly affect the classification as can be seen in Figure 55. A potential solution could be adding more data of the plane background to both ot the training categories or the construction of a simple test stand on which camera and prototype are held in the same place throughout the testing.

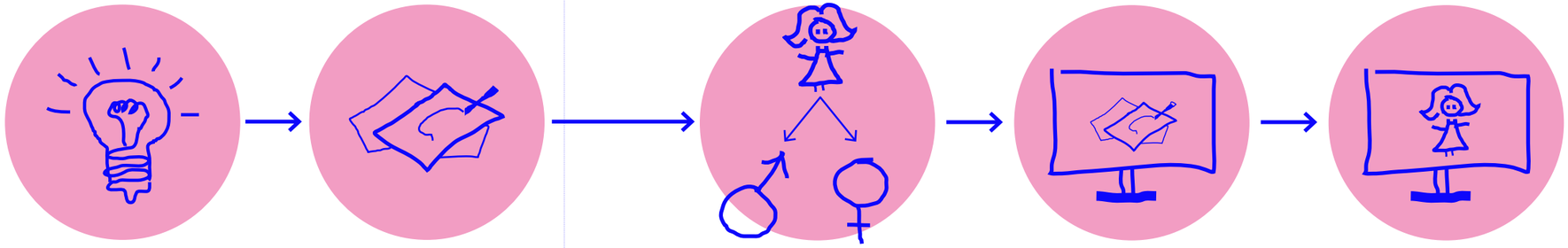


Figure 54 - overview experiment with robo-doll

Reflecting upon the interaction, I was furthermore asking myself, what the added value of the AI was in the process and if a classification exercise could have not created a similar impact. Of course such an impression would have to be further verified in a comparison study, however I experienced a change in confidence when making my decisions based on the measures I created.

However, those measures could easily be adjusted to better accommodate my vision of the design, leaving the question whether the tool steers me towards the right design, or if I steer the tool to generate the design I want. Such difficulties could be potentially avoided by carefully curating a batch of images, that includes texture, colour and other usual design elements, which are then just labelled depending on the context specific categories.

Critically looking at the overall productiveness of this reflexive interaction, I would thus argue that the classification algorithm and the interaction exercises can be potentially of value when aiming for a specific design outcome, such as gender ambiguity in toys. Following up on earlier tests and workshops with this algorithm, I could also see how seamless such a tool could be applied in specific design contexts. It furthermore leaves the freedom to be adapted to individual designers preferences and needs.

The classification measures, combined with the data curation exercise provide the important *auto-confrontation'*, *'clash of expectations'* and *'shift in perspective'*

The final evaluation of productivity will be evaluated in the chapters '*Reflections*' and '*Discussion and Conclusion'*.

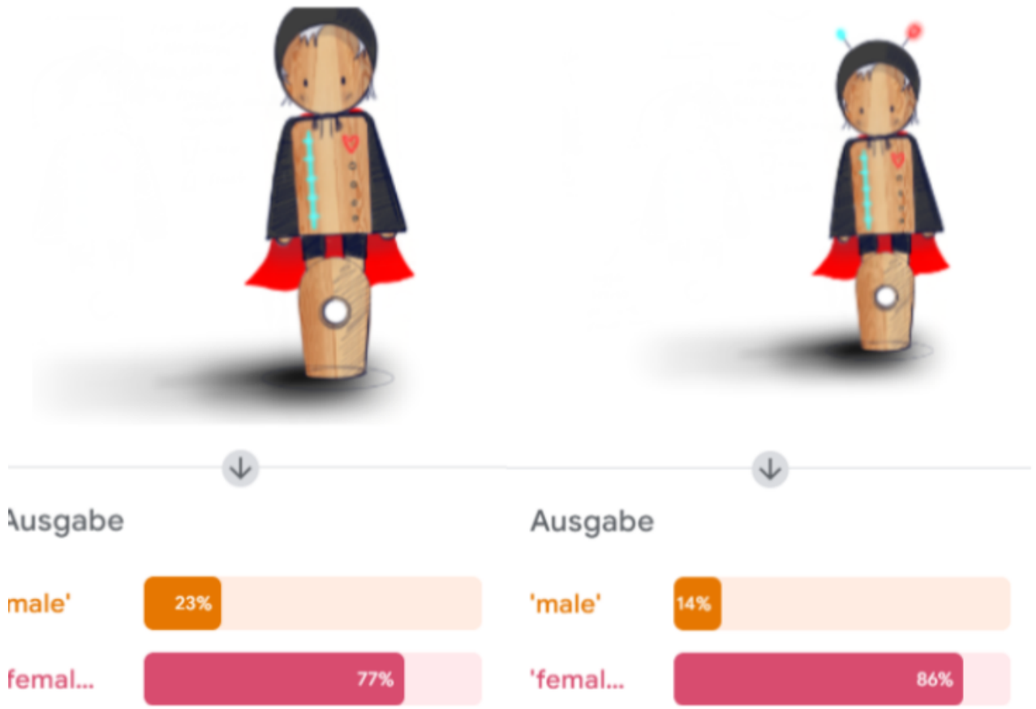


Figure 55- different classification outcome of same design in different camera positions

08.2. PRACTICE B: THE DINO-UNICORN EXPERIMENT

While the aim in the previous experiment was to create an in-between of robot and doll, the second experiment sought to find the creature inhabiting the space between dinosaur and unicorn. Unlike in the previous testing however, a StyleGAN was used instead of a classification algorithm. As such, this idea builds on the early exploration with latent space and latent walk as described in chapter 5, where also the technical aspects are illustrated.

Figure 57 - training data

The idea tested in this experiment is to use a StyleGAN's ability in illustrating images as infinite and related elements in space, to explore the in-between categories and re-frame the collective imaginings of masculine and feminine.

As such this experiment builds on the tactics *'clash of expectations'*, *'shift in perspective'* and *'creating monsters'*.

I started with curating a dataset consisting of unicorn and dinosaur plush images (Figure 57). About 300-400 images were collected and trained in RunwayML as an unlabeled data batch on top of a StyleGAN that was pre-trained on bird illustrations (see Figure 58).

Figure 58- training plushs on birds with RunwayML

The first training took around 4 hours. The resulting images showed similarities with the training data, but were not yet identifiable as plushs (Figure 59, following page). However, I ran a few tests to also experience the interaction with more ambiguous trigger images. While these tests were able to show interesting ambiguous creatures that I would have probably not pictured myself, they either seemed too '*monster-like*' thus not feminine enough, or simply too distantly related to dinosaurs or unicorns and thus not really challenging (Figure 60).

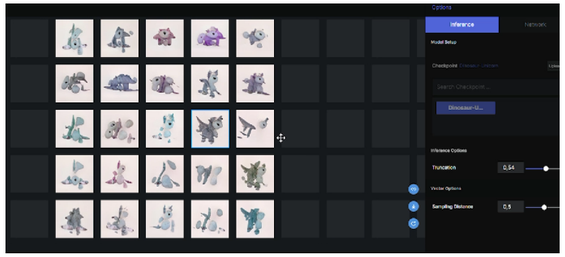


Figure 60 - latent space exploration

In a second iteration the same model was trained an additional 3 hours on the same data, in order to improve the similarity between the training data and the AI generated images. The resulting quality of images was considered high (Figure 61, following page). Further proceeding with the testing, the latent space was again scanned for images depicting the in-between creatures. Although results were perceived as slightly less surprising, compared to the previous round of testing, the images were less ambiguous and easier to make sense of. Searching thus didn't require a long time. By tweaking the latent space's parameter, I could illustrate more or less similarity between neighbouring images.

Figure 59 - first iteration results from latent space

Figure 61 - latent space exploration results from second training batch

Figure 62 - outcome of the first experiment

Secondly, I used the latent walk function of RunwayML t interpolate between images of unicorns and dinosaurs This helped me visualise the vast amount of images sit· ting between the two categories (Figure 63).

Collecting a few images as inspiration material through both, latent space and latent walk, I created a mood boarc for my ideation. I then started sketching out a few of the gathered ideas in little doodles (Figure 64). I chose the design that I perceived as most in-between unicorn and dinosaur, yet somewhat familiar and relatable with the two categories.

With the chosen sketch as the starting point, I started preparing a layout for sewing the prototype. The choice of colours and fabrics was made consciously to underline the playing with categories. More dinosaur-like elements were kept in more feminine colours, while unicron elements were coloured in darker more blue-ish colours Finally the different fabrics were cut out and sewed together (Figure 65). The result still captures my initially sketched idea well.

Figure 63 - interpolation between dinosaur and unicorn

Figure 64 - sketching ideas based on inspiration gathered through latent space and latent walk

Figure 65 - prototyping of unicorn-dinosaur plush

08.2.1. Reflection Practice B:

Looking back at the experiment, I would again rate the overall experience higher in terms of productiveness, then just the images generated by the AI. As such, my observations and experience does not differ much from the early explorations with the StyleGAN as described under '*Early explorations*'. The experience of searching an infinite space of all related images created new awareness about the absurdity of one's categorical thinking. Thus, I found myself forced once more to question my own perception.
While the first test iteration brought surprising and ambiguous ideas, they seemed too far away from anything the kids could relate to. I realised not having any relation to known categories might not have the intended effect of bias awareness. Despite being more relatable though, the images generated by AI in the second, more refined round of training, did not bring the new surprising insight or idea anymore. As a trained designer, I could have easily imagined such a toy between dinosaur and unicorn without the AI. I was furthermore able to draw upon knowledge about AI from my previous experiments, making the curation and collection of data, as well as the training process rather easy. However, it can be assumed that such knowledge is not given for any designer. Unlike the classification algorithm which also allows untrained designers to quickly learn and adapt, StyleGAN training procedures are more complex and time intensive, making additional upfront training potentially necessary for other people to use the tool the same way I did.
This experiment mostly builds on the tactic of '*creating monsters'* and *'clash of expectation',* by visualizing the confronting and surprising *'in-between creatures'.* More detailed reflection, under consideration of aspects like sustainability and ethics can be found in section '*Reflections*'.

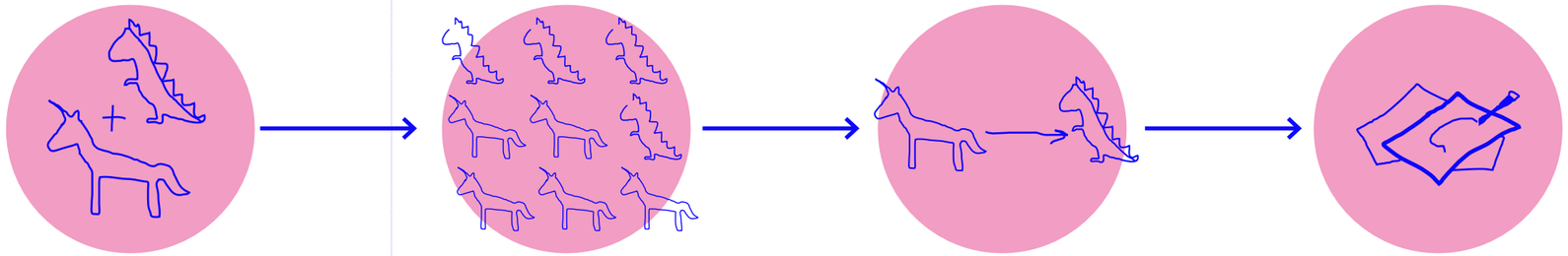


Figure 66 - overview work process experiment 2

08.3. PRACTICE C: THE DRILL–HAIRDRYER EXPERIMENT

In the third experiment, I wanted to highlight the AI's ability to mediate between different categories. As a probability based algorithms, many of the AI outcomes simply represent the average found in the training data. The idea was to diverge from an initial category of drill toys, to something drill-ish, yet gender neutral. Despite not entirely based on a previous exploration, the technical details of a StyleGAN, are described in section 5.

This experiments build on the tactics *'creating monsters'*, *'clash of expectation'* and *'shift in perspective'*.

I started by creating a primary dataset of drills. I then created a second data batch with images of hair dryers. The hairdryer was chosen as the feminine counterpart to drills, as they show high visual similarity, despite reflecting the opposite gender role (Figure 67).

The secondary dataset was slightly smaller than the primary one. Each however had to consist of hundreds of plane object images. To ease the process of data collection, I wrote a little program that helps me augment my data, by rotating, flipping and colouring the existing ones (Figure 68). This program allows me to generate clean data in sufficient amounts in short time and allows me for more control over my data compared to data scraping.

```
[ ]  import albumentations as A
     import cv2
     import imageio
     import imgaug as ia
     import imgaug.augmenters as iaa
     import matplotlib.pyplot as plt
     import random
     from albumentations.augmentations.transforms import ChannelShuffle

[ ]  for i in range(1, 31): #amount of images in folder +1
       input_img = imageio.imread('/content/drive/MyDrive/Data Augmentation/glasvases/{}.jpg'.format(i))
       #HorizontalFlip
       transform = A.HorizontalFlip(p=1)
       augmented_image = transform(image=input_img)['image']
       imageio.imwrite('/content/drive/MyDrive/Data Augmentation/glasvases/augmented/{}_1.jpg'.format(i)

       #Scale and Rotate
       transform = A.ShiftScaleRotate(p=1)
       random.seed(7)
       augmented_image2 = transform(image=input_img)['image']
       imageio.imwrite('/content/drive/MyDrive/Data Augmentation/glasvases/augmented/{}_2.jpg'.format(i)

       #Color Change
       transform = ChannelShuffle(p=1.0)
       random.seed(6)
       augmented_image3 = transform(image=input_img)['image']
       imageio.imwrite('/content/drive/MyDrive/Data Augmentation/glasvases/augmented/{}_3.jpg'.format(i)
```

Figure 68 - programming data augmentation

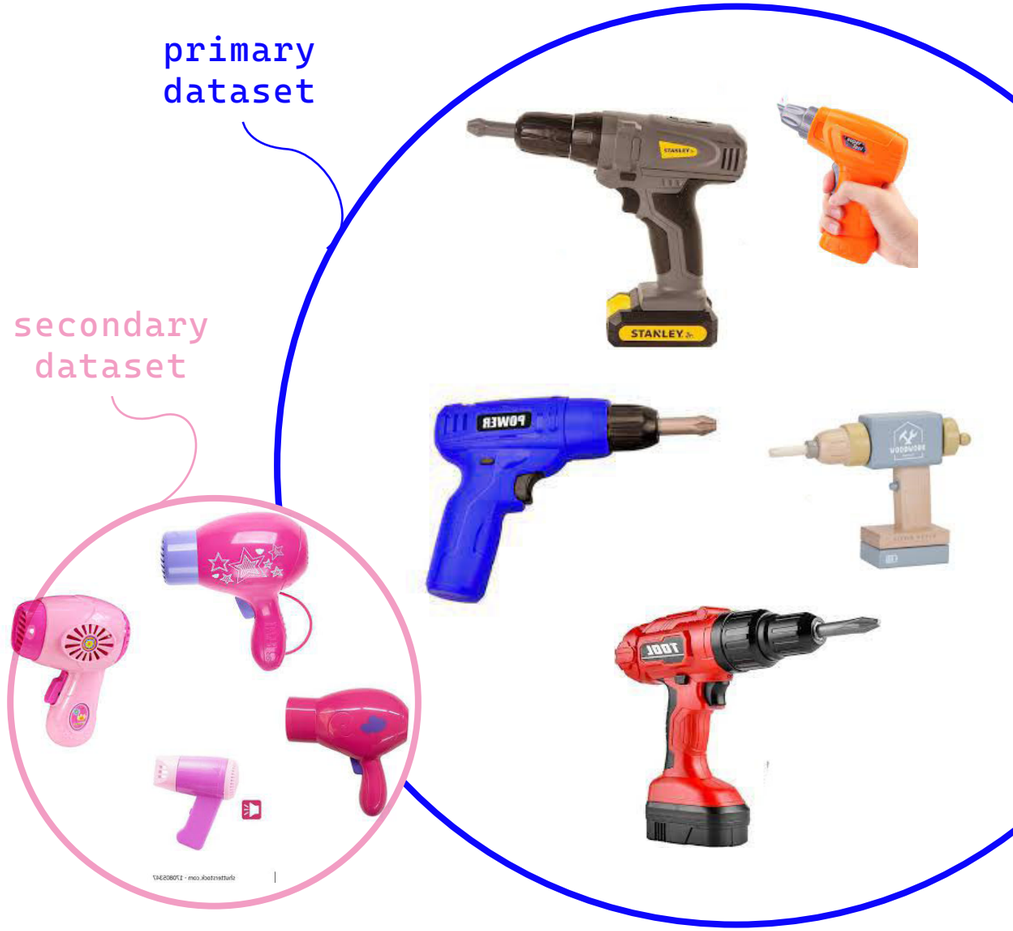


Figure 67 - data preparation for third experiment

Figure 69 - outcome of third experiment

Next, another StyleGAN from RunwayML was trained on the primary data in around 5 hours time. The outcome as can be seen in Figure 70 shows drills, similar to the ones in the primary dataset.

Figure 70 - original image generations based on only primary data

In order to diverge from this initial and traditional idea of a toy tool, I then robotonized the data by training my drill-pre-trained algorithm on my secondary data. However, this additional training round was kept as short as possible. Only another half an hour was trained with the hair dryer data.

This resulted in artefacts that still remind of drills, however they would make drilling impossible (Figure 71). The affordance of how to hold and use it remains, while the context of use feels uncertain. Such an artefact could be imagined as a meta-object, teaching and training certain motoskills, without suggesting a stereotypical context of use.

Figure 71 - robotonized images based on primary and secondary data

Expanding on the idea of confusing the drill with another category of objects, textures or features, I ran a second training round, where I trained images of vases as secondary data on top of the drill-pre-trained algorithm. Vases, so the idea, do represent a lot of feminine characteristics like fragility, and passiveness, which are encoded in their visual appearance.
Vases have furthermore defined contours, other than simple fur or glitter textures. This makes training and readability of the outcome easier, as will be explained under reflection.

The second robotonisation also showed high drill similarity, while making its original function impossible (Figure 72). However, the fragility and femininity of glass could not be illustrated. As such this small exploration was discarded as productive for the overall idea of this experiment.

I ran a third test, this time using a deep style algorithm (Deep Dream Generator, n.d.)(Figure 73). This algorithm requires no long training, and only uses two images as input data. The first image hereby represents the base image, as similar to my primary datasets earlier, while the second one acts as a filter image, aka the secondary dataset. The secondary image however, is not simply mapped on top of the base image. The algorithm furthermore merges the features, while aiming to keep the base image's key structure.

Figure 73 - robotonizing experiments with deep style algorithm

Figure 72 - robotonizing with glass images as secondary data

The resulting images as can be seen in Figure 73, turned out rather unsurprising. They still remind me very much of a drill, rather than some kind of meta object.

Also discarding the third exploration, I went back to my initial testing with the hairdryer, as it proved to be the most productive in terms of re-framing the idea of a toy tool. I thus started further developing my idea of a drill-hairdryer by prototyping with different laser cutted pieces (Figure 74). Those explorations were then translated into a final prototype.

Figure 74 - prototyping different drill-hairdryer ideas

08.3.1. Reflection Practice C:

This work flow, even more than the previous ones, required a bit of experimenting in order to acquire the knowledge needed to get the intended results. As Figure 73 shows, not all of those experiments were as successful. Robotising data requires a careful selection of secondary data. If the images are for instance not showing an object with a defined shape, the initial shapes of the primary data do not simply get diffused, but rather entirely dissolved. However, if the primary data are presented with transparent backgrounds, their shape does not get transformed at all, leaving the simple impression of a filter rather than a re-imagined drill (Figure 76).

Figure 76 - difficulties in robotonizing with dissimilar secondary data

With this toy, as the first one with a more complex three dimensional shape, I furthermore struggled with translating the AI inspiration into a physical 3D object. However, only a few prototyping attempts were needed in order to arrive at an ambiguous toy drill. This experiment reflects the tactic '*creating monsters*' '*shift in perspective*' as well as *'clash of expectation'*. Again this reflection will be followed up in the last chapters of this report.

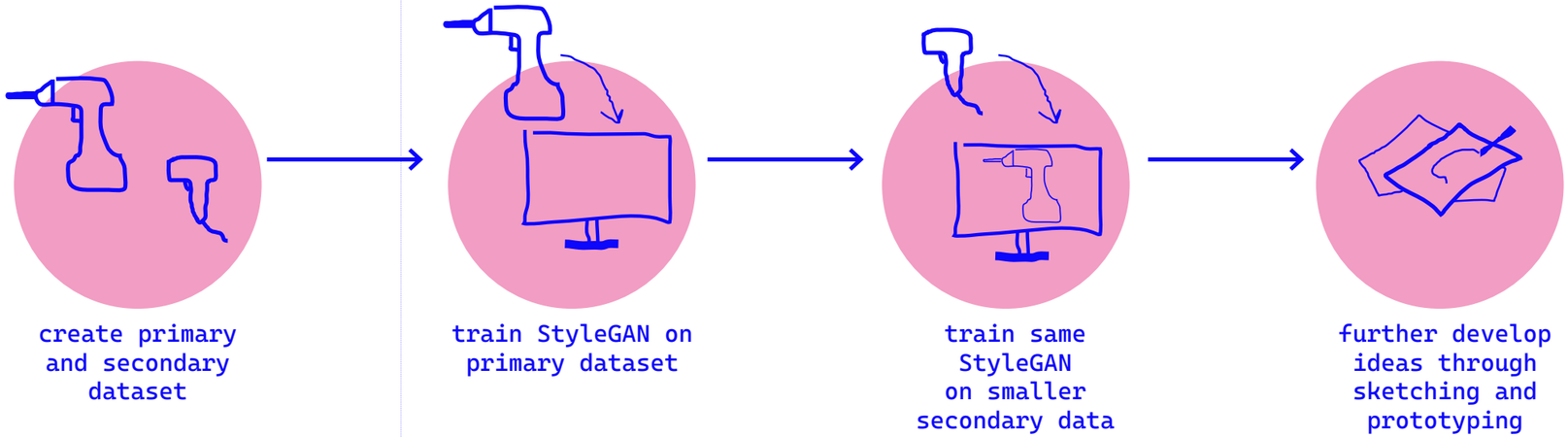


Figure 75 - workflow overview for third experiment

08.4. TYING FINDINGS TOGETHER

While each interaction differed, they all started with the carefully collecting training data and '*encoding of bias*' in the data. Not all off-the shelf AI models require additional training. However, the choice of algorithms that do need extra data and training was made consciously. This data curation exercise is not only creating awareness about the existence of bias in one's way of thinking, it furthermore informs the designer that bias will be passed to the AI. Something that is currently unavoidable, however not always clear for the user of artificial systems.

Despite being meaningful as design activity, the curation of data, as well as the training time required lot's of time in all three scenarios.

While the classification algorithm had a quicker response rate, all algorithms were able to respond to the designers data in one way or the other. This has the effect that the designer is either confronted with his/her bias, or presented with the non-normative in a way that also creates awareness of bias. Although all three algorithms are not producing precise data or measures, the fact that the algorithms can be trained by the designer, creates a personal link to the outcome that is being generated.

Such a personal connection to the AI's generations are potentially relevant in order to increase the '*shock*', the confrontation, surprise and confusion, that is key in sparking reflection.

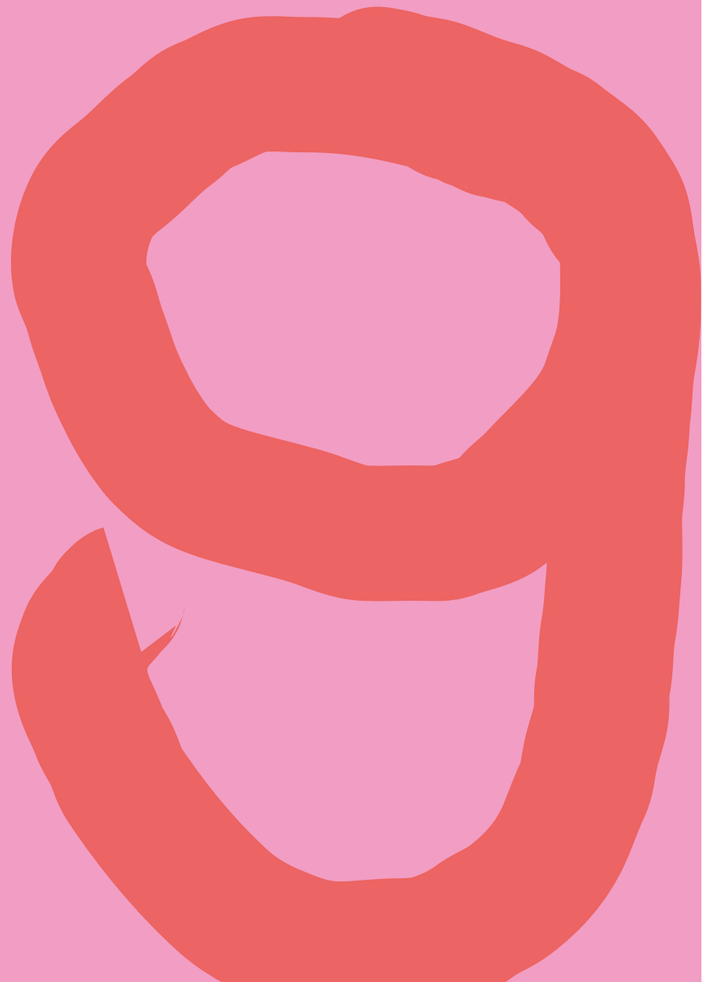
All three experiments showed how AI algorithms can be integrated in the design process. None of the measures or ideas the AI generated were directly used as '*solution*'. The designer always evaluated the AI's creations and stayed in charge of translating the confrontation, inspiration or suggestion into a feasible design solution.

All three experiments also posed new challenges to the designer. AI outcomes were ambiguous, challenging, inconsistent and often confusing and hard to make sense of. It can be expected that additional training for the designer is needed.

Concluding a work in progress definition of reflexive interactions is: '*reflexive interactions are a form of human-machine collaboration, where the AI is responsible for triggering and assisting the designer's process of identifying and challenging bias and collective imaginings, rather than actively proposing the ideal solution itself.*'

Figure 77 - tying findings together

In order to evaluate the design experiments and the proposed reflexive interactions, testing with children as well as expert interviews were conducted. The results are discussed in detail.



EVALUATION & RECOMMENDATION

09.1. PLANNING

The evaluation setup consists of two blocks. While the first evaluation focuses on evaluating the level ambiguity reached in the final artefacts, the second one focuses on more in depth feedback and recommendations regarding the reflexive interactions with AI.

09.1.1. Testing with Children

A testing with children aged 4 months to 4 years is scheduled. The children were presented to the three toys, one after the other. A second researcher was asking the children questions about their perception of the toys. The goal is, to evaluate the level of gender ambiguity the toys are reflecting.

The questions are kept simple. Some of the children are not able to speak yet, thus observation notes are taken as well as notes of the children's answers and discussions. No image or video material is recorded to protect the privacy of the children.

The three main questions are:
What does this toy remind you of?
What name would you give it?
Who do you think this toy is for? / Who would you like to give this toy to?
What could you do with the toy? / What games would you like to play with it?

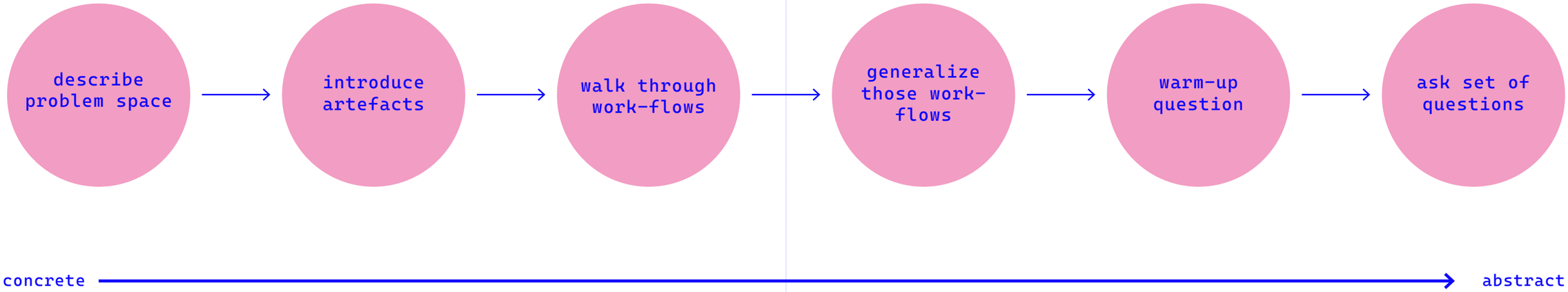
09.1.2. Expert Interviews

A series of expert interviews is conducted in order to discuss the productivity of the proposed reflexive interactions. The area of expertise ranges from traditional AI experts to child toy designers (Figure 79). A total of 7 experts are interviewed. As illustrated in Figure 78, experts are guided from the concrete artefacts, over the more abstract design workflows to more general questions.

More specifically, the interview begins with a description of the problem space, e.g. bias internalisation and perpetuation, and a quick introduction to the research framework and project context. It continues with a presentation of the toy artefacts, as representations and outcomes of the reflexive interactions between designer and AI. Those workflows will then be presented in more detail, finishing with a work-in definition of '*reflexive AI practice*'. This definition will be followed up in the warm up, asking participants to relate to this practice and definition from their own background and experiences. This will allow experts to individually relate to the ideas of this thesis.

Slowly raising the level of abstraction and generalisation, a set of questions (as can be seen in Figure 78) will be asked, in order to discuss the productivity of '*reflexive practices*'. How productivity is defined is also illustrated in Figure 80.

Figure 78 - evaluation setup for expert interviews



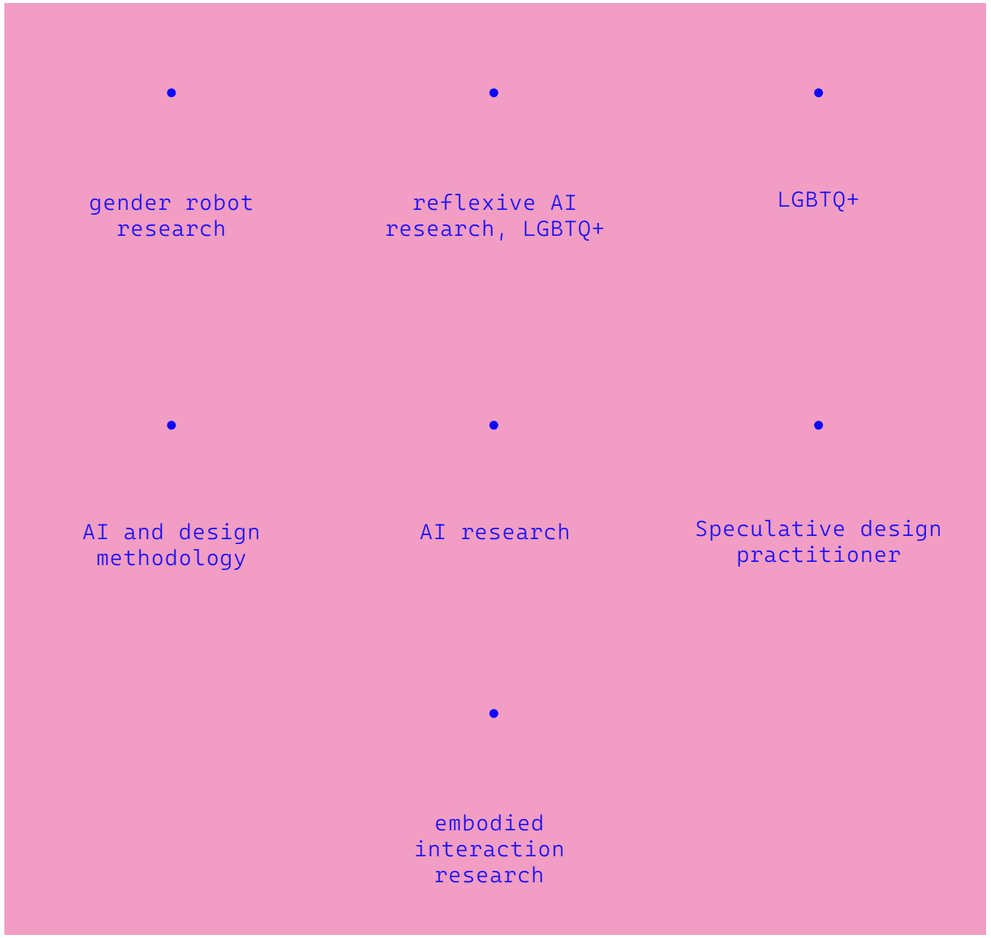


Figure 78 - List of experts

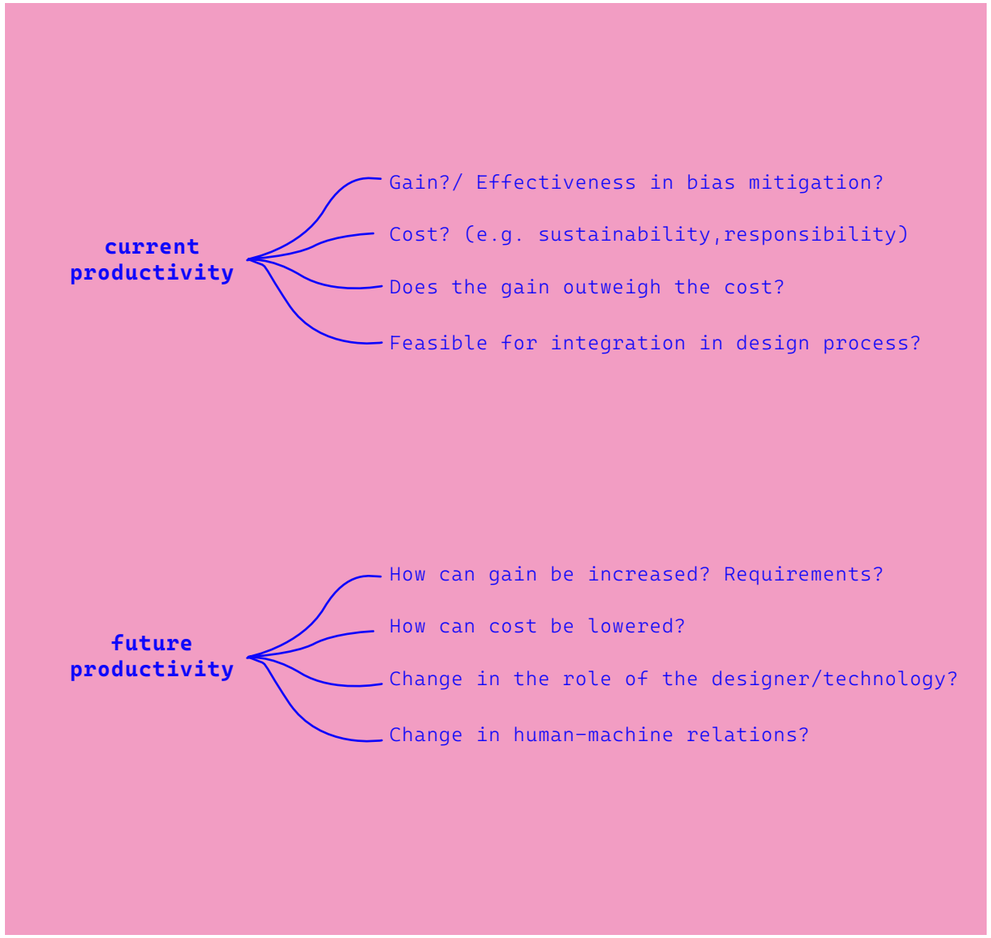


Figure 80 - questions for productivity evaluation

09.2. RESULTS – EXPLORING FURTHER NUANCES

A total of 7 interviews were conducted. Additionally a small testing with 9 children was done. In the following the insights from the expert interviews, as well as the testing with children are synthesised and grouped into: 1) value/gain, 2)cost/limitations and 3) recommendations.

Figure 81 - showing the toys to children

09.2.1. VALUE/GAIN - SURFACING, DISMANTLING AND DE-FAMILIARIZING:

All experts saw potential in the reflexive interactions. Following the understanding that stereotypes are the result of automated cognitive processes, it was widely agreed upon the key role of reflection in the process of identifying and challenging bias. Linking back to Donald Schön (1984), reflection was described as inherent to a good design process. Highlighting the importance of reflection, most experts then understood the AI as an important design assistant that 'interrupts' the automatic and routinized design thinking process.

'[...] *That's the cognitive process that I was mentioning. So if we [make decisions] instantly, we still fall within this like automatic process. So the reflection has to be there at each step, because without that reflection we would never have to reassess.*'

The '*non-human ideas and opinions*' that the AI brought into the design momentum, surprise and confront the designer, forcing them to reflect. '*If you look at the theory of reflective practice, what drives the process is surprise. So it is a sort of 'surprise trigger' [...]. Having your assumptions questioned can also lead to surprising results. [...] An AI that can do that is interesting.*'

AI was not only seen as productive in triggering reflective thought that can enable the designer to surface and identify bias. The experts furthermore saw the power of AI in inspiring and visualising new (gender) imaginings, norms and roles. Starting from the binary/normative was described as the essential starting point in order to enable the designer to dismantle the current binary imaginings. The designer is being confronted with the '*in-between*' and has to form an individual understanding of what this space represents. Reflection, as sparked by the AI, was hereby again the key-element. '*I think it's super cool to start with the binary because it kind of obliges people to question their position. [...] So I feel that it is a fantastic way to start because it obliges you to speculate and reflect and then you can add further layers of complexity to that.*'

It was noted how the understanding of the role of technology in human-machine interactions has fundamentally changed. By *'flipping ideas'* AI becomes a source of meaning, helping the designer to dismantle the binary and defamiliarize with collective imaginings.
One expert describes, how it is not the interest in the technology itself, '*[...] but the effect that the technology has on me, to think about creating multiplicities, creating more fluid ideas of myself.*'

'Well, you basically switch there. So instead of using them as sources of meaning, you use them as like. Starting point for reflection. So the assumption that you make about AI is different. So right now we kind of take it for granted that the AI can solve a lot of problems and then it fails miserably. Especially if you think of social categorization, not just about gender. So first of all, why do we need to know what is the gender of a person? Does it really contribute to anything? But then eventually what you're trying to do with this project is to try to use a I like to kind of assume that AI is biassed. Because it comes from us and we are biassed by default. And use it as a starting point. So you basically are flipping. Like you're just switching the roles.'

The toys themselves were seemingly perceived as ambiguous. The children were not able to immediately identify the toys as anything they know. Different children assigned different names to the toys. The *'robo-doll'* toy was described as *'baby'* or *'robot'*, the *'unicorn- dinosaur'* as *'unicorn'* and *'dragon'*, the *'drill-hairdryer'* used as a water pistol, however no words were used to describe it by the children. The teachers called it '*drill'* and *'hairdryer'*. Not only do these namings indicate that the toys indeed turned out to be more ambiguous than traditional toys, they furthermore stayed familiar to known elements.

'While trying to make sense of something, you are trying to fit it to your framework of understanding. It is like the process of when kids learn new knowledge. There is a process of assimilation and appropriation. So you have to fit it to you, and then you have to use it [...] So you make it your own.'

As discussed in an expert interview, the non-normativity of those toys sparks reflectivity through confrontation with Mary Douglas *'Monsters'* (Douglas, 2002): '*But they might see the differences in with other toys, right? It's more or less like when you have clothes like you have children that, I don't know, you have girls wearing pink clothes and boys wearing blue. And then if you all in a sudden you break this norm then eventually there is this difference that stands there and gets discussed, gets observed, attracts attention [...].'*

However, as highlighted by one of the experts, the perception of the toys by others can be seen as irrelevant, as long as the designer who designed them went through a reflective process. Nonetheless, does this first testing with children show that the designer is able to translate the interaction insights into seemingly less biassed designs.

Overall it was the combination of activities like sorting data and interpreting and reflecting on images and measures, that strengthened the internal reflection the designer had to undergo. Strength was seen in the AI as a facilitator of that process, rather than an autonomous agent who can suggest non-biased solutions.

On a metalevel the project and discussions were perceived as thought-provoking and valuable, beyond the practices itself.

'Very thought provoking, like I was saying, almost talking about this approach is valuable in itself as a sort of meta level of reflexivity.'

09.2.2. COST/LIMITATIONS - THE DILEMMA OF ECOLOGICAL/ECONOMICAL COST VS. SOCIAL GAIN

Despite an overall positive perception of the project's productivity, a few potential limitations and costs were elaborated.
Mostly discussed was the ecological cost of training large amounts of data in a long energy consuming process. Additionally, the training process often happens hidden for the designer on an external computer via a server. The only feedback for the designer, that could potentially signify the amount of energy being used, is the long training time required. However, as mentioned by one of the experts, also traditional design processes come with a cost: *'[...]but there is also a physical cost. [...] Materials have to be extracted and processed and [...] so there is nothing that's cost free. Apart from your own ideas and imagination.'*

This training time however, comes back as an economic cost. While the interactions with the AI are seen as productive in terms of identifying bias in the design process by making the designer reflect, the design process does not become more efficient in terms of time. For a manufacturing industry, like for instance the toy industry, time is still a driving factor for making money.

Two experts furthermore referred to a lack of agency. Being assisted by an AI might in turn lead to the degeneration of the *'design muscle'*, meaning that over time the designer will lose skill in for instance form giving. *'I think in the end the designer might lose the ability to form and like form give actually. There would be a reduction in the skill set of a designer or reliance on the designer to see forms made for them. So like a lessening of your imagination potentially.'*

Additionally, the current structures of power were seen as a limiting factor for a change in collective imaginings, despite the productivity of the reflexive interactions. Usually the people marginalised and discriminated against are not the ones in charge of making decisions, for instance on what toys to sell or what process to change. However, if not in charge of power, the voices of those most in need might be overheard for long, making the transition from one value system to another a really slow paced and long-term process. *'So I would say the benefit is that it's always beneficial to start discussing practices that might be harmful for the people, for the people involved and the cost is that you're working against the system. It is like a given. People don't even realise that. So it's kind of like asking people to wake up all in a sudden from what they have been thought to be the world. And that's difficult. So I*

think the cost outweighs the gain at the moment. But over time then eventually things will balance off.'

An expert also raised concern about an overtrust in AI and how this could potentially lead to the creation of new bias: *'[...] Where you're more biassed to trust the decision of AI over some other considerations. I don't know whether it would lead to new kinds of biases. But again, I guess that's the sort of strength of having this be a reflexive process where. Uh, perhaps it's not so much about the final outcome, but about being critical and asking these kinds of questions during the process.'*

Lastly, it was highlighted that the outcome of this project is not gender neutral toys, but a more reflective design practitioner, who is able to surface and challenge collective imaginings. Whilst the designer thus perceives the created toys as *'neutral'* or *'fluid'* and minimises the amount of gender cues embedded in the toy, the category of the *'neutral'* remains empty for the end-user. Having no concept of what *'the in-between'* represents, the user is praun to project their own bias to fill that empty category. As stated by Søndergaard & Hansen (2018), biases *'are part of any design, sometimes embedded in the design from the beginning, other times as something that happens over time through use'*.

It was discussed how packaging can play a potential role in bringing a trigger for reflection to the user. Packaging would target the parents, who are usually the ones buying the toys for their children. The toy is hereby positioned as the *'in-between'* stereotypical toys (Figure 82). It is highlighted that these stereotypes in toys can affect the identity development of the children. Little name tags for the toys suggest gender neutral names, which in turn could also affect the child's perception of the toy.

Signs of this user-bias could also be observed in the testing with children. The toys could not be identified as a specific object. After a moment of confusion children started calling toys *'unicorns'* or *'dinosaurs'*. However no consensus or agreement on what exactly those toys are could be found. Each child just interpreted the toy as something that related most to their own experiences and interests. Furthermore some peer pressure was observed, where one child declared the *'unicorn-dinosaur'* as a toy for girls. While other children disagreed in the beginning, they eventually ended up calling a girl toy as well.

Figure 82 - packaging of toys

09.2.3. RECOMMENDATIONS/ OUTLOOK - ONE STEP FURTHER, QUESTIONING THE 'IN-BETWEEN'

Discussing how the reflexive interactions and the challenging of collective imaginings could become even more effective, a list of recommendations as synthesised from the interviews, was made.

In terms of ecological cost reduction, the idea of sharing and reusing was discussed. Designers could for instance have access to a shared database, where models have already been pretrained on more context related data. However, additional data collection and training will remain necessary, as it is a crucial part of the reflexive journey. Nonetheless, could more sufficient pre-training and model sharing decrease the amount of training needed.

One expert suggested adding a third layer of reflection. While it is seen as necessary to start breaking with collective imaginings beginning with the normative, the search for the *'non-normative'* can be made more productive by encouraging the designer to question what will be found in the *'in-between'*. Is it gender fluidity? Gender neutrality? Designers will have to individually define this space in order to really reframe the gender categories for themselves.

In terms of terminology and framing of the project, the terms *'collaboration'* and '*reflection'* were discussed. It was argued that a collaboration would be more iterative, a more frequent back and forth between the designer and the AI, where the designer receives input and adjusts. The oxford dictionary (Oxford Languages and Google - English | Oxford Languages, 2022b) defines *'collaboration'* as *'the action of working with someone to produce something'*. Concluding it can be said the term *'collaboration',* especially in the context of human-machine interaction should be chosen carefully and accompanied with further description in order to ensure a common understanding.

Similarly, the term *'reflection'* was discussed. Experts raised the concern that simply calling it *'reflection'* might not be clear enough about the site of transformation. Adjusting it to '*self-reflection'* could in turn highlight that the interactions with AI are meant to challenge the self and make more reflective design practitioners. As such the notion of a self-reflexive exercise would state clearly that the focus is not on the technology, or the design products, but the effect that the technology has on oneself.

DISCUSSION & CONCLUSION

The following chapter reflects and expands on the insights generated through the design experiments, as well as the evaluation with experts and children. Zooming in hereby expands on the value and limitations of those interactions in more detail, while zooming out reflects more broadly on the implications and potential consequences of such human-machine interactions. The chapter concludes with a refined definition of 'reflexive interactions'.

10

DISCUSSION & CONCLUSION

10.1. ZOOMING IN – REFLECTING ON COST-GAIN RELATIONS

As discussed in expert interviews, reflexive interactions through AI come with ecological, economical and a form of social cost while showing potential in mitigating personal and collective bias. Concluding we can say that the interaction with the AI makes the design process more effective in terms of bias mitigation, however it also increases the complexity of it. This might result in the need for special design education.

Higher complexity is also found in the flow of interactions between humans and AI. While used to immediate response, for instance when discussing ideas with a design colleague, the AI requires hour-long training, making natural ways of interacting impossible.

It was also discussed whether the interactions with the AI should be used more situated or as an outside training activity. The generalisation of personal bias may not be useful because it is context specific. While special training sessions outside of the routinized design work would minimise the ecological cost, the designer would lack many situated nuances and insights. Let's take the colour red for example. By experimenting with the classification algorithm in the context of toy design, I was able to make explicit that I associate the colour red with femininity. However, in the context of for instance car design, I associate the colour red more with masculinity, speed and danger. While this example is rather concrete, it is to be expected that many more of those situated insights and associations would be oversimplified and generalised to an extent where they can become useless.

Furthermore, those insights should be personal and not shared. While personal insights can effectively be discussed in groups, as shown in the thingscon workshop, it is crucial for every designer participating in such discussions to undergo the process of reflection him/herself. Staying with the example of the colour red, a designer coming for instance from China, with a communist background, might naturally have a strong association of the colour red with luck, joy and happiness.

Additionally it has to be stated, that the algorithms that were used in the final experiments, are primarily focused on visual information. While alternative algorithms like text-to-image or language models, as explored in the early explorations phase, shown to be usable for reflexive interactions as well, the selection of algorithms used to design the toys is unable to capture bias that is not encoded in visual information. However, the experiments have also shown that designers are able to associate a lot of cultural information to visual elements. Furthermore, de-

signers are known to prefer visual inspiration (Gonçalves et al., 2014), allowing for a more natural design process when using for instance StyleGANs.

A problem could be an *'over-trusting'* in this technology, a phenomenon often observed in partly autonomous cars. Users misinterpret the skill and abilities of the AI and give the technology a level of agency and responsibility it is not made for (Drexler et al., 2018). In accounting for such 'over-trusting' it is crucial to inform the user about the skill and abilities of the technology before use. It is furthermore important to highlight that what the AI is generating, is always just a mirror of one's judgements rather than the 'truth' or anything that can be taken as fact.

Addressing further limitations of reflexive interactions, it is thought of as a more participatory design approach, where the user undergoes the process of reflexively creating objects him/herself. The productivity of reflexive interactions lies in the process of dismantling one's own concept of the world. So essentially, the artefacts created reflect 'neutrality' more for the creator than for those who did not engage in the reflections themselves. As discussed in expert interviews, the user might end up projecting own bias into an otherwise neutral object. Overcoming this projection, a solution could be to have the user, in this case the child, design their own toys. This way, the user would go through the process of surfacing and challenging bias himself.

One of the biggest challenges for reflexive interactions poses the ecological cost associated with the training of the algorithms. However, it can be expected to minimise this cost in the future. While reuse was already discussed in expert interviews, additional adaptations can be made. It has to be highlighted that the algorithms used for these experiments are not designed for designers to use for reflexive purposes. As a result, pre-training the algorithms for instance is not optimised for design content. Training data and images could be chosen that are more relevant for design, thus potentially lowering the amount of additional training time needed.

Additionally, most of the algorithms are designed to perform as precisely and consistently as possible. A classification algorithm for instance is expected to classify images accurately. However, such accuracy is not necessarily needed for sparking situated reflection. As the experiments showed, a styleGAN for instance, produces simply more ambiguous images if trained on less data or for less time. While ambiguity and confusion can be generally seen as something negative, they also carry the potential to trigger reflection by interrupting the designers routinated workflow. As such, it can be speculated that future AI algorithms for reflexive interactions are less cost intensive, because less precision and accuracy, but rather confusion, ambiguity and confrontation is needed. As an additional benefit, such shorter training times would also lower the economical cost. More ambiguous AI generations would also require more design skill, to be translated into actual objects, preventing the 'design muscle' from degenerating.

However, more ambiguous and less precise algorithms might also lead to the projection of further bias into the interpretation of the AI's generations. Hence, future work would have to investigate in what ways AI algorithms could be optimised for reflexive interactions in design contexts.

10.2. ZOOMING OUT– RE-EVALUATING HUMAN-MACHINE RELATIONS

Zooming out, and discussing reflexive interaction with AI in a more holistic way, this sub-chapter looks at potential ethical concerns, AI hype, structures of power and bias perpetuations, as well as potential changes in the role of the designer and the AI.

Looking into ethical concerns regarding the use of AI, most widely discussed is the danger of AI and machine learning reproducing past discrimination (Rainie et al., 2021). However, as discussed and explored throughout this thesis, such exacerbation of human bias can be used as surprise, confrontation and sensitization in order to empower humans to surface and challenge personal and collective bias. Part of the reflexive exercises is especially the conscious encoding of bias in the AI's training data.

Furthermore, the worries like the one expressed in the AI index report by the Stanford Institute for Human-Centred Artificial Intelligence (AI Index 2021, n.d.), is that the main developers and deployers of AI are focused on profit-seeking and social control, and there is no consensus about what ethical AI would look like. What decisions do we want to delegate to the AI? Should systems be designed to either avoid actions that have a significant negative influence on human agency, enabling individuals to make their own decisions, or to step in when it is obvious that human decision-making may be harmful? (Rainie et al., 2021),(Coeckelbergh, 2020)
With the decision for more reflexive instead of normative interactions with AI, a clear decision for more human and less technological agency was made. While this work indicates how such human-machine interaction can be useful to encourage designers to become more critical and reflective design practitioners, the worry of *'profit-seeking'* deployers could still hinder the productivity or even employment of such practices. However, this work shows how understanding bias as a human flaw instead of a technological one, and how reflection and the search for self-understanding in the interaction with the AI can reduce the amount of ethical concern.

Since there is little consensus about what an ethical AI should like, some follow a comparative strategy: *'It's not whether AI systems alone produce questionable ethical outcomes, it's whether the AI systems are less biassed than the current human systems and their known biases'* (Rainie et al., 2021). Following similar ethics, this thesis aimed to challenge existing practices in design and human-machine interaction by exploring alternatives in the ways designers can target bias. However, whether the proposed reflexive interactions are indeed less biassed than current '*normative*' practices is to be proved in future studies.

Amershi et al. (2019) are discussing potential negative outcomes of human-AI interactions. Concern is raised that *'automated inferences are typically performed under uncertainty, often producing false positives and false negatives, AI-infused systems may demonstrate unpredictable behaviours that can be disruptive, confusing, offensive, and even dangerous.' (*Amershi et al., 2019)

While such behaviour can often be seen as negative or even dangerous, this work has shown that confusion, surprise and confrontation are also the sources of reflection, critical thought and re-framing. Additionally AI is called to be *'inherently inconsistent'* (Amershi et al., 2019). Another quality that can be seen as potentially unwanted, however when used on purpose, it can also serve as means to trigger reflection.

Nonetheless, inconsistency and unpredictable behaviour can also erode the user's confidence or, depending on the context, become dangerous. As stated by Cavalcante et al. (2022), one of the two key conditions for meaningful human control in human-AI interaction is tracing: 'in order for a human-AI system to be under meaningful human control, its behavior, capabilities, and possible effects in the world should be traceable to a proper moral and technical understanding on the part of at least one relevant human agent who designs or interacts with the system.' However, since the AI is giving little to no autonomy in the here explored reflexive interactions, which are expected to be performed in addition to traditional design methods that ensure safety, tracing conditions are less relevant.

Another concern that has been raised, is looking into the perpetuation of AI and technology stereotypes through generative models like Dall-e 2 (DALL·E 2, 2022) and Midjourney (Midjourney, n.d.)(Kapoor & Narayanan, 2022). Images used in for instance news articles act like a visual metaphor and not always reflect its content. People reading those articles and seeing those images might be misled. Often those images suggest that AI for instance, has a high level of autonomy and agency, by depicting the term '*AI*' in most contexts as humanoid robots. Being trained on those stock-visualisations of AI, it is not surprising that generative Models – that are being used to generate stock images for news articles– perpetuate or even exacerbate these stereotypical depictions of technology. (Kapoor & Narayanan, 2022) Concluding it can be said, the consciousness of not just the perpetuation of human stereotypes, but also technology stereotypes is needed, in order to establish a meaningful human-machine interaction. However, the means of reflection and conscious data curation, as inherent to the proposed reflexive inter-

actions, are also expected to help the designer recognize and counteract those technology stereotypes.

Reflecting on the overall change in interaction dynamics, a switch in roles is recognized. As experts described it, the assumption about technology is no longer that it solves the problems for you. The focus has moved to the relations between the human and the technology, making it more about the effect of technology on oneself than the technology as such. This changes the way we perceive ourselves and the role of technology. Technology is now more defined in relation to oneself. As described above, AI has given away agency. This can also be seen as a counter proposal to the above mentioned humanoid AI, that can act fully autonomous.

Discussing the changes in the role of the designer is more difficult. As reflection is already a core part of how designers work (Schön, 1984), the interactions with AI are more expected to enhance natural design tendencies than introduce new ones. However, the engagement with training data and abstract algorithms is not standard in every design practice. Changes in the way technology will be incorporated in the design thinking process might change. That the adoption of technologies like AI in the design process is not something of the far future, is proven daily by oio.studio, which incorporated an artificial intelligence as design colleague in their work. However, the posed assumptions would have to be further tested and explored.

As for the future application of such interactions, also more complex categorical discriminations and scenarios are envisioned. While the example of gender representation in toys was an easy to grasp and understand scenario, more complex contexts are expected to also change the cost-gain relations. Imagining rather simple objects like toys in-between gender categories might be a task a trained designer could perform without training an algorithm for several hours. However, envisioning the in-between for instance of different races or ethnicities in contexts like advertisement, hiring, etc. might pose a challenge for designers that can currently not be solved. Applying reflexive interactions in such scenarios might legitimise the high ecological cost.

10.3. CONCLUSION – THE REFINED DEFINITION OF REFLEXIVE INTERACTIONS

To surface and challenge potentially harmful, yet internalised and forgotten norms, this thesis explores and investigates reflexive relations between designer and AI and the effect of unpredictable, surprising and confronting AI behaviour on oneself. A speculative company is created, to provide an immersive context for the research and design of the reflexive human-AI interactions. Within this speculative scope, the problem of gender stereotypes in toys is explored. Based on early explorations - specifically interviews, workshops, generative sessions and testing of different AI models - as well as three introspective research through design experiments, this project concludes that reflexive human-machine interactions – as proposed alternative to normative human-machine interactions – are potentially productive to surface, dismantle and de-familiarize personal bias and collective imaginings.

Answering the questions of '*how to design while being mindful of biases', 'how to integrate such reflections in the design process'* and *'how to spark such reflections through AI'* a set of early technological explorations, literature research, interviews and workshops is conducted. The findings are translated into four design tactics *1)auto-confrontation* - presenting oneself with the traces of own doing -*; 2)shift in perspective* -look at yourself through the eyes of someone else-*; 3)clash of expectations* -break with existing conventions and expectations or create new expectations, that are then destroyed on purpose- and *4) creating monsters* -creation of two excluding categories that can not be put back in place-. These four tactics serve as interaction guidelines and describe different means of triggering and sparking reflection.

In addition, a speculative design-case is set up in the form of a fictional company *'The first gender-fluid child toy company'*. Illustrating the problematic representation of gender stereotypes in toys, this company is aiming to break with the binary ideas of masculinity and femininity as materialised in toys. Presenting itself with problem description, mission and products in form of a website, this company provides a context in which reflexive interactions are explored and communicated. The company serves as a tool for presenting future AI-design practices in a tangible and immersive way, but also as a critique of current gender stereotypes in the design of children's toys.

Situated in the process of gender-fluid child toy design, three introspective design experiments are conducted, in order to explore new relations between human and technology. As a result of these interactions, three toys - each challenging stereotypical and binary ideas of femininity and masculinity as materialised in toys like drills and unicorns - are designed.

The first experiment -the robot-doll experiment- uses a classification algorithm which the designers train on their personal perception of gender by labelling data in '*masculine'* and *'feminine'*. The created measure of one's perception serves as *'auto-confrontation', 'shift in perspective'* and *'clash of expectation'* which in turn forces the designer to reflect on their own tendencies to implement bias in forms or shape, colour, etc..

The second experiment -the dinosaur-unicorn experiment- uses a StyleGAN, trained on images of dinosaur and unicorn plushes. Mapping creatures between dinosaur and unicorn on an infinite grid, the designer can explore the spaces between unicorn and dinosaur and experience the absurdity of binary categorisation. This experience can *'shift perspectives', 'create monsters'* and *'let expectations clash'*.

The last experiment -the dirll-hairdryer experiment- uses a StyleGAN to create meta-objects, that show similarity with the two binary categories of *'drill'* and *'hair-dryer'* while mixing their properties to an extent that causes confusion and irritation about the objects function and context of use. This is achieved by creating and training the algorithm first on a primary dataset of drills, and then only shortly on a secondary dataset of hair-dryers. The tactics that play an important role are '*shift in perspective', 'creating monsters'* and *'clash expectations'*.

As the experiments with three different algorithms suggest, AI can be utilised to help the designer in getting internal insights into personal bias materialisation in design and furthermore assist in exploring design ideas outside of the norm. Not only did the interactions with the AI lead to a more reflective design practice, the outcomes of these design processes furthermore spark reflection through confronting and irritating artefacts, as testing with children suggests.

This thesis also shows that AI's often negatively described behaviour like confusion, bias exacerbation, confrontation and inconsistency, also carries the power to trigger reflective practices that help surfacing and challenge bias.

Compared to normative human-machine practices, the three experiments with AI showed that the role of both the human and the non-human have changed. The technology, instead of searching for solutions, becomes a means to surprise, confront, highlight and explore the 'in-between' and non-normative, helping the designer to be more reflective. The designer in turn becomes responsible for using the reflective practices to surface personal and collective bias and dismantle the binary and normative.

While reflexivity is nothing new in human-machine interactions, error is no longer something that is localised on the technology side. When expectations clash -for instance when the generated outcome is not matching the intended outcome- it is no longer the system that will be fixed, but the human, that is then forced to question own biases, expectations as well as collective imaginings. As such an alternative view on AI ethics and human-machine interaction is illustrated.

However, the productivity of those reflexive interactions depends on the designers ability to reflect and adapt to new tools like classification algorithms and StyleGANs. It is furthermore crucial for success, that the self-reflexive exercises are being situated and adapted to the different design contexts, rather than applied as separate general training exercises, in order to avoid the over simplification or generalisation of biases like colour for instance, that highly depend on the context.

Ecological problems occur due to long algorithm training hours, which are often hidden from the designer by training the algorithms in clouds. Additionally, an over-trust in the model's outcome, or the projection of own ideas into the generated stimuli might arise. New ways to reduce the ecological as well as the economical and social cost have to be found and tested. Future work has to investigate what potential new biases form through the reflexive interactions. Additionally, different design contexts and practices should be explored. More complex biases and contexts might hereby outweigh the ecological cost compared to the social benefit.

This thesis contributes to the field of human-machine interaction. Seeing potential in confusing and bias exacerbating behaviour of the AI, this thesis illustrates alternative alignments in design between humans and non-humans through three situated design experiments. Secondly, this thesis takes a speculative research through design approach, in order to also involve people outside of academia in discussions about future design and AI practises.

Furthermore, this work explores new design practices, expanding the tool kit of the designer with data and algorithms and giving new ideas for human-machine collaboration in the design process.

Finally this thesis concludes with a refined definition of *'reflexive designer-AI interactions'* (Figure 83). Future research and testing is required to validate the potential of such human-machine relations.

> **Reflexive interactions are a form of situated self-reflection exercises, where the AI can help the designers in their process of surfacing, sensitising as well as de-familiarizing and dismantling the personal and collective norms, roles, values through the means of surprise, confrontation, mirroring and exacerbation.**

Figure 83 - refined definition of 'reflexive designer-AI interactions'

10.4. PERSONAL REFLECTION

As mentioned in the very beginning of this report, I had set myself certain ambitions for this project. Reflecting on those ambitions, as well as my overall project journey, I am very happy with this thesis.

Due to my interest in complex topics and a wide variety of disciplines, the project ran the risk of getting out of hand. I am most proud, that all these years of studying have eventually taught me how to follow my interests, while researching and designing in an structured and systematic way. I am very grateful for the input, critique and support of my supervisors, that guided me through this complex project and helped me to not get lost in it.

By looking at myself and my design process through the lens of technology I have learned a great deal about myself as a designer. By setting my own discipline in relation to other disciplines like philosophy, psychology and computer science, I furthermore gained a deeper understanding about design as such.

While I was not able to acquire new skills and knowledge in all domains that interest me, I was nonetheless able to touch upon many, if not all, of those areas through interesting literature, discussions and experiments.

I expanded my repertoire of design research methods and frameworks, by creating artefacts, leading workshops, give interviews and present and discuss my work.

I am very grateful for all the interesting conversations, people and events, that helped me in gaining a more realistic and holistic understanding of my work.

Looking back one of the greatest challenges for me was writing this report. While feeling a bit lost in times, I am proud about my learning progress and the results I achieved.

Engaging with scientific literature and methodologies was not always easy and I needed some time to adjust to the new learning environment. However, I could not be more surprised about how much I have learned since my last graduation project in the Bachelor and I am quite proud to be graduating with a Master of Science now from the Tu Delft.

jump to References and Appendix

REFERENCES

A

Adams W. J. Graf E. W. Ernst M. O. (2004). Experience can change the "light-from-above" prior. Nature Neuroscience, 7 (10), 1057–1058.

Adams, T. E., Ellis, C., & Jones, S. H. (2017). Autoethnography. In J. Matthes,

Agnew, K. (1993) The Spitfire: Legend or History? An argument for a new research culture in design (Journal of Design History, 6, 2, pp.121-130).

AI Index 2021. (n.d.). Stanford Institute for Human-Centered Artificial Intelligence. https://hai.stanford.edu/ai-index-2021

AI Toys. (n.d.). Ani Liu. Retrieved September 17, 2022, from https://ani-liu.com/ai-toys

Almeida, D. B. L. d. (2017). On diversity, representation and inclusion: New perspectives on the discourse of toy campaigns. Linguagem em (Dis) curso, 17(2), 257–270.

Alter, S. (2010). Designing and Engineering for Emergence: A Challenge for HCI Designing and Engineering for Emergence: A Challenge for HCI Practice and Research Practice and Research. In AIS Transactions on Human-Computer Interaction (Vol. 2). https://aisel.aisnet.org/

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. Conference on Human Factors in Computing Systems - Proceedings. https://doi.org/10.1145/3290605.3300233
Artificial Intelligence. (n.d.). In Oxford Languages.

Arzberger, A., Van der Burg, V., Chandrasegaran, S., Lloyd, P. (2022). Triggered using human-AI dialogue for problem understanding in collaborative design. Proceedings of the ASME 2022 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference

Auger, J. (2013, March). Speculative design: crafting the speculation. Digital Creativity, 24(1), 11–35. https://doi.org/10.1080/14626268.2013.767276

B

Beauvoir, D. S., Borde, C., & Malovany-Chevallier, S. (2011, May 3). The Second Sex (1st ed.). Vintage.

Bettany, S., & Woodruffe-Burton, H. (2009). Working the limits of method: The possibilities of critical reflexive practice in marketing and consumer research. Journal of Marketing Management, 25(7–8), 661–679. https://doi.org/10.1362/026725709X471550

Billett, S. (2010), "Subjectivity, self and personal agency in learning through and for work", in Malloch, M., Cairns, L., et al. (Eds), The Sage Handbook of Workplace Learning, Sage Publications, London, pp. 60-72.

Bleeke, J. (2009) Design Fiction: A Short Essay on Design, Science, Fact and Fiction. Near Future Laboratory, 2009.

Blythe, M. (2014). Research through design fiction: Narrative in real and imaginary abstracts. Conference on Human Factors in Computing Systems - Proceedings, 703–712. https://doi.org/10.1145/2556288.2557098

Blythe, M., Andersen, K., Clarke, R., & Wright, P. (2016). Anti-Solutionist Strategies: Seriously Silly Design Fiction. In

Braak, L. (2021, December). Introduction to Probabilistic Classification: A Machine Learning Perspective. Towards Data Science. https://towardsdatascience.com/introduction-to-probabilistic-classification-a-machine-learning-perspective-b4776b469453

Bunn, C. (2022, March 4). Report: Black people are still killed by police at a higher rate than other groups. NBC News. https://www.nbcnews.com/news/nbcblk/report-black-people-are-still-killed-police-higher-rate-groups-rcna17169

Butler, J (1986). Sex and Gender in Simone de Beauvoir's Second Sex. In: Yale French Studies No. 72, Simone
de Beauvoir: Witness to a Century, S. 35–49

Brownlee, J. (2019). A Gentle Introduction to StyleGAN the Style Generative Adversarial Network. Machine Learning Mastery. Retrieved October 4, 2022, from https://machinelearningmastery.com/introduction-to-style-generative-adversarial-network-stylegan/

Boon, B., Baha, E., Singh, A., Wegener, F. E., Rozendaal, M. C., & Stappers, P. J. (2020, September 10). Grappling with Diversity in Research Through Design. https://doi.org/10.21606/drs.2020.362

Bozzon, A. (2021) Lecture about Entity-relationship model, data course
Brockbank, A. and McGill, I. (1998), Facilitating Reflective Learning in Higher Education, Society for Research in Higher Education and Open University Press, Buckingham.

Brynjolfsson, E., and A. McAfee (2014), The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies, W. W. Norton & Company.

Burnap, A., Liu, Y., Pan, Y., Lee, H., Gonzalez, R., & Papalam bros, P. Y. (2016). Estimating and exploring the product form design space using deep generative models. In Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (Vol. 50107, p. V02AT03A013).

C

C. S. Davis, & R. F. Potter (Eds.), The international encyclopedia of communication research methods. Wiley-Blackwell.

Canlı, E. (2018). Queering Design: Material Re-Configurations of Body Politics.

Cavalcante Siebert, L., Lupetti, M. L., Aizenberg, E., Beckers, N.,

Zgonnikov, A., Veluwenkamp, H., Abbink, D., Giaccardi, E., Houben, G.-J., Jonker, C. M., van den Hoven, J., Forster, D., & Lagendijk, R. L. (2022). Meaningful human control: actionable properties for AI system development. AI and Ethics. https://doi.org/10.1007/s43681-022-00167-3

Classification Algorithm in Machine Learning - Javatpoint. (n.d.). www.javatpoint.com. Retrieved October 4, 2022, from https://www.javatpoint.com/classification-algorithm-in-machine-learning#:%7E:text=The%20Classification%20algorithm%20is%20a,number%20of%20classes%20or%20groups.

Coeckelbergh, M. (2020). AI Ethics (The MIT Press Essential Knowledge series). The MIT Press.

Confirmation Bias in UX. (n.d.). Nielsen Norman Group. Retrieved October 2, 2022, from https://www.nngroup.com/articles/confirmation-bias-ux/

Cortese, A. J. (2015, October 2). Provocateur: Images of Women and Minorities in Advertising (Fourth). Rowman & Littlefield Publishers.

Coulton, P. and Lindley, J., G. (2019). More-Than Human Centred Design: Considering Other Things. The Design Journal 22, 4: 463--481.

Cross, N. (2006). Designerly Ways of Knowing.

D

Dahl, J., Y. and Sætnan, A., R. (2009). „It all happened so slowly": On controlling function creep in forensic DNA databases. International Journal of Law, Crime and Justice, 37(3), 83-103.

DALL·E 2. (2022, April 14). OpenAI. Retrieved October 1, 2022, from https://openai.com/dall-e-2/
DALL·E mini by craiyon.com on Hugging Face. (n.d.). Retrieved October 1, 2022, from https://huggingface.co/spaces/dalle-mini/dalle-mini

De Cremer, D., & Kasparov, G. (2022). The ethics of technology innovation: a double-edged sword? AI And Ethics, 2(3), 533–537. https://doi.org/10.1007/s43681-021-00103-x

Deep Dream Generator. (n.d.). Retrieved October 11, 2022, from https://deepdreamgenerator.com

Designs.ai - Creative work done effortlessly. (n.d.). Retrieved September 20, 2022, from https://designs.ai/com

Desjardins, J. (2021, August 26). Every Single Cognitive Bias in One Infographic. Visual Capitalist. Retrieved October 13, 2022, from https://www.visualcapitalist.com/every-single-cognitive-bias/

Diversity, Equity and Inclusion for Embodied AI. (n.d.). DEI4EAI. Retrieved October 2, 2022, from https://www.dei4eai.com/about

Douglas, M. (2002). Purity and Danger: An Analysis of Concepts of Pollution and Taboo (Routledge Classics). Routledge.

Dourish, P., and Bell, G., (2009) Resistance is Futile: Reading Science Fiction Alongside Ubiquitous Computing. http://www.dourish.com/publications/2009/scifi-pucdraft.pdf

Dretske, F. (1994). Introspection. Proceedings of the Aristotelian Society, 94, 263–278. http://www.jstor.org/stable/4545198

Drexler, D. A., Takacs, A., Nagy, T. D., & Haidegger, T. (2018). Handover Process of Autonomous Vehicles – Technology and Application Challenges. Acta Polytechnica Hungarica, 16(9). http://acta.uni-obuda.hu/Drexler_Takacs_Nagy_Haidegger_96.pdf

Duncombe, S. (2007) Dream: Re-imaging Progressive Politics in an Age of Fantasy (New York: The New Press), 182.

Dunne, A., & Raby, F. (2013). Speculative Everything - Design, Fiction, and Social Dreaming. The MIT Press.

D'Ignazio, C., & Klein, L. F. (2020). Data Feminism (Strong Ideas). The MIT Press.

E

Eckert, P. (2014). The Problem with Binaries: Coding for Gender and Sexuality. Language and Linguistics Compass, 8(11), 529–535. https://doi.org/10.1111/lnc3.12113

Edelson, D. C. (2002). Design research: What we learn when we engage in design. In Journal of the Learning Sciences (Vol. 11, Issue 1, pp. 105–121). Lawrence Erlbaum Associates Inc. https://doi.org/10.1207/S15327809JLS1101_4

Enninga, H. (2022, March 21). Are New Technologies Keeping Us Stuck in Old Biases? Newsroom | University of St. Thomas. https://news.stthomas.edu/are-new-technologies-keeping-us-stuck-in-old-biases/

Epstein R. M. (1999) Mindful Practice. JAMA. ;282(9):833–839. doi:10.1001/jama.282.9.833

F

Feine, J., Gnewuch, U., Morana, S., Maedche, A. (2020). Gender Bias in Chatbot Design. In: , et al. Chatbot Research and Design. CONVERSATIONS 2019. Lecture Notes in Computer Science(), vol 11970. Springer, Cham. https://doi.org/10.1007/978-3-030-39540-7_6

Fioravanti, M., Rebaudengo, S. (2022) workshop with oio.studio

Fedor Indutny . (n.d.). Slate's Use of Your Data. Slate Magazine. Retrieved October 2, 2022, from https://slate.com/gdpr?redirect_uri=%2Fblogs%2Ffuture_tense%2F2012%2F03%2F02%2Fbruce_sterling_on_design_fictions_.html%3Fvia%3Dgdpr-consent&redirect_host=http%3A%2F%2Fwww.slate.com

Fulton Suri, J. (2003). The experience of evolution: Developments in design practice. The Design Journal, 6(2), 39e48.

G

Gaver, B., & Bowers, J. (2012). Annotated portfolios. interactions, 19(4), 40–49.

Gaver, W. (2012, May). What should we expect from research through design?. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 937–946). ACM.

Gerber, A., Derckx, P., Döppner, D., & Schoder, D. (2020). Conceptualization of the human-machine symbiosis -A literature review. Proceedings of the 53rd Hawaii International Conference on System Sciences.

Ghajargar, M., Wiberg, M., & Stolterman, E. (2018). Designing IoT systems that support reflective thinking: A relational approach. International Journal of Design, 12(1), 21-35.

Giaccardi, E. (2020). Casting things as partners in design: Towards a more-than- human design practice. In H. Wiltse, ed., Relating to Things: Design, Technology and the Artificial. Bloomsbury.

Goldschmidt, G., (2015). "Ubiquitous serendipity: Potential visual design stimuli are everywhere". In Studying visual and spatial reasoning for design creativity. Springer, pp. 205–214.

Gonc¸alves, M., Cardoso, C., and Badke-Schaub, P., (2012). "Find your inspiration: exploring different levels of abstraction in textual stimuli". In Proceedings of the 2nd International Conference on Design Creativity (Volume 1), pp. 189–198.

Gonc¸alves, M., Cardoso, C., and Badke-Schaub, P., (2014). "What inspires designers? preferences on inspirational approaches during idea generation". Design Studies, 35(1), pp. 29–53.

Green, Jamison. (2004). Becoming a Visible Man. Nashville: Vanderbilt University Press.

Greggor, L., and Hackett, P. M. W. (2017). "Categorization by the animal mind," in Mereologies, Ontologies and Facets: The Categorial Structure of Reality, ed P. M. W. Hackett (Lanham, MD: Lexington Press).

H

Hackett, P. M. W., ed. (2019). Conceptual Categories and the Structure of Reality: Theoretical and Empirical Approaches. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-731-1
Hagendorff, T. (2019). The Ethics of AI Ethics: An Evaluation of Guidelines. arXiv:1903.03425v2

Haraway, D. (1991). A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century. Simians, Cyborgs and Women: The Reinvention of Nature, 149–181.

Haraway, D. (2016) Staying with the Trouble: Making Kin in the Chthulucene. Duke University Press. Durham and London, 2016.

Have I Been Trained? (n.d.). Retrieved October 10, 2022, from https://haveibeentrained.com/?search_text=Child%20toys%20girl
Helyer, R. (2015), „Learning through reflection: the critical role of reflection in work-based learning (WBL)", Journal of Work-Applied Management, Vol. 7 No. 1, pp. 15-27. https://doi.org/10.1108/JWAM-10-2015-003

How Dismantling the Gender Binary Can Help Eradicate Inequality. (2021, January 6). Voices of Youth. https://www.voicesofyouth.org/blog/how-dismantling-gender-binary-can-help-eradicate-inequality

Huang, T. (2022, January 10). Why Computer-Assisted Humans Are The Best Chess Players And What That Means For Technology Operations. Forbes. https://www.forbes.com/sites/forbestechcouncil/2022/01/07/why-computer-assisted-humans-are-the-best-chess-players-and-what-that-means-for-technology-operations/

Höök, K. and Löwgren (2012), J. Strong concepts. ACM Trans. on Computer-Human Interaction 19, 3 , 1–18. doi:10.1145/2362364.2362371

Höök, K., Bardzell, J., Bowen, S., Dalsgaard, P., Reeves, S., Waern, A. (2015) Framing IxD Knowledge , XXII.6 November–December 2015, p. 32.

I

Iraq: Impunity for Violence Against LGBT People. (2022, September 7). Human Rights Watch. https://www.hrw.org/news/2022/03/23/iraq-impunity-violence-against-lgbt-people

J

Jacques Lacan (Stanford Encyclopedia of Philosophy). (2018, July 10). Retrieved October 1, 2022, from https://plato.stanford.edu/entries/lacan/#MirStaEgoSub

Jansson, D. G., & Smith, S. M. (1991). Design fixation.

K

Kazi, R. H., Grossman, T., Cheong, H., Hashemi, A., & Fitzmaurice, G. W. (2017). DreamSketch: Early Stage 3D Design Explorations with Sketching and Generative Design. In the ACM Symposium for User Interface Science and Technology (Vol. 14, pp. 401-414).

Keates, S., Clarkson, P. J., Harrison, L.-A., and Robinson, P. (2000). "Towards a practical inclusive design approach". In Proceedings on the ACM Conference on Universal Usability, pp. 45–52.

Keates, S., and Clarkson, J. (2003). "Countering design exclusion". In Inclusive Design. Springer, pp. 438–453.

Keyes, O. (2018) The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. Proc. ACM HUM.-Comput.Interact., 88, 2 (CSCW): 1-22.

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(6), 1534–1555. https://doi.org/10.1037/0278-7393.25.6.1534

Koch, J., Lucero, A., Hegemann, L., & Oulasvirta, A. (2019, May 2). May AI? Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3290605.3300863

Kuijer, L., & Giaccardi, E. (2018). Co-performance: Conceptualizing the role of artificial agency in the design of everyday life. Conference on Human Factors in Computing Systems - Proceedings, 2018-April. https://doi.org/10.1145/3173574.3173699

L

Lem, S. (1973) Imaginary Magnitude. A Harvester book. New York.

Licklider, J.C.R. (1960), "Man-Computer Symbiosis", IRE Transactions on Human Factors in Electronics HFE-1(1), pp. 4–11.

Lindley, J. and Coulton, P. (2016) Pushing the Limits of Design Fiction: The Case For Fictional Research Papers. In Proc. CHI 2016, ACM (2016),4032-4043

Lindley, J., Coulton, P. and Sturdee, M. (2017) Implications for Adoption. In Proc. CHI 2017, ACM (2017), 265-277

Löwgren, J. (2013, January). Annotated portfolios and other forms of intermediate-level knowledge. Interactions, 20(1), 30–34. https://doi.org/10.1145/2405716.2405725

M

Maher, M. L., & Fisher, D. H. (2012). Using AI to evaluate creative designs. In Proceedings of the 2nd International Conference on Design Creativity Volume 1.

Mann, M., and Matzner, T. (2019). "Challenging algorithmic profiling: The limits of data protection and antidiscrimination in responding to emergent discrimination". Big Data & Society, 6(2), p. 2053951719895805.

Maqsood, T., Finegan, A., & L. Armstrong, H. (2004). Biases and Heuristics in Judgment and Decision Making: The Dark Side of Tacit Knowledge. Issues in Informing Science and Information Technology, 1, 0295–0301. https://doi.org/10.28945/740

Mareis, C. (2012). The Epistemology of the Unspoken: On the Concept of Tacit Knowledge in Contemporary Design Research. Design Issues, 28(2), 61–71. http://www.jstor.org/stable/41427826

Markussen, T., and Knutz, E., (2013) The Poetics of Design Fiction. Proc of DPPI 2013. Newcastle

McQuillan, D. (2018). "People's councils for ethical machine learning". Social Media+ Society, 4(2), p. 2056305118768303.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A., (2021). "A survey on bias and fairness in machine learning". ACM Computing Surveys, 54(6), pp. 1–35.

Midjourney. (n.d.). Midjourney.com. Retrieved November 19, 2022, from https://www.midjourney.com

MoMA | Marcel Duchamp and the Readymade. (n.d.). MoMa. https://www.moma.org/learn/moma_learning/themes/dada/marcel-duchamp-and-the-readymade/

Moore Pervall, T. (2022, September 29). Unconscious Biases That Get In The Way Of Inclusive Design. Smashing Magazine. https://www.smashingmagazine.com/2022/09/unconscious-biases-inclusive-design/

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. IEEE Robotics and Automation Magazine, 19(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Mostaque, E. (2022b, September 2). Stable Diffusion Public Release. Stability.Ai. Retrieved October 1, 2022, from https://stability.ai/blog/stable-diffusion-public-release

Muñoz,J.E (2019). Cruising Utopia

N

Nicenboim, I., Giaccardi, E., Søndergaard, M. L. J., Reddy, A. V., Strengers, Y., Pierce, J., & Redström, J. (2020). More-than-human design and AI: In conversation with agents. DIS 2020 Companion - Companion Publication of the 2020 ACM Designing Interactive Systems Conference, 397–400. https://doi.org/10.1145/3393914.3395912

Nikolas Martelaro and Wendy Ju. (2018). Cybernetics and the design of the user experience of AI systems. interactions 25, 6 (November - December 2018), 38–41. https://doi.org/10.1145/3274570

O

O'Connor, R. (2022, July 20). How DALL-E 2 Actually Works. News, Tutorials, AI Research. Retrieved October 16, 2022, from https://www.assemblyai.com/blog/how-dall-e-2-actually-works/

Olkkonen, M., McCarthy, P. F., & Allred, S. R. (2014). The central tendency bias in color perception: Effects of internal and external noise. Journal of Vision, 14(11), 5. https://doi.org/10.1167/14.11.5

Oxford Languages and Google - English | Oxford Languages. (2022b, August 12). Retrieved October 2, 2022, from https://languages.oup.com/google-dictionary-en/

P

Paasi, A. (2000). Territorial identities as social constructs. Hag-Hagar –International Social Science Review, 1, 91–113.

Perez, C. C. (2021): Invisible Women: Data Bias in a World Designed for Men. Harry N. Abrams

Pfister, H., Jungermann, H., & Fischer, K. (2016). Die Psychologie der Entscheidung: Eine Einführung (German Edition) (4. Aufl. 2017 ed.). Springer.

Pieters, M. & Wiering, M. (2018). Comparing Generative Adversarial Network Techniques for Image Creation and Modification

Preciado, P. B. (2021). Can the Monster Speak: Report to an Academy of Psychoanalysts
Proc. CHI 2016, ACM (2016), 4968-4978.

Prochner, I., & Godin, D. (2022). Quality in research through design projects: Recommendations for evaluation and enhancement. Design Studies, 78. https://doi.org/10.1016/j.destud.2021.101061

R

Rainie, L., Anderson, J., & Vogels, E. A. (2021, June 21). Experts Doubt Ethical AI Design Will Be Broadly Adopted as the Norm Within the Next Decade. Pew Research Center: Internet, Science & Tech. https://www.pewresearch.org/internet/2021/06/16/experts-doubt-ethical-ai-design-will-be-broadly-adopted-as-the-norm-within-the-next-decade/

Raj, Amifa, & Ekstrand, Michael D. (2022). Fire Dragon and Unicorn Princess; Gender Stereotypes and Children's Products in Search Engine Responses. Cornell University - ArXiv. http://arxiv.org/abs/2206.13747

Ravanera, C. (2020, February 11). Working beyond the gender binary. Gender and the Economy. https://www.gendereconomy.org/working-beyond-the-gender-binary/

Rittel, W. , J. , H., & Webber, M. , M. (1973). Dilemmas in General Theory of Planning. Policy Sciences, 4, 155–169.

Rosson, M. Beth. (2017), ACM Digital Library., & ACM Special Interest Group on Computer-Human Interaction. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM.

Runway. (n.d.). Retrieved October 4, 2022, from https://app.runwayml.com/models

S

Sanders, L., & Stappers, P. J. (2013). Convivial Toolbox: Generative Research for the Front End of Design (Illustrated ed.). Laurence King Publishing.

Sargeant, P. K. C. L. R. B. (2022, June 29). Abortion: What does overturn of Roe v Wade mean? BBC News. https://www.bbc.com/news/world-us-canada-61804777

Schmitt, P. (2019). WHY WOULD YOU WANT TO PICTURE IT - ON BEING A VECTOR INSIDE A NEURAL NETWORK. Philippschmitt.Com. Retrieved September 15, 2022, from https://f003.backblazeb2.com/file/studio-ps/Why-Would-You-Want-To-Picture-It-Philipp-Schmitt.pdf

Schweizer Radio und Fernsehen (SRF). (n.d.). Filosofix – Philosophie animiert - Kultur - SRF. Schweizer Radio Und Fernsehen (SRF). Retrieved October 2, 2022, from https://www.srf.ch/kultur/gesellschaft-religion/filosofix

Schön, D. A. (1984, September 24). The Reflective Practitioner (1st Edition). Basic Books.

Shin, T. (2021b, December 14). Real-life Examples of Discriminating Artificial Intelligence. Medium. Retrieved October 2, 2022, from https://towardsdatascience.com/real-life-examples-of-discriminating-artificial-intelligence-cae395a90070

Sofia P. Caldeira, Sander De Ridder, & Sofie Van Bauwel. (2018). Exploring the Politics of Gender Representation on Instagram: Self-representations of Femininity. DiGeSt. Journal of Diversity and Gender Studies, 5(1), 23. https://doi.org/10.11116/digest.5.1.2

Stappers, P. J., & Giaccardi, E. (2017). Research through Design. In M. Soegaard, & R. Friis-Dam (Eds.), The Encyclopedia of Human-Computer Interaction (2nd ed., pp. 1-94). The Interaction Design Foundation.

Stembert, N., & Harbers, M. (2019). Accounting for the human when designing with AI: challenges identified.CHI'19-Extended Abstracts, Glasgow, Scotland Uk—May 04-09, 2019
Sterling, B. (2009) Design Fictions. Interactions. Vol 16. Issue 3.

Søndergaard, M., L., J. and Hansen, L., K. (2018). Intimate Futures: Staying with the Trouble of Digital Personal Assistants through Design Fiction. Proc. DIS '18, ACM, 869–880.
Take a Test. (n.d.). Retrieved October 1, 2022, from https://implicit.harvard.edu/implicit/takeatest.html

T

Tanenbaum, J. (2014). Design Fictional Interactions: Why HCI Should Care About Stories. Interactions, Sept 2014, 22-23. ACM.

Teachable Machine. (n.d.). Retrieved October 4, 2022, from https://teachablemachine.withgoogle.com/train

Thaler, R. H., & Sunstein, C. R. (2021, August 3). Nudge: The Final Edition (Revised). Penguin Books.

ThingsCon 2022 – ThingsCon. (n.d.). Retrieved October 2, 2022, from https://thingscon.org/conference-2022/

Turtle, G. L. (2022). Mutant in the mirror: queer becomings with artificial intelligence. https://doi.org/10.21606/drs.2022.782

U

undefined [Barbie]. (2015, October 8). Imagine The Possibilities | @Barbie [Video]. YouTube. Retrieved September 17, 2022, from https://www.youtube.com/watch?v=l1vnsqbnAkk

undefined [Computerphile]. (2019, June 26). AI Language Models & Transformers - Computerphile [Video]. YouTube. Retrieved October 2, 2022, from https://www.youtube.com/watch?v=rURRYI66E54

undefined [Filmanalyse]. (2022, April 17). Die kapitalistische Ideologie in BIBI & TINA [Video]. YouTube. Retrieved October 2, 2022, from https://www.youtube.com/watch?app=desktop&v=VNowo5m00I

Using Figment with PIX2PIX | Figment. (n.d.). Retrieved October 4, 2022, from https://figmentapp.com/docs/tutorials/pix2pix/

V

Van Alstyne, Greg & Logan, Robert. (2007). Designing for Emergence and Innovation: Redesigning Design. Artifact: Journal of Virtual Design - Artifact. 1. 120-129. 10.1080/17493460601110525.

Van der Burg, V. (2022). Ceci n'est pas une Chaise: Emerging Practices in Designer-AI Collaboration. https://doi.org/10.21606/drs.2022.XXX

Van der Burg, V. (n.d.). The Incredible Algo and I - a contextual design classic [MA thesis].

Vygotsky, Lev (1978). Mind in Society. London: Harvard University Press.

W

Wallendorf, M., & Brucks, M. (1993). Introspection in consumer research: Implementation and implications. Journal of Consumer Research 339e359.

Wang, M. T., & Degol, J. L. (2016, January). Gender Gap in Science, Technology, Engineering, and Mathematics (STEM): Current Knowledge, Implications for Practice, Policy, and Future Directions. Educational Psychology Review, 119–140. https://doi.org/10.1007/s10648-015-9355-x

Westberg J, Jason H. (1994) Fostering learners' reflection and self-assessment. Fam Med. ;26(5):278-82. PMID: 8050644.

Wong, W. L. P., & Radcliffe, D. F. (2000, December). The Tacit Nature of Design Knowledge. Technology Analysis &Amp; Strategic Management, 12(4), 493–512. https://doi.org/10.1080/713698497

X

Xue, H. (2022) Expertinterview

Xue, H., & Desmet, P. M. (2019, July). Researcher introspection for experience-driven design research. Design Studies, 63, 37–64. https://doi.org/10.1016/j.destud.2019.03.001

Z

Zaltman, G. (2003). How Customers Think.

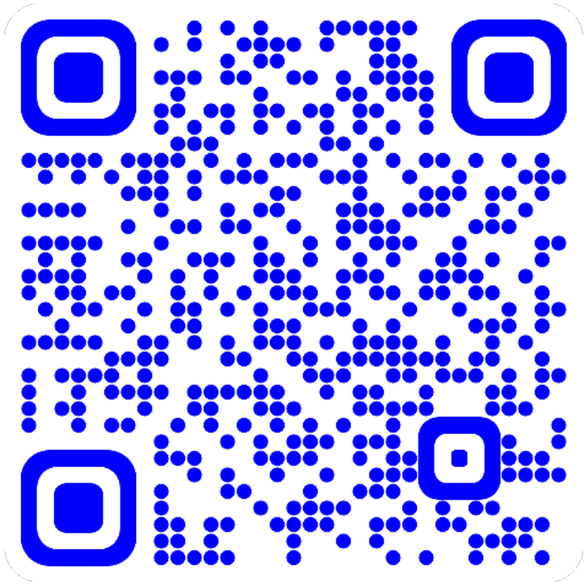
Zijlstra, J., & Daalhuizen, J. (2020). Delft Design Guide (revised Edition). Macmillan Publishers.

Zimmerman, J. & Forlizzi, J. (2008) The Role of Design Artifacts in Design Theory Construction. Human Computer Interaction Institute. Paper 3

Zimmerman, E. (2003) Play as Design: the iterative design process. In: Laurel, B. (Ed), Design Research, MIT Press

öiø / studio. (n.d.). Retrieved October 1, 2022, from https://oio.studio

APPENDIX



Creating Monsters-
crafting gender ambiguous child toys through reflexive designer-AI interactions

Design for Interaction Masterthesis
Faculty of Industrial Design Engineering
Delft University of Technology

Author
Anne Arzberger
Student number: 5382084

Graduation committee
Chair - Prof. dr. Elisa Giaccardi
Faculty of Industrial Design Engineering
Department of Human-Centered Design
Mentor - Prof. dr. Maria Luce Lupetti
Faculty of Industrial Design Engineering
Department of Human-Centered Design
Company - oio.studio
Company mentor - Simone Rebaudengo

project as part of the DCODE Network

December 2023