DOES PRACTICE MAKE PERFECT?
THE COMPLICATED PRACTICE OF DESIGNERS WORKING WITH MACHINE LEARNING AS A DESIGN MATERIAL

Lara Zijlstra
(5668069)

MASTER THESIS
Business Informatics
Utrecht University
April 13, 2021

First Examiner
Dr. Paweł W. Woźniak
Second Examiner
Dr. Almila Akdag

Abstract

Machine learning (ML) is a technology that is undergoing rapid technological innovation. It is expected that this would be followed by a period of design innovation, where designers use the material in new and unexpected ways. And yet, despite increasing interest in ML as a design material, designers seem unable to leverage the material to its full potential. This study aims to characterize the lived practice of designers who work with ML as a design material, in order to determine what is holding them back. I conducted twenty retrospective, semi-structured interviews with design practitioners with experience working on ML-powered products. I analysed the interview transcripts using thematic analysis, and constructed five themes that characterize ML design practice from the data: design process & methods, machine learning as a tool, advocate for the user, need for close collaboration, and real data, real complexities. The findings suggest that despite similarities to non-ML design, designers working in ML face a uniquely challenging practice. This understanding of practice provides the first step towards supporting designers in their practice to fully leverage ML.

Acknowledgements

First and foremost, I want to thank my first supervisor Paweł W. Woźniak. Thank you for your valuable feedback and guidance. Our discussions always kept me sharp and helped me compile the tangles in my mind into this thesis. I also want to thank my second supervisor, Almila Akdag. Thank you for taking the time to provide a fresh perspective and meaningful suggestions on my work. I want to thank my friends and family. Thanks for providing encouragement, proofreading, and much needed distractions. I want to thank the design chapter at ING Advanced Analytics. My interest in ML design was definitely sparked by seeing all the amazing work you do. Finally, I want express my gratitude to every single one of the wonderful designers I got to interview for my thesis. This thesis quite literally would not be there without you. Your stories and enthusiasm for your own work continues to inspire mine.

Contents

Chapter 1

Introduction

This thesis presents a study on the practice of designers who work with Machine Learning (ML) as a design material, with the purpose of achieving a solid understanding of this practice.

1.1 Background

ML is a technology that is receiving a lot of attention as of late (Kuniavsky, M.; Churchill, E.; and Steenson, 2017). Though the technology has existed for a considerable amount of time, ML has really claimed its place in the spotlight in recent years, gaining interest from businesses and researchers alike (Dove, Halskov, Forlizzi, & Zimmerman, 2017; Yang, Scuito, Zimmerman, Forlizzi, & Steinfeld, 2018). Holmquist (2017) poses that ML will continue to become an even bigger part of our everyday lives as intelligence is included in more and more mundane products.

ML is a subfield of Artificial Intelligence (AI), which in turn is a subfield of computer science that focuses on systems that act like humans. As with any technology, design is a considerable part of the ML development process. With the increasing developments in ML technologies, designers are expected to leverage the technology and its capabilities in unexpected ways to create as of yet non-existing products (Dove et al., 2017). However, this does not currently seem to be the case. Rather than coming up with new ideas and being an active part of the development process, designers are just putting the finishing touches on the products (Dove et al., 2017; Yang, Scuito, et al., 2018).

As a result of the lack of design-led innovation in the field of ML, researchers have started to investigate the concept of ML as a design material (Yang, 2018), and the practice of designers working with the material (Dove et al., 2017; Yang, Scuito, et al., 2018).

1.2 Problem Statement

Literature on the practice of designers working with ML as a design material has primarily focused on the challenges these designers have to deal with. Eight challenges of working with ML as a design material have been identified in as of yet. Stembert and Rotterdam (2019) identified five challenges: designers' lack of technical knowledge, data-centered culture, black-box nature of ML, systems changing over time, and personalized solutions. The identification of these challenges is not grounded in practice; they are not based on a study of designers' practices. The same is true for a sixth challenge identified by Browne and Diego (2019): prototyping. Dove et al. (2017) surveyed user experience (UX) designers on the development of ML products and identified a seventh challenge: incorrect mental models. They also reported evidence for the presence of the challenges 'lack of technical knowledge' and 'prototyping'. Yang, Scuito, et al. (2018) interviewed designers with significant experience in the field of ML, identifying the eight and final challenge: lack of design tools and methods specific to ML. They also found evidence for the presence of the 'lack of technical knowledge' challenge. The eight identified challenges of working with ML as a design material will be further discussed later on in this thesis.

The studies that supply evidence for the presence of challenges in practice are based on very small, specific samples. Yang, Scuito, et al. (2018) focused specifically on designers with at least 4 years of experience working on ML. The study therefore might not be representative of designers with less experience. Dove et al. (2017) focused on UX practitioners broadly, also including researchers. Additionally, this study was also limited in terms of depth of the data gathered, as a survey was used.

For all of the challenges identified by Stembert and Rotterdam (2019), except 'lack of technical knowledge', evidence of their presence in practice is missing. The presence and understanding of these four challenges specifically, and for the other four to a lesser extent, is still mainly based on academic understanding of practice, rather than on the actual work of designers: their lived practice. This is a substantial issue, as this approach has been shown to result in a gap between lived practice and the academic understanding of practice (Gray, Stolterman, & Siegel, 2014). Particularly for research aimed at improving practice, it is extremely important to have a sufficient understand of lived practice (Stolterman, 2008). In other words, in order to successfully tackle the problem of designers being unable to leverage ML capabilities to their full potential, it is necessary to reduce the gap between lived practice and academic understanding by constructing a thorough understanding of the lived practice of designers that work with ML as a design material.

To summarize, this study will address the following problem statement: *There is an insufficient understanding of the lived practice of designers who work with ML as a design material, especially in terms of the challenges faced and the strategies used to battle these challenges.*

1.3 Purpose

The aim of this study is to characterize the lived practice of designers who work with ML as a design material. In doing this, the study addresses the gap between the academic understanding and the lived practices of designers. It will determine which challenges of working with ML as a design practice are present in the lived practice of designers, as well as strengthening the understanding of those challenges. In turn, this study is a potential first step towards a solution that supports designers in leveraging ML capabilities to their full potential.

1.4 Research Questions

In order to address the problem statement, the following research question has been defined:

> *RQ: How can we characterize the lived practice of designers who work with machine learning as a design material?*

In order to help answer this question and focus the answers towards closing the gap between academic understanding and lived practice, a series of sub-questions have been defined. Eight challenges of working with ML as a design material have been previously identified. However, not all of these challenges are sufficiently grounded in practice. Thus it is not certain that these are indeed an exhaustive list of challenges that designers face in their work. Therefore, the following sub-question has been defined:

> *SQ1: Which challenges of working with ML as a design material do designers experience in practice?*

The findings for this question will on one hand confirm or contradict the presence of challenges from the academic understanding, and on the other hand potentially identify other challenges beyond the existing set.

For some of the challenges identified, the literature provides an academic understanding of potential approaches for handling these challenges. Yang, Scuito, et al. (2018) provide some approaches from the practice of experienced UX designers for facing the challenges 'lack of technical knowledge' and 'data-centered culture'. However, for the other challenges (and potentially unidentified challenges), no approaches for handling the challenges of ML as seen in practice have been identified. Therefore, the following sub-question has been defined:

> *SQ2: How do designers face the challenges of working with ML as a design material?*

The intention of this question is to identify how designer's confront the challenges they experience in their day to day work. The focus is on the challenges that were actually identified in lived practice. Additionally, for the challenges already identified in literature, the objective is also to compare the findings from actual practice to the academic understanding of approaches towards the challenge.

Though few approaches to the challenges of working with ML can be found, literature does provide some strategies for allowing designers to familiarize themselves to lesser well-known design materials. Some strategies can be applied to design materials generally, while others are specific to ML. Fass, College, and Groves (2019) organized a participatory workshop where designers created physical representations of ML. Luciani, Lindvall, and Löwgren (2018) suggest using 'curated collections' to familiarize designers to the possibilities of ML. Both strategies are created from an academic understanding, and not grounded or found in the practice of designers. In order to determine which strategies are used in practice, the following sub-question has been defined:

> *SQ3: What strategies are used to facilitate familiarization with ML as a design material?*

This third and final sub-question concerns the strategies used by designers but also how those strategies are applied. Compared to SQ2, this question is focused on general strategies applied in practice, rather than on specific solutions to the challenges of working with ML as a design material.

1.5 Scope

There is not yet a clear definition for what is considered working with ML as a design material. The concept broadly refers to to designers working on products that are powered by or contain ML technologies. Yang, Scuito, et al. (2018) and Dove et al. (2017) explicitly mention User Experience (UX) designers. However, this is one of the job titles that refers to an increasingly diverse range of roles. Therefore, in the context of this study, designer refers to any type of role concerned with the human-centered design of technology. Including, but not limited to, UX designers, interaction designers, service designers, and design managers.

Designers work with ML as a design material when they work on products that contain ML technologies, or ML products. As with designers, there is no distinctive boundary that determines what is and what is not a ML product. Designers working on a recommendation feature in a larger product are just as much working with ML as a design material, as designers who work on special purpose ML systems. As long as the product that designers work on contains ML technologies, it is considered working with ML as a design material in the context of this study.

1.6 Outline

In order to create a clear understanding of the context of the study, a literature review was conducted. Chapter 2 provides an overview of findings from this review. Next, Chapter 3 describes the research method used to answer the research questions. The collected data was analysed using a reflective thematic analysis. The findings of this analysis are presented in Chapter 4. Chapter 5 summarizes the answers to the research questions. Finally, the interpretation and implications of the findings are discussed in Chapter 6, together with the limitations and recommendations for further research.

Chapter 2

Related Work

In order to answer the research questions posed in the previous chapter, it is important to first gain an extensive understanding of the concepts that underlie the notion of ML as a design material. Firstly, various perspectives on the concept of design as practice are discussed in Section 2.1. This is followed by an in-depth discussion of the theory of design that this thesis adheres to in Section 2.2. This section also includes an explanation of the concept of design material and more specifically, ML as a design material. The next section, 2.3, describes the concept of design innovation, which is closely related to the idea of design materials. The view of design innovation is also specifically applied to ML in this section.

Aside from the concepts that underlie ML as a design material, it is also vital to review the current understanding of how designers work with ML. In Section 2.4, the eight previously identified challenges of working with machine learning are discussed in detail. This is followed by an overview of potential strategies for approaching work with design materials in Section 2.5. Section 2.6 discusses the current state of research on the practice of designers that work with ML. The final section of this chapter provides a summary and highlights the gap in literature that this thesis will address.

2.1 What is Design?

Design is a difficult concept to define. It is possible to look at design as a perspective, approach, or set of methods. Additionally, design is a large and rapidly evolving field. There is a significant amount of approaches to, and method for design and it would be impossible to discuss all of them. This section will discuss some long-standing perspectives of design as an activity, that have informed many of these approaches and methods. For a comprehensive overview of design perspectives, approaches, and methods, see van Boeijen, Daalhuizen, and Zijlstra (2014).

The simplest way to look at design is to see it as a term with a set definition. However, some say that "a definition carved in granite would neither be feasible nor indeed welcome" (Galle, 2009, para. 5). Nevertheless, not everyone agrees and some definitions have been proposed. For example, Coyne (1995); Lowgren (1995) and Stolterman (1999), who define design as creating something that does not yet exists, and Nelson and Stolterman (2013), who describe design as a type of inquiry that encompasses various activities and processes.

However defined, the concept of design and in particular design as an activity can be approached from various perspectives. Dorst and Dijkhuis (1995) and Ralph (2010) highlight two points of view or 'paradigms' of design processes. Within one paradigm, design is seen as a rational process, while within the other paradigm, design is seen as a set of improvised and creative actions. Fallman (2003) proposes three separate accounts on what design is: a conservative, a pragmatic, and a romantic account. The conservative account shows many similarities to the the rational paradigm, and the pragmatic account is very similar to the action-centric paradigm. While the romantic account shows some similarities to the action-centric paradigm, it provides a different view of design that moves beyond the two paradigms. In what follows, these three approaches will be further elaborated on.

2.1.1 Rational Paradigm

The rational paradigm, or conservative account, considers design to be a scientific discipline that uses theory and techniques known from the more traditional natural sciences. This perspective is very much based on the theories of problem solving introduced by Simon (1992). According to Simon (1992), designing is a process of problem solving, with designers selecting the best way to solve a problem by applying systematic knowledge. This also implies that design only begins once a problem arises.

Hodgkinson (2004) defined three characteristics that summarize the rational paradigm. The first of these characteristics is that designers work from pre-defined objectives and restrictions when they are designing. The problem, goals, and constraints of the prospective solution must be known in advance (Simon, 1992). Within the rational paradigm, it is important that the problem that a designer is working on is very explicitly detailed in some kind of specification, such as a requirements specification for software engineering (Lowgren, 1995).

The second characteristic is that the process of designing artifacts is driven by plans. If there are no plans, there is no action. In case of unexpected circumstances, the design process must be replanned before it continues. Design processes are evaluated by comparing the executed actions and the plans. Design is problem-solving, driven by plans with a designer working towards a goal by following a plan limited by the constraints of the situation (Newell & Simon, 1972).

The third characteristic is that the design process is understood through a pre-determined sequence of phases. The design process is structured and can be described in terms of distinct, methodological steps (Lowgren, 1995). The phases and their order is planned in advance. The sequence should eventually lead to the goal as composed at the start of a design process, but should also take place within the boundaries of the projects pre-defined constraints (Suchman, 1987). A previously commonly used example based on the rational paradigm is the waterfall model (Royce, 1987).

Because of it's methodological nature, the rational paradigm suggests that is it possible to express design work in terms of what design activities have to be performed in what order, and other guidelines (Fallman, 2003). This implies that design skills can be easily transferred to inexperienced designers (Lowgren & Stolterman, 2007).

The rational paradigm poses that the design process is not based on the personal skills and knowledge of the designer (Lowgren, 1995). Instead, the most important part of a design process is the method. The The quality of a design is determined only by how good a designer is at following an arbitrary set of steps (Fallman, 2003).

2.1.2 Action-Centric Paradigm

The action-centric paradigm, or pragmatic account, was a direct response to the rational paradigm, which does not take into consideration the critical obstacle of linking a process and a task in a specific design situation. The action-centric paradigm is centered around the notion of design being about engagement in a specific design situation. The situation that design takes place in refers to both the reason for the design and the context in which the design process is taking place (Lowgren & Stolterman, 2007). According to this paradigm, design is carried out in this specific context which is already populated by many objects, individuals, and processes with their own characteristics (Fallman, 2003). Design is a constantly interpretative processes with the designer taking the role of the interpreter of the situation that they are working in (Coyne & Snodgrass, 1991). As one of the founding fathers of the action-centric paradigm puts it, design is a reflective conversation with the materials of the design situation at hand (Schön, 1983).

As for the rational paradigm, Hodgkinson (2004) also defined three characteristics of the action-centric paradigm. The first characteristic is that designers work from their own creativity and current emotions when designing. Schön (1983) states that when designers design, they do not separate their thinking from their doing. This perspective can also be found in the ethnomethodological view of action. According to this view, *"the organization of situated action is an emergent property of moment-by-moment interactions between actors, and between actors and the environments of their actions"* (Suchman, 1987, p.179). Both views on action imply that design is guided by creativity (Love, 2000).

The second characteristic it that the process of designing artifacts is spontaneous. Schön (1983) describes how designers work through continuous reflection-in-action. A designer constantly maneuvers between framing a problem, taking actions to improve the problem situation, and evaluating those actions. As designers will adapt their framing and actions to the result of their evaluation, the next step is not known in advance. Rather, the decision on what to do next is made impromptu based on the designers reflection.

The third characteristic is that the design process does not adhere to predetermined stages. Schön (1983) views each design situation as an unique, ambiguous, and often conflicting situation. That combined with the view of design as reflection-in-action shows that the design process cannot adhere to a specific structure, as each situation and the work done on designing for that situation differs. Therefore, there is no sequence of stages that can be used to describe and understand a design process. Consequently, designers have to face the issue of knowing how to approach this variety of tasks. According to Schön (1987), this is the 'artistry' and 'essence' of being a designers.

As opposed to the rational paradigm, the knowledge that is required for successful reflection in action cannot be easily captured in methods. Instead, for the action-centric paradigm, learning happens through doing. Inexperienced designers learn from experienced designers or 'coaches' by working through examples (Schön, 1987).

Compared to the rational paradigm, the designer takes a more central role in the process of designing. As mentioned before, knowing how to reflect in action is the artistry and essence of being a designers. The designer is a 'self-organizing system' that knows both how to create and how to reflect on their own creation (Jones, 1992). As Schön (1987) puts it: *"In contrast to analysts or critics, designers put things together and bring new things into being"* (p.42).

2.1.3 Romantic Account

The romantic account suggests that design is closely related to art, the same way the rational paradigm is related to science (Coyne, 2001). Louridas (1999) gives a description of the relationship between design and art that meets this perspective: *"This relation makes design what it is: design is not just about the creation of useful artefacts, it is equally about the creation of beautiful artifacts"* (p.5).

The romantic account shows similarities to the action-centric paradigm in the sense that both perspectives highlight how creativity and imagination guide the design process. The process does not follow a predefined method, and neither is it based on reflection, but rather on a specific designer's own taste. The outcome of the romantic design process is judged on aesthetics (Schön, 1987). However, there are also contrasting views to the action-centric paradigm.

According to the romantic account, the design process is somewhat mystical. It cannot and should not be entirely explainable, as thinking about what happens when they design might cause designers to become worse at designing (Fallman, 2003). Jones (1992) describes the design process as a 'black box', the important part is the outcome, exactly how the outcome comes to be is not of interest.

Out of all three perspectives, designers are held in the highest regard in the romantic account. *"They are seen as imaginative masterminds equipped with almost magical abilities of creation"* (Fallman, 2003, p.226). Designers are seen as artists, in the same way the rational paradigm sees designers as scientists.

2.2 Design as Reflective Practice

This thesis will adhere to the action-centric design paradigm, particularly the theory of design as reflective practice. The general view of reflective practice was developed by Donald Schön as he observed professionals at work in different occupations, including design. According to Schön (1983), all professional practice is similar to design, and requires this essential know-how of approaching unique situations. He started to explicitly apply his view of reflective practice to design situations in particular, drawing on examples from architecture design education (Schön, 1983).

Central to Schön's theory is the concept of the design situation. This concept is twofold: on one hand, it refers to the context in which the design activities are carried out, but on the other hand, it also refers to the reason for initiating the design process (Lowgren & Stolterman, 2007).

2.2.1 A Reflective Conversation

The main point of Schön's theory is that design is a reflective conversation with the materials of the design situation (Schön, 1983, 1987, 1992). When faced with a design task, a designer interprets the task and the current situation. The designer then reflects on the situation in order to determine the next step. In other words, a designer determines what he is designing while designing it. This is called reflection-in-action; the design process is a continuous succession of reflection-in-action events.

At the core of reflection-in-action lies another concept: knowing-in-action. This concept refers to unconscious activities and implicit knowledge that designers use in their everyday work, as they go through reflection-in-action. This know-how that designers display in action is often different from their own characterization of their expertise.

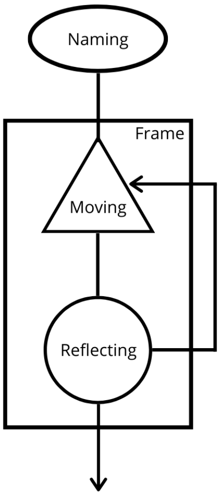

Figure 2.1: Graphic representation of the four activities, adapted from Valkenburg and Dorst (1998)

Valkenburg and Dorst (1998) clarify the process of reflection-in-action by distinguishing four activities (Figure 2.1). The first one, 'naming', refers to when designers name relevant factors in their situation. The designers then move on to the next activity: 'framing'. They frame a situation in a certain way, based on their knowledge-in-action. The third activity, 'moving', refers to the actions a designer takes towards a solution to the design problem. These moves expose the knowing-in-action. The response to a move, also called an action, can produce an unexpected outcome, leading to surprise. In turn, this surprise leads to reflection on the action, the third activity. This reflection-in-action can cause a designer to reconsider their knowing-in-action, and try and frame the situation in a different way, causing the process of moves and reflection to repeat.

2.2.2 Design Material

Another central concept to Schön's theory is design material. The unique situation that a designer faces is *"a material one that is apprehended, in part, through active, sensory appreciation"* (Schön, 1992, p.4). This is the case whether a designer is drawing on a sketch pad or operating a computer. Through the active sensory appreciation of the design situation, designers construct and reconstruct the matter and links between matter that they are working with: their design material.

By going through the activities of a design process, and engaging in a reflective conversation with the material of their design situation, designers are able to envision new solutions using their design material (Louridas, 1999).

2.2.3 Machine Learning as a Design Material

Machine learning (ML) is *"programming computers to optimize a performance criterion using example data or past experience"* (Alpaydin, 2014, p.3). It is a subfield of Artificial Intelligence (AI), which refers to systems that can perform tasks as if they have human intelligence (Russel & Norvig, 2016).

ML systems are also AI systems. However, AI can also encompass intelligent systems other than ML. Though they are not the same, what is referred to as AI today is mostly narrow AI, which is capable of solving only a specific set of problems, mostly problems that can be addressed using ML. ML is starting to become increasingly more popular, with more and more new services and products using the technology in novel ways (Dove et al., 2017).

Computational technologies such as ML can be more than just a way to implement ideas, they can be expressive design material on their own (Landin, 2005). The notion of ML as a design material has emerged in literature (Dove et al., 2017; Yang, 2018; Holmquist, 2017), driven by the vision of reflective conversations with the material leading to design innovation.

Various authors have discussed ML as a design material. For example, Girardin and Lathia (2017) argue that design for ML experiences is different from more traditional design materials, because it is driven by emergent technological opportunities. Similarly, Luciani et al. (2018) describe ML as a material that is "more unpredictable, emergent, and alive than traditional ones" (p. 1). This perspective is also echoed by Van Allen (2017), who states that ML is a completely different context for design compared to more traditional design, as ML systems are not visual. Additionally, they focus on complex behaviour, and are not necessarily fixated on task completion. As ML starts to play an increasingly bigger role as a design material, designers will have to learn to work with its limitations and opportunities (Holmquist, 2017).

2.3 Design Innovation

Design materials, particularly technology-based ones, tend to originate from technical innovation. It is expected that this technical innovation is then followed by a period where designers come into play and start imagining new and unexpected ways to leverage the technology (Yang, Zimmerman, Steinfeld, & Tomasic, 2016; Dove et al., 2017). This process of designers coming up with new ways to use the material is known as design innovation. Dove et al. (2017) use the technology of music players to provide examples of design innovation. As music player technology matured and allowed for smaller device sizes, design brought forth a variety of MP3 players. Technological progress in the fields of network and internet speed led to the possibility of streaming media, which led designers to design streaming services such as Spotify and Apple Music. In each of these examples, a technological advance sparks a whole new series of applications for designers.

The notion of design innovation following technical advances is also echoed by Louridas (1999) view on design as bricolage. Bricolage is a French verb that can be loosely translated into tinkering. According to this concept of design as bricolage, engineers and designers fulfil different roles in the process of innovation. Engineers create new technologies with new potential: the means. Designers do not create new things, but construct new ideas using the means. This process of constructing new ideas happens through reflective conversation with the design material.

2.3.1 Innovation in Machine Learning

ML is a field that is currently experiencing mostly technical innovation. ML has been a topic of research for more than 50 years. There are plenty of textbooks (e.g. Russel & Norvig, 2016 and Goodfellow, Bengio, & Courville, 2016), introductory articles (Domingos, 2012), online resources (Google AI, n.d.), and even resources specifically targeted towards designers (Hebron, 2016). With ML technologies becoming more prevalent in the products we use everyday, some are even speculating that ML is the new UX (Brownlee, 2015).

Because the technical innovation in the field of ML is continuing to advance, it would be expected that design innovation would follow suit. Designers should be coming up with new and innovative ways to leverage ML technologies in unexpected ways. However, this does not seem to be the case. Dove et al. (2017) report that designers seem to be struggling with identifying opportunities for unfamiliar applications of ML. This process of identifying opportunities is vital for design innovation, which does not seem to be taking place in the field of ML at this point. Yang, Scuito, et al. (2018) describe similar findings on design innovation not taking place in ML right now.

The lack of design innovation in the field of ML could be due to the nature of the material. ML, compared to more traditional design materials, is a newer and more volatile design material (Luciani et al., 2018). There are various characteristics of ML that make it such a challenging design material to work with compared to more traditional design materials. These challenges of working with ML as a design material will be discussed in the next section.

2.4 Challenges of Working with Machine Learning

There are various challenges that make ML a difficult material to work with, and might contribute to the lack of design innovation in the ML field. Stembert and Rotterdam (2019) identified and described five challenges: lack of technical knowledge, data-centered culture, black box nature of machine learning, system changes over time, and personalized solutions. This list can be extended by the challenges of prototyping (Browne & Diego, 2019), incorrect mental models (Dove et al., 2017), and existing design methods (Yang, Scuito, et al., 2018). Most challenges are closely related and thus show some overlap.

2.4.1 Lack of Technical Knowledge

Both Stembert and Rotterdam (2019) and Dove et al. (2017) identify designers' lack of technical ML understanding as one of the major challenges of working with ML as a design material. Various studies report that designers have a very limited understanding of the technical aspect of ML (Yang, Scuito, et al., 2018; Fass et al., 2019; Dove et al., 2017). Specifically, when asked about ML, designers do not tend to discuss technical details or important concepts such as ML's relationship to data, statistical analysis, and ground truth (Dove et al., 2017). This causes designers to encounter issues in the understanding of what ML can do in their field of design (Yang, 2018), and issues in successfully applying common design methods to ML projects (Bratteteig & Verne, 2018).

Interestingly enough, though designers acknowledge their lack of technical knowledge, Yang, Scuito, et al. (2018) found that they do not feel like it interferes with their ability to design with ML. Additionally, while their technical understanding of ML is lacking, designers are considerate of the non-technical implications of the technology, such as implications on privacy, identity, and reliability (Fass et al., 2019).

It is interesting that a lack of technical knowledge is brought up as a challenge for designing with ML, as the main perspective on the need for technical understanding for design innovation is that it is not really required at all. Dove and Fayard (2020) state that technical knowledge is not necessary, as design is a collaborative practice and ML understanding can be provided by data professionals. This perspective is echoed in studies of design practice which show that both experienced and inexperienced designers leave the technical understanding of ML to engineers. Dove et al. (2017) report that rather than working from a technical understanding, designers use examples of ML applications to understand where and how ML can be applied. Yang, Scuito, et al. (2018) describe similar findings with designers explaining ML in terms of abstractions in the form of examples rather than technical terms. However, Dove et al. (2017) also acknowledge a drawback of understanding ML only in terms of examples: designers might start to see ML as a magical solution for any problem, due to a lack of knowledge on the limitations of the technology.

Rather than a requirement for technical knowledge, there is a need for a more basic understanding of the material, that allows designers to see potential in ML. Holmquist (2017) states that anyone working with ML, including designers, do not need to be experts on the topic of ML, but must understand the possibilities and limitations of the technology. If designers do not understand the limitations of ML, their designs are likely to fail (Holmquist, 2017). This perspective is echoed by Bratteteig and Verne (2018), who state that it is essential for designers to have a basic understanding of what is possible and impossible to achieve with ML. According to Yang (2017), if designers' understanding of ML is adequate, they can not only conceive new ways to apply ML, but also envision new ways to use data more generally. Van Allen (2018) explains how an understanding of ML provides designers with a unique view on the material that can lead to new approaches for using ML, as well as collaborations and new design methods.

One way of improving designers' knowledge on ML is through design education. Yang, Scuito, et al. (2018) suggests two approaches: including ML literacy -or more generally the language of data science- in design education, and promoting collaboration between designers and data scientists in education. Van Der Vlist et al. (2008) experimented with the embodied intelligence method to teach ML to design students. By using a physical embodiment of a ML system, the design students gained an understanding of ML and were able to built a neural network in three days by themselves.

2.4.2 Data-Centered Culture

In a human-centered design process, designers generally start by exploring a problem context and extracting user needs, and then continue the process to design a product or feature that addresses this user need. In other words, the problem setting and problem solving are separate processes. For projects related to ML, it is also important to maintain a human centered design process (Stembert & Rotterdam, 2019). However, this does not currently seem to be the case. Creating a separation between problem setting and problem solving is very complicated in ML processes (Yang, 2017). Most new ML products or features emerge from available data rather than from a user need (Girardin & Lathia, 2017). These data-driven design processes lack purpose and therefore often fail to address users' needs (Colborne, 2016; Yang, 2017).

Dove et al. (2017) report that a majority of ML projects is led by data rather than design, due to the unbalance between engineers and designers, with designers sometimes only joining towards the end of a project. According to Fass et al. (2019), design is seen as an end goal, rather than something that a project can be led by. Additionally, the complexity of ML technologies and designers' aforementioned struggles to understand it only aggravate the dominance of data in the design process.

Both Dove and Fayard (2020) and Stembert and Rotterdam (2019) mention how especially in ML projects it is vital for designers to ensure that both the user and the context are taken into account. Yang, Scuito, et al. (2018) describe how designers working with ML recognize that they work in a data-centric environment. Additionally, they understand that the only way to facilitate concern for the user and context is to embrace this data centered culture, for example by adapting the data science vocabulary. Other more elaborate strategies to incorporate the human-centered perspective to ML design are proposed as well. Baumer (2017) suggests human-centered algorithm design with three possible strategies: theoretical, participatory, and speculative. Another approach by John Morley & Associates incorporates the human-centered design thinking approach with a data-centered ML pipeline (Schmarzo, 2017).

2.4.3 Incorrect Mental Models

Another challenge is the way thatML works contradicts the mental models of both users and designers, due to the fact that ML systems are unpredictable. Riedl (2019) describes how the way ML works is very different from the expectations of users without technical training. When interacting with people, humans can reason what another human will do and why. This is not the case for interactions with ML systems. As opposed to systems based on heuristic rules, ML does not always think or behave the way a human would, and may come up with a completely unexpected solution that does not align with common sense (Dove & Fayard, 2020). More so, ML systems might even display behaviors that are not just considered confusing, but even disturbing or dangerous (Amershi et al., 2019). Because of this unpredictability of ML systems, user expectations are not met. The unexpected behavior can lead to negative user experiences (Dove et al., 2017; Amershi et al., 2019; Kocielnik, Amershi, & Bennett, 2019). Fass et al. (2019) also reported findings on the incorrect mental models of users. Designers acknowledged how ML "projects an illusion of accuracy and efficiency" (Fass et al., 2019, p.5), neglecting the uncertainty that can also be associated with the technology. Fass et al. (2019) found that designers are aware of this mismatch between users' mental models and the way that ML works and acknowledge the need for design methods that address this issue.

However, ML also challenges designers' mental models. As mentioned before, ML systems deal with large and dynamic datasets, and learning over time. This is different from the material that designers typically work with (Dove & Fayard, 2020). Girardin and Lathia (2017) describe another issue surrounding mental models that designers experience when working with ML. The qualitative research adopted by designers is used to construct conceptual models about a product's users. These conceptual models are used during the design process in order to design a product that brings value to the users. However, in the end the conceptual model used for design is generally not implemented into the statistical model of the ML product. Girardin and Lathia (2017) highlight that the goal of collaboration between designers and data engineers should not be for the conceptual model to become the statistical model, but translating the conceptual model in a way that makes sense to the statistical model.

2.4.4 Black-Box Nature of Machine Learning

Another challenge of ML is that it is really complicated to explain how a ML system does what it does (Holmquist, 2017). ML systems are seen as black boxes that produce an output, but it is not transparent to users how this output comes to be. This challenge is closely related to the previous challenge of users' mental models. The way ML systems are composed makes it hard to explain exactly what the mechanics of the system are, even by the person who has written the algorithm and initially trained the system. A common example is a Google Translate neural network that created its own language in order to translate from languages that it was not trained on (Wong, 2016).

The real challenge for designers here is to design purposeful ML systems. A lack of transparency in ML systems can cause users to distrust the system and vice versa (Riedl, 2019). Designers need to consider these implications of ML's black-box nature. They need to make sure that the systems they design are meaningful; systems that address the user needs without coming across as creepy (Dove et al., 2017; Dove & Fayard, 2020).

There are two promising approaches that address this challenge: explainable ML and interpretable ML. Although the terms are often used interchangeably, it is important to make a distinction, as they each refer to different concepts (Lipton, 2018). Interpretability refers to "the degree to which a human can understand the cause of a decision" (Miller, 2019, p. 8). Some ML models are simple enough that they are intrinsicly interpretable. While for more complex models, interpretation is performed after model training, this is known as post hoc interpretability (Molnar, 2021).

Harvard Business Review Analytic Services provides a definition of the concept of explainable AI that reads: "machine learning techniques that make it possible for human users to understand, appropriately trust, and effectively manage AI" (Harvard Business Review Analytic Services, 2019, p. 3). The field of explainable AI aims to create explainable models that still maintain high quality levels (Adadi & Berrada, 2018).

The main difference between the two approaches is that interpretable AI is concerned with cause and effect of the model in general, while explainable AI is concerned with specific parameters and their effects on the model. Despite the differences between the two approaches, in some way they both allow users to understand how a ML system does what it does.

The major challenge of making an understandable ML model is that the model needs to be understood by humans. In other words, a complex model needs to be translated into a human understanding. Design professionals can make an important contribution to this challenge as the process of communicating a complex model to a human requires Human-Computer Interaction expertise as well as ML competence (Adadi & Berrada, 2018).

There have been various efforts negate the effects of the MLs black-box nature through the use of explainability and interpretability. For example, Lim and Dey (2010) propose a toolkit for context-aware systems that improves the explainability and interpretability of these systems. Ribeiro, Singh, and Guestrin (2016) introduced LIME, an explanation technique that can be used to explain classifier models. Sokol, Hepburn, Santos-Rodriguez, and Flach (2019) explain this technique by proposing a framework that allows anyone to build their own custom explainer of black-box models.

2.4.5 System Changes over Time

As the name suggests, ML systems learn. As the datasets that a system learns from grow bigger over time, the behavior of the system can change. If the system learns from biased or incomplete data, the resulting system change could even be negative (Bratteteig & Verne, 2018).

This dynamic relationship to data is different than what designers are generally familiar with (Dove & Fayard, 2020). Rather than design for a single system behaviour as they are used to, designers have to think ahead and account for a system that might change its behaviour in the future (Stembert & Rotterdam, 2019). Designers also need to consider that a system needs to collect new data in order to improve its behavior and tailor it to the user. Within these systems, user interactions become sources of data themselves (Girardin & Lathia, 2017). This comes with new difficulties: if a user has to explicitly provide data in order to train a system, it could be a barrier for using the system (Holmquist, 2017).

While the outlook of interactions adapted to a specific user sounds ideal, these adapted interactions often come at the expense of usability (Yang et al., 2016; Yang, 2017). For example, the adaptive behavior of ML systems directly conflicts with one of the most universal UX principles: consistency (Stembert & Rotterdam, 2019). Another ramification of interactions changing over time that needs to be considered is evaluation metrics. Common evaluation metrics are not suited to interactions or interfaces that change over time (Stembert & Rotterdam, 2019). Therefore, for ML systems that change over time, evaluation metrics need to move beyond optimization metrics and should also consider more used-centered measures (Kirsch, 2017). This issue is also reiterated by Cramer2017, who state that it is common for ML metrics to not necessarily be good representatives of user success.

2.4.6 Personalized Solutions

With ML, it becomes possible to create personalized components, such as functionalities, interfaces, or content, for example, for specific users or user groups. Some even see this personalization as the next great thing for UX (Kuang, 2013; Brownlee, 2015). This range of potential solutions presents a challenge to designers, as they have to consider the changing and new elements that users will interact with (Stembert & Rotterdam, 2019).

Though using ML to provide personalized content is nothing new anymore, ML can also be applied to deliver personalized interfaces to a user (Yang et al., 2016; Spradlin, 2015). Whereas designers are used to creating a single best interface for all users, adaptive interfaces might select a best solution from a series of interfaces based on user-generated data. As described, classic evaluation metrics for interfaces are not suited to a variety of interfaces and should be extended to include user-centered measures as well as optimization criteria. Additionally, the challenges of explainability and transparency also come in to play for adaptive interfaces in particular, as it should be clear to a user why he is presented with a particular interface.

A potential approach towards this challenge is Algorithm Driven Design (Vetrov, 2017). The approach suggests that designers should work together with algorithms in a way where an algorithm can automate part of the routine design process, such as generating basic user interfaces (UIs) or performing A/B tests. Designers can then focus on designing the behavior of the system and final touches to design. The biggest downside to this approach is that without competent tools, the automation is not possible.

2.4.7 Prototyping

A big part of most design processes is creating prototypes: an early model of a product meant to test a concept. Prototyping ML products is complicated (Browne & Diego, 2019). The main concept of a prototype is to create the prototype in such a way that it is just enough to test a concept with. ML challenges this idea of prototyping; a functional ML prototype requires a lot of data and is therefore a large commitment (Yang, 2018). It's complicated to consider certain aspects of ML that are vital to achieving a good result, such as the statistical approach used, or the training data set, early on in the design process (Bratteteig & Verne, 2018). Creating realistic prototypes is particularly challenging in newer fields where no data sets or algorithms are available of yet (Dove et al., 2017; Yang, Scuito, et al., 2018).

The main implication of not being able to sufficiently prototype a ML product is that ideas cannot be tested on a small scale. The technical validity and whether a product is accepted in a context remains unknown until after implementation (Yang, 2018).

For interface design or interaction design, plenty of tools are available for designers to prototype and test a web interface or simulate smartphone interaction. This is not the case for designing with ML. Dove et al. (2017) report this lack of ML prototyping tools, and also explains how this prevents designers from researching the impact of false positives and false negatives in ML products.

Some interactive machine learning platforms can potentially be used for prototyping, such as Wekinator (Fiebrink & Cook, 2010), and Google Teachable Machine Learning (Google AI, 2017). However, these platforms require technical skills or are too limited for sufficient prototyping respectively. Malsattar, Kihara, and Giaccardi (2019) introduce ObjectResponder, a tool that allows designers to combine objects recognized by a smartphone camera to responses. The limitation of this tool is that it can be used only to prototype object recognition products.

Browne and Diego (2019) suggest using Wizard-of-Oz prototyping for the development of ML products, calling it "the ideal method for prototyping machine learning experiences" (p.5). With this prototyping method, participants are made to believe that they are interacting with a functional system, while they are actually interacting with a human who is emulating the functionality of the system. Of course this method would still require at least a small data set and algorithm to set the abilities and limitations of the 'system'. Van Allen (2018) experimented with another potential ML prototyping tool based on wizard of oz prototyping: the Delft AI Toolkit. The toolkit is a working prototype of an open source visual authoring system that allows designers to create ML prototypes through simple dragging and dropping.

2.4.8 Design Methods and Tools

As mentioned before, a major challenge of working with ML as a design material is a lack of prototyping tools specific to ML. This is part of a larger problem of existing design methods. Bratteteig and Verne (2018) address how the challenges previously introduced in this section, such as lack of ML understanding, system changes over time, and prototyping, pose challenges to a common design method: participatory design.

Existing design methods are useful to designers working with ML only to a certain extent (Yang, Scuito, et al., 2018). Dove et al. (2017) report similar findings on a lack of methods and tools specifically tailored to the challenges of designing with ML. Van Allen (2017) also touches on the uncertainty of classic design methods such as journey maps and content strategy working for ML design.

The only way to address this challenge is by developing new design methods specifically for ML. This statement is echoed by Van Allen (2017), who states that in order to design successful ML products, it will be necessary to redesign approaches to design. Dove et al. (2017), Yang, Scuito, et al. (2018), and Van Allen (2018) also highlight the need for the design of tools and methods specific to designing with ML.

There are some steps being made on the path towards ML-specific design methods. Zhou, Qi, Gong, and Sun (2019) propose the ML-Process Canvas. The ML-Process Canvas is a tool that helps designers to consider the user, ML system, and scenarios throughout the entire design process. Zhou, Sun, Zhang, Liu, and Gong (2020) propose the Material Lifecycle Thinking (MLT) design method, which considers ML as a continuously changing material that has an iterative lifecycle. The method consists of a cycle of five main steps, focused on reflecting on ML's entire lifecycle and involving the co-creators of the user experience during this lifecycle.

In terms of tools, Dove and Fayard (2020) designed a set of cards that prompt conversation on ML challenges to be used during ideation. Each card can be used to ask a specific question about the use of ML, and the use of the cards as a whole promotes reflection on design with ML.

2.5 Familiarizing With Design Materials

ML is not the only challenging design material that designers have to tackle. This section offers an overview of strategies that have been used to explore design materials. First, general strategies for acquainting designers to design materials are discussed. Then, approaches for immaterial design materials like ML are reviewed. Finally, strategies specific to working with ML are introduced.

2.5.1 General Strategies

One way of allowing designers to become more familiar with design materials is through the use of 'sensitizing concepts'. Sensitizing concepts are concepts used by design researchers to produce knowledge in research-through-design (Zimmerman, Stolterman, & Forlizzi, 2010). They are artifacts that can be used to guide innovation. The goal of the artifact is to sensitize other designers to new possibilities of a design material. As Yang, Banovic, and Zimmerman (2018) state, *"they help designers grasp and feel the new design materials."* (p.2). Another approach somewhat similar to sensitizing concepts is the use of 'boundary objects'. Boundary objects are artifacts that sit between two areas of expertise. They can be used to support communication and harmony between two perspectives (Star, 2014).

Another way to get designers to create novel things with a new design material is through a process known as matchmaking (Bly & Churchill, 1999). In matchmaking, designers start by identifying the capabilities of the technology that they are working with. They then work through this list of capabilities and spot activities related to the capabilities. The designers then determine domains related to the activities, and users related to the domains. This process can help designers to see their material in a new light.

An example of a newer design material which required designers to do some exploring is haptics. Moussette and Banks (2011) explain how designers can only master the material and sensitize themselves to its capabilities by working with the material. In order to facilitate this, they created a series of simple haptic devices. Following up on this work, Moussette (2012) started a challenge of building at least one haptic device every single day to improve engagement with the material. He described this design marathon as a reflective conversation with haptics that led to a series of 'working strategies' for this design material.

2.5.2 Strategies for Immaterial Design Materials

A type of material that is especially tricky to design with is immaterial material. With a physical design material, designers can actively explore the real material by 'playing' with the material. This allows them to gain a tacit understanding of what is possible with the material. The previous example of haptics illustrates this active exploration. With immaterial materials, designers cannot interact with the material in the same way as physical materials, limiting their tacit understanding (Buxton, 2007).

A common example of an immaterial design material is technology. Technology is a very complicated design material as it is in a sense a material without qualities. This results in a design process that is less constricted by the design material (Lowgren & Stolterman, 2007). Ozenc, Kim, Zimmerman, Oney, and Myers (2010) studied how to support designers working with this material, specifically for the the design of UI controls. They conducted two participatory workshops. These workshops identified a need for a tool that supports designers in the process of designing new UI controls. This tool should allow for refinement and continuous communication of this refinement with developers. Additionally, as designers use their bodies to devise and communicate new gesture-based controls, the tool should also allow communication through gestures.

Sundström et al. (2011) proposed the use of 'inspirational bits' as a way for designers to get to know software as a design material. Inspirational bits are *"a rough way of seeing the technology that allows us to look at it, feel it, and experience it over time and space, exposing all or some of the properties of a material"* (p.1569).

The previously introduced concept of boundary objects can also be applied to software as a design material. Boundary objects can be used to improve the collaboration between designers and developers. With boundary objects, developers can become the voice of the design material, allowing designers to have a reflective conversation with their material through the developer (Ozenc et al., 2010).

Another example of an immaterial material is experience (as in, the experience part of user experience) (Buxton, 2007). For this design material, various tools and methods have been proposed to make designing with the material easier. Buchenau and Suri (2000) suggests 'experience prototyping': an approach that allows both designers and users to become familiar with an intended user experience. Ehn and Kyng (1991) suggest for designers to 'sketch' real life scenarios using cardboard and paper to simulate realism. Odom et al. (2012) propose that designers try to replicate both the physical and social context of an experience, and ask potential users to act out scenarios that involve new experiences. They call this approach 'user enactments'.

2.5.3 Strategies for Machine Learning

Some suggestions have been made for strategies that make it easier for designers to work with ML as a design material specifically. Fass et al. (2019) present a participatory workshop where designers create physical representations of models of the effects of ML. Working with physical materials allowed designers to think and communicate about ML in a new way. The workshop led to a series of designerly abstractions that can be used as boundary objects between data science and UX to promote new possibilities for designing with ML.

Considering the characteristics of ML that make it a challenging design material, Luciani et al. (2018) suggest the use of 'curated collections': annotated portfolios of exemplary examples (exemplars). These curated collections can be used to familiarize designers to a design situation for future design. Luciani et al. (2018) also provide an example of a curated collection, and show how this collection can be used to improve the design of an air traffic control system.

2.6 Practice of Designers Working with Machine Learning

In HCI research, the understanding of professional practice is often based on academic interpretation of practice (Gray et al., 2014), rather than on the actual work of design practitioners (Stolterman, 2008). This is an issue as it causes a gap between what Gray et al. (2014) calls the 'projected practice community' and actual practice. Especially for research aimed at improving design practice, it is important to have a good understanding of actual design practice. Research that is not grounded in this understanding is likely to fail (Stolterman, 2008). Daly, Adams, and Bodner (2012) echoe this perspective and explains how design is characterized by more than current definitions; it is also defined by the people who engage in it. Inquiries into how design professionals work serve a purpose beyond understanding design, they show how design is practiced.

There is quite a bit of work on the challenges and potential solutions of designing with ML, as evident from the previous sections. However, the practice of the designers that actually work with ML as a design material in everyday life has not received a great deal of interest as of yet. This is particularly interesting because the practice of related roles such as data scientists or engineers, collectively called 'data workers', in the field of ML have been thoroughly investigated (e.g. Muller et al., 2019, Kim, Zimmermann, DeLine, & Begel, 2016, Passi & Jackson, 2018).

Though there is little research on the practice of designers working with ML, there is some. Dove et al. (2017) conducted a study on how designers work on developing new ML products and services. A survey on 43 UX practitioners, including designers and researchers, identified practitioners' struggle in areas such as ML knowledge, the purposeful use of ML, and ML as a design material. While this study highlights some problematic areas for designers working with ML, it focuses specifically on the creation of new services, rather than the overall practice of designers in the field of ML.

Yang, Scuito, et al. (2018) conducted a series of interviews with a focus on practice with design practitioners (including UX designers, UX researchers, and design managers) with at least four years experience in designing ML-enhanced products. The main findings of the study were centered around three topics: participants' lack of understanding of ML capabilities, differences in the design process associated with ML, and the data centered culture that participants admit to working in.

The sample size of the interview was 13 UX practitioners, including designers, researchers, and design managers. All practitioners had at least 4 years of experience in the field of ML, with most of them having over 10 years of experience. The specific sample begs the question whether the findings of the study are also generalizable to all UX practitioners, in particular those with less experience.

2.6.1 Collaboration with Data Scientists

As mentioned before, the roles and practice of data workers has been a topic of interest in research. The collaborative practices of designers and data workers specifically have not been studied as of yet. However, some researchers studying design practice or data worker practice have commented on it.

Both Dove et al. (2017) and Yang, Scuito, et al. (2018) report how designers and data workers work in close collaboration on ML products. There is however a contrast between types of organizations that impacts this collaboration. There are large data-driven organizations where designers have unrestricted access to data scientists, and ML innovation is actively promoted. On the other hand, at some organizations the access to data scientists is much more limited. Designers who work in the second category have less access to exemplars and are therefore less likely to come up with new ways to use ML, sticking to common applications such as recommender systems or intelligent reminders (Yang, Scuito, et al., 2018).

Girardin and Lathia (2017) discuss how UX designers work with data scientists on products that are based on a feedback loop. In other words, products that change over time based on user data gathered during use. They describe how in data-driven organizations, designers work much more closely together with data scientists and software developers, compared to traditional organizations. This close collaboration is necessary for systems with feedback loops, since a holistic view is required to successfully build and better these systems. They conclude by saying that it is the joint responsibility of data scientists and designers to construct ML-based experiences that bring everyday value to the user.

2.7 Summary

ML is a unique design material. Its immaterial and emergent nature make it more complicated compared to traditional materials. For most technological design materials, a shift from technical developments to design-led innovation can be observed. In this process of design innovation, designers are expected to work with the design material, and envision new and unexpected ways to leverage it. However, within the ML field this shift does not seem to be taking place.

The reason for the lack of design innovation is not entirely clear. Though there is quite a thorough academic understanding of the challenges of ML as a design material, and means to deal with these challenges, little work has been done on the lived practice of designers that actually work with the material in their day to day work. The same can be said for strategies for familiarization with ML as a design material. The lack of work on the lived practice of designers working with ML is concerning, as it is fundamental to understand their current practice, before steps can be made towards improving it.

The current characterizations of practice are limited, and based on specific samples of practitioners. For example, they focus specifically on the creation of new products, or on experienced designers working with ML. By creating an exhaustive understanding of the current practice of designers who work on ML products in their daily work, this study will lay the groundwork for improvements of this practice that allow designers to fully leverage ML as a design material.

Chapter 3

Method

The purpose of this research is to investigate the practice of designers with a broad range of experience in working on ML-powered products. In order to achieve a comprehensive overview of designers' practice, I chose to conduct retrospective, semi-structured interviews. I decided upon this research method mainly based on two considerations. Firstly, a retrospective research method was preferred as it allowed participants to discuss completed projects from beginning to end, covering the entire design practice. Interviews are commonly used as a research method to collect data about the past from interviewees' memories (Seaman, 2008).

The second consideration was the depth of data gathered by interviews. Previous studies on the practice of designers working with ML have used surveys and interviews (Yang, Scuito, et al., 2018; Dove et al., 2017), both being methods that allow for retrospective data gathering. The main goal of this study is to characterize practice. Practice is a complicated matter that might be difficult to capture in numbers or short textual descriptions, the type of data commonly gathered from a survey. Interviews are used often to characterize practice, and to acquire the deeper insights associated with the matter (Lazar, Feng, & Hochheiser, 2017). Therefore, interviews are the preferred research method.

The interview were semi-structured, as this allows for more focus on the interviewees' point of view (Bryman, 2015). Since the focus of the current study is designer's practice, it makes sense to emphasize their perspectives in the interview. Additionally, semi-structures interviews have the advantage of allowing unexpected information to be voiced (Seaman, 2008), which is helpful in the context of identifying new challenges and approaches to working with ML.

In addition to the interviews, I decided to use a pre-interview survey to gather background information from the participants prior to the interview. However, this survey served primarily to support the data gathered from the interviews, not as its own point of data collection.

This chapter describes the process of interviewing as performed by this study. The process can be broadly split into three phases: preparing the interviews, conducting the interviews, and analysing the data gathered. Each phase will be discussed separately in its own section.

3.1 Preparation

The first phase of this study was to undergo the necessary preparations for the interviews. Firstly, I decided that the narrative of the interviews would be centered around a specific example project selected by each participant. By starting from a specific instance in the past, participants were able to give specific, detailed answers. Additionally, discussing a nearly completed project allowed the interviewees to draw conclusions about their work that they might not have came to had they still been midway into their project. Finally, by walking through the example project, designers brought up challenges in their practice themselves, rather than be led by the challenges included in the interview guide. The practice of having interviewees walk through an example project to structure an interview has been used before in studies of design practice, for example Yang, Scuito, et al. (2018), and Newman and Landay (2000), and studies of data science practice, for example Zhang, Muller, and Wang (2020), and Muller et al. (2019).

As part of the process of preparing the interviews, the following three activities were performed concurrently: producing the pre-interview survey, formulating an interview guide, and recruiting participants. Each activity is discussed separately in the following three subsections.

3.1.1 Pre-Interview Survey

I decided to conduct a short survey prior to the interviews to gather participant background information. By gathering background data prior to the interview, the information from the survey could be used to prepare for the interview. Moreover, participants could fill out the survey quickly and in their own time, preventing the collection of background data to take away time from the actual interview.

This pre-interview survey served a dual purpose. First of all, having a general view of a participant's background before the start of the interview helped the interviewer to ask the most relevant questions for that particular participant. Second, the background information acquired through the survey helped to put the participants experiences into context. In turn, being appreciative of a participant's background can help in interpreting their results correctly. For example, it could be expected that a designer who expresses very little familiarity with ML has a different experience designing products with the material, compared to a designer who is very familiar with the concepts. In conclusion, awareness of the participants' background helps to draw the correct inferences.

The survey was created and distributed through Google Forms. The contents of the pre-interview survey can be found in Appendix A. The survey roughly consisted of two parts, which will be discussed separately.

The first part of the survey contained basic demographic and background questions. The participants were asked about their age, educational background, and experience working on ML products. These questions served to provide a general idea of a participant's background. The survey also contained a set of questions about the example project that the participant has selected for the interview. A reminder was included to emphasize that the questions had to be filled out according to the situation of the specific project. Participants were asked about their role during the project in question, as well as the size of the organization that they worked for at the time of the project. These questions generated a basic understanding of the participant's example project.

The second part of the pre-interview survey concerned the participant's familiarity to common concepts both from the fields of design and ML. The purpose of this part especially was to gauge a participant's knowledge and experience with ML, in order to ask the appropriate questions, as well as to provide context for the interpretation of results.

Participants' familiarity with a total of seven concepts was measured using a scale of 1 to 7, with 1 being 'I know nothing about them', 4 being 'I have heard of them but have no hands-on experience', and 7 being 'I work with them in my day-to-day work'. This scale was adapted from Yang et al. (2018). The scale was selected as it captures not just knowledge of the concept, but also a participant's experience in working directly with or on the concept.

Concepts related to both design and ML were included, despite the focus of this part being able to estimate the participants' familiarity with ML. This exposed the expected differences in designer's familiarity to their own field of design, compared to the field of ML. The concepts related to design that were used in the survey are: 'wireframes', 'personas', and 'user flows'. The ML related concepts that were used in the survey are: 'model training', 'model validation', and 'training set, validation set & test set'. Additionally, the concept 'correlation & causation' was included to gauge participants familiarity with statistics, which strongly underlies ML. The concepts 'user flows', 'training set, validation set & test set', and 'correlation & causation' were directly adapted from Yang, Scuito, et al. (2018). The list of concepts was extended with the other concepts based on discussions with a data scientist and designer, to more accurately capture the whole fields of design and ML.

3.1.2 Interview Guide

An interview guide was formulated to be used during the interviews. The interview guide can be found in Appendix B. The interviews were semi-structured. Therefore, the interview guide must be seen as a overview of topics that could be discussed during the interview, rather than a rigid question and answer format. It must also be noted that not all topics or questions were expected to be addressed in each interview, due to both time constraints and participants' inexperience with some of the topics. The topics addressed in an interview depended mainly on the topics that were brought up by each interviewee. Additionally, it must also be noted that the semi-structured nature of the interviews meant that new questions or follow up questions could be generated based on the interviewees responses that are not part of the interview guide.

The interview guide broadly consists of three segments. Each interview started with an introduction to the study and an explanation of the goal of the interview. This was followed by practical information surrounding the interview, such as timing and confidentiality. At this point, interviewees had the opportunity to ask questions about the provided consent form, before signing said form.

The narrative of the interviews was centered around an example project as selected by the interviewee. The second segment of the interviews concerns this project. After signing the consent form, the interviewee was asked to provide a description of the example project they had selected. This description served as a natural start to get the interview going. Additionally, it provided some context to the experiences of the designer in question. To prevent designers from sharing experiences related to only their example project, interviewees were reminded at this point to bring up anything from other projects or experiences related to the topics discussed during the interview. Additionally, the interviewer also asked pointed questions about the interviewee's broader experiences around certain concepts as they came up in the interview.

After describing the selected example project, the interviewee was asked to walk the interviewer through their work on the project, starting with the initiation of the project. Project initiation is one of two topics that was included based on the challenge of data centered culture as identified in the previous chapter. It is included in the first segment of the guide, as opposed to all other challenge-based topics, because the initiation of a project would chronologically be discussed at the start of the example project.

The remaining challenges of working with machine learning as a design material as identified in the previous were all included as topics in the third segment of the interview guide. The second topic derived from the challenge of data centered culture, data collection, is included in this segment. I decided to include probes for the eight challenges as they were topics that were expected to arise throughout the interview based on findings from previous studies. Additionally, a probe for the topic of familiarization to ML as a design material was included to address the third research question.

The exact probes included for each of the nine topics are different, but there were overarching themes. For example, the probes often inquired about how a challenge impacted the designers' practice, and how they dealt with the challenge in question in their work. For some of the topics, such as design methods or prototyping, interviewees were asked what approaches they used, why, and how those approaches worked for them. For the probes regarding ML knowledge, designers' self-identified level of knowledge was explicitly not included in the wording of the question as to allow the interviewees to express their opinion on their level of knowledge freely.

The topics included in the third segment of the interview guide were addressed as they came up during the interview. For example, if the interviewee brought up struggling with understanding ML in their project, the questions related to the topic ML Knowledge could be brought up. The topic of ML knowledge was always directly followed by the topic of familiarization to ML as a design material, to play into the natural progression of knowing about ML, to applying ML.

If any topics did not naturally come up during the interview, and sufficient time was left to address them, they were addressed at the end of the interview. At this point, the topics with the most relevance to the interviewees work were addressed first, as it was expected that the interviewees would have more experiences to share about those particular topics.

The final topic included in the interview guide is further challenges. This topic was included to allow interviewees to share any challenges they faced in their work on ML projects, that were not discussed discussed in the interview at that point. This allowed for the discussion of challenges that arose outside of the example project, or broader challenges that the interviewees have experienced across their work.

In some cases, all challenges had been discussed and the interview ended there. In other cases, the challenges brought up at this point led to further questions. The interviews ended when the interviewees had no more challenges left to discuss, or the allotted time for the interview ran out. After the interview ended, there was the opportunity for interviewees to ask questions or address any concerns with the interviewer. Additionally, interviewees were asked whether they would like to be presented with the results of the study. Finally, the interviewees were thanked for their time and effort in contributing to the study.

3.1.3 Participant Recruitment

The process of recruiting participants for this study consisted of two major steps: determining the inclusion criteria for participants, and the actual process of recruiting relevant participants.

Inclusion Criteria

In order to aid the selection of relevant participants, the inclusion criteria presented in table 3.1 were used.

Table 3.1: Summary of inclusion criteria

Inclusion Criteria	
Role	Must be self-identified designers working on the specified product type
Product type	Must have experience in working on products that contain or are powered by ML technologies.
Experience	Must have at least 1 (nearly) completed project or design case in the field of ML
Language	Must be comfortable with either English or Dutch.

First of all, participants were expected to fulfil some kind of designer role on ML products. This includes classic titles such as UX designers, product designers, or design managers, but also any other role doing the work of a designer on an ML product under a different title.

Additionally, participants were required to have experience in working on products that contain or are powered by machine learning. This includes special-purpose ML products, as well as ML-based features in larger products. For example, a designer working on a recommendation system that is part of a large news website is also sufficient.

Participants were expected to have at least one completed (or almost completed) project or design case working on an ML product. This was required to ensure participants had enough experience to discuss during the interview, as well as had an example project to guide the narrative of the interview.

Finally, as interviews were to be conducted in either English or Dutch, the participants had to be comfortable with speaking one of these two languages.

Recruitment Procedure

Participants were recruited through four channels: the AI x Design Slack community, LinkedIn, my personal network, and snowball sampling through participating designers. Figure 3.1 gives an overview of how many people were initially approached through each channel. In the AI x Design Slack community, potential participants were approached through private message after responding to a recruitment message posted in a public channel. On LinkedIn, potential participants were identified by mentions of ML (or AI) and design on their profile. These potential participants were then approached through cold messaging. Potential participants in my personal network and from snowball sampling were approached through email.

The potential participants that expressed interest after the initial approach received a comprehensive informational email. This email can be found in Appendix C. Aside from the goal and practical information surrounding the interview, the email also contained a criteria check for participants to make sure they fit the profile as stated in the previous section. From the 41 potential participants that were initially reached out to, 27 expressed interest and received this informational email.

If the participants fit the expected profile and were willing to participate, they would receive a scheduling email, which can be found in Appendix D. In this email, a date and time for the interview was set, and participants were asked to fill out the pre-interview survey. From the 27 participants that received the informational email, 21 participants received a scheduling email. Unfortunately, one interview was cancelled, resulting in a total of 20 interviews scheduled in the end.

A visualization of the recruitment process can be seen in Figure 3.1. This figure also shows the distribution of participants across the four channels.

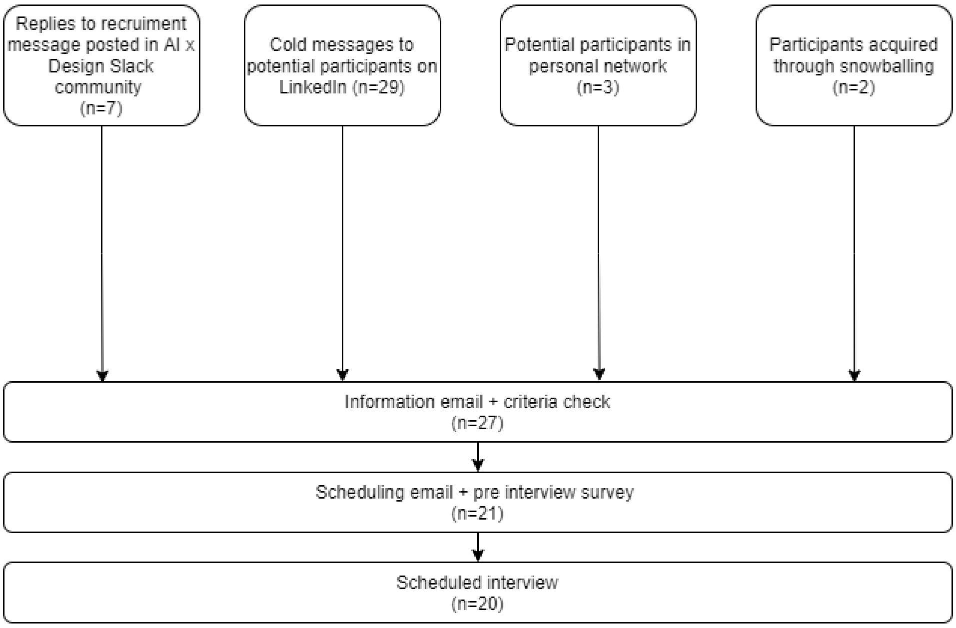

Figure 3.1: Overview of the recruitment process and sourcing of participants

3.2 Conducting Interviews

As participant recruitment took place, the first interviews were conducted simultaneously. All interviews were conducted remotely through Google Meet video conferencing software. Due to an unexpected interruption, the interview with P05 had to be ended early, but was completed the following day. Interviewees P08 and P09 requested to be interviewed together, as they had worked on their selected project together.

Interview audio was recorded using OBS Studio. The audio files were manually transcribed into text files. Intelligent verbatim transcription was used, as the purpose of the interviews was to extract the meaning that the interviewee was trying to convey, not the exact letter of what was said.

The following section describes the participants that were interviewed for this study.

3.2.1 Participants

A total of 20 participants was interviewed. Hennink, Kaiser, and Marconi (2017) suggests that between 16 and 24 participants are needed to reach meaning saturation in qualitative research, with a more exact estimate for a required amount of participants depending on the saturation parameters of the research. As the current study contains both values that indicate a small sample for saturation (purpose of capturing themes, homogeneous population), as well as values that indicate a large sample for saturation (conceptual codes, emerging codebook), an intermediary sample size of 20 is expected to result in sufficient saturation.

Table 3.2 provides an overview of all participants. As seen in the table, the majority of participants completed an education in the field of design. However, other backgrounds were also represented with some participants having a background in engineering and computer science, among others. Aside from P02, all participants described themselves as fulfilling the role of a designer in their respective example project. P02 was a researcher performing design as part of their research, and therefore still met the inclusion criteria presented in section 1.3.1. Aside from P03, all participants had at least one year of experience in working with ML products or ML features in larger products, with most participants indicating either 1-2 or 2-4 years of experience. The size of the organisations that the participants work at, ranges all the way from smaller companies with less than a hundred employees, to large corporations with over 10,000 employees.

The last column of the table contains a short description of each participant's example project. This description was adapted from the interviewees own description of their example project during the interview. Though there is some overlap in types of products that the interviewees selected as their example project, the interviews still cover a diverse range of projects.

Table 3.2: Summary of participant background based on the pre-interview survey. The last column, interview example project, features a short description of the project described during the interview. This column is adapted from the interview transcripts.

ID	Education	Role	Exp	Org Size	Example Project
01	Design, HCI	Product designer	1-2 yrs	<100	Image recognition feature for UI tool suite
02	Design, AI	Researcher	2-4 yrs	1,000-10,000	Co-learning autonomous agent
03	Design	Experience designer	<1 yr	<100	Fan experience for visually impaired people
04	Design	Design manager	1-2 yrs	100-1,000	Recommender for streaming platform
05	Law	Product designer	2-4 yrs	>10,000	Prediction tool for programmatic advertising
06	Design	Experience designer	1-2 yrs	100-1,000	Computer vision object counting application
07	Design	Experience designer	2-4 yrs	>10,000	Image making experiences
08	Design	Data designer	2-4 yrs	>10,000	Intelligent sleep health monitoring

Continued on next page

Table 3.2 – continued from previous page

ID	Education	Role	Exp	Org Size	Example Project
09	Computer science	Data designer	2-4 yrs	>10,000	Intelligent sleep health monitoring
10	Design, business	UX designer	1-2 yrs	>10,000	Enterprise search
11	Mechanical engineering	Product designer	1-2 yrs	100-1,000	Object recognition for human animal interaction
12	Design	UX designer	8-10 yrs	1,000-10,000	Topic model generator
13	Engineering, education	Service designer	4-6 yrs	>10,000	Intelligent assistant for process optimization
14	Design	UX designer	1-2 yrs	<100	Topic model generator
15	Computer science, journalism	AI designer	2-4 yrs	>10,000	Recommender for streaming platform
16	Computer science, information science	UX designer	2-4 yrs	1,000-10,000	Machine learning platform
17	Design	Service designer	2-4 yrs	<100	Intelligent text summarization for public organisation
18	Design	Product designer	1-2 yrs	<100	Intelligent coaching chat bot
19	Design	Design manager	1-2 yrs	<100	Proactive diary for patients with respiratory issues
20	Design	Product designer	1-2 yr	>10,000	Prediction tool for financial markets

3.3 Analysis

The third and final phase of this study consists of the analysis of the data gathered in the previous phase. For the process of analyzing the data, thematic analysis was used. Thematic analysis is a method for qualitative data synthesis, in which recurring topics, or themes, in data are identified, analysed, and reported (Braun & Clarke, 2006). According to Braun and Clarke (2016), thematic analysis is a suitable method for research questions related to experiences or accounts of practice.

3.3.1 Thematic Analysis

There are various approaches to performing a thematic analysis. This study follows the reflexive thematic analysis approach. In reflective thematic analysis, themes are conceptualizations of meaning-based patterns, rather than summaries of data regarding certain topics. What sets reflective thematic analysis apart from other approaches is that the aim is not to create summaries of data that minimize the influence of the researcher. Instead, the aim is to create a coherent interpretation of the data, which is grounded in the data but created through the active involvement of the researcher (Braun, Clarke, Hayfield, & Terry, 2019).

Before starting the first phase of the process, it is important to consider exactly how thematic analysis will be performed. I followed an inductive approach to coding the data, meaning codes and themes were established iteratively from the data, rather than predetermining codes up front. Inductive coding allows me to focus on the experiences described by the participants, rather than approach the data from a predetermined set of assumptions. It must be noted that the analysis cannot be purely inductive, as it is guided somewhat by the concepts investigated in the literature review. However, the coding of data is primarily guided by what is present in the dataset, not the preconceived concepts from the literature review.

Themes can conceptualize both semantic and latent meaning. The research questions of this study partially refer to concepts that can be captured from the semantic, or surface level, of data. For example, participants' accounts of their design process, or descriptions of design methods used. However, as the focus of interest is also on the lived practice, or the experience of participants, it is also important to consider the latent meanings of data. In other words, the conceptual meaning that underlies the explicit patterns present. Therefore, the data was coded on both a semantic and latent level.

Braun and Clarke (2006) detail an approach to reflective thematic analysis that consists of six phases, a summary is provided in Table 3.3. A comprehensive overview of the process of each phase can be found in Braun and Clarke (2016).

Table 3.3: Summary of the phases of performing thematic analysis as described by Braun and Clarke (2006)

Phase	Name	Process
1	Familiarization with data	Read interview transcripts and take notes on initial impressions
2	Generating codes	Code interview transcripts. A process that takes place over multiple iterations
3	Constructing themes	Develop initial themes from patterns captured across codes
4	Revising themes	Check how well the themes support the data. Return to previous phases if necessary
5	Defining themes	Write theme descriptions that capture the centralizing concept of the theme
6	Reporting findings	Reporting the themes. Must be seen as a final stage of analysis where revisions can be made to the set of themes

I started by familiarizing myself with the data by reading each interview transcript at least ones. I coded the interview transcripts digitally using Nvivo. During the process of coding, I focused specifically on described processes and experiences. Coding took place iteratively through multiple passes, going over each interview transcript multiple times until the coding sufficiently captured the content of the dataset. As I was coding the transcripts, I would also review the codes I had at the time, and merge or split codes if necessary.

Once I was content with the set of codes, I started developing initial themes by grouping related codes. These themes were revised multiple times. Throughout the process of revising themes, I discussed my candidate themes with my supervisor. Additionally, I checked the themes against the dataset to see whether there was good fit with the data. I continued this process until a sufficient fit was achieved.

I then wrote a description for each theme. For a few themes, I was unable to write a sufficient description, due to the disjointed and weak nature of the themes. For this reason, these themes were dropped.

I performed an extra step of analysis on codes related to the design process and design methods, to see those codes not just as isolated events in the process, but to also consider the sequence in which they take place. In order to analyze the process-related codes in order, I mapped out coded process steps or methods in a spreadsheet. I then colour-coded similar steps, looking for patterns across participants. The comparison can be found in Appendix F.

I wrote an initial report on the themes. Based on the reporting and discussion with my supervisor, I decided to revise the themes once more into five complete and overarching themes. In the end, I reported on this final set of five themes: 'design process & methods', 'machine learning as a tool', 'advocate for the user', 'need for close collaboration', and 'real data, real complexities'. The report is presented in the next chapter and includes an in-depth description of each theme, as well as excerpts from the interview transcripts that represent the theme.

Chapter 4

Findings

This chapter discusses the results of the pre-interview survey, as well as the five themes as identified from the thematic analysis. In section 4.1, the the results of the concept familiarity questions that were part of the pre-interview survey are discussed. This is followed by separate sections detailing each of the five themes. Quotes lifted from the interview transcripts are included in the discussions of the themes. The wording and grammar of these quotes has not been edited.

.

4.1 Familiarity with Machine Learning

Participants were asked about their familiarity with certain concepts related to ML or design on a scale of 1 to 7. Due to the sample size of this study, no statistical testing was performed on this data. Definitive conclusions about the differences in familiarity to ML and design would require further research. However, this is not the focus of the current study. The results in the form of median for each concept as presented in Table 4.1 serve mainly to put further findings into context.

Table 4.1: Medians of the concept familiarity questions of the pre-interview survey, split by concepts related to ML, and concepts related to design.

ML Concept	**Median**	**SD**
Correlation and causation	4	1.50
Model training	5	1.30
Training set, validation set, and test set	5	1.52
Model validation	5	1.45
Design Concept		
User flows	7	0.81
Wireframes	7	1.39
Personas	7	0.95

As seen in the table, the medians for concepts related to design are all 7, compared to the medians for concepts related to ML, which range from 4 to 5. This suggests that for this study's sample, there is a difference in designer's familiarity between ML and design. As expected, designers seem to be more familiar with design.

4.2 Design Process & Methods

The first theme generated from the data was built around participants' descriptions of their work on a process level. All participants described their work on their selected example project in terms of broad process steps or stages, as well as specific methods they used throughout their process.

Throughout participants' descriptions of their process, four common stages in a particular order stood out. Figure 4.1 shows an overview of the four stages. The four stages that were commonly present in projects described by the participants were research, ideation, prototyping, and validation. The latter two stages were mentioned as taking place in small iterations.

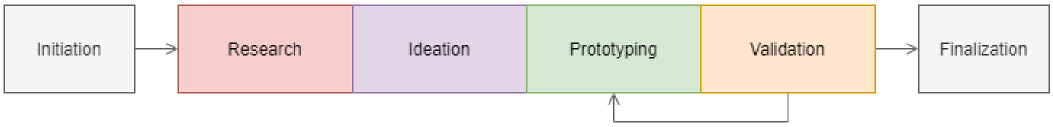

Figure 4.1: Graphical representation of the broadly the design process described by the participants. Also included are the themes that relate specifically to one or more of the stages. *Influence roadmap is included twice for readability, but refers to the same theme.

It must be noted that there were inherent differences in the process and methods described by the interviewees, purely due to when they started working on a project, or the nature of the project they described (e.g. strongly research based versus heavily visual design based). In some cases, one or more of these stages was not discussed. In other cases, more steps were mentioned that took place in between the four stages. The stages included in the visualization are the ones that were present in some form in the majority of the participants' accounts of their process.

Most participants described a single project cycle, meaning the process depicted in Figure 4.1 must be seen as part of a larger development project. Initialization and finalization were mentioned by the participants to identify the start and end of their project, but not necessarily as stages of the process. Additionally, the way projects were initiated or finalized differed a lot between participants. Therefore, they are included in the visual, but not described in detail in this section.

4.2.1 Research

A majority of participants mentioned starting their process with research. The research conducted often had more than one purpose.

> *"I felt it was really important, both because the research helps me understand the product, but it also let's me understand what's going on for other people, and from all the research I could make a plan about what needed to be fixed"* (P14)

The research practices described by the participants included classical user research, identifying pain points and user needs, or learning about the user's current work processes. At the same time, designers described using research to learn about ML in the product they are working on, or to learn about the domain their project takes place in.

> *"So we have to talk with the users, we have to do some qualitative research before anything to understand how people interact with this, or how people are aware of these systems that we are designing." (P04)*

In line with these goals, the research methods used were often qualitative. More than half of the participants mentioned conducting this type of research, such as interviews, contextual inquiries, and diary studies, during this stage of their projects.

4.2.2 Ideation

Nearly half of the participants described doing some form of ideation following research. This was mostly the case in projects where designers were working on features of products that did not exist yet.

> *"I created a work session in which data scientist and designers have to together understand or see the potential of AI in the solutions. And also for example we started to involve also the commercial person, because they started to sell a lot of artificial intelligence products that weren't necessarily using artificial intelligence technologies."* (p06)

There were quite a lot of differences between the types of ideation mentioned by the participants. Whereas some participants described a more loose form of ideation, for example in the form of workshops, others mentioned more structured approaches, such as requirements gathering.

> *"I also was making workshops with co-creation sessions, those being if you are a programmatic advertising specialist, what would you do? The general question about the whole workshop is around not me creating stuff and asking for your opinion, but for you as the user to create something that will change your life."* (P05)

In terms of specific methods recounted by the participants, ideation workshops were mentioned multiple times. Some participants described facilitating those workshops, involving users, data workers, or other relevant parties. Sometimes the workshops were very general brainstorming sessions, while others were more focused co-design sessions, where the collaborators were invited to come up with specific ideas. Overall, the participants characterized these ideation workshops as successful endeavours.

4.2.3 Prototyping

A large majority of participants detailed doing some form of prototyping. They described different levels of prototypes, ranging from simple interface mock ups without any interactivity, to fully functional, high fidelity prototypes.

"I would say it was a prototype that was evolving. So I had, I would save, from the original for evolved versions of, of the prototype, to really iterate and capture all of the requirements, to address the issues that have been surfaced." (P16)

The process of prototyping as detailed by the participants took place multiple times in short iterations. With the prototype evolving from those simple mock ups, to high-fidelity prototypes.

"So what I did really is actually draw visuals. So you know, some of the designer in me comes out, it's like, well, let's draw what's going on. So let's draw the user journey out." (P13)

Aside from completed prototypes, designers also described the use of methods that produce other visual artifacts during this stage. For example, user flows, personas, sketches, and wireframes were mentioned. Though these might not be prototypes in their own right, they are artifacts that support the designer in the process of creating prototypes.

4.2.4 Validation

Most participants described following an iteration of prototyping with the validation of said prototype.

"And after a few round of internal feedback and putting the setting right. I started taking it to the user and talking to them. Getting feedback on it, and that was very helpful, to see how they think about it. And then I figured, more gaps, some more requirements that we needed to work on." (P11)

The majority of prototypes were validated through user testing. However, several of participants mentioned other types of validation, such as checking the technical feasibility of their work with data workers, or gathering feedback on prototypes from clients. Similar to prototyping, designers describe validation as an iterative process.

"And then every time we would deliver something, I would run another set of user testing to learn about how are users using it, improvements." (P11)

Participants described a variety of research methods that they used during the validation stage. The research methods used here are a combination of qualitative and quantitative. In terms of the qualitative methods, some participants described doing A/B tests, and sending out questionnaires to users. For the quantitative methods, around half of the participants mentioned conducting user tests with a prototype.

4.3 Machine Learning as a Tool

This theme illustrates the way participants characterize ML. While in literature the concept of ML as a design material has emerged, participants discussed ML as one of the tools in their problem-solving toolkit, rather than as a material.

"I see machine learning and data science as a means to an ends. So there is a problem, and machine learning could be the solution to that problem." (P20)

Several participants described seeing ML as one of the tools that can potentially be used to solve a user problem. However, it is also described specifically that it is a tool that should only be used when it is an appropriate solution to user problem, in other words, when it is meaningful.

> *"This amount of effort, this amount of technology is as powerful as the use case behind it is. So I think that's the biggest factor here is to find the right use case for making AI work in a meaningful way, and a valuable way"* (P05)

The concept of meaningful ML use was present throughout the interviews. Several participants emphasized the importance of using ML in a meaningful way. They characterized meaningfulness in relation to the user: meaningful ML solutions solve user problems. They highlighted the importance of ML solving user problems, but also mentioned how it is important to be mindful that the use of ML does not cause any additional effort on the users' part.

> *"The danger is you just start a project, which is just a gimmick. And then everyone says, yes, it works, but what's the impact?"* (P17)

Several participants commented on how they often see a different approach to using ML, where ML is used without it directly solving a user problem. In these cases, ML is not a tool or material, but rather an end goal. A number of participants commented negatively on seeing ML as an end goal, recounting how adding ML to a product without considering whether it solves a user problem is the wrong approach.

The notion of ML as a tool in the designers toolkit was also apparent in the particular way participants talked about ML.

> *"And so understanding that, what had to go into the model, and then what came out of it, with pages and pages of different charts and interactive text things. It was understanding what went into it and what comes out."* (P14)

Rather than approaching it from a technical perspective, and talking about ML in technical terms, several participants talked about ML in terms of input and output. As far as their discussions on input went, they talked about needing to know and understand what data goes into the model. When talking about output, the participants user-centered perspective became apparent in the way they discussed it mostly in terms of the output that is presented to the user.

This is in line with the idea of ML as a tool to solve user problems with, as the input and output are much more important to a problem, than the inner workings of the technology are. Closely related to the way designers talk about ML, is the way they seem to understand the technology.

> *"I don't think all the designers need to really understand the technical processes. But you need at least to understand the machine learning potential, the basics, to understand which technology is best, suits the problem you have the best"* (P15)

Throughout the interviews, several participants mentioned how their technical understanding of ML is quite limited, they only need to know the fundamentals. However, some also emphasize how it is very important to have a good understanding of the boundaries of the technology. In terms of ideation, they need to know what is and is not possible in order to come up with new ideas. In terms of working with existing models, they need to know what the model can and cannot do to make decisions surrounding, for example, prototyping.

Rather than following the traditional, technical understanding of ML, designers have formed their own designerly understanding of the technology. This understanding is more concerned with the potential of ML. In other words, where ML can be used as a tool meaningfully. Aside from the way participants described understanding ML, they also discussed ML knowledge more broadly during the interviews. Their discussion of ML knowledge, and particularly the value of this knowledge, are in line with the notion of ML as a tool.

> *"I think it was important to understand the concepts like you said. I didn't work with them directly but I think it's an important aspect of designing it." (P14)*
>
> *"Actually, before this project this kind of topics weren't like very interesting to me. I was always aware of these kind of situations and all of the things that are happening in the world and it .. amazing for me but it wasn't like interesting. But after this project and shows me how flexible is working with this kind of features in our product that we used."* (P03)

Several participants specifically stated that they think it is important for designers working on ML products to have a basic level of ML knowledge. Additionally, quite a few participants shared that they are interested in learning more about ML. A couple mentioned specifically how they initially were not very interested in ML, but upon starting to work with the technology realized that they want to learn more about it.

> *"I think that my knowledge around the machine learning, data science, artificial intelligence is pretty basic. ... But the amount of knowledge that I have gives me an opportunity to think broader than just a designer about the constraints, the tech opportunities that we have on hand, something that we can really use out of the box without any additional effort involved."* (P05)

The concept of ML as a problem solving tool is especially present in participants' discussion of the value of ML knowledge. Several participants mentioned how they feel it makes them a better designer; they feel like they have an advantage compared to designers without ML knowledge. Additionally, a handful of participants discussed how having a good ML knowledge, in particular of the possibilities and limitations, helps them to see more opportunities. In other words, ML knowledge helps designers to see appropriate contexts where they can use ML as a tool, in a meaningful way.

Participants described not only the value of ML knowledge, but also the way they go about acquiring that knowledge.

> *"I have tried to take a machine learning course on Udacity, but that didn't work out"* (P02)
>
> *"Especially the data science stuff, I'm interested in that and actually my partner happens to be a data scientist so when I had questions it was really easy to be like, question about how that works."* (P14)

Rather than focusing on a formal education and technical specifics, participants describedr various informal ways that they use to improve their machine learning knowledge. Various participants described asking personal relationships to help them understand certain ML techniques, while others mentioned learning through personal projects. These approaches are in line with the designerly understanding of ML, and go to show how designers' interest does not seem to be primarily with the technical details of ML, but rather the possibilities that the technology offers.

4.4 Advocate for the User

This theme illustrates the main role participants described seeing for themselves in the ML projects they worked on. The primary role participants discussed is being an advocate for the user perspective. While the importance of this perspective is clear to the participants, many remarked how the main focus often seems to be on the technical perspective in the projects they described.

> *"And I feel in this case I'm more a translator, or more someone that is a mediator between the user needs and the engineer needs."* (P01)

Over half of the participants mentioned feeling like they have to advocate for a more human, or user-centered perspective. What is really interesting is how multiple participants discuss their position in relation to users and their data worker colleagues. They described seeing themselves as positioned in between these two groups, translating the language of user needs to the technical language of their colleagues, and vice versa. These participants describe themselves as a bridge, or integrator between the two groups.

What is also clear from the majority of participants is that they see a very minimal role for themselves on the technical side of things. Over half of the participants described very little involvement in the technical, data-centered side of the projects they worked on.

> *"But how it works in detail, to be honest I don't care about that very much. If it works, it works. I trust a data scientist or engineer to have taken care of it. How they did it, I don't really care about that."* (P20)

Multiple participants stated they did not feel they needed to be involved on the technical side of things, because it was not their expertise. Additionally, some participants explicitly reported feeling content with their minimal involvement in the more technical side of their projects, citing lack of interest or, again, their lack of expertise as the main reason.

Participants specifically mentioned having little involvement in data collection and the creation of algorithms. They mentioned how they rely on their data worker colleagues to take care of the technical side of the process. The minimal technical involvement of participants shows again how they are primarily concerned with the user perspective.

Though several participants recounted minimal involvement in data collection, others cited they were involved specifically to fulfil their main role: to represent the needs of the users.

> *"Well, I think I don't manage the data myself, but I'm in constant alignment with the people who do it. Because we need to know how to create the experience. How we made the product, the interaction, that they need to collect the data in a way that is the middle point between user needs and what the product needs."* (P19)

Multiple participants specifically discussed involvement on the account of being able to better understand the user than their technical colleagues might be able to. Additionally, some participants discussed being involved in data collection on a level of domain understanding. For example, as part of their research they interview users or subject matter experts and therefore help clarify the context that the data is gathered from. Once again, the focus remains primarily on the user perspective.

The need for a more user-centered perspective is clear from the accounts of several participants describing how design is often seen as an afterthought. Several participants detailed how they felt design was not a priority in the ML projects that they worked on, but merely a finishing touch to a technology that is already taken care of. Specific instances that demonstrate this perspective happen throughout all stages of the development process, on all levels.

> *"Other times you'll get a team that's like, UX is here to make the mock ups so the developers can put pretty pictures on top of our amazing technology."* (P10)

On a general level, several participants, most of them UX designers, described how their colleagues see their work as no more than creating user interfaces. The UX designers are reduced to UI designers, and are not expected to be involved with any other parts of the development process. The implication here is that the product is good to go, and only needs visual final touches from designers before being fully completed. .

> *"And there was one project where I was introduced to fix things after the engineering team had already implemented a big chunk of the product. I think what happens in that case is, the friction between you and the engineers, because they've invested in this, time. People don't see the vision, they don't see the whole piece. They probably just implemented something kinda quickly and as a designer you come in fresh and with your user research or whatever it is you're doing to understand what the root problems are, how you of course correct things."* (P12)

Another example of being seen as only visual designers was shared by multiple participants. They explained how they were brought on to do a mostly visual redesign of an already existing project. Designers are not involved from the start, but only when the technology is already up to a certain level. This once again illustrates how design is not a priority in these projects.

> *"The company, before I came on, had tried to launch that anyway, as a beta product. So they invited a handful of people to use it. The feedback was pretty negative. Everyone used it basically one time, and then was like, this is too hard I don't wanna do it."* (P14)

A handful of participants share several accounts of being brought on to projects that had bad user experience, or starting research and unearthing UX-related problems in the products. This goes to show that the lack of design intervention early on can have negative consequences.

> *"What was not given the opportunity here, is we were not able to use design thinking, or design strategy. And also, when I say that, I mean, the design structure around how do you do problem definition."* (P13)

A very specific problem that is a direct result to only being involved later on in projects is the lack of design thinking in the stage of problem setting. A few participants describe noticing the lack of design thinking at this point in the process, particularly the lack of difference between knowing the problem and finding the solution.

Quite literal evidence for the notion that design is not a priority is how multiple participants described having a hard time to get design prioritized over technical progression. They describe how it's challenging and frustrating to be on a different page then technical leadership or colleagues, and continuously having to advocate for the perspective of the user.

Conversely, a number of participants shared positive accounts where they were able to get the user perspective prioritized on the roadmap of a product. Participants describe how prioritization is primarily handled by product managers, and how they often have a close relationship with product management and are therefore able to help shape decisions on prioritization. Their work in getting design prioritized is another way in which the designers are able to represent the user in their projects.

> *"So it was like, to me it was great. I came in right as that conversation with the team was being identified. And they turned to UX, they turned to me do the prototyping to kind of... We have a pain point, we think we can solve it, we think it can be useful for our customers. But we really need UX to do prototyping, to visualize it, to see it, before we really invest in the engineering piece. So it was really perfect."* (P12)

Nearly half of participants commented positively on being involved in early on in projects, either at the initiation, or early enough that there is still some leeway in the direction of the project. They feel fortunate to be involved early on as it allows them to bring in a user-centered perspective from the start. More broadly it allows them to steer the project in a design driven direction. Additionally, being involved early on also helps the designers to grasp a more complete grasp of the ML aspects of the product.

> *"But what will often happen is, there is a very defined process for how features arrive in product, and that is something that has been followed for a number of years, and is something where design contributes at various stages of that and offers everything that they have to input."* (P07)

> *"And, for example, the process would be, if I see that the user is taking too much time answering something on the bot. And this is something that I will bring to my boss and say, you know, maybe we need to check the content there"* (P18)

As mentioned, participants describe having some influence on the roadmap through their close relationships with product managers. They are able to influence the roadmap especially when they have conducted research and have the data to back up their suggestions for next steps.

4.5 Need for Close Collaboration

This theme describes the efforts required to create a close collaboration between designers and data workers. Several participants explicitly acknowledged the importance of this close collaboration, and how it's a two way street: both parties need to put in the effort. Throughout the interviews, the participants described various ways that they try and make collaboration with data workers easier.

"One thing I've been having at work for the past couple of years is to train, to educate people about, the power of design, beyond just colours and how things look and lay out. But try to communicate with them, the power of design thinking, the power of for example the double diamond model." (P11)

"My colleague, who's the data scientist and machine learning specialist, she'll actually kind of, if we're talking about a specific, I don't know, machine learning algorithm, she'll actually explain that to us. And so this is what happens, this is what I'm doing. This is good, the outputs we get and the inputs we need. So that we get a general sense of what's going to be happening there, and how that's going to influence our work as well." (P18)

Some of participants discussed using education to improve collaboration. They describe both the education of their colleagues, as well as their own education. They detail trying to educate their data worker colleagues on the value of design, for example through workshops, or by joining meetings that are not design-related, but that they feel could benefit from a design perspective. Designers also try to educate themselves on ML, in order to better understand their colleagues' work. They close the gap between themselves and data workers even further by using their colleagues to improve their knowledge. Over half of the participants described asking data worker colleagues to explain ML concepts to them. Multiple participants also mentioned using data workers to fully understand how the products that they work on work, especially when they just start working on the product.

Participants discussed issues in communication between themselves and data workers that could be addressed specifically by improving ML knowledge. Participants framed the struggle in communication in terms of languages; they feel they speak a different language from their data workers colleagues.

"Because it was hard to communicate with [the developer]. When he was trying to work with this kind of features and he was trying to explain it to us. But none of us would understand it. So the lack of expertise, not expertise, the lack of knowledge that we had about this kind of project was wasting a lot of our time during the project." (P03)

"And I see that it's very important to a UX working a lot with IT, to know all that for example, the glossary. I think that's absolutely necessary. If you don't have the glossary, you can not understand, what are they doing and what are they talking about." (P06)

Participants describe how they feel that they need to learn the data workers' language in order to fully understand what's going on. Specifically meetings with data workers were mentioned as a point of struggle, with designers being unable to participate in the conversation. Some of the participants talked about communication in close relation to understanding of ML. They need to know common terms and concepts in order to follow and add something to conversations.

"I think another thing that we all realized in the team, that having tangible screens, and step by step flows, brings the team and common understanding to the point, we don't go off the engine discussing something that everybody might interpret it differently when reading the text." (P16)

Rather than tying to speak the language of the data workers, a handful of participants brought up a reverse approach: using designers' visual skills to improve communication. Some describe creating visuals specifically with the purpose of communication, while others discuss using common visual artifacts, such as mock ups or design flows, to support conversations between designers and data workers. It is interesting too see how designers use their unique visual skills to help them navigate the ML landscape.

Aside from fostering mutual understanding of each others fields, and in turn improving communication, participants also describe efforts to close the gap by actively involving data workers in the design process.

> *"I also did a lot of co-design sessions with the engineering team. So I would come with a general idea of what I thought it looked like, and then to really extract out of them what in their minds it needed to look like"* (P10)

A few participants mentioned facilitating co-design sessions with data workers. The sessions have a dual nature, they help to come up with new ideas to improve the product, but they also help to the mutual understanding and communication. A couple of participants also described how they would bounce ideas of data workers, or vice versa, in general conversations.

> *"Because as a designer you can make very different suggestions about how it would be better to create an AI in this field, but you are not clear at most times about the feasibility check. So how real is this model to become available for the cost, for the budget that we have, is it really something that we can build?"* (P05)

Participants also actively involve data workers in the design process by using them to check the feasibility of their ideas. Nearly half of the participants mentioned needing data worker colleagues to check whether their ideas are possible, as they lack the technical judgement required to decide this for themselves. Sometimes this meant getting a general understanding of what is possible, at other times bringing up a specific idea and checking whether that would be feasible.

4.6 Real Data, Real Complexities

This theme illustrates how the data-centered nature of ML puts significant limits on the prototyping and validation stages of the design process. Participants repeatedly brought up how the close relationship between ML and data causes complexities especially in those stages.

Participants emphasize how prototyping is especially well suited to determine the quality of a ML product beyond just it's technical accuracy. However, at the same time they acknowledge that data is such a substantial part of the experience of ML products, that does need to be taken into account.

> *"So AI, you're probably quite familiar with the terminology, they have basically their training set and their testing set. But there are also a lot of external factors eventually in the user experience to take into account. Like how does the user see it, how is it being delivered to them. And that's where my prototyping comes in."*(P08)

Several participants brought up the need to prototype ML products. This is another instance of the designer representing the human perspective, as prototyping, and in particular testing those prototypes, covers the entire user experience, going beyond just technical accuracy. Some designers address specifically how having realistic interactions in prototypes is especially essential to ML products, as interactivity is key to how the user will eventually experience the product. They express how they are able to learn a lot about how the user interacts with the product and unearth real challenges by just having potential users interact with realistic prototypes for a couple of minutes.

> *"So when you just sort of click a text input and it gives you all these outputs, it doesn't really connect with like, the user didn't really have to think about what they're looking for and then if the results are what they expected to get when they had typed in whatever. Since it was all kind of hypothetical."* (P14)

Aside from making prototypes realistic in terms of interactivity, several participants also indicated that it is best to create prototypes with data that is as close to real data as possible. Using fake data, such as dummy data, can make the prototyping experience less realistic. A couple of participants described how they felt their validation results were negatively impacted by the use of fake data.

Although the participants expressed a preference for using real data in prototypes, this is not always an option.

> *"So, a lot of times developing these intelligent systems, we don't have the data available, we don't have a clue of what the data looks, so we need to fill in the gaps in some way."* (P09)
>
> *"So I can only test with fake data, and that really limits me in getting accurate feedback"* (P20)

Some participants describe the general lack of availability of data throughout the entire design process, which especially impacts prototyping and testing. Other describe particular issues with data regulations related to security and compliance, which limits the use of data in prototypes and especially in testing with external users. Both issues limit designers particularly in how much realism they can add to prototypes and in turn validation.

Only a few participants described prototypes that use the same data that the final product would use. Due to the issues surrounding data availability and regulations, many participants described alternative solutions. Depending on the data that is available for use in prototyping and validation, they described different levels of realism in terms of the data used in their prototypes.

> *"So what I did is I actually used a data set that the data science team had used. So I used a real data set, it was definitely not... I mean it was dummy data in the sense that there was nothing you could drill in on. But I made sure that in the mocks I had, the things that were being derived were actually things that were being derived."* (P12)

A common alternative to real data mentioned by several participants is using one specific example from real data throughout the prototype. Though this approach does of course come at the cost of interactivity, participants were positive about it in consideration of its constraints.

"We found some website that was like, top 100 movies of all time, so then I went to Wikipedia, downloaded all of their HTML files, which of course comes with PNGs and image files and other file types. So that allowed us to test not only the search capabilities, but also the sync capabilities, to test the report, the settings page. ... So that gave us a bit of a little testing ground to work with some sort of real data, even though obviously you're not going to have movies in there." (P10)

Another approach used by several participants who had no access to real data was to use dummy data. This data is similar to real data in structure, but not grounded in actual data that will used in the product. While participants were positive about certain aspects of this approach, such as the ability to test certain interactions such as producing reports, they were concerned with the lack of realism impacting testing results.

"We took some users from the office, this person always arrives at 7 am. We're going to put a picture of them, we're going to open the door, and we leave a message. Welcome message, no message, no pictures. We started that way. No technology." (P06)

A handful of participants indicated that they used a Wizard of Oz approach to prototyping. They were positive about being able to quickly test mostly small parts of an interaction, without needing help from technical colleagues. They did mention downsides of the approach as well: a real machine learning model can act differently than your Wizard of Oz protocol, so you cannot be sure that the experience your delivering is fully realistic.

What becomes apparent from the way participants talk about data in prototypes is that the type of data used in prototypes is not a design choice. The general consensus is to aim for realism both in terms of the data included. However, the exact choice of data can also depend on the nature of the prototype and validation. Due to the issues surrounding data availability or data regulations, designers are limited in how they can take those considerations into account. As illustrated by the approaches described by the participants, it may be necessary to make a trade off between data and interactivity realism.

Chapter 5

Discussion

This chapter discusses important aspects of the findings presented in this study. First, the findings will be interpreted in relation to the wider context of ML design practice. Then, implications of the findings for research and design practice are discussed. Finally, limitations of the current study are provided.

5.1 Data Complicates Design

The theme 'real data, real complexities' highlights a challenge that is caused by MLs dependency on data. This theme demonstrates how the data-centered nature of ML limits the creation of prototypes and the user validation of those prototypes. The issues with prototyping ML projects present in these theme are in line with earlier findings on ML design practice by Browne and Diego (2019).

The theme 'real data, real complexities' establishes a two-sided discussion on prototyping ML products; on one hand designers realize the importance of using prototypes for user testing beyond technical accuracy, but at the same time they also acknowledge the importance of data and technical accuracy to the overall experience. The latter issue has been pointed out before by Bratteteig and Verne (2018), who state that parts of experience such as the statistical approach used, or the training data set, can be hard to incorporate into prototypes.

Contrary to past research, that poses that the fidelity of prototypes does not impact usability results (Walker, Takayama, & Landay, 2002), 'real data, real complexities' shows that designers need highly realistic prototypes in order to get accurate results. The descriptions of prototyping ML in this theme are in line with Yang (2018), who reported that the limitations of ML prototyping prevent designers from fully uncovering the validity and usability of ML products in their real context before release. Though designers acknowledge the need for realistic prototypes, they also describe issues in accessing and using real in prototypes, mainly due to the unavailability of data, or strict regulations concerning data security. Dove et al. (2017) and Yang, Scuito, et al. (2018) also describe issues surrounding data availability in prototyping for ML design practice.

The theme 'advocate for the user' also emphasizes a challenge of ML design practice that is related to MLs dependency on data. The data-centered culture in the field of ML that stems from this dependency makes it difficult for designers to bring a user-centered perspective to bear onto ML projects. In the theme, designers describe experiencing design being an afterthought rather than a priority. Past research also supports the notion brought up in this theme; Fass et al. (2019) report similar findings, describing how design is seen as an end goal in the development process of ML products.

The data-centered culture in the field of ML is also apparent in the way designers describe their collaboration with data workers in the theme 'need for close collaboration. This theme illustrates that designers are aware of the need for a close collaboration with data workers in order to succeed on ML projects. This sentiment is echoed by Girardin and Lathia (2017), who state that collaboration between designers and data workers is a condition for successful practice in the field of ML. The current study also shows that there are still issues in collaboration between the two parties, this is in line with the findings by Yang, Scuito, et al. (2018).

The themes 'advocate for the user' and 'need for close collaboration' both highlight phenomena that are not unique to ML design. The lack of design prioritization is also present in non-ML projects. Past research suggests that developers see designers' role solely as the creation of UIs to put on top of technology (Xu, 2012). The close collaboration and corresponding friction between designers and data workers also stretch beyond the field of ML to software development more broadly (Ferreira, Sharp, & Robinson, 2011). However, both issues can be aggravated by the data-centered culture in the field of ML.

5.2 Meaningful Use of Machine Learning

Previous authors have repeatedly brought up the concept of ML as a design material (Yang, Scuito, et al., 2018; Dove et al., 2017). Contrarily, the theme 'ML as a tool', reveals that designers working with ML see it as a tool, rather than a material. This study is the first one to introduce this concept. This implies the presence of a gap between of what Gray et al. (2014) calls the projected practice and lived practice.

The underlying concept of ML is a tool is that ML should be applied meaningfully. Although the concept of ML as tool is novel, the underlying view is present throughout past research. Yang (2017) state how it is one of the primary roles of UX designers in ML projects to promote the purposeful use of ML to solve user problems. Ceconello, Spallazzo, and Sciannamè (2019) argue that through design, the field of ML can be moved from complicated systems that do not address user problems, towards design innovation in the form of meaningful ML products. Luciani et al. (2018) affirm that innovation in the field of ML is fully dependent on the meaningful human use, rather than technical continuation.

Throughout the theme 'ML as a tool', it becomes clear that designers define meaningful ML use in relation to the user: ML is meaningful if it addresses a user problem. The importance of the user perspective is clear to ML designers. This is also apparent in another theme: 'advocate for the user'. In this theme, designers describe how the main role they see for themselves in ML design projects is to be an advocate for the users. This corresponds to previous research on design practice: Dove et al. (2017) also report on designers' wish to make work in the ML field more user-centered. However, the same study also argues that UX designers should also consider a statistical perspective, and more importantly the interplay between these two perspectives. In contrast, the theme 'advocate for the user' suggests that designers feel they do not need to be involved in most of the technical aspects of ML projects. Only data collection was mentioned as a technical aspect that designers felt they could contribute to, but again their main contribution would be to represent the needs of the user better than data workers are able to.

5.3 No Need for Machine Learning-specific Design Methods

Past research suggests that common design processes and methods such as participatory design and journey maps are not well suited to design with ML (Bratteteig & Verne, 2018; Yang, Banovic, & Zimmerman, 2018). More generally, Van Allen (2017) highlights the need to create ML-specific design methods to successfully design ML products.

In contrast, the first theme introduced in this study, 'design process & methods', suggests that there are few differences between ML projects and non-ML projects in terms of design process and methods. In the theme, four stages of the ML design process are introduced. These four stages show overlap with existing, popular approaches to design processes, such as the User Centered Design Process (International Organization for Standardization (ISO), 2019), or the British Design Council Double Diamond Process (Design Council, 2005).

Furthermore, the theme also describes the use of various qualitative and quantitative methods throughout ML design processes. The use of both types of methods accords with the previous findings on ML design practice, as does the use of traditional UX methods such as user studies and usability testing (Yang, Scuito, et al., 2018). The specific methods mentioned in the theme also match design methods described in general design practice studies (e.g Gray, 2016), design method taxonomies (e.g.Alves & Jardim Nunes, 2013), and reference books (e.g.van Boeijen et al., 2014). This further supports the notion that on a process and method level, ML design does not differ much from non-ML design.

5.4 Implications

The themes 'design process & methods' and 'ML as a tool' present implications for research on ML design practice. Several authors have previously suggested that the ML design practice research community should work on the creation of design methods specific to ML. The theme design process methods suggests that designers are doing fine with methods that are not created specifically for ML. Rather than focusing on the creation of methods specific to ML, research on ML design practice could instead address the challenges of working with ML that designers are experiencing.

The concept of ML as a tool being present in this study but not in previous ML design practice research implies the presence of a gap between the academic understanding of practice, and the lived practice. As the presence of such a gap prevents researchers from creating solutions that improve the practice, it is vital that the ML design practice research community works towards closing the gap first.

The remaining three themes, 'advocate for the user', 'need for close collaboration', and 'real data, real complexities' have clear implications on practice. All three themes highlight challenges that are aggravated by MLs dependency on data, which is inherent to the technology and therefore not subject to change. Rather than eradicating the challenges altogether, designers will have to learn how to work around them. However, working around these challenges does require additional time and effort on the designers' part, it may be that this is preventing the designers from leveraging ML to its fullest potential.

The findings of the current study present one avenue that could be pursued to address the lack of design prioritization in practice: education on the value of design. By educating their colleagues and especially leadership roles on the value of design, designers will no longer have to be the only ones advocating for the user. This may help them devote more time to innovation.

The difficulties in collaboration between designers and data workers could also be alleviated through education. By designers educating themselves on the technical aspects of ML, and data workers educating themselves on design, the two groups might be able to speak the same language and work together more easily.

The issues experienced when prototyping and validating ML products presents an extra problem: designers cannot fully validate their ideas before release. This may slow down or even inhibit innovation from taking place. A potential solution could be the creation of prototyping tools specific to ML, that consider the need for realistic data and data-based interactions. Additionally, the findings of this thesis show that designers already use a plethora of approaches to prototyping and validating ML products. Design practitioners could benefit from sharing their ideas among themselves, for example in ML design communities.

5.5 Limitations

There are a number of methodological limitations to the present research that should be considered. First, the retrospective interviews that were used as the main method of data collection are subjects to biases. For example, my bias as an interviewer might have caused me to unconsciously formulate questions as to receive answers that confirm to my views. I have tried to mitigate this as much as possible by creating an extensive interview guide with neutral probes up front.

Second, selective sampling was used to recruit participants. Participants' presence in AI design interest groups, or mentions of AI or ML on their LinkedIn profiles, could be indicative of a heightened interest in the subject. The sample might not be representative of designers that do work with ML, but are not specifically interested in the technology. This points towards a larger limitation. The results are based on the experiences of a small sample of design practitioners. The experiences shared might not be universal for all practitioners. Additionally, the perspectives of data workers and associated roles are not represented in this study. In order to understand the practice of designers working on ML more completely, more studies with diverse samples should be performed.

Third, the results of the interviews were analysed by means of thematic analysis. The thematic analysis was performed by a single researcher. Therefore, the results of the analysis might be impacted by researcher bias. In order to combat this, the analysis was continuously revised based on discussions in order to minimize this bias. It should be considered that it is also the strength of reflective thematic analysis that a coherent representation of the data is created through active involvement of the researcher. My hope is that researcher bias beyond this active involvement has been mitigated enough to create this coherent representation.

Chapter 6

Conclusion

This chapter will summarize answers to the research questions of this study. First, the subquestions are answered. Then, the main research question is answered. Interesting opporunities for further research are provided. The conclusion is resolved with a summary of this thesis.

6.1 Challenges of Working with Machine Learning

The purpose of this research question was to both confirm or contradict the presence of the eight previously identified challenges, as well as idenfity new challenges beyond the existing set. The current study confirms the presence of the challenge 'data-centered culture'. The issues with the data centered culture prevent designers from bringing a user-centered perspective into ML projects, even though they consider this their main purpose. This study also confirms the presence of the challenge 'prototyping'. The main cause of this challenge is MLs dependency on data. Designers use various approaches to prototyping with different levels of realism.

The findings of this study partially confirm the presence of the challenge 'lack of technical knowledge'. While the findings indicate that designers indeed have limited technical knowledge, it also shows that they are not necessarily experiencing direct problems because of it. Designers mostly experience negative results of their lack of technical knowledge indirectly as it causes them to struggle to communicate well with data workers. The lack of technical knowledge seems to be an aspect of a broader, previously unidentified challenge: collaboration with data workers.

This study also identifies a second new challenge: meaningful ML use. Designers describe how they see ML as a tool, that they can ideally choose to use when it can be meaningful. However they also how ML is currently often seen a as a goal. Instead of being used meaningfully, ML is seen as a must-have, or a gimmick.

This study somewhat contradicts the presence of the challenge 'design methods and tools'. Designers describe using common methods and do not mention struggling with the use those methods.

The current study does not provide consistent evidence for the presence of the following previously identified challenges: 'black-box nature of ML', 'system changes over time', 'personalized solutions', and 'incorrect mental models'.

6.2 Facing the Challenges

Throughout the five themes, designers describe approaches for how they deal with most of the challenges. Only for the new challenge 'meaningful ML use', participants do not describe a clear approaches to how they face this challenge.

Although the way designers deal with a challenge is of course different for every single one, there are overarching . Designers describe how addressing challenges in ML design practice really needs to be a team effort. They often describe working closely with product managers, data workers, or developers in order to face design challenges.

Designers also use their design skills to work around the challenges of designing with ML. They use their skills as a designer to come up with creative solutions to problems, this is for example clearly present in the creative approaches designers take towards prototyping. Additionally, designers use their visual skills to create visuals that support them in facing a challenge.

6.3 Familiarization with Machine Learning

Designers mention one main strategy that they use to familiarize themselves with ML: they learn about the possibilities and constraints of ML technology. Although designers have limited technical knowledge, they acknowledge the importance of knowing the boundaries of the technology. Only when they have a good understanding of what is and isn't possible with ML, they are able to see where and how ML can make a meaningful impact.

6.4 The Lived Practice of Designers Working with Machine Learning

The five themes presented in this research highlight different characteristics of designing with ML that make it a unique practice. Rather than working with ML as a design material, designers view ML as a tool that should only be used meaningfully; to solve user problems. However, in the field ML is still often seen as a goal or gimmick instead. The lack of focus on design is present throughout the development process, even though designers try to actively advocate for the user perspective as much as possible.

MLs strong dependency on data aggravates common issues for designers in software development, and also introduces entirely new challenges. A common challenge that is aggravated in ML projects is the collaboration with data workers. Designers have to devote more effort into collaborating smoothly. A more unique problem is that MLs data dependency combined with data limitations makes it hard for designers to create realistic prototypes and perform realistic user tests. This impacts designers' ability to determine the quality of products before release.

Though the practice of designers working with ML is unique, there are also similarities to non-ML design. Especially in terms of the design process and methods used, working with ML is not so different from working with other design materials. To conclude, designers who work with ML face a uniquely challenging practice. A lot of what makes ML so challenging to work with is inherent to the nature of the technology, and therefore something designers will have to learn to work with.

6.5 Future Work

The current study identifies various areas that could benefit from further research. First, this study identified a gap between projected practice and lived practice, especially concerning the way designers view ML as a tool rather than a material. Further work on the lived practice of designers should be conducted to minimize this gap. It would be particularly interesting to further investigate the concept of ML as a tool, as this study is the first one to introduce the concept. It would be interesting to see if other studies are able to identify a similar sentiment across the population of designers working with ML.

Second, though data collection was brought up as a specific part of the ML development process that designers can contribute to in the current study, there does not seem to be further evidence for this idea. The way designers can be involved in and positively influence data collection processes is an interesting avenue for further research.

Third, both the current study and previous studies show that designers collaborate closely with data workers. However, as of yet no work has been done specifically investigating this collaboration in the field of ML. An interesting direction for further research to pursue would be to investigate this collaboration, both from the perspective of designers and data workers.

Finally, though prototyping and validation is an important part of the design process, and various studies have identified that it is a problem area, no work has been done on the challenge specifically as of yet. As prototyping and validation are phases of the design process that depend particularly on the use of tools and methods, it would be interesting to pursue the creation of those tools specifically for ML design.

6.6 Summary

Although ML undergoing rapid technological innovation, designers seem unable to leverage the material to its full potential. This thesis presents a interview study conducted with design practitioners who have experience working with ML to characterize their practice. Through thematic analysis of the interview transcripts, I constructed the following five themes: 'design process & methods', 'ML as a tool', 'advocate for the user', 'need for close collaboration', and 'real data, real complexities'. The findings show that while ML design practice has similarities to non-ML design practice, there are challenges that make designing with ML uniquely complicated. I expand on these findings to determine what these challenges are, and how designers face them in their daily practice.

References

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138–52160. doi: 10.1109/ACCESS.2018.2870052

Alpaydin, E. (2014). *Introduction to Machine Learning* (3rd ed.). MIT Press.

Alves, R., & Jardim Nunes, N. (2013). Towards a taxonomy of service design methods and tools. In *International conference of exploring services science* (pp. 215–229). doi: 10.1007/978-3-642-36356-6$_1$6

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... Fourney, A. (2019). Guidelines for Human-AI Interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1–13). ACM. doi: 10.1145/3290605.3300233

Baumer, E. P. (2017). Toward human-centered algorithm design. *Big Data and Society*, *4*(2). doi: 10.1177/2053951717718854

Bly, S., & Churchill, E. F. (1999). Design through matchmaking: technology in search of users. *interactions*, *6*(2), 23–31. doi: 10.1145/296165.296174

Bratteteig, T., & Verne, G. (2018). Does AI make PD obsolete? Exploring challenges from Artificial Intelligence to Participatory Design. *Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial*, *2*. doi: 10.1145/3210604.3210646

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. doi: 10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2016). Thematic Analysis. In E. Lyons & A. Coyle (Eds.), *Analysing qualitative data in psychology* (2nd ed., pp. 84–103). doi: 10.1007/978-981-10-2779-6$_1$03 $-$ 1

Braun, V., Clarke, V., Hayfield, N., & Terry, G. (2019). Thematic Analysis. In P. Liamputtong (Ed.), *Handbook of research methods in health social sciences* (1st ed., pp. 843–860). doi: 10.1007/978-981-10-5251-4$_1$03

Browne, J. T., & Diego, S. (2019). Wizard of Oz Prototyping for Machine Learning Experiences. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems* (pp. 1–6). doi: 10.1145/3290607.3312877

Brownlee, J. (2015). *Apple Finally Learns AI Is The New UI.* Retrieved 2020-06-01, from `https://www.fastcompany.com/3047199/apple-finally-learns-ai-is-the-new-ui`

Bryman, A. (2015). *Social Research Methods* (5th ed.). Oxford University Press. doi: 10.1007/978-3-319-65442-3$_4$

Buchenau, M., & Suri, J. F. (2000). Experience prototyping. *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, 424–433. doi: 10.4018/978-1-4666-4623-0.ch011

Buxton, B. (2007). *Sketching User Experiences: Getting the Design Right and the Right Design* (1st ed.). Morgan Kaufmann. doi: 10.1016/B978-0-12-374037-3.X5043-3

Ceconello, M., Spallazzo, D., & Sciannamè, M. (2019). Design and AI: prospects for dialogue. *Convergências - Revista de Investigação e Ensino das Artes*, *23*. Retrieved from http://convergencias.esart.ipcb.pt/?p=article&id=350

Colborne, G. (2016). *Interaction design in the age of algorithms.* Retrieved 2020-05-29, from https://www.cxpartners.co.uk/our-thinking/interaction-design-in-the-age-of-algorithms/

Coyne, R. (1995). *Designing Information Technology in the Postmodern Age.* MIT Press.

Coyne, R. (2001). *Technoromanticism: Digital Narrative, Holism, and the Romance of the Real* (1st ed.). MIT Press. doi: 10.1080/01972240252818243

Coyne, R., & Snodgrass, A. (1991). Is designing mysterious? challenging the dual knowledge thesis. *Design Studies*, *12*(3), 124–131. doi: 10.1016/0142-694X(91)90020-W

Daly, S. R., Adams, R. S., & Bodner, A. M. (2012). What does it mean to design? A qualitative investigation of design professionals' experiences. *Journal of Engineering Education*, *101*(2), 187–219. doi: 10.1002/j.2168-9830.2012.tb00048.x

Design Council. (2005). *What is the framework for innovation? Design Council's evolved Double Diamond.* Retrieved from https://www.designcouncil.org.uk/news-opinion/what-framework-innovation-design-councils-evolved-double-diamond

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, *55*(10), 78–87. doi: 10.1145/2347736.2347755

Dorst, K., & Dijkhuis, J. (1995). Comparing paradigms for describing design activity. *Design Studies*, *16*(2), 261–274. doi: 10.1016/0142-694X(94)00012-3

Dove, G., & Fayard, A.-L. (2020). Monsters, Metaphors, and Machine Learning. In *Proceedings of the 2020 chi conference on human factors in computing systems* (pp. 1–17). doi: 10.1145/3313831.3376275

Dove, G., Halskov, K., Forlizzi, J., & Zimmerman, J. (2017). UX design innovation: Challenges for working with machine learning as a design material. *Conference on Human Factors in Computing Systems*, 278–288. doi: 10.1145/3025453.3025739

Ehn, P., & Kyng, M. (1991). Cardboard computers: Mocking-it-up or hands-on the future. In J. Greenbaum & M. Kyng (Eds.), *Design at work: Cooperative design of computer systems* (pp. 169–195). Lawrence Erlbaum Associates.

Fallman, D. (2003). Design-oriented Human-Computer Interaction. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 225–232). doi: https://doi.org/10.1145/642611.642652

Fass, J., College, R., & Groves, E. (2019). Making Machine Learning Tangible for UX Designers. *CHI' 19 Extended Abstracts*.

Ferreira, J., Sharp, H., & Robinson, H. (2011, aug). User experience design and agile development: Managing cooperation through articulation work. *Software - Practice and Experience*, *41*(9), 963–974. doi: 10.1002/spe.1012

Fiebrink, R., & Cook, P. (2010). The Wekinator: a system for real-time, interactive machine learning in music. In *Proceedings of the eleventh international society for music information retrieval conference.*

Galle, P. (2009). *Philosophy of design: an introduction.* Retrieved from `https://kadk.dk/en/cephad/philosophy-design-introduction`

Girardin, F., & Lathia, N. (2017). When user experience designers partner with data scientists. *AAAI Spring Symposium - Technical Report*, *SS-17-01* -, 376–381.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.* MIT Press.

Google AI. (n.d.). *Learn From ML Experts at Google.* Retrieved 2020-06-04, from `https://ai.google/education/`

Google AI. (2017). *Teachable Machine Learning.* Retrieved 2020-05-29, from `https://teachablemachine.withgoogle.com/`

Gray, C. M. (2016). "It's more of a mindset than a method": UX practitioners' conception of design methods. In *Conference on human factors in computing systems* (pp. 4044–4055). doi: 10.1145/2858036.2858410

Gray, C. M., Stolterman, E., & Siegel, M. A. (2014). Reprioritizing the relationship between HCI research and practice: Bubble-up and trickle-down effects. In *Proceedings of the conference on designing interactive systems: Processes, practices, methods, and techniques, dis* (pp. 725–734). doi: 10.1145/2598510.2598595

Harvard Business Review Analytic Services. (2019). *Lessons from the Front Lines of Business* (Tech. Rep.).

Hebron, P. (2016). *Machine Learning for Designers.* O'Reilly Media. doi: 10.1007/978-3-642-28661-2

Hennink, M. M., Kaiser, B. N., & Marconi, V. C. (2017). Code Saturation Versus Meaning Saturation: How Many Interviews Are Enough? *Qualitative Health Research*, *27*(4), 591–608. doi: 10.1177/1049732316665344

Hodgkinson, G. (2004). Teaching Design Thinking as Practice. In *Proceedings of edmedia 2013–world conference on educational media and technology* (pp. 1520–1524).

Holmquist, L. E. (2017). Intelligence on tap. *Interactions*, *24*(4), 28–33. doi: 10.1145/3085571

International Organization for Standardization (ISO). (2019). *Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems (ISO 9241-210:2019).*

Jones, J. C. (1992). *Design methods* (2nd ed.). New York: Wiley & Sons.

Kim, M., Zimmermann, T., DeLine, R., & Begel, A. (2016). The emerging role of data scientists on software development teams. *Proceedings - International Conference on Software Engineering*, -, 96–107. doi: 10.1145/2884781.2884783

Kirsch, A. (2017). *Explain to whom? Putting the User in the Center of Explainable AI* (Tech. Rep.). Retrieved from `https://hal.archives-ouvertes.fr/hal-01845135`

Kocielnik, R., Amershi, S., & Bennett, P. N. (2019). Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems. In *Proceedings of the 2019 CHI conference on human factors in computing systems.* Retrieved from `https://doi.org/10.1145/3290605.3300641` doi: 10.1145/3290605.3300641

Kuang, C. (2013). *Why a New Golden Age for UI Design Is Around the Corner.* Retrieved 2020-06-01, from `https://www.wired.com/2013/08/design-and-the-digital-world/`

Kuniavsky, M.; Churchill, E.; and Steenson, M. W. . (2017). *designing the user experience of machine-learning systems.pdf.*

Landin, H. (2005). Fragile and Magical-Materiality of Computational Technology as Design Material. In *Proceedings of the 4th decennial conference on critical computing: between sense and sensibility* (pp. 117–120).

Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research Methods in Human-Computer Interaction* (2nd ed.). Morgan Kaufmann.

Lim, B. Y., & Dey, A. K. (2010). Toolkit to support intelligibility in context-aware applications. In *Ubicomp'10 - proceedings of the 2010 ACM conference on ubiquitous computing* (pp. 13–22). doi: 10.1145/1864349.1864353

Lipton, Z. (2018). The mythos of model interpretability. *Communications of the ACM*, *61*(10), 36–43. doi: 10.1145/3281635

Louridas, P. (1999). Design as bricolage: Anthropology meets design thinking. *Design Studies*, *20*(6), 517–535. doi: 10.1016/s0142-694x(98)00044-1

Love, T. (2000). Philosophy of design: A meta-theoretical structure for design theory. *Design Studies*, *21*(3), 293–313. doi: 10.1016/s0142-694x(99)00012-5

Lowgren, J. (1995). Applying design methodology to software development. In *Proceedings of the conference on designing interactive systems: Processes, practices, methods, and techniques, dis* (Vol. 23-25-Augu, pp. 87–95). doi: 10.1145/225434.225444

Lowgren, J., & Stolterman, E. (2007). *Thoughtful Interaction Design* (New ed.). MIT Press. doi: 10.7591/9781501739088-002

Luciani, D. T., Lindvall, M., & Löwgren, J. (2018). Machine learning as a design material: A curated collection of exemplars for visual interaction. In *Proceedings of norddesign: Design in the era of digitalization, norddesign 2018* (pp. 1–10).

Malsattar, N., Kihara, T., & Giaccardi, E. (2019). Designing and Prototyping from the Perspective of AI in the Wild.
doi: 10.1145/3322276.3322351

Miller, T. (2019, feb). *Explanation in artificial intelligence: Insights from the social sciences* (Vol. 267). Elsevier B.V. doi: 10.1016/j.artint.2018.07.007

Molnar, C. (2021). *Interpretable Machine Learning.* Retrieved from `https://christophm.github.io/interpretable-ml-book`

Moussette, C. (2012). *Simple Haptics Sketching perspectives for the design of haptic interactions* (Unpublished doctoral dissertation).

Moussette, C., & Banks, R. (2011). Designing through making: Exploring the simple haptic design space. In *Proceedings of the 5th international conference on tangible embedded and embodied interaction, tei'11* (pp. 279–282). doi: 10.1145/1935701.1935763

Muller, M., Lange, I., Wang, D., Piorkowski, D., Tsay, J., Vera Liao, Q., . . . Erickson, T. (2019). How data science workers work with data. *Conference on Human Factors in Computing Systems - Proceedings*, 1–14. doi: 10.1145/3290605.3300356

Nelson, H. G., & Stolterman, E. (2013). *The Design Way* (2nd ed.). MIT Press. doi: 10.7551/mitpress/9188.001.0001

Newell, A., & Simon, H. A. (1972). *Human Problem Solving.* Prentice Hall. doi: 10.2307/2063712

Newman, M. W., & Landay, J. A. (2000). Sitemaps, storyboards, and specifications: A sketch of web site design practice. *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, 263–274.

Odom, W., Zimmerman, J., Davidoff, S., Forlizzi, J., Dey, A. K., & Lee, M. K. (2012). A fieldwork of the future with user enactments. In *Proceedings of the designing interactive systems conference* (pp. 338–347). doi: 10.1145/2317956.2318008

Ozenc, F. K., Kim, M., Zimmerman, J., Oney, S., & Myers, B. (2010). How to support designers in getting hold of the immaterial material of software. In *Conference on human factors in computing systems* (Vol. 4, pp. 2513–2522). doi: 10.1145/1753326.1753707

Passi, S., & Jackson, S. J. (2018). Trust in data science: Collaboration, translation, and accountability in corporate data science projects. In *Proceedings of the ACM on human-computer interaction* (Vol. 2). doi: 10.1145/3274405

Ralph, P. (2010). Comparing two software design process theories. In *Lecture notes in computer science* (pp. 139–153). doi: 10.1007/978-3-642-13335-$0_1$0

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135–1144). doi: 10.1145/2939672.2939778

Riedl, M. O. (2019, jan). Human-centered artificial intelligence and machine learning. *Human Behavior and Emerging Technologies*, *1*(1), 33–36. doi: 10.1002/hbe2.117

Royce, D. W. W. (1987). Managing the Development of large Software Systems. In *Proceedings of the 9th international conference on software engineering* (pp. 328–338).

Russel, S., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach* (3rd ed.). Pearson.

Schmarzo, B. (2017). *Design Thinking: Future-proof Yourself from AI.* Retrieved 2020-05-29, from `https://infocus.delltechnologies.com/william-schmarzo/design-thinking-future-proof-yourself-from-ai/`

Schön, D. (1983). *The Reflective Practitioner: How Professionals Think in Action.* Basic Books.

Schön, D. (1987). *Educating the Reflective Practitioner.* Jossey-Bass Publishers.

Schön, D. (1992, mar). Designing as reflective conversation with the materials of a design situation. *Knowledge-Based Systems*, *5*(1), 3–14. doi: 10.1016/0950-7051(92)90020-G

Seaman, C. (2008). Qualitative Methods. In F. Shull, J. Singer, & I. Sjøberg (Eds.), *Guide to advanced empirical software engineering.* Springer.

Simon, H. A. (1992). *The Sciences of the Artificial* (3rd ed.). MIT Press. doi: 10.7551/mitpress/12107.001.0001

Sokol, K., Hepburn, A., Santos-Rodriguez, R., & Flach, P. (2019). bLIMEy: Surrogate Prediction Explanations Beyond LIME. In *33rd conference on neural information processing systems.*

Spradlin, L. (2015). *Meet Project Phoebe: A moonshot concept for mutative design.* Retrieved 2020-06-02, from `https://medium.com/project-phoebe/meet-project-phoebe-a-moonshot-concept-for-mutative-design-88d997f7ff14`

Star, S. L. (2014). Categories and cognition: material and conceptual aspects of large scale category systems. In S. J. Derry & C. D. Schunn (Eds.), *Interdisciplinary collaboration: An emerging cognitive science* (pp. 167–186). Psychology Press.

Stembert, N., & Rotterdam, H. (2019). Accounting for the Human When Designing with AI-Challenges Identified. In *CHI'19-extended abstracts,*.

Stolterman, E. (1999). The design of information systems: Parti, formats and sketching. *Information Systems Journal*, *9*(1), 3–20. doi: 10.1046/j.1365-2575.1999.00044.x

Stolterman, E. (2008). The nature of design practice and implications for interaction design research. *International Journal of Design*, *2*(1), 55–65.

Suchman, L. A. (1987). *Plans and Situated Actions: The Problem of Human Machine Communication.* (2nd ed.). Cambridge University Press. doi: 10.2307/2073874

Sundström, P., Taylor, A. S., Grufberg, K., Wirström, N., Belenguer, J. S., & Lundén, M. (2011). Inspirational bits: Towards a shared understanding of the digital material. In *Conference on human factors in computing systems - proceedings* (pp. 1561–1570). doi: 10.1145/1978942.1979170

Valkenburg, R., & Dorst, K. (1998). The reflective practice of design teams. *Design Studies*, *19*(3), 249–271. doi: 10.1016/s0142-694x(98)00011-8

Van Allen, P. (2017). Reimagining the goals and methods of UX for ML/AI. *AAAI Spring Symposium*, 431–434.

Van Allen, P. (2018). Prototyping ways of prototyping AI. *Interactions*, *25*(6), 46–51. doi: 10.1145/3274566

Van Der Vlist, B., Van De Westelaken, R., Bartneck, C., Hu, J., Ahn, R., Barakova, E., ... Feijs, L. (2008). Teaching Machine Learning to Design Students. In *International conference on technologies for e-learning and digital entertainment* (pp. 206–217).

van Boeijen, A., Daalhuizen, J., & Zijlstra, J. (2014). *Delft Design Guide: Design Methods.* BIS Publishers.

Vetrov, Y. (2017). *Algorithm-Driven Design: How Artificial Intelligence Is Changing Design.* Retrieved 2020-06-02, from `https://www.smashingmagazine.com/2017/01/algorithm-driven-design-how-artificial-intelligence-changing-design/`

Walker, M., Takayama, L., & Landay, J. A. (2002). High-Fidelity or Low-Fidelity, Paper or Computer? Choosing Attributes when Testing Web Prototypes. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *46*(5), 661–665. Retrieved from `http://guir.berkeley.edu/prototypefidelity` doi: 10.1177/154193120204600513

Wong, S. (2016). *Google Translate AI invents its own language to translate with.* Retrieved 2020-05-30, from `https://www.newscientist.com/article/2114748-google-translate-ai-invents-its-own-language-to-translate-with/`

Xu, W. (2012). User Experience Design: Beyond User Interface Design and Usability. In I. Nunes (Ed.), *Ergonomics - a systems approach.* IntechOpen.

Yang, Q. (2017). The Role of Design in Creating Machine-Learning-Enhanced User Experience. *AAAI Spring Symposium - Technical Report*, 406–411.

Yang, Q. (2018). Machine Learning as a UX Design Material: How Can We Imagine Beyond Automation, Recommenders, and Reminders? *2018 AAAI Spring Symposium Series*, 467–472.

Yang, Q., Banovic, N., & Zimmerman, J. (2018). Mapping machine learning advances from HCI research to reveal starting places for design innovation. *Conference on Human Factors in Computing Systems*, 1–11. doi: 10.1145/3173574.3173704

Yang, Q., Scuito, A., Zimmerman, J., Forlizzi, J., & Steinfeld, A. (2018). Investigating how experienced UX designers effectively work with machine learning. In *Proceedings of the 2018 designing interactive systems conference* (pp. 585–596). doi: 10.1145/3196709.3196730

Yang, Q., Zimmerman, J., Steinfeld, A., & Tomasic, A. (2016). Planning adaptive mobile experiences when wireframing. *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, 565–576. doi: 10.1145/2901790/2901858

Zhang, A. X., Muller, M., & Wang, D. (2020). How do Data Science Workers Collaborate? Roles, Workflows, and Tools. , *4*(May), 1–23.

Zhou, Z., Qi, Z., Gong, Q., & Sun, L. (2019). ML-process canvas: A design tool to support the UX design of machine learning-empowered products. *Conference on Human Factors in Computing Systems*, 1–6. doi: 10.1145/3290607.3312859

Zhou, Z., Sun, L., Zhang, Y., Liu, X., & Gong, Q. (2020). ML Lifecycle Canvas: Designing Machine Learning-Empowered UX with Material Lifecycle Thinking. *Human-Computer Interaction*, 1–25. doi: 10.1080/07370024.2020.1736075

Zimmerman, J., Stolterman, E., & Forlizzi, J. (2010). An analysis and critique of research through design: Towards a formalization of a research approach. In *Proceedings of the 8th ACM conference on designing interactive systems* (pp. 310–319). doi: 10.1145/1858171.1858228

Appendices

Appendix A

Pre-interview Survey

Thank you for taking the time to fill out this survey! It won't take more then 5 minutes, but will save us some time during the actual interview. In my email to you, you will find a participant ID. Please make sure you provide the right participant ID, so I can connect your answers to you.

Your participant ID

Your age

In which field(s) did you complete an education? (multiple answers allowed)

- Design
- Computer science
- Information science
- Human-computer interaction
- Cognitive psychology
- None
- Other...

How many years of experience in designing for ML products/features do you have?

- Less than one year
- 1-2 years
- 2-4 years
- 4-6 years
- 6-8 years
- 8-10 years
- Over 10 years

Important note Please answer the following questions on the project (or design case) we will use to guide the conversation during the interview, i.e. not your current situation (unless it is the same of course).

What best describes your role during the project?

- UX designer
- product designer
- design manager
- interaction designer
- experience designer
- service designer
- other...

What is the approximate size of the organization that you worked for during the project?

- $<$100 employees
- 100-1,000 employees
- 1,000-10,000 employees
- $>$10,000 employees

Participants were asked to indicate their familiarity with the following concepts on a scale of 1 (I know nothing about them), 2, 3, 4 (I have heard of them but have no hands-on experience), 5, 6, 7 (I work with them in my day-to-day work).

- Correlation and causation
- Model training
- User flows
- Training set, validation set test set
- Wireframes
- Personas
- Model validation

Appendix B

Interview Guide

Goal
I want to create an understanding of how designers actually work and compare that to the existing academic understanding of practice of designers in the field of ML. The goal for today's session is to generate knowledge on how you work as a designer on ML products or features.

Project Description

- Can you give a brief description of the project you selected?

Project Initiation

- Explain how this project was initiated
- Explain when you got involved in the project
- What are your thoughts about when you got involved in the project?

Project Process

- Walk me through your work on this project
- What was your next activity in this project?
- Why did you do that activity (at that point)?
- How did doing the activity work for you?

Further Challenges

- Are there any other challenges or issues you've experienced working on ML products as a designer that we haven't discussed yet?

The following topics were discussed as they came up or at the end if there was sufficient time left.

Data Collection

- Explain how you were involved in data collection for this project
- What are your thoughts about how you were involved in data collection?
- Tell me about relating data collected to domain knowledge in your project

ML Knowledge

- Explain how your level of ML knowledge impacts your work
- What are your thoughts about your current level of ML knowledge?
- Explain why you think it's important for a designer to have some ML knowledge?
- Tell me about about things you do to learn about ML

Familiarization to ML as a design material

- Tell me about about things you do to think of new ways or places to apply ML technologies?

Design methods

- Describe the overarching design methodology you used for this project
- Explain why you decided to use this methodology
- Tell me about how this methodology worked for you
- What design methods did you use for this project?
- Explain why you decided to use this method
- Tell me about how this method worked for you

Prototyping

- Describe what kind of prototypes you created during the project
- Explain why you chose this method of prototyping/this type of prototype
- Tell me about how this prototype worked for you
- Explain how you used data in these prototypes
- Tell me about how using data this way worked for you

Mental models

- Tell me about how the incorrect mental models that many have of ML impact your work a a designer
- Describe how you deal with this challenger

System changes over time

- Tell me about how the fact that you're designing something that changes over time impacts your work as a designer
- Describe how you deal with this challenge

Personalization

- Tell me about how the personalization opportunities that ML provides impact your work as a designer
- Describe how you deal with this challenge

Black box nature of ML

- Tell me about wow the black box nature of ML impacts your work as a designer
- Describe how you deal with this challenge

Appendix C

Recruitment Email

Hello [name],
Thank you so much for expressing interest in being a participant for interviews for my research! In this email you will find some additional information.

Goal
I am studying the practice of designers who work on machine learning enhanced products in order to close the gap between the academic understanding of practice, and the actual work of designers in this field. Therefore, the goal of the interview is to generate knowledge on your perception of your practice as a designer.

Profile
In order to be a good fit for my thesis interviews, you must meet the following two criteria:

- Are a designer (UX, product, interaction, design manager etc) with experience in working on products that use ML technologies (including ML-based features in larger products)
- Have at least one (nearly) completed design case in the field of ML to discuss during the interview to guide the conversation

If you are not sure whether you meet (one of) these requirements, please let me know and we can figure it out together.

Confidentiality
The focus is on your experience of how you work, not what you work on, it is not necessary to disclose anything confidential about the project you are describing beyond a very general description (for example: a health coaching app, an intelligent feature in a social media app). Results (in the form of direct quotes) will be used in my thesis, which will be published in the Utrecht University theses archive. However, I will ensure that results cannot be linked to your person.

Practicalities
Because I don't want to waste your time during the actual interview, I will to ask you to fill out a short survey (5 min) in advance. This survey contains some basic demographic questions and questions on your familiarity with certain concepts. I will send you the survey before the interview. The interview itself will take 60-75 mins and will take place through Google Meet during a time that is convenient for you. The interview can be conducted in English or Dutch, based on your preference. The audio of the interview will be recorded.

Scheduling an interview
If you would like to participate in an interview, please send me 3-5 potential timeslots that would work for you (+ if you are not on CET time, please also include your timezone). I will get back to you to schedule an interview.

Thanks in advance for your effort! If you have any further questions, please feel free to contact me.

Kind regards,
Lara Zijlstra

Appendix D

Scheduling Email

Hello [name],

I'm glad you would like to participate in my research! [timeslot] works best for me. I'll schedule a calendar meeting and include a Google Meet link there.

I want to ask you to think of one project or design case where you worked on a ML product that we can use during the interview to guide the conversation. I'm interested in your experience in general, but this one project will serve as a kind of red thread throughout the interview. I also want to ask you to fill out this quick survey before the interview. Your participant ID for this survey is [participant ID].

If you have any further questions, please let me know!

Kind regards,
Lara

Appendix E

Codebook

The codebook describes the initial sub-themes that make up the final five themes presented in this study. For each sub-theme, the following information is provided: the number of files (interview transcripts) that the theme is present in, the number of references (segments that were coded with that theme), and the theme description.

Design Process & Methods

Sub-theme	**Files**	**Ref**	**Description**
Design methods used	19	127	Participants describe using a large variety of research methods. Some designers discuss having different kinds of workshops that facilitate design collaboration with users or with data worker colleagues. For example, having brainstorming sessions or doing a Google design sprint. They also discuss many methods that they use to produce visual artifacts, such as building user journeys, and making wireframes. Designers are neutral about the methods used when discussing them.
Research methods used	16	77	Participants describe doing a lot of research as part of their design work for ML products. A plethora of different research methods were mentioned in the interviews, ranging from qualitative methods, such as interviewing and diary studies, to quantitative methods, including surveys and A/B testing among others. Research takes place throughout the entire process, for example starting with identifying user needs, doing user testing throughout the development process, and using tools such as Hotjar after launching the product. It is interesting how the described methods are common in non-ML projects as well, and how the designers remain neutral when discussing the research methods used. It seems ML and non-ML projects do not differ much in regards to methods used and there do not seem to be outspoken issues with the discussed methods in ML products specifically.

Machine Learning as a Tool

Sub-theme	Files	Ref	Description
Meaningful ML use	11	27	Participants emphasized the need to use ML as a potential problem solving tool, not as a cherry on top. Some designers explained how they see ML as one of the tools in their toolkit, that can and should only be used when it's an appropriate solution to a user problem. It is only a powerful tool when it's used at the right time and place. They explain how it should only be used to make the user's life easier, and not require any extra effort. The correct approach is to start with a user problem, and use ML if it can solve the problem appropriately, not to start thinking from how ML could be applied. Some designers are adamant that ML should not be applied just for the sake of having ML, but in a way that's meaningful to the user.
Value of ML knowledge	14	41	Participants acknowledged the need to have a decent understanding of ML in order to do good work on ML projects;. they feel it makes them better designers. They emphasize how it is not necessary to have a deep understanding of the technical properties, but find it very important to understand what is and is not possible with the technology. Some designers explain how being aware of the possibilities allows them to identify opportunities where ML can make a valuable impact. They admit to not initially being particularly interested in learning about ML, but working on ML projects has made them realize just how valuable the knowledge can be, and sparked their interest to learn more.
Specific approach to learning	15	45	Participants describe what they did to improve their ML knowledge. Some designers explicitly mentioned that they do not seek out formal education, or tried it and failed. Instead, they mention learning through personal relationships, being able to ask for support from their friends or partners. They learn through doing, by doing personal projects and dabbling in coding. Some designers mention actively working on acquiring knowledge, feeling the need to understand how it works. For example, looking up terms used by colleagues on the internet after meetings. Some designers also mention trying to learn about ML in their own language; by creating visual artifacts that help them understand it, but also be able to educate others about ML. What is interesting about the ways designers acquire ML knowledge is two things that stand out: designers actively look to acquire knowledge, in an informal way. This is interesting because knowing how designers learn, they can be helped to learn in a way that fits them well.

Designerly understanding	13	22	From the interviews it becomes clear that designers have a very particular way of talking about and understanding ML. Acknowledging that their technical understanding of ML is limited, designers talk about ML not in technical terms, but in terms of input and output. For example, the data the machine needs, but especially the output that is presented to the user. Additionally, they describe not needing to understand the technical aspects of ML, but instead understanding the material by understanding what it can and can't do, as that is relevant to their work. Rather than following the traditional, technical understanding of ML, designers have formed their own way of understanding the material. This is interesting because by being aware of the way designers understand the material, it is easier to support them in improving their understanding.

Advocate for the User Perspective

Sub-theme	Files	Ref	Description
Specific Approach to Learning	13	29	Participants describe what they did to improve their ML knowledge. Some designers explicitly mentioned that they do not seek out formal education, or tried it and failed. Instead, they mention learning through personal relationships, being able to ask for support from their friends or partners. They learn through doing, by doing personal projects and dabbling in coding. Some designers mention actively working on acquiring knowledge, feeling the need to understand how it works. For example, looking up terms used by colleagues on the internet after meetings. Some designers also mention trying to learn about ML in their own language; by creating visual artifacts that help them understand it, but also be able to educate others about ML. What is interesting about the ways designers acquire ML knowledge is two things that stand out: designers actively look to acquire knowledge, in an informal way. This is interesting because knowing how designers learn, they can be helped to learn in a way that fits them well.
Design is an afterthought	12	26	Participants discuss how design thinking on a general level seems to be missing in the development process. Some designers mentioned how their data worker colleagues think their sole purpose is to design screens to put onto existing technology. This is also reflected by those who mention being brought on late in the process to do a mostly visual redesign of an existing product. They also mention that those existing products tend to have rather bad user experience and are not very well designed. Some designers discuss the lack of design thinking at the start of projects, in the problem definition phase. They discuss struggles in getting their work prioritized over that of data workers. Overall, it is clear that design does not seem to be the first priority in ML projects, only coming in once the technology is taken care of.

Represent the user perspective	17	64	Participants indicated that they feel their role in ML projects is to be the voice of the user. They emphasize the importance of considering the user perspective and the context in which the user interacts with the product, and their responsibility to represent those parts of the experience. Some designers self-identify as a 'translator' or an 'integrator' between the user perspective and the technical perspective. A translation that goes both ways; allowing users to understand technical processes and terms, but more importantly, sharing the user voice in a way that's understandable to their data-minded colleagues. They see this role for themselves in all parts of the process, including the more data-minded activities such as data collection. Even though they are working on an ML project, the designer's classical objective to represent the user does not change.
Design should influence product roadmap	15	39	Participants commented positively on being involved in determining the roadmap of a project. Some designers describe positive experiences of being involved in project initiation, or being involved very early on when little is set in stone. They feel 'fortunate', or 'lucky' to be involved early on, as it allows them to help steer the project in a design-driven direction, or helps them to fully understand the context of the project, for example. Some designers describe how prioritization in projects is handled primarily by product managers, however, designers often work together with the product manager and are therefore able to help shape decisions on prioritization. Designers use user research data to push their agenda. Either at the start of projects, by initiating projects that address a user problem or pain point, or using data driven proof of the necessity to prioritize functionalities to get their work prioritized over engineering efforts.

Space to Collaborate

Sub-theme	Files	Ref	Description
Create the space to collaborate	17	61	Participants acknowledge that it is important that designers work closely with their data worker colleagues. They mention various things they try and do to bridge the gap between themselves and their data worker colleagues. Some designers are trying to educate their teams on the power and value of design. Likewise, they also put effort in educating themselves on ML, and achieving the basic understanding that is required to successfully communicate with data workers. Some designers mention using their data worker colleagues to improve their understanding of ML in general or the product that they are working on. They also try to involve data workers in their process, for example passively by scheduling meetings to keep them updated, or actively by hosting co-design sessions. Whether it is through education, close communication, or involvement in design, designers actively work on creating a space to collaborate closely with their colleagues. Awareness of how designers currently approach this can help to facilitate even closer collaboration.
Need for a shared language	12	34	Participants acknowledge the importance of being able to communicate well with their data worker colleagues. However, communication remains a struggle, due to the simple fact that they are not familiar with the vocabulary used by their colleagues. Not being able to follow conversations prevents them from asking critical questions, for example. Some designers talk about having to make an effort to overcome this issue. They either try to improve the communication by trying to speak the language of data workers, learning about common ML terms and concepts. Others try to get the data workers to speak their language; through creating visuals such as mock ups to engage in the conversations. Regardless of whose language is used, it is vital that it is a shared one.

Data Limits Prototypes

Sub-theme	Files	Ref	Description
Data limits prototyping	9	21	Participants discuss encountering difficulties in prototyping ML products related to the data-centered nature of ML. In order to properly prototype ML, you need the data. Designers describe data not being available when they need it. It is not within their expertise to create data or algorithms, so it is not a problem that they can solve themselves. Additionally, designers have to be mindful of sensitive data and strict regulations. Sometimes these prevent them from testing prototype with external users, but at other times the regulations are so strict that they cannot even use real or seemingly real data in prototypes or mock ups at all. It is these practical considerations that determine how data is used in prototypes, not the designers' judgement.
Use real data to understand real complexities	16	58	Participants discussed the importance of creating a prototyping experience that is as realistic as possible. Realistic in terms of both the data included in the prototype, as well as the interactivity of the prototype. Designers discussed how this realism is extremely important to ML products, as you are in a sense also testing the accuracy of the model through the eyes of the user. Having a realistic prototype that users can actually interact with, helps designers to get good feedback and discover the real challenges of the product. There's different ways designers describe creating realistic prototypes, depending on a variety of factors. Some are able to use a fully functional model with real data in their prototypes. Others had to use a sample of the dataset, or used one specific example. If the designers did not have access to the data, they opted for Wizard of Oz or dummy data. However, especially the latter causes problems in terms of getting useful feedback. Though a realistic prototype is highly regarded, designers do acknowledge the difficulties that it brings along.

Appendix F

Comparison of Process Level Codes

Comparison of process level codes												Legend
p01	Research	Data gathering	Training model	Prototype	User testing	**Repeat**	Developed produ	Questionnaires				**Research**
P02	Storyboards of context	Requirements	Prototyping	Test	**Repeat**	Final product						**Ideation**
P03	Researching user needs	Brainstorm, ideat	Prototyping	Test or feedback	**Repeat**	Showcase						**Prototype/Visualize**
P04	Diary study	Card sorting	Prototype	Validate prototype								**Validation**
P05	Research, interviewing, shadowing etc	understanding domain	Co-creation, worl	Prototyping	Testing							**Release/Final Step**
P06	Research	Co creation workshops	Gather needs and opportunities	Feasibility check with Dws	Final solution concept	Get feedback	DWs do their stu	Mock ups	Development	Testing		
P07	Understanding the domain	Fake mock ups, visualize it	Sharing learnings									
P08/p09	Preparing devices for data tracking	Give out devices, start data collection	Build dashboards	Make visualizations								
P10	Research	Mock up	Check w developers	test internally	**Repeat**	Finalized mock ups	Release					
P11	Research	Creating personas	Prototyping	Getting feedback, user testing	**Repeat**	Release						
P12	Research	Co design / white-boarding with dw	Prototyping	Feasibility check	Validation	**Repeat**						
P 13	Research	Ideation	Paper prototyping	Validation	**Repeat**							
P14	Product audit	Research	Wire framing	validation, internal testing	Share with team	Mock ups	Final mock ups	Hand of to development team	Testing			
P15	Creating hypotheses	Crease prototype for algo testing	Initial interviews	Diary study	validate algo prototype							
P16	Research	Iterative prototyping at the same time	testing prototypes	**repeat**								
P17	Map current processes	ideation	Prioritization	Feasibility check w AI expert	project hand off							
P18	Brainstorm		Requirements	Research at the same time	Concept ideation	Prototyping	Testing					
P19	User research	Prototype	user testing	**Repeat**	Release	Questionnaire						
P20	Understand context/domain	Research	Design sprint	Research	Fail							
	Mock ups	Validate										