Co-funded by the
Erasmus+ Programme
of the European Union

Exploring Ethics and Human Values in Designing AI

MASTER THESIS
to obtain the Erasmus Mundus Joint Master Degree

in Digital Communication Leadership (DCLead)

of

Faculty of Cultural and Social Sciences
Paris Lodron University of Salzburg

Technical Faculty of IT and Design
Aalborg University in Copenhagen

Submitted by

Reza Arkan Partadiredja

11934294
hey@rezaarkan.com
Vognporten 14, Albertslund 2620, Denmark

Jannick Kirk Sørensen (Aalborg University Copenhagen)
Christopher Frauenberger (Paris-Lodron Universität Salzburg)
Leah Lievrouw (University of California, Los Angeles)
Ramesh Srinivasan (University of California, Los Angeles)

Department of Communication Studies

Salzburg, 31 July 2021

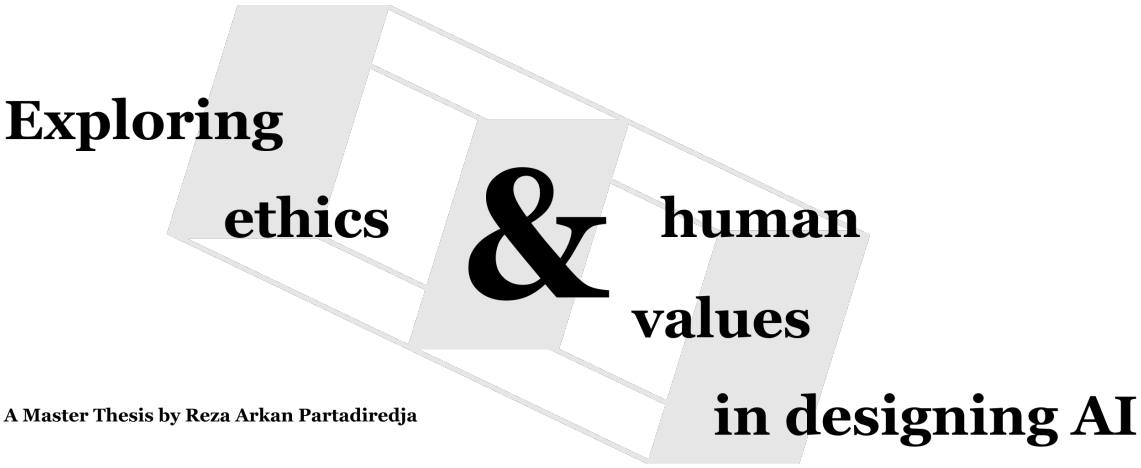
Table of Contents

List of Figures

List of Tables

For those who have guided me,

And for those who matter the most.



A Master Thesis by Reza Arkan Partadiredja

Thank you.

Executive Summary

With the increasing concern over the risks of AI and the proliferation of ethical principles, design plays a potentially important role as it has operationalized applying values and considering ethics throughout history. However, designers face unique challenges in designing AI due to its distinct materiality. While some investigations have been conducted into how designers design AI, not many inquiries have seemingly been made on how designers apply values and consider ethics in designing AI despite emerging studies that argue on how design could theoretically play a role. Based on this, this project asks the question: *how do design(ers) apply values and consider ethics in designing AI?*

To explore this inquiry, eight interviews with both AI designers and developers were conducted to give insights on how their AI design processes are, what role design played in the process, the challenges they faced in designing AI, their thoughts on the implication of AI for societies, how they see AI ethical values and their attempts in applying them to their actual practice. Subsequently, a workshop session with five designers were held to generate as many ideas as possible to how they might apply values in designing AI then following it up with a reflection on whether, how, and why these ideas should be applied in practice.

The findings show emphasis on human-centered and participatory approaches to apply values and consider ethics in designing AI. However, these efforts are hindered with inherent challenges on top of other factors that complicates how values and ethics can be translated into practice. Nonetheless, participants express the essential role and contribution of designers in the development of AI. These findings lead to the notion that a paradigm shift in design practice within the context of AI may be required. In further synthesis, a framework was proposed to reframe AI in different perspectives: (1) *AI as is*, (2) *AI as a design material*, and (3) *AI as a sociotechnical system*. From these reframings, ideas and further questions were generated.

As a result in exploring how design(ers) apply values and consider ethics, the insights from the findings of both interviews and the workshop were synthesized along with the literature highlights which produced further ideas and questions that can serve as basis for further endeavours into the inquiries on the intersection between AI, design, and ethics.

1. Introduction

1.1. Background

> *"**A designer is first and foremost a human being.***
>
> *Before you are a designer, you are a human being. Like every other human being on the planet, you are part of the social contract. We share a planet. By choosing to be a designer you are choosing to impact the people who come in contact with your work, you can either help or hurt them with your actions. The effect of what you put into the fabric of society should always be a key consideration in your work."*
> (Monteiro, 2019, p. 19)

Stumbling upon the sobering quote above was a good reminder. While I was never formally educated in design, I was a designer by profession. I spent the earlier part of my career working for the technology industry, designing digital products for millions of users across South-East Asia. The more familiar I became with design, the more I realized its importance for the benefit of businesses and companies through creating better experiences for its users. However, this view of design felt myopic. The more I realized that, the more I craved for perspectives in a wider sense.

From the many perspectives shared throughout my Master's education, I eventually found my interest gravitating towards artificial intelligence (AI). Starting with a critical review suggesting the interplay of design and datafication (Partadiredja, 2020), to creating an illustrative experiment poking at the phenomena of widespread human-like AI generated contents (Partadiredja et al., 2020), to receiving the opportunities to engage with relevant experts. The accumulation of experiences eventually led me to an intersection where design is situated in an increasingly datafied society and artificial intelligence becomes ever more prevalent. *What of design, what of its role, and what of the societal concerns?*

It is in this starting point that I began to come across the many risks and societal concerns in relation to the increasing adoption of artificial intelligence (Cave & ÓhÉigeartaigh, 2018; Crawford & Joler, 2018; Floridi et al., 2018; Neff, 2016; Turchin &

Denkenberger, 2020), which sometimes paints an otherwise somewhat gloomy future where AI runs supreme. In response, a proliferation of ethical principles focused on mitigating the grief consequences of AI have surfaced in recent times (Fjeld et al., 2020; Floridi & Cowls, 2019; Hagendorff, 2020; Jobin et al., 2019). Amidst this "principle proliferation" (Floridi & Cowls, 2019), many have then explored how these high-level documents can relate to practice (Mittelstadt, 2019; Morley et al., 2019; van de Poel, 2020). But more specifically, how do these principles relate to the design practice in designing AI?

This line of thought was based on an established design tradition of operationalizing the processes of applying values and considering ethics (Devon & van de Poel, 2004; Friedman, 1996; Monteiro, 2019; Shilton, 2013) and, in this sense, should extend to the role that designers supposedly play in designing AI. However, for design itself, AI is a relatively new material and comes with its unique set of use qualities that designers may find exceptionally challenging (Bratteteig & Verne, 2018; Holmquist, 2017; Stoimenova & Price, 2020; Yang et al., 2020).

In this regard, some investigations have been conducted to give empirical evidence and insights into how designers design AI (Dove et al., 2017; Girardin & Lathia, 2017; Liao et al., 2020; Yang et al., 2018) and there is increasing work on the novel ways of working with AI as a design material (Amershi et al., 2019; Koch et al., 2019; Subramonyam et al., 2021; van Allen, 2018; Zimmerman et al., 2020). Meanwhile, not many inquiries seem to have been made into how designers apply values and ethics considering emerging studies that explore how they could theoretically play a role in addressing unintended consequences (Stoimenova & Kleinsmann, 2020) and embedding values into AI systems (Dignum, 2017; Umbrello, 2019; van de Poel, 2020). Nonetheless, in the face of these unique challenges in designing AI, how then do designers actually apply value and consider ethics?

Therefore this thesis seeks to explore how design(ers) apply values and consider ethics in designing AI. And in this, the answers I am looking for may not be a straightforward one. In fact, I have pretty much set my expectations that I may have set myself up with a rather complex topic. Regardless, I am motivated as stars seem to have aligned on my interests and hope that this will be both a memorable work for my future endeavours and a humble contribution to the nascent discussions on AI, design, and ethics.

1.2. Research Question

Having laid out the background, the research question of this exploratory project boils down to: *How do design(ers) apply values and consider ethics in designing AI?*

1.3. Research Relevance and Significance of Study

As elaborated in the background, there is an increasing concern over the risks of AI (Cave & ÓhÉigeartaigh, 2018; Floridi et al., 2018; Neff, 2016; Turchin & Denkenberger, 2020) and in response is the proliferation of ethical principles (Fjeld et al., 2020; Floridi & Cowls, 2019; Hagendorff, 2020; Jobin et al., 2019) in which these high-level documents need to be translated into practice (Morley et al., 2019; Whittlestone et al., 2019).

Design itself is a field that has operationalized applying values and considering ethics throughout history (Devon & van de Poel, 2004; Friedman, 1996; Monteiro, 2019; Shilton, 2013), however they face unique challenges in designing AI due to its distinct materiality (Bratteteig & Verne, 2018; Holmquist, 2017; Stoimenova & Price, 2020; Yang et al., 2020).

While some investigations have been conducted into how designers design AI (Dove et al., 2017; Girardin & Lathia, 2017; Liao et al., 2020; Yang et al., 2018), not many inquiries have seemingly been made on how designers apply values and consider ethics in designing AI despite the emerging studies that argue on how design could theoretically embed values into the design of AI systems (Dignum, 2017; Umbrello, 2019; van de Poel, 2020). This project emphasizes this research gap and seeks to make a contribution in exploring how designers apply values and consider ethics in designing AI.

1.4. Scope and Limitations

Artificial intelligence is a relatively fuzzy term that is commonly used to categorize a system capable of learning from its environment (Haenlein & Kaplan, 2019). As it is, in essence, a technology to enable certain solutions, its implementation may differ significantly depending on the context that it is situated in, resulting in a wide range of AI systems with potentially significant differences between their features. For the scope of this study, AI was kept as a general term denoting a system's capability to learn and evolve through engagement with its environment. This decision was a practical one, as

limiting AI within specific contexts or certain industries could prove challenging to conduct within the given timeframe of this project.

Values can also be viewed from different angles and have different aspects. As this project primarily relates closely to AI ethical principles, the values referred here are limited closely to moral values. On top of that, this research project does not aim to provoke the discussions revolving around ethics but rather simply aims to explore how designers engage with them in the process of designing AI through empirical investigations.

2. Theoretical Background

This chapter aims to establish a literary overview on design, AI and its socioethical concerns and all other related areas to situate this research. To build a foundation for the research question, the first section attempts to gather literature to form an argument on designing AI alongside the socioethical implications that the technology brings. The second section then establishes the current overview of values and ethics in design and brings it to the context of designing AI.

2.1. Primer on Designing AI

2.1.1. **Artificial Intelligence**

Artificial intelligence can sometimes be considered a broad umbrella term. Despite popular and mainstream recognition, its descriptive meaning can oftentimes be fuzzy, and in this regard, Haenlein & Kaplan (2019) suggests that AI systems to be commonly defined as *"a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation."* In more elaborate terms, the European AI High-Level Expertise Group (2019) defines AI as *"software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions"*. Taking inspiration from both definitions, I will refer to AI in this work loosely as a system which can learn and evolve from external data. Another key distinction of AI that I take from both definitions is the emphasis both definitions have on data, the process of learning from the data, and producing an output.

As the name suggests, much of its novelty pertains to the notion of its *intelligence*, of which it can be used to further categorize different types of AI. Drawing from Kaplan and Haenlein (2019), AI can be categorized into three stages based on their subsequent level

of capabilities: (1) narrow-artificial intelligence, (2) artificial general intelligence, and (3) artificial super intelligence (Figure 1).

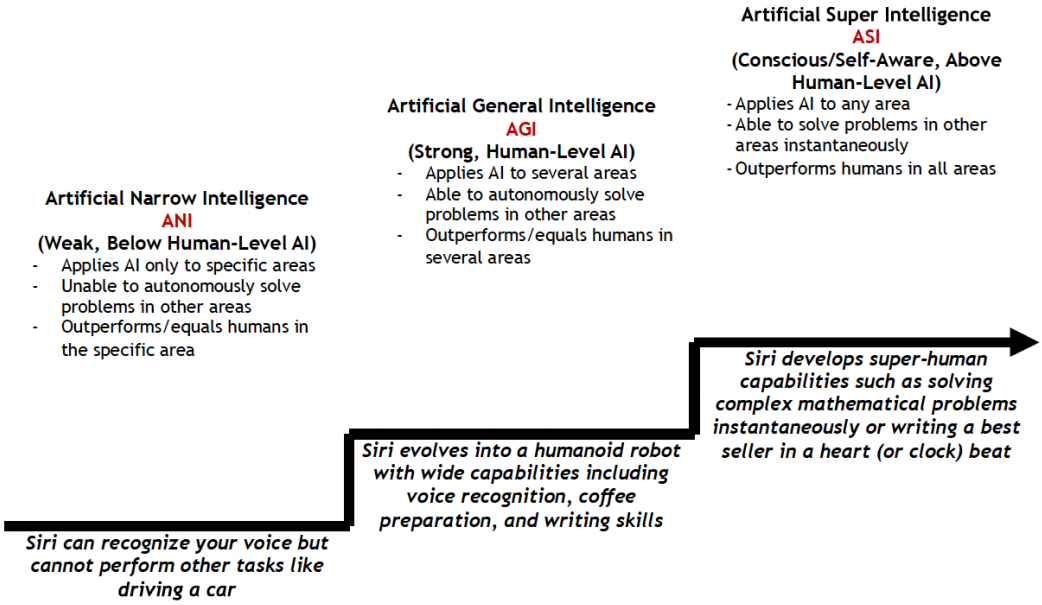


Figure 1. AI based on their capabilities (source: Kaplan & Haenlein, 2019)

From current design perspective, narrow-artificial intelligence best describes almost all of the AI systems that users interact with in their daily life (Stoimenova & Price, 2020) as it encompasses AI systems that are still tied to specific areas (e.g. *"Siri can recognize your voice but cannot perform other tasks like driving a car"*) whereas artificial general intelligence and artificial super intelligence relate to more futuristic concepts of AI not yet common to our everyday lives. In this sense, all of what is considered as AI in this body of work falls under the category of narrow-artificial intelligence.

2.1.2. Designing AI

What is design? As a word, design exists as both noun and verb, implying the nature of both the thing and the action. However, its definitive meaning that is supposedly tied to both the craft, the artifact, and the profession can sometimes be rather elusive. In part, this may be due to the constantly changing environments of designs, such as the evolution of technology, society, and industries; all of which are, in turn, influenced by design itself (Krippendorff, 2005). As such, design does not exist in a vacuum — and neither does its definition.

As this research focuses on exploring the design perspective on how ethical considerations and values are applied to AI, it is important to lay a definition of what is meant by design. Moreover, this section seeks to add basis on the supposedly influential role that design can offer in the creation process of technologies. Therefore, the aim of this section is not just to give a working definition of design(ers) that is beneficial for the research but to illustrate the growing responsibilities (and complexities) of design throughout the course of history and how it should be an important part of designing AI.

2.1.2.1. Design as a noun

As a noun, the history of design can be seen from the development of artifacts. Within much of the 20th century, design was typically tied to industrial design where it concerned itself mostly with physical artifacts, furniture, fabric, industrial appliances, and (industrial) architecture; concerns of materiality were primarily limited to its physical dimensions (Krippendorff, 2011). As society becomes increasingly digital, however, the concerns eventually grew beyond physical materiality (Buchanan, 1998; Höök & Löwgren, 2021; Krippendorff, 2011).

The rapid technological and societal changes in relation to the artifacts can be accounted for the growing scope and challenges of design (Krippendorff, 2005), as designers find themselves encountering new problems pertaining to novel materiality (or immateriality), giving birth to specialized domains such as interaction design and human-computer interaction (Rogers, 2012). To illustrate from the lens of artifacts, Krippendorff's *trajectory of artificiality* (2005, 2011) shows a progressing history of the growing design problems and considerations corresponding to 6 stages of artificiality (see Figure 2)

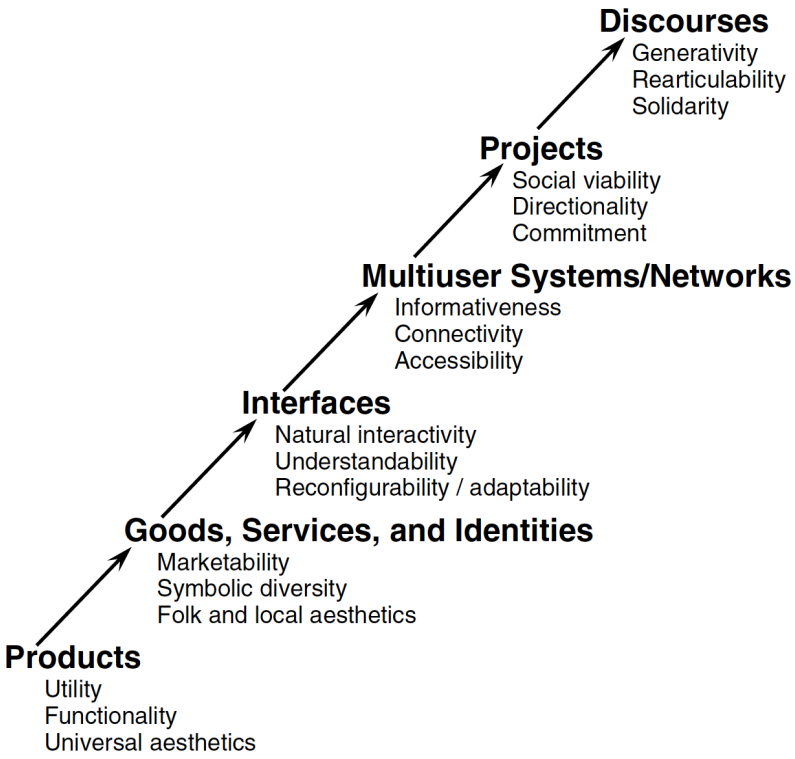


Figure 2. The trajectory of artificiality (source: Krippendorff, 2011)

In this trajectory (Figure 2), the concerns of designers have evolved throughout history to adapt to reflect the properties, materialities, and complexities around the artifacts (Krippendorff, 2005, 2011). For *products*, for example, the emphasis revolves solely around utility, functionality, and aesthetics while disregarding the situated effects of cultural diversity and instead assumes a universal rationality towards users (Krippendorff, 2005, 2011). The third stage of *interfaces* emerged as a result of an increasingly digital world where screens serve as an integral point of interaction for users as the internal workings of computers become obscured and less relevant. As a case, the design of the Xerox Star (1989) helped set the modern standard for personal computing interfaces with concepts like *windows* and visual metaphors like *the desktop*. This presents a trajectory in which design has evolved throughout history reflecting the changes to the artifacts it is intertwined with.

Buchanan's (1992, 1998) *four orders of design* also presents the idea of the evolving nature of design and its broadening scope of discipline. In Golsby-Smith (1996), the four orders are used to describe the domain expansion of design, throughout the subsequent scopes of: *words and symbols*, *objects*, *strategic decision making*, and *culture system*.

Whereas the first and second domain puts the designers' focus solely on the artifact, the third and fourth domain sees designers take into account the situated context of the artifacts, inevitably widening the considerations of design to include processes, people, communities, and cultures (Buchanan, 1992; Golsby-Smith, 1996). Both the *trajectory of artificiality* (Krippendorff, 2005, 2011) and the *four orders of design* (Buchanan, 1992, 1998; Golsby-Smith, 1996) suggest the increasing complexities of design alongside the development of the artifacts.

Throughout the advances of technology, the artifacts continuously evolves prompting design to work with conditions that are continuously changing, resulting in the need to invent and reinvent the *interface* for new ways of interaction through the use of novel metaphors, affordance, and signifiers (Johnson et al., 1989; Lovejoy, 2021). With the increasing pervasiveness of digital solutions in everyday life, the *interface* goes beyond graphical, becoming invisible, natural, and everywhere (Rogers, 2012). Considering such advances, rethinking the ways in how humans embody interaction with the increasingly computerized artificial environment becomes a necessity (Höök & Löwgren, 2021).

With AI as an artifact, the rethinking of design is necessary to account for its idiosyncratic properties (Holmquist, 2017; Yang et al., 2020). As a design material, one of its most distinct properties is its adaptive nature, in which the system is able to learn and evolve based on its engagement with its environment (van Allen, 2017; Yang et al., 2020; Zimmerman et al., 2020). Although user interfaces still play a crucial role, designing interactions with AI entails designing something increasingly intangible and invisible, which in itself is another unique property (Dove et al., 2017; van Allen, 2017). However, in most cases where explainability is concerned, designers must also consider the appropriate way to give a thorough explanation how the AI can arrive at such conclusions (Cramer & Kim, 2019; Liao et al., 2020). On top of that, there is a distinct dependence on continuous streams of data for some AI systems to function which designers must also consider (Holmquist, 2017).

In practical reality, it should also be taken into account that not all AI are of equal complexity. As Yang et al. (2020) elaborates in their proposed conceptual framework, AI systems can be differentiated based on the degree of *output complexity* (the possibilities of the output, ranging from few to infinite) and *system capability* (the capability of the system, from fixed to evolving) and thus be categorized into levels of AI design complexity.

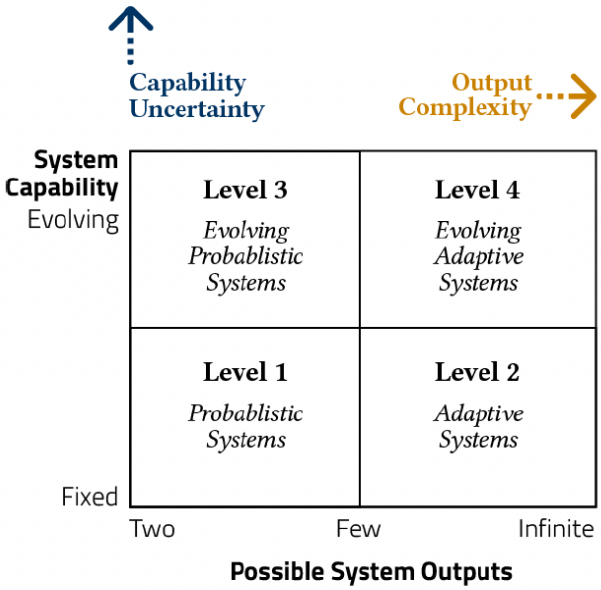


Figure 3. Levels of AI design based on its complexity (source: Yang et al., 2020)

These unique properties leads to some distinct challenges tied to the design of AI systems such as; challenge of designing for a learning system that requires data from active engagement with users (Girardin & Lathia, 2017; Holmquist, 2017), the challenge of expressing and prototyping AI ideas throughout the process (Bratteteig & Verne, 2018; Dove et al., 2017; van Allen, 2018; Yang et al., 2018), and the challenge of grasping the capabilities and limitations of AI technologies for both designers and users (Bratteteig & Verne, 2018).

Against some of these challenges, novel design considerations accounting for the properties of AI are emerging. For prototyping, workflow ideas such as *model-informed prototyping* (see Figure 4) are proposed to illustrate how designers can account for evolution in which actual outputs of AI models are fed into the interface designs, making exploration and evaluation of design choices possible (Subramonyam et al., 2021). Another prototyping tool that could potentially be useful is the Delft AI Toolkit[1] proposed by van Allen (2018), a visual tool to simulate the behaviour flow of an AI system through the changing of input data.

[1] See https://github.com/pvanallen/delft-ai-toolkit for the repository

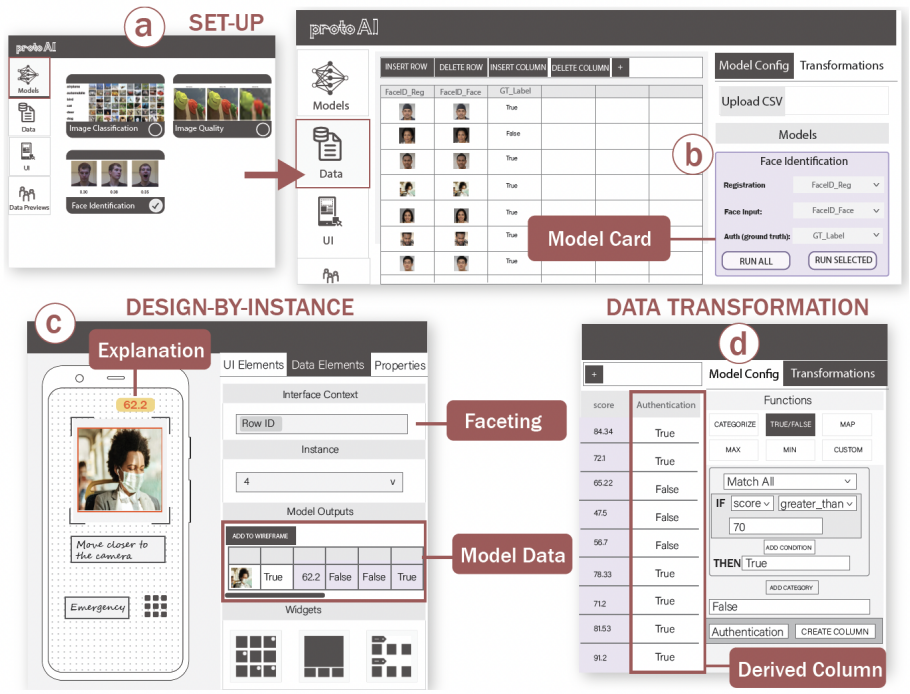


Figure 4. Model-informed prototyping (source: Subramonyam et al., 2021)

Meanwhile other efforts are aimed towards explainability and AI literacy for designers and users alike (Liao et al., 2020; Long & Magerko, 2020) and, to account for the distinct interactions between humans and AI, Amershi et al. (2019) proposes a human-AI interaction guideline consisting of baseline design considerations for AI systems. Furthermore, van Allen (2017) suggests a departure from human-centered design as the designing of AI requires an expansion to consider the greater ecosystem where, instead, the center of design revolves around the system and its outcomes.

Moreover, as the thinking behind the design of technologies can no longer exclude the interconnectedness considerations of the humans in which it is catered for and the pluralities of cultures and environments in which it is situated (Höök & Löwgren, 2021), designers of artificial intelligence can expect to find themselves designing interfaces and systems for pervasive solutions with materials that are increasingly intangible (Rogers, 2012; van Allen, 2017) with capabilities of learning and adapting (Holmquist, 2017; Yang et al., 2020) for people and cultures that are increasingly diverse (Buchanan, 1998), thus potentially leading to increasingly complex situations (Höök & Löwgren, 2021) and 'grand challenges' (Stephanidis et al., 2019).

2.1.2.2. Design as a verb

As a verb, the history of design can be seen from its progress as action. A peek into studies concerning design can reveal the many shades in which it has been defined or understood as an activity (Buchanan, 2001). Design studies have also been concerned with the nature of design practice itself, its epistemology, its relation to science, and the sciences of other disciplines (Buchanan, 2001; Cross, 2001; Krippendorff, 2005; Stolterman, 2008). Nonetheless, definitions are an important starting point to contextualize this body of work.

Throughout Buchanan's work (2001), the basic definition of design is the human power involving the activity of conceiving, planning, and making products. From a more practical setting, Spool (2013) simply defines design as the *rendering of intent*. On the other hand, national organizations that have benefitted from a thriving design industry also refrain from giving a rigid definition of design as an activity (Benton et al., 2018; Danish Design Centre, 2018), opting instead for loose definitions of design as *"a systematic, creative process"* (Danish Design Centre, 2018, p. 1) to provide value and produce various outputs centered on human experiences and behaviour (Benton et al., 2018).

In these definitions, the implications lay in the notion that everyone can be a designer through intentional actions of conceiving and rendering that resemble the broad definition of designing (Spool, 2013). While the differences can be reduced in a way that professional designers rely on an established set of competencies and methods (Krippendorff, 2005) and undergo rigorous training (Kolko, 2011), a recent snapshot of reality from the technology industry sees the increasing demands of professional designers (Maeda, 2017, 2018, 2019) with emerging arguments on the value that design brings (Benton et al., 2018; Buley et al., 2019; Sheppard et al., 2018) alongside the popular adoption of *design thinking*, an approach to problem-solving based on the design mindset packaged for relatively general consumption (Brown, 2008; Kolko, 2018). This suggests the relevance of specialized craftsmanship in design and alludes to the idea that professional designers must embody both mindset and craftsmanship (Kolko, 2011; Krippendorff, 2005).

When paired with a preceding word with contexts pertaining to artifacts, design can also refer to differing strands of the discipline. For example, the field of human-computer interaction, which emerged from the intersection of psychology and computing (Grudin,

2008), contextually applies the design approach to exploring the interaction between two intelligent beings (Winograd, 2006). Interaction design, on the other hand, tends to focus on shaping digital artifacts and creating spaces of action to give structure and form to human environments and activities (Löwgren & Stolterman, 2004, p. 171). While definitions are hard to explicitly pinpoint, there are characteristics that can be useful to denote the action of design. Three characteristics summarized from literature in design studies will be referenced to give an outline of design.

First, design draws a clear distinction from the sciences and engineering (Cross, 2001; Kolko, 2010; Krippendorff, 2005). Whereas the sciences are often focused on analyzing existing things and *how they are*, design is focused on the creation of new artifacts and *how they should be* (Cross, 2001; Krippendorff, 2005; Simon, 1988). Whereas the scientific focus concerns the universal and the existing, design deals with the specific, intentional, and non-existing (Stolterman, 2008). While problem-solving is part of design, it is often approached intuitively with iterations based on continuous reflection (Schön, 1984) rather than with the objectivity and technical rationality prevalent in the engineering approaches (Krippendorff, 2005) used to solve well-formulated *tame problems* (Cross, 2001; Rittel & Webber, 1973).

Second, in contrast to *tame problems*, the nature of the problems designers face are typically of what Rittel and Webber (1973) defines as *wicked problems*: ill-defined problems inherently intertwined with other problems. In this regard, wicked problems are naturally found in situations where values, people, culture, and society are involved and considered (Buchanan, 1992; Friedman & Kahn Jr, 2003; Krippendorff, 2005). Public designs in cities, for example, illustrate complex situations where the outcome may unintentionally improve the lives of some at the cost of others and where problem-solving can be intentionally driven by stakeholders to achieve the interest of certain groups (Krippendorff, 2005; Winner, 1980).

Third, in facing problems of great uncertainty, design can be seen as an approach of finding clarity in the chaos or organizing complexity (Kolko, 2010). This is typically done through abductive reasoning and sensemaking processes. In this, intuitive methods honed through a reflective practice are often employed (Cross, 1982). Similar to most abductive reasoning and sensemaking processes, the purpose of this design approach can then be as a way to creatively produce new knowledge through inference (Fischer, 2001; Kolko, 2010).

From these characteristics, these designerly ways of doing and thinking can function as a guiding composition of what design is and what AI designers are; further implied by the recent work in the intertwining between AI and design (Dove et al., 2017; Girardin & Lathia, 2017; van Allen, 2017; Yang et al., 2018; Zimmerman et al., 2020). To conclude with a useful definition for this research, the working definition of *design* (action) can be then understood as the activities in which designerly ways of thinking and doing and *designers* (subject) as those with the responsibility to enact such activities to create *designs* (object) of AI products, services, and systems.

2.1.3. Socioethical Implications of Artificial Intelligence

Moving to current times, recent advances have enabled narrow-artificial intelligence systems to be integrated into a variety of systems. In Europe alone, AI startups have raised EUR 3.6 billion in 2017, an increase of almost three times more than the previous year, operating in financial, health, marketing, advertising, business intelligence, and automotive industries (European Commission, 2018). Europe is not alone as China and the United States are making significant progress in what some dubs as 'the AI Race' (Castro & McLaughlin, 2021) despite significant risks that may arise from this competitive narrative (Cave & ÓhÉigeartaigh, 2018).

Despite the increasingly widespread adoption of AI (Cam et al., 2019; MIT Technology Review Insights, 2020), the ability to mitigate its associated risks is still at its infancy (Morley et al., 2019). A common example of AI gone rogue is of Microsoft's infamous Tay chatbot which had unintendedly turned into a racist and misogynistic within 24 hours of it interacting with users on Twitter (Ballard et al., 2019; Lee, 2016; Neff, 2016; Stoimenova & Price, 2020; Yang et al., 2020). While this may just be a relatively harmless example, Turchin and Denkenberger (2020) have catalogued catastrophic global AI risks envisioned throughout works, such as the potential destructive capability enabled through mass production of advanced military drones and the gradual displacement of human autonomy and responsibility.

Alongside the increasing adoption of AI is the proliferation of documents aimed at providing guidance regarding AI systems. Many documents take the form of ethical principles that aim to provide normative guidance in the design of AI systems (Fjeld et al., 2020; Jobin et al., 2019) — there are Ethics Guidelines for Trustworthy AI from the European Union (2019), Ethically Aligned Design by the IEEE (2019), and AI principles

by Microsoft (n.d.), Google (2018), and IBM (2021) respectively, just to illustrate a few examples. However, the mere existence of these principles alone are not enough (Hagendorff, 2020; Mittelstadt, 2019; Morley et al., 2019; Whittlestone et al., 2019).

In response, the work of Morley et al. (2019) attempts to bridge the gap between ethics and practice through an extensive typology of existing guidelines, methods, and applied tools onto a set of ethical principles. While the work of Morley et al. (2019) is multidisciplinary in a sense that it compiles guidelines and tools from different fields, it is ultimately developed for the practically-minded machine learning (ML) community to navigate the challenges and dilemmas in designing AI. However, the pursuit of ethically good AI calls for an interdisciplinary collaboration that extends beyond the ML community. What is then the role of the design? How are designers addressing ethical concerns in their designing of AI systems? To what extent are ethical considerations and values applied into the design process?

2.2. Applying Values and Considering Ethics in Designing AI

Artifacts have values (van de Poel, 2020; Winner, 1980). Values are embedded in artifacts (Friedman, 1996) whether through intention or realization (Umbrello, 2019; van de Poel, 2020), through the way in which they are arranged (Winner, 1980), or laden through every conscious design decision (Shilton, 2013). Arguably, this puts the designer with a great amount of responsibility when it comes to designing the technologies of our everyday life (Monteiro, 2019). And designers should be aware of this.

> *"**Design cannot avoid ethical questions.** And, finally, because improvements must be understandable and decidable by those affected, not imposed by lone designers or authorities who are not acknowledged by the community in question, **artifacts must make sense to most, ideally to all of those who have a stake in them.**"* (Krippendorff, 2005, pp. 25–26)

The aim of this section is to give a brief outlook of how design has developed ways throughout time as a means to embody values within artifacts and mitigate socioethical risks.

2.2.1. Values and Ethical principles

As briefly touched upon, the creation of ethical principles can be a way to minimize the risks of AI; as is seemingly the case with the recent proliferation of these documents in the face of disruptive effects brought by the prevalence of AI systems (Floridi & Cowls, 2019). Thankfully, the proliferation of these ethical principles have not gone unnoticed and we can now benefit from the critical summarization provided by recent literature (Fjeld et al., 2020; Floridi & Cowls, 2019; Hagendorff, 2020; Jobin et al., 2019). As these summarizations show, a significant amount of the AI ethical principles share some common ascribed values (Fjeld et al., 2020; Floridi & Cowls, 2019; Hagendorff, 2020; Jobin et al., 2019).

In Hagendorff (2020), 22 AI ethical principles are analyzed and compared. Similarities in some of the values between different sets of principles were found. On top of that, Jobin et al. (2019) observes a global convergence around 5 ethical principles (transparency, justice and fairness, non-maleficence, responsibility and privacy) despite a divergence in how these principles are interpreted. Moreover, Fjeld et al. (2020) beautifully illustrates this landscape of ethical AI principles, denoting where significant commonalities between organizations lie (Figure 5).

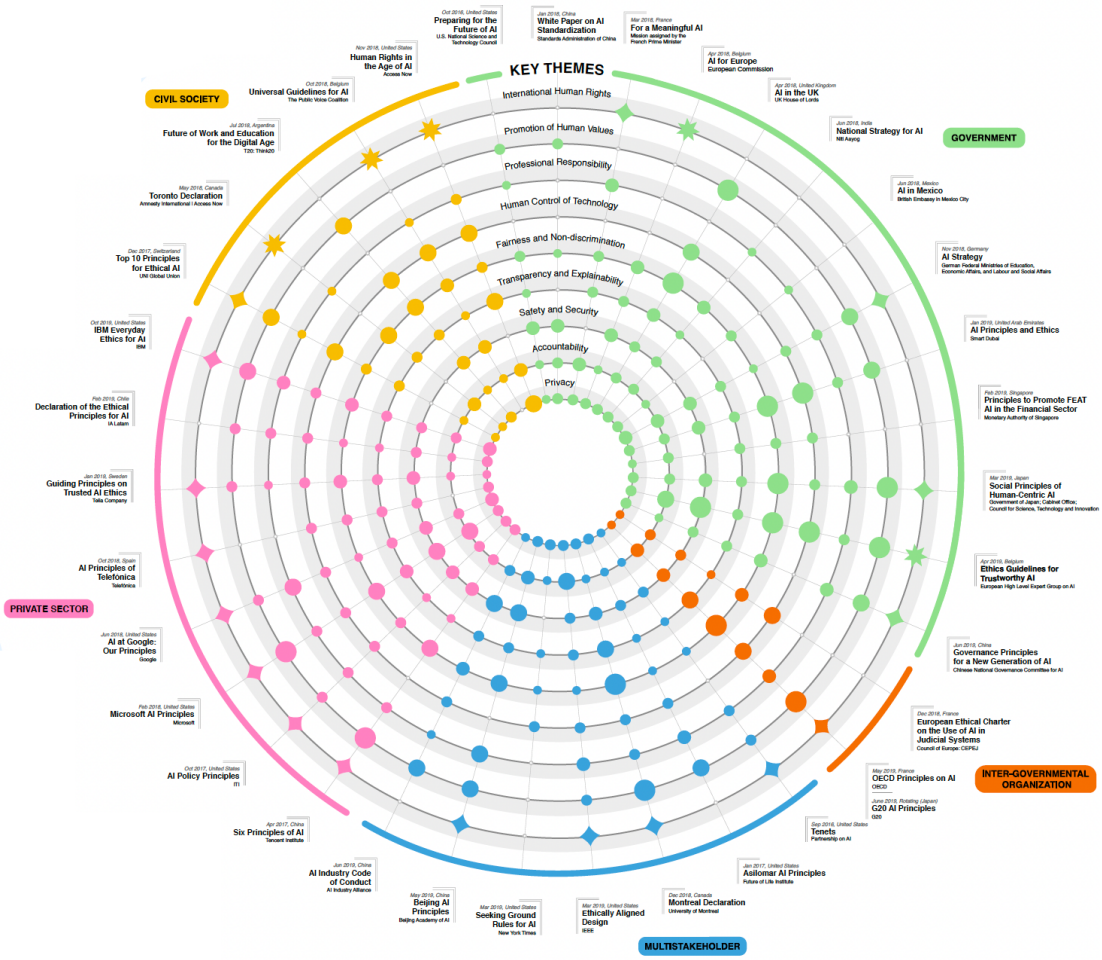


Figure 5. The AI ethical principles landscape (source: Fjeld et al., 2020)

In an attempt to solve the problem of 'principle proliferation', Floridi and Cowls (2019) conducted detailed analysis on existing high-profile documents of ethical guidelines to find that they indeed have a high degree of overlap, which serves as a basis for their proposal of the five core principles for ethical AI: *beneficence*, *non-maleficence*, *autonomy*, *justice*, and *explicability*.

From these examples, we can establish that AI ethical principles refer to high-level documents where certain normative (and oftentimes common) values are ascribed to provide guidance applicable to the creation of AI systems (Fjeld et al., 2020; Floridi & Cowls, 2019; Hagendorff, 2020; Jobin et al., 2019; Whittlestone et al., 2019). In this sense, the values contained in these principles can (and perhaps should) be able to be translated into practice (Hagendorff, 2020; Morley et al., 2019; Whittlestone et al., 2019) and be embedded into the design of AI systems (Dignum, 2017; Umbrello, 2019; van de Poel, 2020). However, a common issue pertaining to the proliferation of these principles is the

challenge in bridging between the abstract (values) and the tangible (artifacts) in designing AI (Hagendorff, 2020; Morley et al., 2019).

2.2.2. Applications in design practice

While the task to reconcile the abstract and the tangible in designing AI can be challenging, design is a discipline supposedly used to facing such complex wicked problems (Buchanan, 1992; Rittel & Webber, 1973). Design has also established the importance of putting human perspectives at the front and center of its processes, shown from ample evidence in both theoretical and practical works throughout history (Devon & van de Poel, 2004; Friedman, 1996). This can benefit the recent calls for a human-centered approach to AI systems (AI HLEG, 2019). Moreover, the embedding of values within designed artifacts has been a recurring focus (Friedman, 1996; Shilton, 2013; van de Poel, 2020) and the calls for design ethics have been echoed numerous times (Devon & van de Poel, 2004; Dignum, 2017; Friedman & Kahn Jr, 2003; Monteiro, 2017), most irrespective of the development of AI technologies.

As elaborated in previous sections, design itself seems to have an established tradition in considering human perspectives (Buchanan, 1998; Krippendorff, 2011; Muller & Kuhn, 1993; Rogers, 2012; Spinuzzi, 2005), which extends to embedding values and considering ethics into designed artifacts (Cummings, 2006; Devon & van de Poel, 2004; Friedman, 1996; Monteiro, 2017; Shilton, 2013). Participatory design is one such approach consisting of various techniques and methods in which, in essence, requires the active participation of relevant stakeholders into the design process, including them in activities and decision-making (Muller & Kuhn, 1993; Spinuzzi, 2005). Through this participatory approach where stakeholders are actively involved, their values become embodied in practice (Iversen et al., 2012).

Value-sensitive design is another approach in which it explicitly emphasises for the values embedded into the artifact through a tripartite analysis through conceptual, empirical, and technological investigations (Friedman, 1996; Friedman et al., 2002). Shilton (2013, p. 376) have also reported on what she calls as *value levers*: "*practices that pried open discussions about values in design and helped the team build consensus around social values as design criteria.*"

On top of that, there is also a growing awareness that algorithmic advances alone are insufficient considering the systems are designed to interact with and around humans

and thus the many resounding call for human-centered perspectives on AI and machine learning (Gillies et al., 2016; Harper, 2019; Ramos et al., 2019; Riedl, 2019). In this regard, we will then shift the focus to the various ways that ethical concerns have been considered and values have been embedded in the design of AI systems where some earlier literature concerning the problematic outcomes of AI in design can be traced back to the field of human-computer interaction (Höök, 2000; Norman, 1994); the work of Norman (1994) from over 20 years ago has already envisioned some of the challenges with AI, namely on user agency and privacy. Höök (2000) elaborates on another set of challenges to consider; partly that AI systems may violate usability principles.

In more recent works, Umbrello (2019) explores value-sensitive design as a methodology in the design and development of AI systems which argues how explicit values can be translated into design requirements and vice versa. Turned into the form of a game, value-sensitive design is also employed by Ballard et al., (2019) in creating 'Judgement Call', a game to surface ethical concerns. More recently, van de Poel (2020) elaborates on how AI as a sociotechnical system can embody values and makes an argument to constantly monitor their consequences and to undertake continuous activities to redesign the system as a whole or its components in ensuring the intended values are realized. The participatory approach to design has also been utilized to explore ethical concerns and values related to AI (Liao & Muller, 2019). Alternatively, there are emerging efforts that focus on addressing the unintended consequences in designing AI (Stoimenova & Kleinsmann, 2020).

A scan of works on designing with societal impact in mind reveals various ways in which ethical considerations and human values can be integrated into design practice. Some of these methods and frameworks can be found on a number of online repositories, with practical examples such as the *layers of effect* method[2], cards to encourage positive behavioural change[3], and writing down ethical contracts[4]. Specific to designing AI, prompt cards seem to be a relatively common solution to trigger ethical discussions early in the design process (Ballard et al., 2019) with another example provided by the AI

[2] https://www.designethically.com/toolkit

[3] https://www.artefactgroup.com/work/#tools

[4] https://www.ethicsfordesigners.com/

agency 33A.ai[5]. These can be useful approaches to clarify design intentions front and center, but its effectiveness in addressing the evolving nature of AI has yet to be seen.

Aside from that, design guidelines catering specifically for human-AI interactions have been formulated (Amershi et al., 2019). Similarly however, it can be argued that some of this approach still lacks the novelty to fully account for the uniquely inherent risks of AI (Yang et al., 2020). Established approaches, such as participatory design, face similar difficulties when put in the context of designing AI (Bratteteig & Verne, 2018). All of this seems to suggest that designing AI remains a uniquely challenging task for designers (Bratteteig & Verne, 2018; Dove et al., 2017; Stoimenova & Price, 2020; Yang et al., 2020).

Recent observations show how designers have been tackling challenges that are unique to AI (Bratteteig & Verne, 2018; Dove et al., 2017; Liao et al., 2020; Yang et al., 2018). In a survey, Dove et al., (2017) outlined challenges faced by UX designers in understanding AI, expressing AI ideas, and enunciating purposeful use of AI (in relation to ethics) in which they further suggest new research directions potentially beneficial for design. Bratteteig & Verne (2018) specifically analyzes participatory design approaches in designing AI systems and recounts similar challenges. Liao et al. (2020) gathered AI designers and explored the opportunities and challenges specific to designing explainability solutions for AI. Yang et al. (2018) conducted interviews with UX designers and provided insightful findings into the processes currently established in designing AI systems.

From these, a challenge that I would like to emphasize is simply this evolving nature of AI capable of outputs that adapts to the situations as commonly shown by the literature (Bratteteig & Verne, 2018; Höök & Löwgren, 2021; Stoimenova & Price, 2020; van de Poel, 2020; Yang et al., 2020).

This evolving nature enunciates a stark difference between AI in training and in normal use, as for AI, use is essentially training (Bratteteig & Verne, 2018). The changing nature of output that can be produced are great hindrances to prototyping (Bratteteig & Verne, 2018; Dove et al., 2017; Yang et al., 2020), a seemingly core tenet to validate concepts common in design practice, and essentially challenges the notion of securing a degree of user agency over the artifact through means of active participation in the design

[5] https://www.33a.ai/ethics

process (Bratteteig & Verne, 2018). And in practice, it seems that few designers have actually accounted for the evolving capability of AI after it is deployed (Yang et al., 2018).

Of course, for the most advanced AI systems, the capability to evolve and produce an almost infinite amount of possibilities implicates the way values can be embedded into the design (Umbrello, 2019; van de Poel, 2020). As noted by van de Poel (2020), what sets AI apart from other sociotechnical systems is that the capability to adapt based on its interaction with its situated environment may undermine the embodied values, thus rendering the intended values unrealized. How then do the established practices of design in applying values and considering ethics play a role in AI development? Moreso, if this implication is indeed the case, how do designers already involved with AI projects apply values and consider ethics in their designing of AI systems?

3. Methodology

Design approaches are theorized to be rather effective in terms of creatively producing practical new knowledge and it has an ability to make sense of complexity through efforts in synthesis and reframing (Cross, 1982; Kolko, 2010). Its logic is neither deductive, in which it seeks to validate theories, nor inductive, in which it builds substantial observations to formulate theories (Fischer, 2001), but rather it seeks to creatively produce new knowledge through abductive inference (Fischer, 2001; Kolko, 2010). Therefore, I am inspired to operationalize this exploratory research as a design project and ascribe myself to the designerly way of doing typically associated with design practice and design research (Buchanan, 2001; Cross, 1982; Kolko, 2010; Schön, 1984) in order to synthesize practical results through the reconciliation of both data and literature.

The primary reference in operationalizing this project is derived from Kolko (2010) in his elaboration on abductive reasoning and sensemaking as the drivers of design synthesis as a methodological approach. Abductive reasoning in itself is an approach to infer an explanatory hypothesis to the phenomenon in question and can be viewed as a creative means to open up new knowledge or change the semantics of a conceptual system (Fischer, 2001). On the other hand, sensemaking is a methodology of disciplining diversity and complexity without reducing it through homogenization (Dervin, 1998).

In Kolko (2010), sensemaking is operationalized in design synthesis as a way of "making sense of chaos" through active effort from the designer in finding patterns and forging connections. In this sense, there are some similarities that can be drawn with the systematic approach of grounded theory common to social science research (Compton & Barrett, 2016).

For this research project, the primary source of insights were both the literature and the findings. Abductive reasoning concerns us in presenting a "phenomenon" to be understood (Fischer, 2001). The literature, as laid out in the previous chapter, illustrates the phenomenon established through relevant arguments and past works in the research area. To complement this, interviews and a workshop session were conducted to gather empirical data as a basis of information on the reality of the phenomena through the cases of conveyed experiences from practitioners. Insights from both literature and findings

were then combined, synthesized, and reframed in an attempt to formulate ideas and questions that could be beneficial for further discussions of the field.

3.1. Interviews and Workshops

To complement the assorted literature and gather empirical evidence on how designing AI systems with consideration to applying values and ethical concerns is approached, data are gathered from two different ways:

1. Qualitative Interviews, a reflective approach to understand in how and why designers were applying values and considering ethics in designing AI
2. Nightmare Scenario Workshops, a generative approach to explore in how and why designers might consider ethical considerations and values in designing AI

For the first step of data gathering, qualitative interviews with experienced AI practitioners were conducted to establish a general idea of how and why practitioners were already taking into account ethical considerations and values in their process of designing AI through a reflection of their professional experiences. For the step, a generative approach was taken through the conduct of an ideation workshop with designers to imagine the ways in which values can be applied in the process of designing AI and why it should be applied. In both phases, inquiry towards why certain ways to apply values and ethical considerations are also made to potentially discover underlying themes of challenges and (re)solutions.

3.1.1. **Qualitative Interviews**

Qualitative interview is chosen as it is a powerful method to produce knowledge from the lived experiences of subjects (Kvale, 2008); in this case, practitioners with professional experiences in designing AI. In general, the qualitative interviews cover the topics of how their AI design processes are, what role they played in, the challenges they faced in designing AI, their thoughts on the implication of AI for societies, how they see AI ethical values and their attempts in applying them to their actual practice. Insights produced from the interviews themselves are used as the basis to complement existing literature that already describes the challenges of designing an ethical AI.

3.1.1.1. Thematizing the Interview

As the nature of this study is rather exploratory, an open approach with little structure to the interview is typical (Kvale, 2008). Instead, the area of the issue needs to be charted. In this case, this means breaking down the research questions into relevant parts that can be used to construct an interview guide.

Starting with the deconstruction of the research question, 3 themes can be surfaced that are relevant to the project. First, the word 'design(ers)' alludes to both the field of design and the designers as the subject of this research. Second, the phrase 'apply values and consider ethics' refers to the action that this research is primarily concerned with. Finally, the setting that question is situated in is indicated by the prepositional phrase of 'designing AI', entailing the specific contexts related to such a process. These deconstructions were then used to thematize the interview as the basis in forming the interview guideline which will be discussed in the following section (Table 1).

Table 1. Thematizing the interview from the research question

Research Question	Deconstruction	Interview Themes
How do design(ers) apply values and consider ethics concerns in designing AI?	'design(ers)'	Design(ers) throughout the process
	'apply values and consider ethics'	Applying values and considering ethics
	'designing AI'	The process of designing AI

3.1.1.2. Forming the Interview Guideline

To account for the diverse backgrounds and experiences of the interviewee, the interviews were semi-structured to allow room for flexibility in exploring the topic (Kvale, 2008). The themes developed earlier were then used as the basis to formulate a general outline of the interview questions (Table 2). The arrangement of the themes and the questions were done with the intention of putting the question of values and ethics last. I refrain from bringing up the issues of human values and ethics upfront to avoid it affecting the way the process of designing AI is told and instead let the participants share their experiences in a candid manner first. The interviews were then loosely structured

to start with the process of designing AI, then identifying the role and contributions related to design(ers), and finally inquiring on how values were applied and ethics were considered in the aforementioned process.

Table 2. Forming the interview guideline from the themes

Research Question	Interview Themes	Interview Questions
How do design(ers) apply values and consider ethics concerns in designing AI?	The process of designing AI	What and how is the process of designing AI?
		What are the challenges in designing AI and how are they resolved?
	Design(ers) throughout the process	What are the roles of designers and how can design contribute in designing AI?
	Applying values and considering ethics	What are the societal implications of AI?
		What are the ways that ethics can be considered and how is it applied in your practice?
		What are the challenges in applying ethics into practice?
		What is the role of AI ethical principles in your practice?

The questions for *designers* and *developers* were respectively differentiated in their delivery. As this work mainly focuses from the *designers'* perspective, extra attention was given when talking to *developers* on how they collaborate with designers and how they see the role of design in working with AI. This is an effort to better triangulate the role and contributions that design can bring to AI-related works.

3.1.1.3. Sampling and Recruitment

To explore the topic of how design(ers) consider values and ethical concerns in designing AI, the interviews were conducted with a purposive sampling (Bryman, 2016) of two types of practitioners with varying degrees in experience, scope, and organizational environment in designing AI. These two types will be referenced throughout this body of work as *designers* and *developers*.

For the practitioners defined as ***designers***, they encompass design practitioners with varying degrees of experience with AI. Similar to previous studies (Dove et al., 2017; Girardin & Lathia, 2017; Liao et al., 2020; Yang et al., 2018) practitioners typically have a formal design education and/or have professional design roles at the organization that they are employed in. In essence, these are designers that conduct the design approach that have been characterized thoroughly in the earlier theoretical background, referencing a distinct emphasis on the designerly ways of thinking and doing as elaborated in the previous sections (Cross, 2001; Kolko, 2010; Krippendorff, 2005; Schön, 1984; Stolterman, 2008).

For the practitioners defined as ***developers***, they encompass engineers with varying degrees of experience with AI. These are practitioners with formal engineering background and/or have technical roles and responsibilities at the organization that they are employed in. This category of practitioners usually approach the creation of AI using a more 'rationalist approach' which puts more emphasis on understanding the internal workings of intelligent agents (Winograd, 2006) through technical means of doing, such as programming the algorithms and developing the AI models.

The purpose of having two types of practitioners for this qualitative interview varies. As this research is approached from a design perspective, interviewing *designers* is essential. By interviewing them, insights can be produced on how designers see AI, their process of designing AI which implies where design can contribute, and how they consider ethical concerns in their processes. On the other hand, interviewing *developers* can complement (or contrast) the design perspective through insights of how they develop AI, how they collaborate with designers and the ways that they see design can contribute, and to how they translate high-level ethical concerns into practice.

Despite the increasing adoption of AI technologies, finding practitioners with relevant experience was still a challenge. Considering the situation, there were 3 ways in which interviewees were recruited:

- From with personal networks
- From referrals by interviewees or personal networks
- From professional social medias, such as LinkedIn

Identifying whether these practitioners were AI practitioners in their own right was a challenge. For participants recruited through personal networks, their experiences were briefly conveyed through a brief chat with me. For practitioners that are referred

by other interviewees or personal networks, their experiences were described by those that were giving them reference. For those recruited through professional social media, a scan of their public resume and work portfolio was done.

This pragmatic approach was to ensure a higher number of data but may impact the diversity of participants' backgrounds and is then a limitation of this research. From that approach, 8 experts were recruited for the interviews (Table 3).

Table 3. List of interview participants

No	Profile	Experience[6]	Organization[7]	From
Designers				
P1	Chief Executive Officer and Co-founder AI design consultancy firm • Designed AI concept solutions for industry clients	> 10 years	AI design consultancy firm with 1 - 10 employees	LinkedIn
P2	Senior Designer Large international tech company • Designed the user interface of an AI prediction tool	> 10 years	International technology company with 100000 - 200000 employees	Referral
P3	AI Design Researcher PhD candidate in the field of AI and Design • Researching about a design approach to predict the unintended consequences of AI	5 - 10 years	European educational institution	LinkedIn
P4	Experience Designer International design agency • Designed a recommender system and a chatbot	< 5 years	International design agency with 5000 - 10000 employees	Personal network
Developers				
P5	Chief Technical Officer AI development consultancy firm • Lead the engineering and data team to develop AI solutions for enterprises	> 10 years	AI development consultancy firm with 50 - 100 employees	LinkedIn
P6	Chief Technical Officer AI development consultancy firm • Directly work with clients to develop AI solutions for enterprises and governments	> 10 years	AI development consultancy firm with 100 - 150 employees	Referral

[6] Experience was approximated through either the interview or the interviewee public professional profile

[7] Organization profile was approximated through public and official sources. In cases where it was not available, it was approximated through platforms such as LinkedIn

P7	Data Scientist Large international tech company • Designed and developed AI models for an antivirus software	5 - 10 years	International technology company with 100000 - 200000 employees	Personal network
P8	Machine Learning Engineer Fellowship at an AI institution • Designed interactive AI products and developed machine learning algorithms	< 5 years	AI institution with 300 - 500 members	Referral

3.1.1.4. Conducting the Interviews

The interviews were conducted in the order of P4, P7, P1, P2, P5, P3, P8, P6 based on their schedule availability. Throughout the subsequent sessions, the interview guideline went through small iterations. While the questions remained the same, the adjustments were more on the delivery and structuring the flow of the session.

The interviews were mostly conducted over telecommunications software Skype[8] and were recorded with consent from the interviewees. The recordings were then processed by the transcribing software Descript[9] to generate a draft transcription of each session. By using Descript, the draft transcriptions were then used as the basis for synthesizing the insights. Further details on the synthesizing process are elaborated in the Synthesizing section below.

3.1.2. Nightmare Scenario Workshop

To explore how design(ers) might consider ethical considerations and values in designing AI, a generative workshop followed by a reflective retrospective approach was conducted. The main objective for the workshop is for designers to generate as many ideas as possible to how they might apply values in designing AI then following it up with a reflection on whether, how, and why these ideas should be applied in practice.

The approach was mainly inspired by a combination of brainstorming workshops, design fiction, and took some elements from participatory design (Bleecker, 2009; Holtzblatt & Beyer, 2014; Liao & Muller, 2019; Muller & Kuhn, 1993). The workshop was intended to be conducted only with *designers*. However, due to constraints, participants with non-design backgrounds were also included in the session.

[8] https://skype.com/

[9] https://descript.com/

3.1.2.1. Designing the Workshop

Although different in its implementation, the workshop takes inspiration from the work of Liao & Muller (2019) where they used participatory design fiction workshops to study how they might embed values in the design of AI systems. With a generative mindset in mind, the workshop was designed to encourage divergent thinking for the participants. This means that in many stages of the workshop the participants were free to creatively explore as many spontaneous thoughts as possible without much restriction. However, at later stages of the workshop, the participants were encouraged to consider the practicality of their ideas. This was an effort to nudge the participants away from immensely impractical ideas and focus more on generating creative ways that are somewhat feasible. These considerations of practicality tie in to the end of the workshop where participants were asked to reflect on whether these ideas were applied in practice — *if not, why do they think this is the case?* This last part is a crucial part of the workshop to uncover knowledge on the realities of applying values and considering ethical concerns in designing AI.

As can be seen from Table 4, the workshop was divided into 5 sessions.

Table 4. Nightmare scenario workshop structure

Overarching Question	**Sessions**	**Objective**	**Rundown**	**Duration**
How do design(ers) apply values and consider ethics concerns in designing AI?	Introduction	Warm up for the session and prepare participants in the context of AI	Introduction to the workshop	5 mins
			Brainstorming	5 mins
			Quick Presentation	3 mins
	Step 1 - Defining values of a "good AI"	Surface the values that designers find relevant to an AI that enables great experience and ensures good for society	Instructions for participants	2 mins
			Brainstorming	7 mins
			Quick Presentation	4 mins
	Step 2 - Violating the values through	Explore the worst scenario possible situations and ways for AI systems to go wrong	Instructions for participants	2 mins
			Brainstorming	10 mins

	nightmare scenarios		Quick Presentation	5 mins
	Step 3 - Fixing / preventing the violation	Generate ideas and potential ways for design(ers) to consider ethical concerns and values in their practice	Instructions for participants	2 mins
			Brainstorming	10 mins
			Break	5 mins
			Brainstorming	4 mins
			Quick Presentation	10 mins
	Retrospective	Reflection for participants on whether these ideas are applied in practice which may lead to interesting afterthoughts and new perspectives	Retro and discussion	12 mins
			Comments and feedbacks	2 mins
			Closing	2 mins

The introductory step serves as a warm-up. In this stage, the participants were asked for brief introductions followed by a short brainstorm to get familiar with the context of AI. To do so, participants were tasked to individually define what AI is to them. They were given the freedom to elaborate AI in technical terms, describe its characteristics, give examples, or explain it in any way it suits them. Aside from being a warm-up for the participants, this stage was beneficial to the analysis as it gives a way to contextualize the perspective of each participant through the way they understand AI.

The first step of the workshop (Figure 6) tasks participants to gather their thoughts on what values makes an AI good — both in terms of delivering a great experience and serving to benefit society. As a way to stimulate this, participants were also encouraged to think of great AI that they have come across and synthesize the embodied values that they see within those examples.

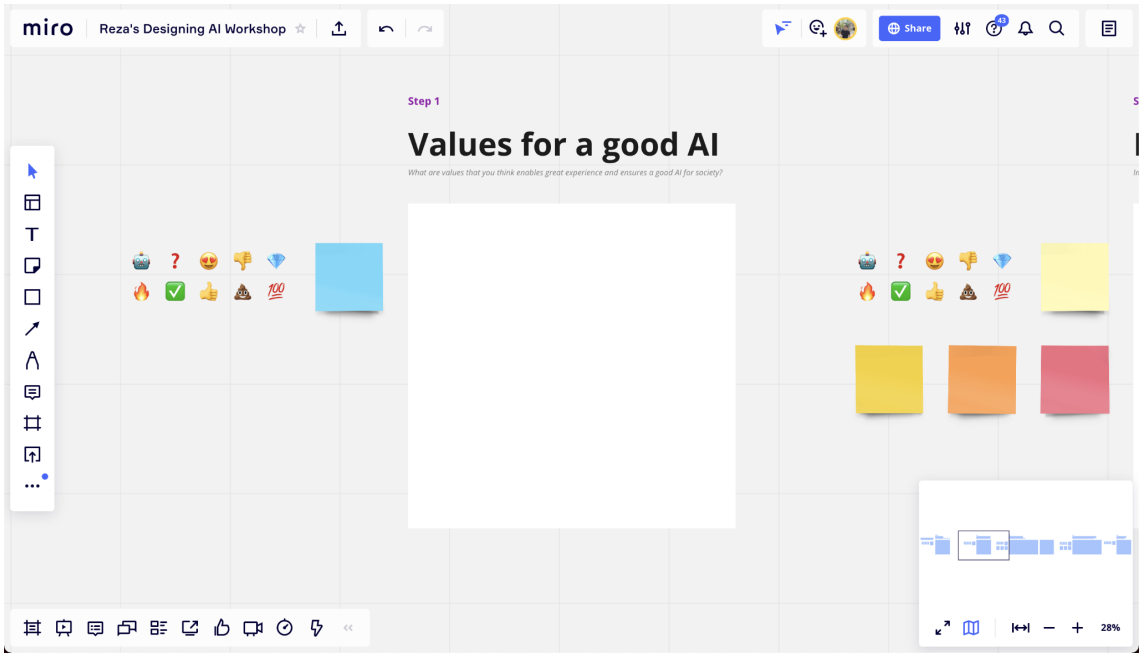


Figure 6. Step 1 of the nightmare scenario workshop

For the second step of the workshop (Figure 7), participants were tasked to dream of *nightmare scenarios* — the worst possible scenario imaginable that can happen. To do so, participants were asked to reference any of the values generated from the previous step, then think of ways for imaginary AI systems to violate those values and imagine situations that can lead to these hypothetical nightmares. At the end of the step, participants were then asked to choose a minimum of 3 *nightmare scenarios* of their liking in preparation of the final step of the workshop.

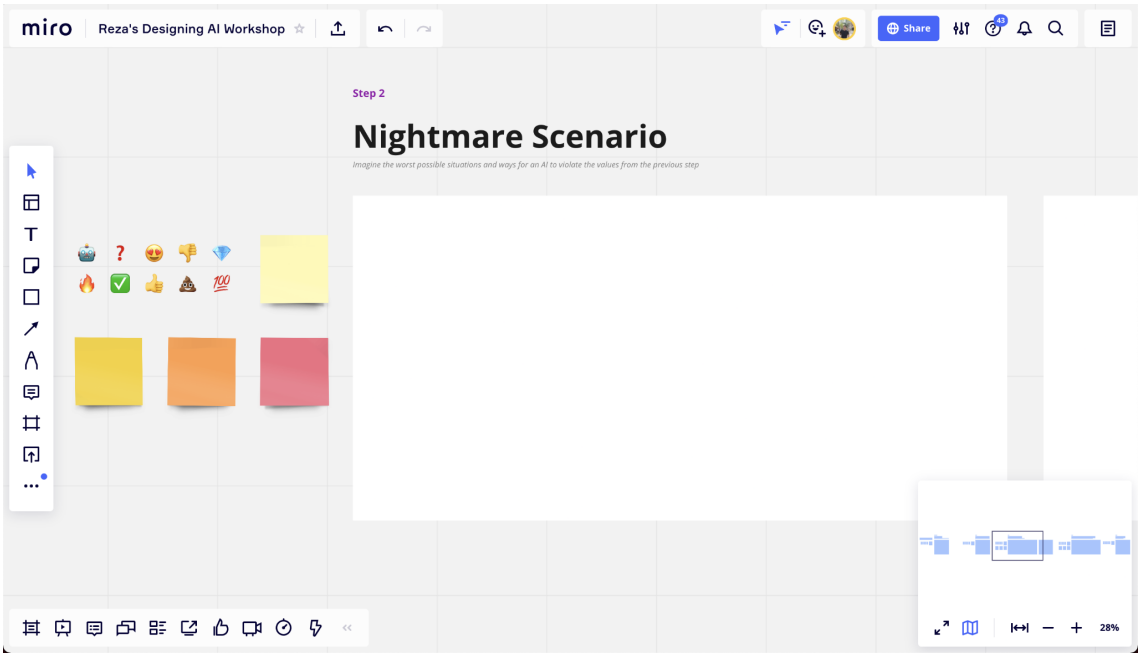


Figure 7. Step 2 of the nightmare scenario workshop

Having chosen a set of *nightmare scenarios*, participants were then asked to envision themselves being in the position of designers with the power over these problematic imaginary Ais (Figure 8). *How might they fix these problems*? More importantly, *how might they prevent these nightmares from happening*? Given the thought, the participants were then tasked to come up with practical ideas as an answer to their chosen AI nightmares.

Last but not least, participants were asked for a quick debrief to internalize this experience. They were tasked to reflect then share their thoughts on how this workshop relates to their daily practice. In retrospect, *what is being done in reality? Are these ideas applied in practice? How applicable are these ideas? If not, why?* This last step was conducted to capture interesting afterthoughts and new perspectives while also serving as a way to close the session.

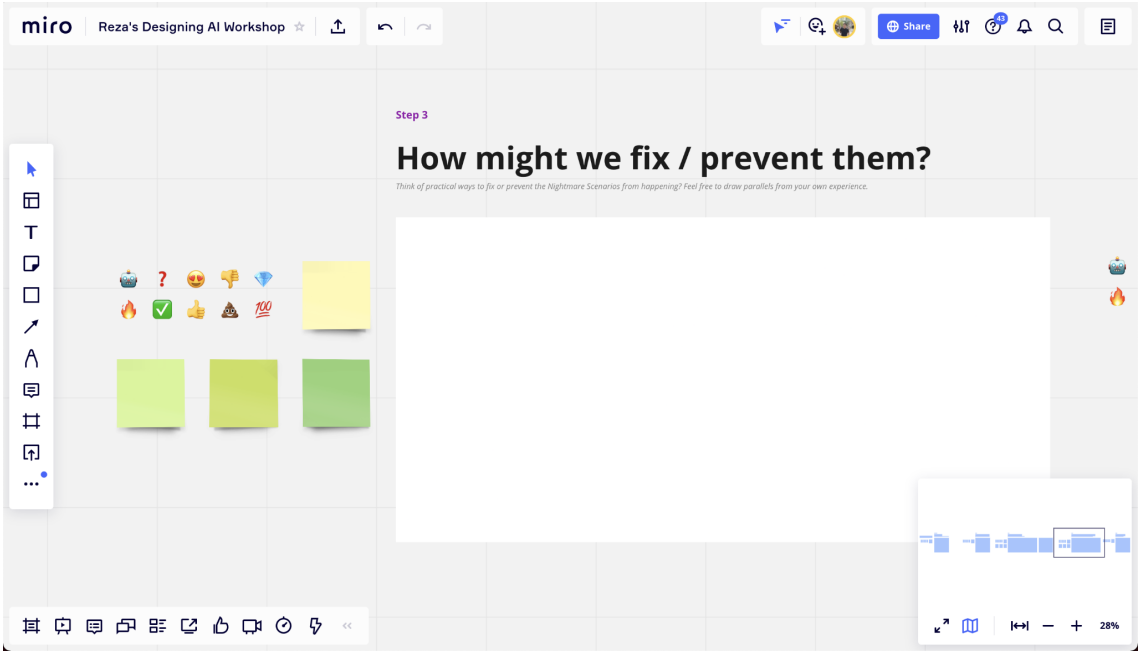


Figure 8. Step 3 of the nightmare scenario workshop

3.1.2.2. Sampling and Recruitment

To achieve the goal of this phase, the workshop was intended to be conducted with primarily *designers*. Similarly with the qualitative interviews, *designers* here are defined as practitioners with either a formal design education and/or have professional design roles at the organization that they belong to as similarly implied by previous investigations (Dove et al., 2017; Girardin & Lathia, 2017; Yang et al., 2018) and is characterized by the designerly approach elaborated within the theoretical background. For this workshop, two types of *designers* were recruited:

1. Designers experienced with AI projects
2. Designers with minimal experience or have no experience with AI projects

Having these two types of designers with contrasting levels of familiarity with AI was an effort to better stimulate the workshop. Designers experienced with AI projects can draw a lot from their expertise but may have an unconsciously limited perspective for generating new ideas due to their familiarity with the landscape. On the other hand, designers without prior experience with AI projects may not be familiar with the properties of AI as a design material (its capabilities, limitations, and how it works for example) but may give a fresh unbiased perspective to the workshop.

Participants were recruited through a personal network of AI designers from an AI design community[10] that I am a part of. They were recruited through a public announcement within the communication channel of the community. Through the announcement, interested community members were directed to a pre-workshop form to elaborate their background, indicate their experience with AI and their schedules.

However, in reality, it was difficult to get participants of relevant experience for this workshop partly due to the seemingly lack of interest and perhaps the timing of the session. I would assume that Zoom fatigue[11] amidst the pandemic would also be a contributing barrier as some of the participants that expressed interest in joining did not show up for the session. The lack of highly relevant participants may affect the results of the analysis in that they may not be able to represent the intended inquiry on professional AI designers and is thus a limitation of this study. Nonetheless, the workshop was successfully conducted with 5 participants (Table 5).

Table 5. List of workshop participants

No	**Profile**	**Experience[12]**	**From**
W1	UX Designer + Researcher Leading a collaborative project on AI ethics	Highly familiar with AI and has professional experience in designing AI	AIxDesign Community
W2	Data Scientist + Commercial Manager Data scientist that has shifted to commercializing AI and automation offerings	Highly familiar with AI and has professional experience in developing AI	AIxDesign Community
W3	Arts Researcher PhD candidate researching on how technology can be appropriated towards social justice and equity in the arts	Vaguely familiar with AI and has no experience related to AI	Personal network
W4	Graduate Student in Design Master in Digital and Interaction Design	Highly familiar with AI and is studying the design of AI	AIxDesign Community
W5	Graduate Student in Design Master in Strategic Product Design	Highly familiar with AI and has experience in designing AI	Personal network

[10] https://www.aixdesign.co/

[11] https://news.stanford.edu/2021/02/23/four-causes-zoom-fatigue-solutions/

[12] Experience was approximated from a pre-workshop form that participants submitted

3.1.2.3. Conducting the Workshop

The workshop was conducted online by using both Miro[13] and Google Meet[14] simultaneously. Miro is a visual collaboration tool that has the features to run an online workshop while Google Meet was used as a telecommunication medium to communicate with the 5 participants. As most of the participants have full-time commitments the workshop was conducted on a workday during the after work hours at 17:30 - 19:00 CEST. This timing was suitable as some participants were located in the US and that this schedule would translate to 8:30 - 10:00 PDT.

While the workshop was conducted according to plan, small adjustments in the timings of each step was made. A participant was also late to the workshop and another had to leave before the last step of the workshop. However, aside from that all 5 participants were fully engaged and the workshop seems conducive. The resulting Miro board can be seen below (Figure 9).

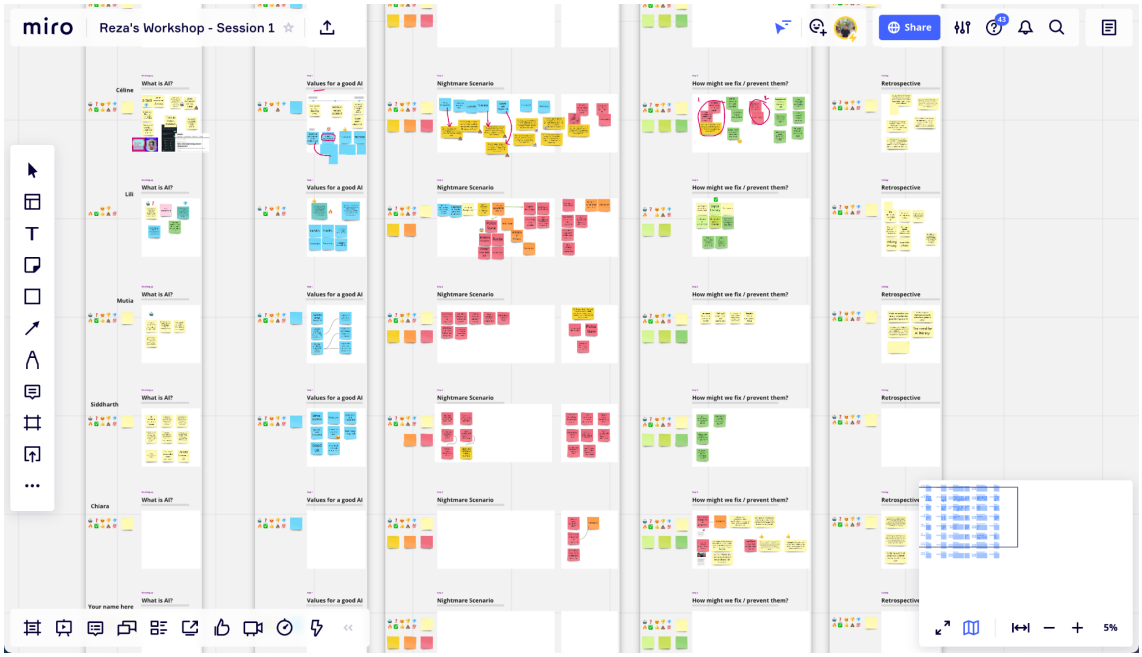


Figure 9. The Miro board after the nightmare scenario workshop

[13] https://miro.com/

[14] https://meet.google.com/

3.2. Synthesizing

> *"Design is always about synthesis — synthesis of market needs, technology trends, and business needs."* (Jim Wicks, Director of Motorola's Consumer Experience Design accounted in Kolko 2010)

From both the interviews and the workshop data, design synthesis will be conducted. As explained earlier, design synthesis is essentially an abductive sensemaking process (Kolko, 2010). Throughout forging of new connections and finding new meanings, synthesis is fundamentally a means to apply abductive reasoning and infer conclusions by applying a variety of sensemaking techniques (Kolko, 2010; Naumer et al., 2008). In this sense, this may be similar to what Kvale (2008) refers to as bricolage from the social sciences, an eclectic form of analysis for interviews beneficial in generating meaning for qualitative data that lack an overall sense at first reading.

As an overview, the ways in conducting the design synthesis were:

1. Noting down elaborations, perceived patterns and tensions, and interesting remarks from both the interviews and the workshop as a tacit way to elicit and describe the insights
2. Open coding on the interview transcripts to surface distinct concepts
3. Affinity diagramming as a way to consolidate the results of the open coding together along with the workshop data to surface themes and form categories and subcategories
4. Combining insights from both the findings and the literature and reframing the results through a proposed framework

3.2.1. **Eliciting Insights**

The first process is to elicit insights from both the recordings of the qualitative interviews and the result from the workshops. In analyzing the interview, the draft transcription provided by the automatic transcribing software Descript was the starting point. The draft was then corrected through a re-listening of the recording and through the draft, initial notes were jotted down as personal reference to signify potentially

interesting remarks. I particularly looked at the interviewees' answers, elaborations, and other interesting remarks made throughout the session.

After all the transcriptions were corrected, a second listening of the recording alongside thoroughly reading each line of the transcript was conducted as a process of open coding. For analysis of interviews, open coding is usually done first to surface distinct concepts that can be grouped into categories later on (Bryman, 2016; Kvale, 2008). In this stage of initial coding, there is emphasis to generate as many new concepts and ideas as possible (Bryman, 2016).

To conduct open coding, the transcriptions were imported to the free open-source coding tool Taguette[15] and through a line-to-line analysis of the text highlights were made on segments that were interesting. Through a form of *meaning condensation*, a technique to compress statements into concise forms while retaining the main idea (Kvale, 2008), the highlighted text segments were then condensed into sticky notes and were placed as *insights* on the digital whiteboard Miro to give an overview visualization (Figure 10).

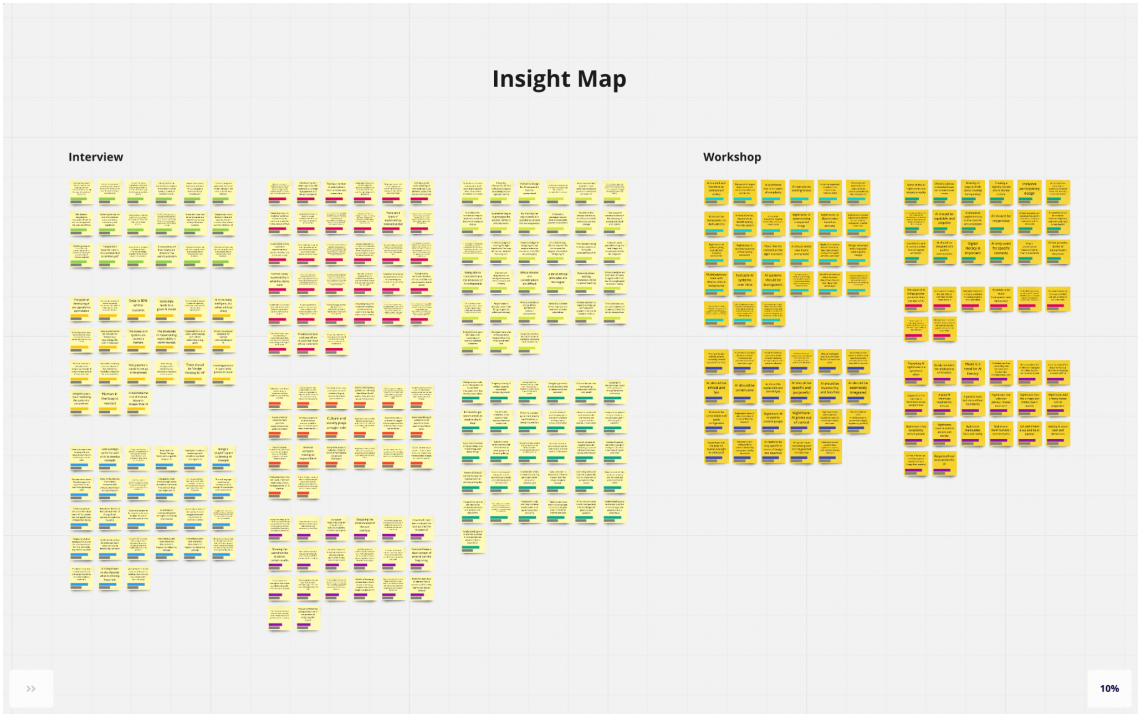


Figure 10. Insight map of data from both interviews and workshop

[15] https://www.taguette.org/

In analyzing the workshop results, similar procedures were undertaken. The notes, scribbles, and illustrations made by participants from the workshop were analyzed and went through a similar process of meaning condensation to form sticky notes of insights. The sticky note insights from both the interviews and the workshop were then grouped based on participants. To conclude, the result of this phase was a map of elicited insights from all the data which were then used for the affinity diagramming discussed in the next section.

3.2.2. **Affinity Diagramming**

Having the insights from both the interviews and the workshop mapped out, affinity diagramming sessions will be conducted. Affinity diagramming is a common activity in design practice originating from ethnography practice (Plain, 2007) usually conducted to make sense of the data through the clustering of insights (Dam & Siang, n.d.; IDEO.org, n.d.), grouping them in a way where each cluster results in describing a certain issue (Dam & Siang, n.d.; Holtzblatt & Beyer, 2014). In a sense the goal of this affinity diagramming technique may be similar to what is commonly referred to as selective or thematic coding (Compton & Barrett, 2016), the process of revealing core categories by focusing on the most common codes and what is usually judged as the most revealing about the data (Bryman, 2016).

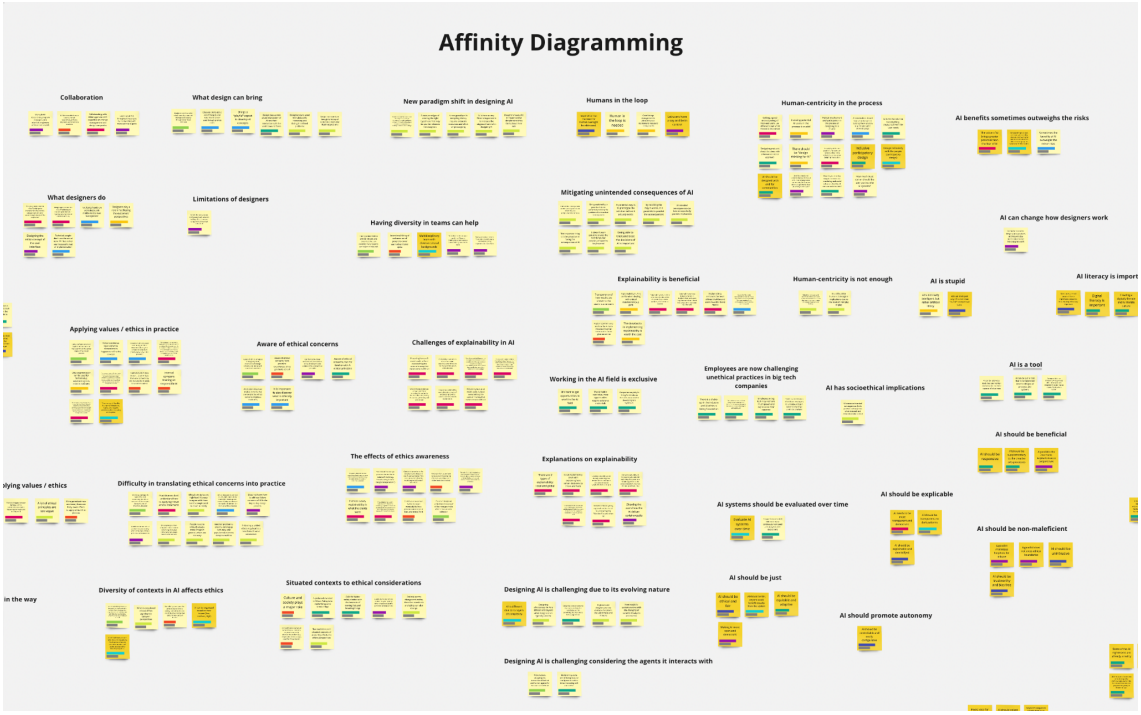


Figure 11. First round of affinity diagramming

The first session of the affinity diagramming (Figure 11) managed to cluster the insights into groups each representing a distinct idea or concept such as ‘having diversity in teams can help’ or ‘human-centricity in the process’. While this was useful to accentuate commonalities and contrasts between insights, it resulted in too many categories. As it is common for these processes to involve iterations (Bryman, 2016; Holtzblatt & Beyer, 2014), a second session of affinity diagramming to further cluster the groups was then conducted.

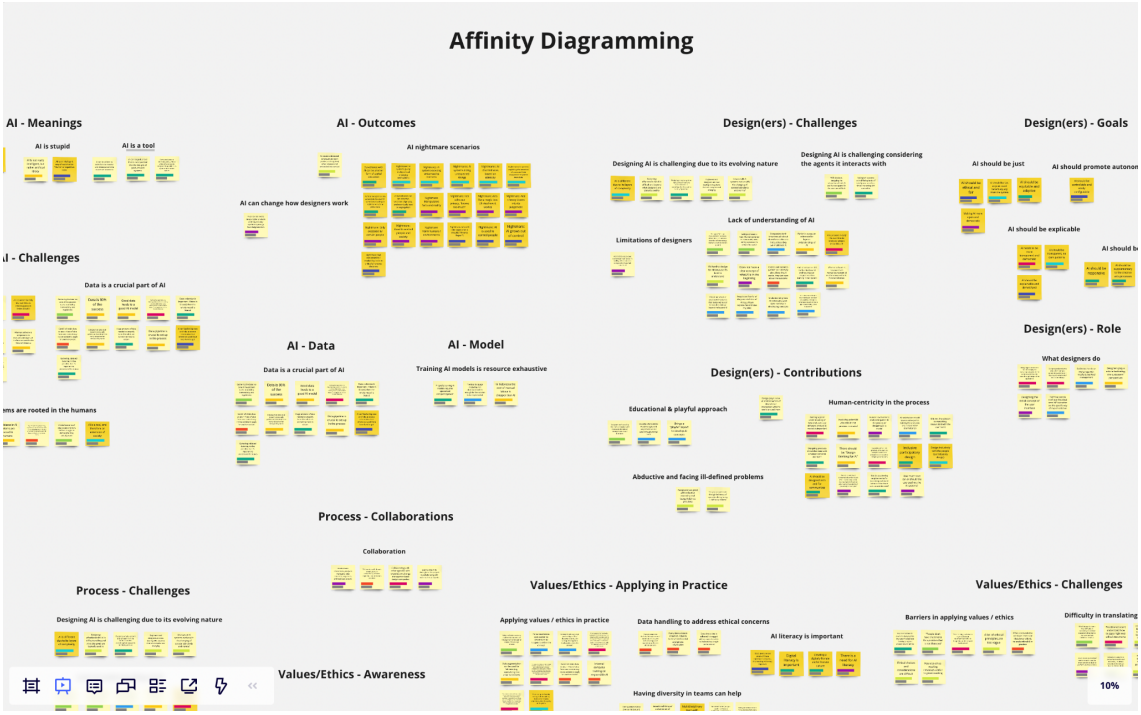


Figure 12. Second round of affinity diagramming

The second round of affinity diagramming (Figure 12) was sufficient enough to result in categories and subcategories, as seen in Table 6 below, each containing a general aspect of the bigger issue (see Table 12 in the Appendix for more details). These categories and subcategories conclude the findings from the data and thus became the basis for further analysis and used as the unit of synthesis to combine insights taken from the existing body of literature.

3.2.3. Insight Combination & Reframing

In reconciling between the findings and the literature, the approach of *insight combination* was adopted. As part of the abductive sensemaking paradigm conveyed by Kolko (2010), insight combination is a method to establish pairings between what is considered as design insights and design patterns; respectively *design insights* refer to combination of an observation and knowledge while *design patterns* refer to the trends and repeated elements that appear in the design (Kolko, 2010).

Table 6. Categories and subcategories as the result of affinity diagramming

Subcategory	Scope
Category 1 — The foundationals of AI	
1A — What is AI	The meanings and definitions of artificial intelligence.
1B — Why AI	The reason and goal behind the utilization of AI.
1C — The data	Data as a technical part of AI.
1D — The model	Models as a technical part of AI.
1E — The output	The outputs of AI systems.
1F — The challenges of AI	The general challenges revolving around the implementation of AI systems.
1G — The outcomes of AI	The perceived outcomes (potentially) brought by the implementation of AI technologies.
Category 2 — The process of designing AI	
2A — The process	The processes pertaining to the approaches and the processes in designing AI.
2B — Collaborations throughout the process	Identifying collaborators and ways of collaborations between differing roles in the process of creating AI.
2C — Challenges during the process	General difficulties and challenges encountered during the process of designing and developing AI.
Category 3 — Design(ers) in the process of designing AI	
3A — The goal and role of design(ers)	The goal and role of designers in the process of creating AI systems.
3B — Design(ers) contributions	Ways in which design have or can potentially contribute to the development of AI systems.
3C — Challenges for design(ers)	Design challenges and difficulties faced by designers specific to designing with and for AI technologies.
Category 4 — Applying values and considering ethics in designing AI	
4A — Awareness of ethical considerations and its effects	The awareness towards ethical considerations and its subsequent effects.
4B — Reasons of applying values and considering ethics	The reasons as to why values should be applied and ethics should be considered for AI systems.
4C — Applications of values	Existing and potential applications of how values are applied and

Subcategory	Scope
and considering ethics in practice	ethics are considered in the day-to-day practice of creating AI systems.
4D — Barriers in applying values and considering ethics in practice	The challenges and barriers that creates difficulty in applying values and considering ethics in the practice of designing AI.
4E — Complexities in applying values and considering ethics	The complexities rooted in the diversity and situated nature of values and ethics.

For this project, the 'design insight' was an analogous reference to the findings of both the interviews and the workshop session while the 'design pattern' was used in referring to the summarization of established literature on the research topic. This summarization was created from the basis of the collected works discussed in the theoretical background.

Table 7. Literature highlights

Highlights	Literature Examples
AI presents many new opportunities and risks	• Castro & McLaughlin, 2021 • Cave & ÓhÉigeartaigh, 2018 • Floridi et al., 2018 • Turchin & Denkenberger, 2020
Designing AI poses new challenges that are inherent to its unique nature	• Bratteteig & Verne, 2018 • Holmquist, 2017 • Höök & Löwgren, 2021 • Stoimenova & Price, 2020 • Yang et al., 2020
Designers needs to and are coming up with new ways to consider the novel challenges in designing AI	• Amershi et al., 2019 • Dove et al., 2017 • Girardin & Lathia, 2017 • Lovejoy, 2021 • Stoimenova & Kleinsmann, 2020 • Subramonyam et al., 2021 • van Allen, 2017, 2018 • Yang et al., 2018 • Zimmerman et al., 2020
The development of AI can benefit from having human-centered perspectives	• Gillies et al., 2016 • Harper, 2019 • Ramos et al., 2019 • Riedl, 2019
Ethical principles are made in response to the inherent risks of AI	• Fjeld et al., 2020 • Floridi and Cowls, 2019 • Hagendorff, 2020 • Jobin et al., 2019

Translating AI ethical principles into practice	• Mittelstadt, 2019 • Morley et al., 2019 • Whittlestone et al., 2019
Design theoretically presents the many ways in which values and ethics can be embedded in practice	• Devon & van de Poel, 2004 • Dignum, 2017 • Friedman, 1996 • Shilton, 2013 • Umbrello, 2019 • Van de Poel, 2020

As the shared affinities between the two tables above (Table 6 and Table 7) were synthesized and combined, *reframing* is conducted. In design synthesis, reframing is known as *'a method of shifting semantic perspective in order to see things in a new way'*, allowing the exploration of novel associations and hidden links (Kolko, 2010, p. 23). For the result of the reframing, I propose a framework to frame further discussions for the synthesis of both findings and literature. These frames were developed iteratively and are based on how my understanding of the topic has evolved throughout the project and represents the growing complexity professed by the differing lenses in which one can choose to see the problematics of AI.

Table 8. Insight combination between findings and literature

Reframing	Findings Categories	Literature Highlights
AI as is	The foundationals of AI	AI presents many new opportunities and risks
AI as a design material	The process of designing AI	Designing AI poses new challenges that are inherent to its unique nature
		Designers needs to and are coming up with new ways to consider the novel challenges in designing AI
	Design(ers) in the process of designing AI	The development of AI can benefit from having human-centered perspectives
AI as a sociotechnical system	Applying values and considering ethics in designing AI	Ethical principles are made in response to the inherent risks of AI
		Translating AI ethical principles into practice

		Design theoretically presents the many ways in which values and ethics can be embedded in practice

The following chapter will describe in detail the findings of both the qualitative interviews and the workshop session. Further elaborations on the framework and discussions surrounding the topic will be delivered in the Synthesis chapter.

4. Findings

This chapter will describe the findings within each category. Consequently, the chapter will be divided into 4 sections; starting with an overview of AI in general, to describing the process of designing AI, to by how design(ers) are within the process, and lastly followed by an elaboration on applying values and considering ethics in the practice of designing AI.

Table 9. Categories and subcategories of the findings

Category	**Subcategories**
Foundationals of AI	1A — What is AI 1B — Why AI 1C — The data 1D — The model 1E — The output 1F — The challenges of AI 1G — The outcomes of AI
The process of designing AI	2A — The process 2B — Collaborations throughout the process 2C — Challenges during the process
Design(ers) in the process of designing AI	3A — The goal and role of design(ers) 3B — Design(ers) contributions 3C — Challenges for design(ers)
Applying values and considering ethics in designing AI	4A — Awareness of ethical considerations and its effects 4B — Reasons of applying values and considering ethics 4C — Applications of values and considering ethics in practice 4D — Barriers in applying values and considering ethics in practice 4E — Complexities in applying values and considering ethics

4.1. The foundationals of AI

"AI is not really intelligent, but rather artificial idiocy" is probably something you would not have expected to hear from the Chief Technology Officer of an AI development firm (P6). However, as P6 illustrates, the term *artificial intelligence* may have varying definitions and carry different meanings reflecting individual experiences in contrast to what is commonly established. In this regard, P6 refers to how AI may not be as intelligent as people like to think in which they refer to systems that typically learns from data

reflecting human behaviour and is thus, potentially, a cumulative of human decisions both smart and stupid.

Similar to more popular definitions, others define AI as technology that is able to simulate intelligence (W3, W5) capable of smart decisions through patterns and statistical inference (W2). Others see it as merely a tool to be considered as part of the design process (P8) that is distinct in a way that it's dependence on the data gathered makes it an extension of society (as illustrated by W1 in Figure 13 below). An observation by P1 mentions that while the term "AI" is widespread, it is rarely used by developers as they often opt for more technically accurate terms. In contrast, P4 reports the lack of comprehension commonly found from their interactions with Indonesian clients where the layman explanations of its features are preferred.

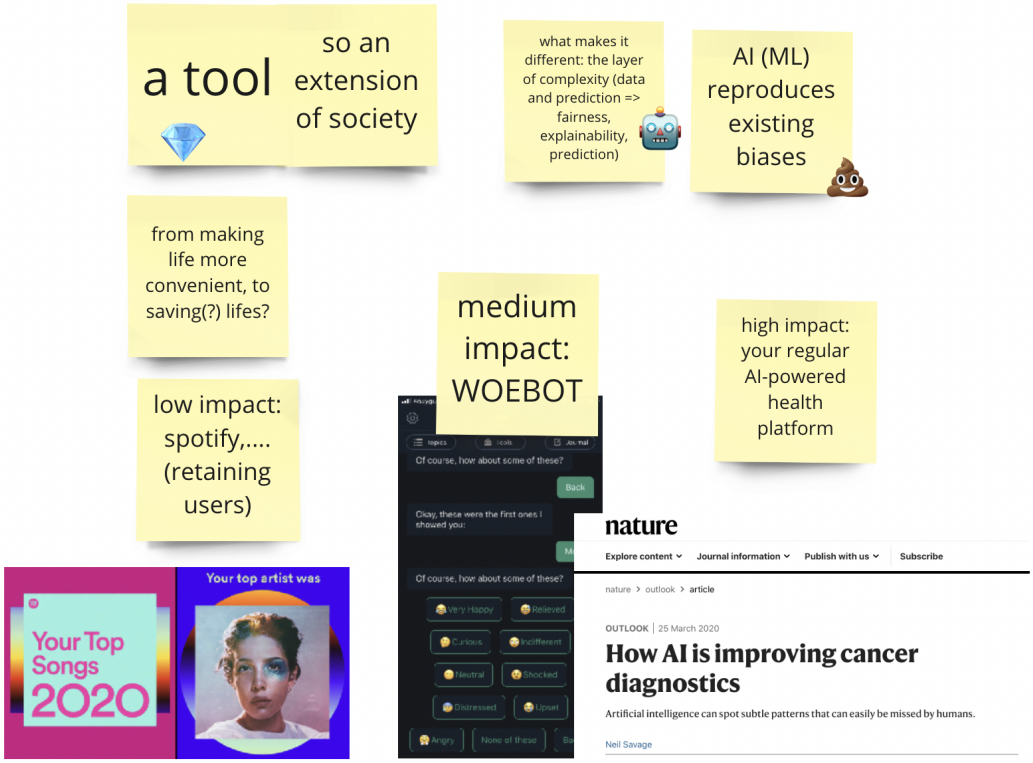


Figure 13. AI as defined by participant W1

Regardless of the differing range of definitions and meanings ascribed to AI, there seem to exist commonalities in its appeal of usage. For example, both P1 and P6 recall examples of previous experiences where AI is implemented to optimize certain processes. On the other hand, P1, P4, W3, W4 all recount the automation benefits that AI

can bring. While it may obviously differ depending on particular use cases, the value of implementing AI seems to converge on either optimization (P6), automation (P6), and novelty (P1, P3, W1).

A common and integral aspect of AI is in its data. From their experiences, P4, P5, P6, P7, and P8 have stated the importance of data for AI. For example, as both CTOs of AI development firms, P5 and P6 typically account for the feasibility of designing AI solutions through data factors such as its availability and quality. On top of that, P4 has reflected on their experience that the gathering of relevant data is a challenge that they have encountered numerous times, a similar challenge for P5, P6, and P8.

> *"And you know, and oftentimes large datasets aren't that publicly available. Unless you're… No, except for specific types of problems, these data sets are available, but if you want to do like a very… If you want to create a product for a very kind of specific niche, that data is probably not available unless you're already a company who is kind of gathering that very specific kind of data."* (P8)

This requirement for data can also be seen as a capability as W2 notes that AI systems can feed on big data to extract hard-to-get information. For W1, it is through this integral need for data itself that AI can be seen as an extension of society. It is without doubt that according to the participants that data is a crucial part of AI and perhaps lend credibility to the expression by P6 that *"data accounts for 80% of the success [for AI systems]"*.

The data gathered and collected over time are typically used as inputs to train AI systems, either initially as training data or over time depending on the requirement (P5, P6, P8). The models here are typically implied as the ways in which an AI learns from the data given (P5, P7, P8). In this case, as P1, P5, and P8 have eluded, the results of an AI system depends on the sophistication of the model implemented, with the more refined models being enablers of more advanced capabilities. However, as a slight note, the challenge related to training AI models is the significant amount of computational resources that it requires (P3, P8).

While the data and models of AI systems can be identified as distinct concepts uniquely tied to AI, the outputs of said system can vary significantly. According to P8, there are systems that are inherently AI in nature with its primary function rooted in machine learning, such as the voice-assistant Alexa, and there are systems that

implement AI capabilities without fully relying on it. Meanwhile, W1 differentiates its outputs based on the impact that it manifests, with examples from low impact such as recommendation stations, medium impact as an AI agent for mental health, and to high impact with the example of AI systems deeply embedded in public healthcare. One thing that is seemingly common is the distinctly evolving nature of its outputs (P3, P8) and the layers of complexity associated with its process of producing an output (W1, P5, P6). Either way, many participants are inclined to believe that the outcomes of AI technologies are rather disruptive to society (P1, P2, P3, P4, P6, P8, W1, W2, W3, W4, W5).

Alongside its disruptive effects lie the perceived implications of AI systems, in both its most exciting and terrible forms. For P1, one of the impacts that AI technologies will bring is centered on its capabilities to efficiently automate significant amounts of manual labour while it can also be seen as an enabler in creating new value propositions. As P2 imagines, the prospect of having a highly intelligent personal AI assistant was thought-provoking enough to trigger discussions on what it means to be human in a world riddled with artificiality.

> *"I think I'm both fascinated about this, but it also worries me, but perhaps that is not the biggest issue… Can you trust what AI gives you if it's to a question you don't understand why the answer is like it is. Can you act on that when it's really an emergency matter or what happens to the world if we don't understand why we do stuff anymore? Plus if we do it because something tells us it's the best way to solve a problem, I think it's important to find out how to keep sense and meaning for humans when using these kinds of tools… "* (P2)

There are even more AI innovations, both strange and convenient, that may already have set into motion implications for our not-so-distant future. In a reflective example, W5 enjoys the benefits brought by the convenience of Google Photos but wonders to what extent their images are being used while W3 mentions Berenson[16] the primitive critic robot conceived by Denis Vidal and Philippe Gaussier that may give birth to the strange

[16] See https://www.vice.com/en/article/aenq45/robot-art-critic-berenson

(perhaps futuristic) vision of AI criticizing human works of art. And as P3 curiously asks, what is Microsoft going to do with its patent to revive dead loved ones as chatbots?[17]

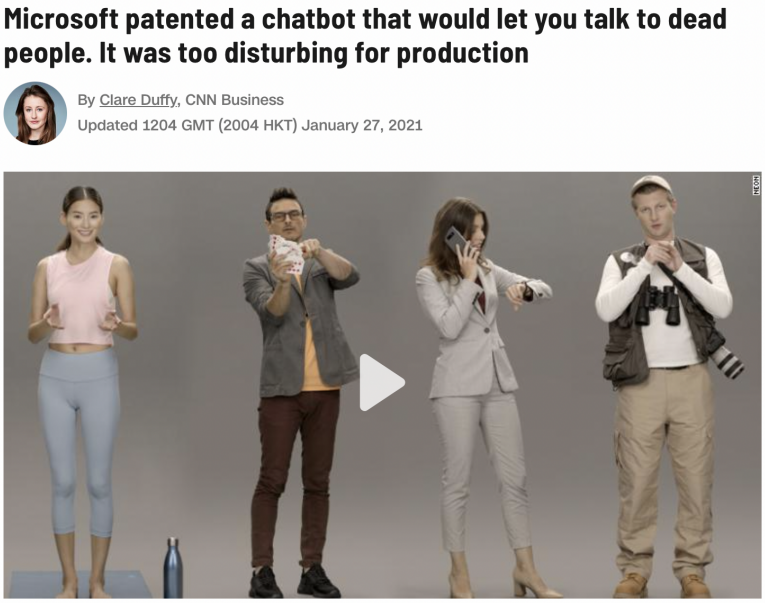


Figure 14. Related news on the topic mentioned by P3 (source: CNN)

In dreaming up the nightmares made possible by AI, workshop participants came up with interesting scenarios with equally horrific consequences. To illustrate some of these imagined nightmares:

- W1 dreams of a scenario of AI being completely unreliable and doing the exact opposites of what people want,
- W3 gave a reasonably familiar narration of an AI facial recognition system evolving to become a tool for discrimination and segregation for the eventual police state (see Figure 15),
- W5 gave some ideas of AI being deceitful, deceptive, harmful, while being both controlling and incomprehensible,
- And ultimately W2 takes inspiration from both Black Mirror[18] and The Terminator[19] franchise in envisioning an AI system given too much power and control over humanity.

[17] See Figure 14 and https://edition.cnn.com/2021/01/27/tech/microsoft-chat-bot-patent/index.html

[18] A reference to the film series Black Mirror (see https://en.wikipedia.org/wiki/Black_Mirror)

[19] Referencing Skynet (see https://en.wikipedia.org/wiki/Skynet_%28Terminator%29)

With bleak realization, the workshop participants later realizes that the seeds to some of these “nightmares” have already been planted in reality.

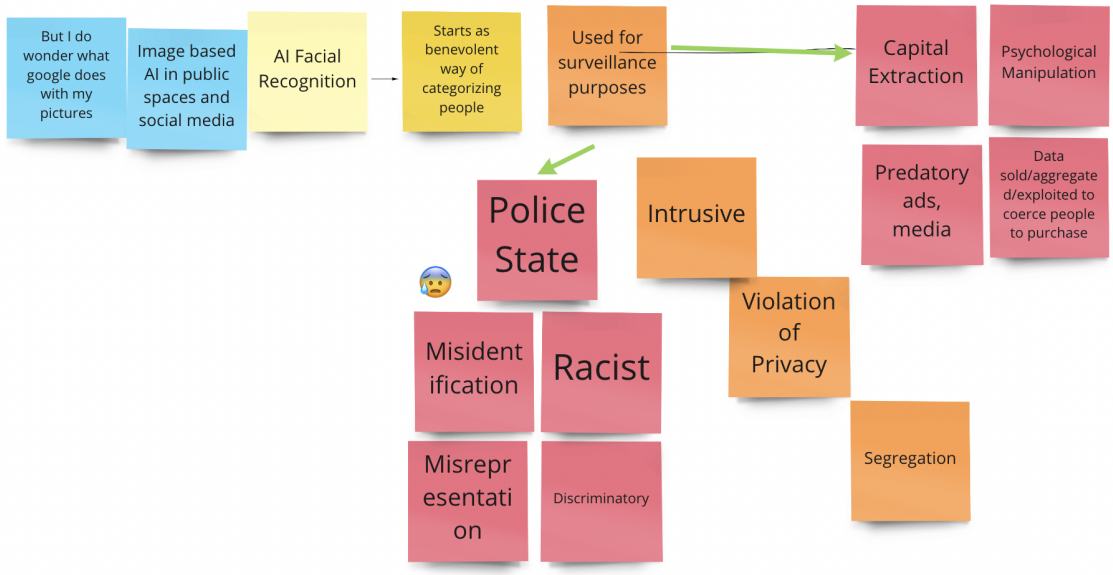


Figure 15. AI nightmares as dreamt up by W3

Despite the potential and perhaps inevitable implications, most participants can still see the many benefits enabled by AI. In that, some believe that these advances are here to stay (as shown in Figure 16) and that ways of figuring out ways to mitigate and minimize its negative consequences are paramount (P3, W2, W4, W5).

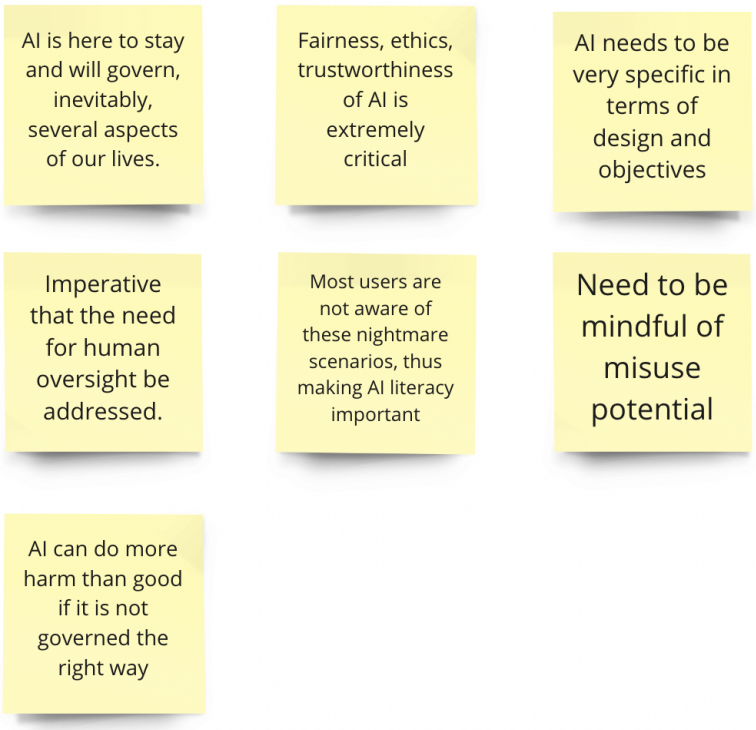


Figure 16. End of workshop reflections by W5

To summarize, some key insights can be inferred from both the qualitative interviews and the workshop sessions regarding the general view towards AI as it is:

- As a technology, AI can be perceived in various light and hold vastly different meanings based on the differing experiences of the participants.
- Data is crucial in that its availability and quality is a typical challenge in the development of AI systems and that it significantly affects the output created by such systems.
- Models are the ways in which the AI learns from data and that the level of capability for the system is usually associated with it.
- While AI can be utilized in many ways and have a different range of usages, it is typically perceived to be implemented in a way where its outputs evolve in tandem with the collected data over time and there are layers of complexity associated with the process of producing an output.
- The increasing effect of AI technologies on everyday lives can be perceived as inevitable and that there is a resounding agreement in the need to minimize its negative consequences.

4.2. The process of designing AI

As with any other design practice, the process of designing AI relies on situational contexts and can vary from case to case. From the interviews, practitioners approach the designing activity with different considerations in mind. From the qualitative interviews, the practitioners that are designers by profession (P1, P2, P4) approach designing AI with a human-centered approach, just like how they would with any other project. This generally means that they would go through the steps of understanding the users' needs and problems to generate the potential ways to solve it. To illustrate the process, P4 accounts of their experience as a consulting experience designer in the process of designing a recommender system for a large technology company:

> *"So we do foundational research and then make user journeys. We identify there are several types of user journeys depending on the context. And what we do is that we craft the ideal recommendation system. We created a storyboard for it and then we made a really big workshop attended by big stakeholders. [...] by actually doing a workshop and getting feedback together, we grounded more into like this is what the company can actually do now and then create a roadmap."* (P4)

The consideration in using AI as a material integral to the designed solution differs between practitioners. This difference may simply be related to the organizations in which they belong to. As P1 heads an AI design firm, they are more focused on seeing the ways in which AI can solve the problem right from the beginning. Moreover, the clientele that they are working with are typically organizations that are already interested in implementing AI. In most cases, they immediately start by identifying user needs and map opportunities where AI specifically can bring benefit. Similarly, P5 belongs to an AI development agency and so the need to use AI was clear from the very beginning.

On the other hand, P2 and P4 are not part of organizations focused on AI. For them, the idea of using AI technology was something that came up further down the process; it was a necessity that was identified after some foundational studies. For example, the decision to build a recommender system only became apparent to P4 as the need for it was eventually identified after initial research and in later stages of collaborations with the team:

> *"For the travel company one, it never started as an AI specific project. It started as exploratory research, and that's what they commission us for. We did foundational research on their traveling journey and when the research was done we shared that to the whole company and okay, there was a business unit that was interested to actually utilize the results. And then one of the team was saying 'we are creating a product, and maybe we need a recommendation ecosystem. Our recommendations are so messed up. Can you help create the ideal way based on your foundational research?'"* (P4)

It is worth considering that while AI is very distinct in its materiality, the processes most participants are concerned with remain relatively similar with the design process of digital technologies in general. As their process started out without any prior intention of designing for AI, the process applied by P4 was not unique to AI technologies, despite facing uniquely challenging problems. In their process to design interfaces, P2 made additional efforts to research best practices to display outcomes predicted by AI but otherwise seem to approach the project like any other work. Additionally, the AI design process that P1 has implemented successfully for its clients was adapted from the Google Design Sprint:

> *"[talking about the origins of the AI design process] … And so there I already used some design thinking, workshop methods, card sorting systems and I saw that within a couple of hours they were actually able to develop first concepts for companies. And so I used this technique or methodology like a design process. At the same time, the Google Design Sprint came out and I just merged that …"* (P1)

In terms of involvement and collaboration, the designers have indicated that they have worked closely with technical developers. For example, P2 shares their experience working with computer scientists and researchers, P4 closely works with engineers, product managers, and regulation experts. What is interesting to note is that having close collaborations with more technical minded practitioners has helped the designers learn and understand more about AI.

> *"I think I got an idea about how it was working with some of the people that created the code and had other perspectives on that and learned a lot from that during the process."* (P2)

In comparison to the processes, the developers (P5, P6, P7) describe a more technical approach in designing AI. The examples of P5 and P6 are elaborated as they tend to have experience in developing customer-facing AI solutions. For P5, an initial step to help clients in designing for an AI solution is to have initial workshops for brainstorming different solutions, discovering potential areas of interests, and considering the availability of the data. For these workshops, P5 would often collaborate with design agencies *"that does that way better than we do."*

It is interesting to note that while the processes of designers P1 and P4 are similar in conducting initial workshops, P5 specifically raises the issue of data availability early on, implying data as a crucial element of AI perhaps distinct to workshops to design solutions for other technologies and perhaps P5's inclination towards the technical side of AI. In relation to this, P4 reflects that it would have been better for them to consider the inputs of an AI system (data) in the conceptualization workshops with stakeholders instead of just focusing on just the outputs.

> *"I think what was missing is that we never really approach it as an AI project. What's missing is… So we get the output like, okay, this is what the system can recommend, but we never… Considered the inputs and what data that you actually need to be able to recommend this. So I feel like it can be more ideal if we were also mapping the inputs."* (P4)

On the other hand, P6 would begin with discussions with clients on the business value chain to identify where AI could bring the most value in terms of business. This ties in with P6's common goal of implementing AI as means to automate and optimize for cost reductions. Having identified where AI could bring the most value, feasibility analyses are then conducted to consider factors such as the availability and quality of existing data alongside the complexity of the algorithm needed to be implemented to train the model. Only after that would a pilot project then be developed as proof of concept.

A common pattern in their process is that they would often give focus to technical feasibility such as the availability of data (P5, P6), the kinds of data that is to be collected (P7), or the resources needed to train the models (P8) in the earlier stages of the process. However, it is interesting to note that despite differences, they all imply that human-centricity and involving users within the process is an important aspect in designing AI (P5, P6, P7, P8).

> *"But, I think for the most part, if you're designing products, you should be approaching that from a human-centered perspective, really understanding what are the needs of your users and designing the product around that instead of trying to shove AI as a solution to everything."* (P8)

To contrast this perspective, P3 foresees that a human-centered design approach of doing participatory workshops is not enough and suggests that a new paradigm in design practice may be needed. This may be related to a common challenge encountered in the process regarding the difficulty in comprehending the materiality and capabilities of AI (P3, P6) and communicating its distinct nature effectively to users (P2, P5).

As a design researcher highly familiar with this issue, P3 alludes to a gap in operationalizing the design processes for AI contexts and are currently exploring multiple ways to "front-load" the predictions and consequences of AI models before they are being developed. If successful, P3 sees that this novel approach could help in testing the outcomes of AI systems without having to spend substantial costs and resources in prototyping.

Another seemingly common issue encountered in the process is the lack of understanding of AI capabilities. P1, P4, P6, and P8 gave examples from their experience in dealing with organizations with relatively minimal knowledge of AI. This becomes problematic when key stakeholders possess deeply misinformed opinions on the capabilities of AI and often cite robots in science fiction as reference (P6). In another case, lack of understanding may come from the designers themselves as P2 assessed that they did not necessarily have a thorough understanding of AI from the beginning when given the task to design a forecasting project, but was fortunate enough to be able to learn from colleagues with relevant expertise.

From all the qualitative interviews, some key insights can be inferred from the data. To summarize:

- Designing AI can be tackled with the usual human-centered approach as explained through first-hand accounts from P1, P2, and P4 with supporting evidence from P5.
- All of the design process primarily revolves around conceptualizing and building the solution through a preconceived notion of AI requirements (data) while not much is reported in accounting for the evolving capabilities of AI in the earlier stages of the design process.
- Participatory or human-centered approaches alone may not be enough, as stated by P3. The seeming lack of novelty applied in designing AI despite its very distinct properties indicates a gap in operationalizing the design process for AI.
- A seemingly common challenge in the process is on understanding AI and the extent and limitations of its capabilities.

4.3. Design(ers) in the process of designing AI

Based on the participants' various experiences, role and contributions of design(ers) in the process of designing AI varies. As elaborated in the previous section, some of the processes started out through exploratory research to identify user needs. In this stage, the potential value gained through adopting AI technologies is likely explored and evaluated with the help of designers (P1, P4). When the choice to implement AI has been confirmed, designers are then tasked to conceptualize the solution through participatory workshops with stakeholders (P1, P4, P8). Design expertise comes in later on to conceptualize and envision the interface for interaction between the AI system and its users (P2).

> *"I was again the one to come up with an initial design concept for the UI [...]. I mean, there was of course a lot of thinking before I started this product on what it should be able to do and how it could show these kind of things and this kind of information. And so my role was to find out how we could put this together and make this a tool where people can explore these data in a good and meaningful way."* (P2)

These perceived roles and contributions of designers in working with AI are also elaborated by some of the developers interviewed. For example, P5 is a CTO at an AI development agency and therefore works with many clients to develop AI solutions. In the process of developing such solutions, design agencies are often called to help in conceptualizing the solution. Design agencies are called in for their expertise in conducting early-stage workshops and *"figuring out what kind of applications is beneficiary"*. Aside from that, P5 notes on how design plays an important role in building an interface as a bridge between the AI system and the end user.

> *"And we're also collaborating with a few design companies helping us to facilitate these workshops. So, there's some really good companies out there that does that way better than we do. So they work very well in the brainstorming part of it, the initial brainstorming, figuring out what kind of applications is beneficiary."* (P5)

Coming from a similar developer profile, P6 likewise gives their support in the importance of design in the process of creating AI systems as they have also enthusiastically proclaimed the need of a "design thinking for AI" multiple times throughout the interview. They then recount an experience of deploying a human-centered and participatory approach in designing an AI system to optimize logistical delivery in which the initially skeptical and resistant dispatcher employees were involved.

Within these tasks and responsibilities that designers have taken on throughout the process, some of the ways in which design contributes can be highlighted. While telling the stories of conducting workshops, P1 recounts a common fear plaguing the many participants in coming to the workshop, rooted in their inexperience with AI. In facing these scenarios, P1 employs their AI cards created for a participatory approach to bring a more playful atmosphere, easing the participants into understanding AI allowing them to comfortably come up with new concepts by the end of the session. Considering the problem of the common lack of AI understanding mentioned in the previous section, perhaps design can be seen as contributing to the ways in which these concepts can be better communicated.

"So we have over 60 cards. On each card is a specific AI technology. And usually then on the backside, there are also some cases and we sorted them into categories. So we have, I think, seven categories. And so that gives a little bit of an overview of… These are all AI technologies and then those categories, there is a little bit the sense of, okay, here is, like, computer vision or here is like 'I can talk' or I don't know… Some prediction and stuff like that and sort of put them a little bit into categories." (P1)

Having collaborated with AI developers and heard about the constant challenges in acquiring data that are consistent in their formatting, P4 sees that problem as an avenue in which designers can contribute to. In an example, P4 mentions that designers could help in this problem by understanding the consequences of designing input elements, such as forms or calendar widgets, and setting a constant standard in their formatting between different use cases.

For developers, the common response to what design can contribute is typically in their human-centered approach, bridging the gap between the technology and the users (P4, P5, P8). Though, in a more fundamental perspective, P3 sees that the essence in which design can contribute is through the established tradition of abductive thinking and facing ill-defined complex problems.

"And what I think one of the things that design can offer and designers are good at, well, that's a very big statement, but I think design, we're very good at what is called abductive reasoning, the notion of synthesizing information, and dealing with ill-defined problems. [...] So what my dissertation is doing, I'm taking principles of design and I'm just applying it to the development of AI. So I obviously think there's a lot of things that design can contribute, but I think if we have to boil it down to one thing is what I mentioned is dealing with ill-defined problems and being able to create a solution that deals with that, that addresses that because the engineering is… It's not that they're not good at it, rather they've never operationalized it, which we have 60, 70 years of history of us doing that. So, I think that's our biggest... No, not advantage, but biggest contribution that we can give." (P3)

Nonetheless, there are some challenges related to design and encountered by designers as conveyed by some of the participants. First, there is the challenge of

designing that is capable of learning from its environment and changing its behaviours accordingly (P3, P8). According to P3's assessment, there is a gap in operationalizing design practice to accommodate constant change after deployment.

> *"So essentially, the example with the hammer, when you see a hammer, there's a handle. So, you know, it fits my hand. I know I can pick it there, and that's an affordance. And when you do it with products and even services, you sort of can create and have a higher level of control over the affordances, essentially of how something can be used. It can, of course, be used in different ways. Uh, for instance, a hammer can be used to smash someone's head, but that's obviously, you know, you can say 'This is not my responsibility. My responsibility ends here and this was someone deciding to smash someone else's head is entirely up to their own volition.' That's something that we, you know, can have a very clear cut line. With AI we don't have that, because you cannot really design for it, insisting that certain affordances can be designed, but because when we talk about AI, most people actually mean machine learning. And with machine learning, the thing learns based on data that you kind of feed it. Of course you train it on data, you have control, but then once it's released into the wild, it starts behaving in different ways."* (P3)

Aside from that is the common challenge in fully grasping the capabilities of AI (P2, P3, P4, P8). As a developer with some background in design education, P8 elaborates that the intricacies of AI may be difficult to grasp for most designers.

> *" … Having worked deeply in the sort of mathematics behind it and the sort of computational theory behind machine learning, I think that's a really difficult space to kind of bring design into this because that sort of research field is evolving really fast. And prior to just like a lot of… Kind of like background knowledge and understanding of how these systems work on a mathematical and computational level. That's kind of like, yeah, that that'd be a difficult area for designers to come in."* (P8)

From all the qualitative interviews, some key insights on the role and contributions that design(ers) can bring in the process of designing AI can be inferred from the data. To summarize:

- Designers primarily contribute at the earlier stages of the process, taking on roles related to conceptualizing the solution concepts and crafting the user interfaces.
- While a common perception on the perceived contribution that design can bring is the human-centered approach, design can also play a role in making AI comprehensible for inexperienced audiences. Furthermore, at its core design is a field with an established history of employing abductive reasoning and facing ill-defined problems.
- There are some challenges related to design and encountered by designers as conveyed by some of the participants, mainly due to the properties of AI being hard to grasp and comprehend.

4.4. Applying values and considering ethics in designing AI

> *"I think that kind of intersection between interaction design and AI is really interesting to me because you can have these autonomous systems, which think for themselves and behave on their own. And just like the way you kind of train AI as the way they behave, you can design how they interact with the world or the way they interact with people. And that kind of all sounded very fascinating to me. And it seems pretty critical for designers to be involved in that intersection because there's a lot of ways you can create technology that exploits people. A lot of people are fearful of AI and technology in general because of the big data revolution that's happened. A lot of companies are using that to take our data and profit off of it with real repercussions without limitations on how they can invade our privacy. So, yeah, I just felt designers should be more involved in that space."* (P8)

Indeed. As briefly mentioned in the first section of this chapter, most participants are aware of or can perceive the implications of the disruptive effects in relation to AI systems. In this section, focus is shifted towards the potentially negative consequences and how values are applied and ethics are considered in the process of designing AI. But

first, let me elaborate on what, both interview and participants, imply as values that constitute a "good" AI.

Throughout the earlier steps of the workshop, participants are tasked to think of good AI examples that they are familiar with and analyze the values that they think are embodied within it. The values that they come up with can then be assumed as something that represents their subjective expectations of a good AI. Further elaborations were explained in the Methodology chapter.

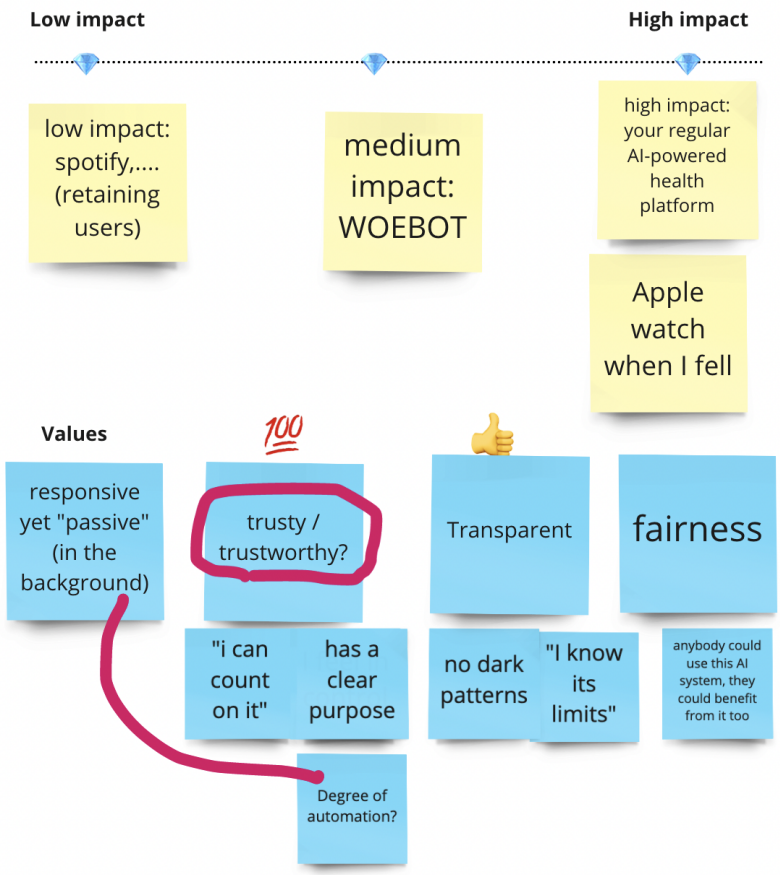


Figure 17. AI values important to participant W1

As illustrated by Figure 17, the workshop participants came up with some values that reflect existing AI ethical principles: trustworthy (W1, W2), transparent (W1, W3), fairness (W1, W3), benefits people's lives (W5), minimizes loopholes for misuse (W5), controllable (W2), and explainable (W2). Aside from these, other values seem to be related to the expectation of the interaction with AI systems such as responsive (W1, W3) and convenience (W2, W5).

On the other hand, interview participants were probed to gather their thoughts on the ethical discussions surrounding AI. In consequence, some of the discussions with participants revolve around certain values such as human control (P1, P2, P4), explainability (P2, P5, P6, P7), fairness (P4, P6, P7, P8), and data privacy (P4, P7). In the case of P3, they decided to discuss the topic of ethics in a broader term rather than having it revolve around certain values.

There are various ways in which participants relate these values and ethical considerations to their practice of designing AI. For the workshop participants, they were asked to imagine practical ways of fixing or preventing the nightmares that they have dreamed up from happening. The results range from what I see as abstract ("making AI more open and democratic" by W5) to some relatively practical ideas ("prioritizing multicultural and non-partisan databases" by W4). Some ideas echoed the statements of the interviews such as: having a multicultural and diverse team (W1, W3) and having a participatory and human-centered approach to designing the AI (W1, W3, W4). Some other ideas seemed relatively novel: creating a platform that analyzes AI systems where users can participate in "fixing" the flaws (W4) and perhaps allowing users to gain a share of profits for personal data that they have willingly given up (W1).

To complement these ideas, the interview participants gave some examples of how they have applied values and considered ethics in their practice of designing AI. As P1 deals with clients from multiple backgrounds and perspectives, they see that a crucial element is bringing ethics into the discussions at the earlier phases of the development. For this, they have created AI ethics cards which can be used to prompt discussions. These cards are designed as a means to systematically frame the ethical concerns into productive discussions. Coupled with their AI design sprint, they claim that this approach can help in translating abstract ethical discussions into concrete actions.

> *"So what we have is 18 cards, so 18 different ethical topics. They're not always on the top of your mind, there's too many. So the cards help you. What happens in the AI Design Sprint is the team develops the first concept and then has an ethical discussion. And ethical discussion means it's not a discussion, it's really an organized, systematic way to to consider ethical perspectives. [...] So for example, if I'm an employee in HR and now it's about process automation, and now we have an AI concept within my HR department. Then I would go through those cards and see,*

okay… Which AI ethical cards would be relevant for this concept and maybe it's three cards or five cards. And then I would make a decision to see which of those cards is most important for me. I pick one card and that opens up an area and then I would use a post-it and really write down okay, within this field, what is really important for me in this small or ethical field relating to this concept and then they would describe it. So the purpose is to find out what is ethically important for them." (P1)

P4 shares similar circumstances in that they face multiple clients with potentially differing values. Throughout their experiences with different clients, they have at various times conducted participatory sessions with stakeholders on raising awareness and uncovering ethical concerns using methods from DESIGN ETHICALLY[20], an online toolkit composed of various tools and frameworks. Some of the methods that they have used are specifically: dichotomy mapping, layers of mapping, and confusion matrix.

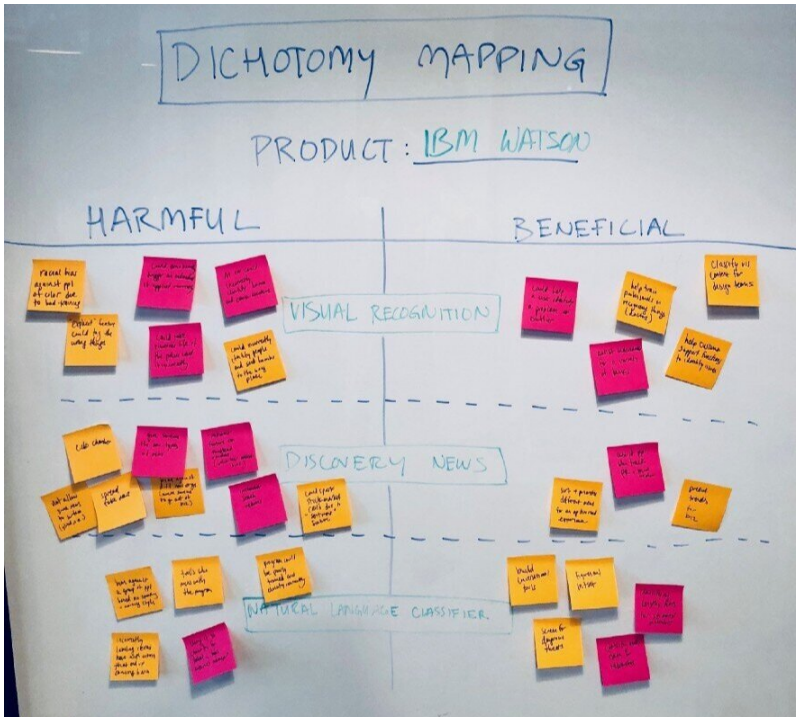


Figure 18. Illustration of dichotomy mapping (source: designethically.com)

Coming from the same organization, P2 and P7 gives an insider look at how values are applied and ethics are considered from an international company. Albeit based on

[20] https://www.designethically.com/

different branches in different continents, both P2 and P7 mention receiving training sessions on responsible AI as a prerequisite before working on AI-related projects. However, in terms of applying things in practice there are considerable differences. As a designer, P2 had to conduct their own research on their own initiative to discover the best practices in designing for AI while P7, a data scientist, was prescribed some standards of operation in developing AI. For example, P7 was required to tag every data that they are working with, affecting the duration that the data can be kept. P7 was also considerably fortunate in that a dedicated data governance team exists to ensure the data they are working with are up to a certain ethical standards. In addition to that, P7 maintains their own initiative to be conscious in their sampling and usage of data.

> *"I consider myself lucky. Just because as a data scientist I didn't have to do the data governance. So there are data engineers, their team is very good at scrubbing the data. So we can't really see the file names. We don't really know the machine names. We don't really know which company it is. Is it IBM? Is it other companies? We don't know and we don't care to be honest. Because it is scrubbed, we don't care so we don't have bias on who the customer is. So that's my benefit. I don't have to do that myself, it is completely done by the data governance team."* (P7)

Shifting to the developers that have worked with multiple companies, both P5 and P6 demonstrated a great deal of technical knowledge in applying certain values, explainability and fairness respectively. In explaining explainability, P5 gave a thorough explanation on how this is implemented through extensive metadata annotation and including this as part of the clients' scope of development. Their commitment to AI ethics are reflected in their newly released organizational code of conduct. This has worked to a degree that P5 have previously rejected to work with certain clients should their objectives be conflicting with their code of conduct.

> *"… But we have actually had some cases where we said, well, this particular solution that you are aligned with is not really providing the necessary value or it is not according to our own code of conduct. And then we leave the job to somebody else."* (P5)

In striving towards a fair AI system, P6 explained certain methods such as using synthetic data and implementing data augmentation to reduce biases from an otherwise biased dataset. However, enforcing this to the clients is another issue as P6 mentions that the best they can do is to advocate the benefit of implementing certain elements that support fairness and explainability but at an additional cost. The decision to implement is then left for the clients to decide.

From another perspective, P8 made an observation that ethical concerns are rarely considered in the day-to-day of fellow machine learning engineers while P3 remains cautious of 'ethics-washing' as they believe that actual ethical problems are difficult to solve in a way that they are very much situated within their societal contexts which then also pose the risk of developing into a new form of colonialism where certain values are exported, promoted, and enforced to different parts of the world.

> *"I think this is primarily going to be decided on a country by country basis. And of course, we are going now into a very difficult debate on whether Western values are the best values. For me, European, not so much Western, but European values are the best thing I think. But I'm a European I've... I've lived with these values. I like my privacy. I think this is the best thing for me. But we have to be very conscious about the fact that this could be like a new version of colonialism right. So we are exporting our values. I really don't like that."* (P3)

Applying values and considering ethics in practice comes with many challenges that may contribute to the reluctance in prioritizing ethical concerns. When it comes to AI ethical principles, some have alluded to their contents being too abstract and vague to be translated into practice (P1, P3, P4, P5). In this case, P1 narrates an experience of working with academic researchers where they had very high expectations of AI ethics that eventually became a barrier, stifling the development of solutions that they actually need. P1 describes this as being unable to drag the ethical principles down "from the clouds" and perhaps decide on some slight-yet-necessary compromises in their technical implementations.

Having experienced similar encounters, P5 mentions that their firm has been trying to bridge the gap between high-level abstract statements and what it actually means in practice for data scientists and engineers. They've made progress in this regard as they

can give examples of how high-level statements can be translated to requirements that affect the code or the workflow of engineers. In contrast, P3 sees that deriving practical solutions from ethical principles are rather difficult partly due to their vagueness.

> *"Ethical considerations are very difficult. Ethical choices are very difficult. Sure on the like super high level of ethics we, regardless of whether you're in China, you're in Indonesia or in Bulgaria, or in the Netherlands, you will all agree on what's good or not. Very high level. I think there were also ethical principles that came out from China. And if you look through the principles, they are very much the same as the principles of the European Union or whatever. And that's a problem because they're so vague. And when they're so vague, you cannot really agree on anything."* (P3)

Another challenge is the apathy encountered by P4. While at a glance this may seem like another case of practitioners not being able to relate ethical concerns into their practice, it is in fact rooted in the reality that, for some, ethical concerns are simply not a priority. Simply put, in the words of P4, *"people don't have time to fix problems that are not there yet."* This is echoed by P8, where they observe that their colleagues perceive ethics usually as an afterthought rather than a starting point.

Adopting new practices based on ethical considerations may also require additional effort and resources. On implementing explainability for AI systems, P5 and P6 elaborates how resource intensive it is not just in terms of additional computational processing but also requires additional effort in figuring out the best way to present to the users without it being too overwhelming to understand. Adopted practices based on the privacy-focused GDPR regulation has also impacted work for P7 in that they now have a very limited time to certain data and thus limiting the time they have to work on new features. If the data expires before the development is finished, then they are typically due for some setbacks.

Aside from the challenges, there are reported complexities related to the application of values and ethical considerations. Starting first with the perspective that not all AI should be considered equally, in a sense that different AI systems pose different sets of risks and imply the need for a different approach (P3, P4, P7, W1). With this, P7 wonders that this may be a factor in considering ethics as they compare the practical, and perhaps less prone to ethical concerns, nature of working in the cybersecurity industry compared

to working in the urban planning field where even the smallest of decisions are rife with complex socioethical dimensions. This is echoed by P3 and W1 where they see that the application of AI in public healthcare would have significantly different consequences than, say, AI in the private gaming industry (W1).

The organizational setting can also be a factor that adds further complication. As P2 reflects, they see that even when a company has done its best to ensure compliance to ethical standards there can be possibilities when their products and services are used by third-parties outside their control to conduct unethical activities. From their perspective in Silicon Valley, P8 also notes the lack of business incentive for ensuring ethical conduct is discouraging especially in an environment built on the foundations of "move fast, break things".

While there is a rising awareness and advocacy from workers for more ethical designs, it can also be the case that there exists a power asymmetry between employees and employers vast enough to contain (perhaps shut down) such rallying cries (P8). On this, P8 gave the example of Google's controversial dismissal of AI ethicist Timnit Gebru[21]. Relying on the individual conscience to advocate for change towards ethical practice may not be the best option as P3 likens this to the situation of government whistleblowers, such as Daniel Hale, who are immediately put into danger without any guarantee of protection. Fortunately, P8 also mentions the progressing efforts of tech giant workers in unionizing to advocate for more power to decide the conduct of the company.

The inherently challenging capability of AI to learn described in previous sections may also be the ultimate complication in applying values and considering ethics. As P3 argues, participatory approaches to these matters are not enough considering that there is a natural limit on the human mind that cannot possibly foresee all the unpredictable and unintended consequences of AI. To further complicate matters, P3 also reminds us that our standards of good and bad also change over time. Then how can we design what is good and bad for a system capable of change when the standards of good and bad itself changes over time?

Despite its challenges and complexities, participants also have stated reasons as to why applying values and ethical considerations should be pursued. In its most basic form, the objective may simply be the effort to find answers to the overarching concerns on the

[21] www.theverge.com/2020/12/3/22150355/google-fires-timnit-gebru-facial-recognition-ai-ethicist

risks of AI and its implications as illustrated in the first section of this chapter. As workshop participants have explicitly expressed concerns in the realization that some of the seeds to AI nightmares have already been planted (W3), the call for more ethical considerations in AI becomes more than justified (W1, W2, W3, W4, W5), a sentiment vocally shared by some interview participants as well (P4, P6, P7, P8).

However, striving for certain values can also be beneficial for the business. For example, P6 elaborates that explainability helps and will continue to be an important aspect when dealing with stakeholders such as high-ranking government officials as they often would inquire how certain critical decisions are made. This is echoed by P7 according to their experiences in the urban planning context, as both fairness and explainability goes hand-in-hand in making sure of ethical decisions. As P1 mentions, in some cases it makes productive sense to consider ethical standpoints earlier in the process rather than having to deal with the impacts and disgruntled employees after the fallout. Ethical considerations can also be used to raise awareness to consequences that can harm the business as elaborated by P4.

To surmise briefly this generous report on the application of values and the consideration of ethics in designing AI, here are key insights that can be inferred from both interview and workshop participants:

- Most participants from both interviews and the workshop elaborates on the importance of applying values and considering ethics in designing AI, citing values, such as privacy, trustworthy, fair, explainable, and so on, that reflect values within some AI ethical principles.
- There are many potential ideas and existing implemented examples of applying values and considering ethics in the process of designing AI. A main essence in the ideas or the mentioned examples is the importance of keeping users involved and its interests accounted for in the process typically through a participatory approach.
- There are barriers and challenges in practice that are hard to overcome such as the difficulty in translating high-level statements into practical implementations, the seemingly lack of effort to prioritize in considering implications upfront, the resource extensive implementation for fairness and explainability in solutions, and privacy regulations that may get in the way of initiatives.

- There are further complications that make applying values and considering ethics hard to implement such as the range of actors involved with the AI solutions, the different variety of AI system that carries different risks, the diverse and situated nature of human values that are far from universal, and simply the fact that what society considers as good and bad changes over time.
- Despite challenges and complexities, most participants can elaborate reasons in applying values and considering ethics in designing AI either as answers to some of the overarching concerns towards AI but also as potential added value beneficial for business.

5. Synthesis

To synthesize the findings of both the qualitative interviews and the workshop sessions, a framework will be employed. The framework will be implemented to consolidate the findings together with the highlights from the literature and frame them as distinct lenses to view the discussions of designers applying values and considering ethics in designing AI. The framing combining both the findings and literature highlights can be seen in Table 10 below.

As previously elaborated in the Methodology chapter, *reframing* is often a technique commonly employed in design synthesis to see things in a new way and explore the novel ideas that it might create (Kolko, 2010). The process of creating this framing was an iterative process that happened throughout the research project which reflected my growing understanding of the topic. With this new understanding, I have then proposed to *reframe* the lenses in which AI can be seen in 3 ways, each signifying a differing emphasis and, with it, complexities that entail in applying values and considering ethics.

First, through seeing *AI as is*, I propose that AI can be seen as its most basic form, a technical artifact. Through this lens, AI is seen from a distinctly technical perspective consisting of *data*, *model*, and *outputs*. With seeing *AI as a design material*, the technical artifact is then put into motion within a process, a design approach, and is thus treated as such a design material possessing unique properties. In this, the AI is seen as an intangible material that designers work with. Finally, I conclude with seeing *AI as a sociotechnical system*, broadening the scope of AI systems to consider its social underpinnings accounting for its interactions with the humans surrounding it and the institution it is situated in, thus emulating the confluence of all complexities in relation to AI.

Table 10. AI framings with summary of findings and literature highlights

Summary of Findings	**Literature Highlights**
Frame 1 — AI as is	
• AI can be perceived in various ways that hold different meanings • Data is a crucial part of AI systems, models affect the level of capability for the AI systems, and there	• AI presents many new opportunities and risks

is greater complexity for the outputs as AI systems learn over time • The increasing effect of AI on everyday lives is inevitable and there is a resounding agreement to minimize its negative consequences	
Frame 2 — AI as a design material	
• Designers mainly contribute at the earlier stages of the process and not much is gathered on accounting for the evolving nature of AI • On top of bringing human-centered approaches, design can contribute in other ways • A human-centered approach in designing AI may not be enough • A common challenge in the process and for designers is on understanding AI and accounting for its capability to evolve	• Designing AI poses new challenges that are inherent to its unique nature • Designers needs to and are coming up with new ways to consider the novel challenges in designing AI • The development of AI can benefit from having human-centered perspectives
Frame 3 — AI as a sociotechnical system	
• Most participants sees the importance of ethical AI with some emphasizing on values that are reflected in some AI ethical principles • Many ways and examples of applying values and considering ethics emphasize a human-centered participatory approach • There are challenges hard to overcome such as translating ethical statements into practice • There are further complications to consider such as the diverse and situated nature of both AI solutions and human values • Striving to apply values and consider ethics in designing AI is not only as an answer to concerns but can be beneficial for businesses	• Ethical principles are made in response to the inherent risks of AI • Translating AI ethical principles into practice • Design theoretically presents the many ways in which values and ethics can be embedded in practice

With the 3 frames devised in the table above, this chapter will be divided into 3 sections to accommodate discussions for each of the framings: (1) *AI as is*, (2) *AI as a design material*, and (3) *AI as a sociotechnical system*.

5.1. AI as is

Simple things first, framing AI solely as a technical artifact. This framing is a simplistic point of view in which AI is seen as a technology consisting of 3 main components: *data*, *model*, and *output* as seemingly distinct components found in basic explanatory concepts of AI (Information Commissioner's Office, 2021; Microsoft, 2018). In this regard, emphasis is put on those 3 distinct components and the importance they have on an AI system. By adopting this simple framing as a concise starting point, focus towards the distinct parts of the AI artifact may help identify unique challenges and further prompt relevant considerations of applying values and considering ethics at the level of a technical artifact and its components.

As many have voiced in the findings, the foundation of any AI is in its data (P4, P5, P6, P7, P8). In this regard, a lot of the discussion revolves around the collection of data, the availability of the data, and the quality of the data. Meanwhile, the *model* represents the different algorithmic ways in processing the data to infer predictions (Information Commissioner's Office, 2021; Microsoft, 2018). Models can be created from certain datasets and be trained over time as new data is gathered (Information Commissioner's Office, 2021; Microsoft, 2018). The predictions inferred by the model can then be used by the system in whichever ways it sees fit, whether to be shown to the user as is the case for some forecasting apps as is the case for P2, or to be used as a basis for decision-making integral to enabling certain features (Information Commissioner's Office, 2021; Microsoft, 2018). In this regard, I chose the term *output* rather than *prediction* to imply a more general sense of what AI can impact.

Tying back to Haenlein and Kaplan's (2019, p. 1) definition of AI as *"a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation"*, perhaps a more layman explanation of the components could simply be: *data* relates to from what and where AI learns, *models* are how AI learns, and *outputs* are what AI does based on its learnings. An illustration of this can be found in the diagram below.

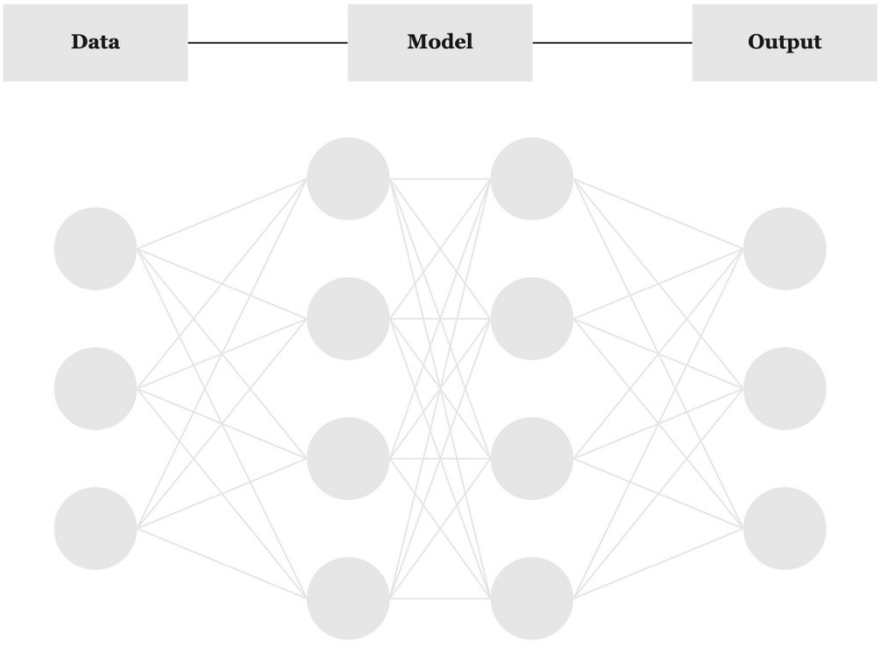


Figure 19. AI as is (from popular depictions of machine learning)

What are the opportunities for designers to help with the data and model?

In the designing of AI, designers often start with a user-centered perspective in a sense that they gather insights on what the desired *output* would be (P1, P2, P3, P4, P5, P8). But as outlined, *data* and *models* are important components that contribute to the way in which AI systems can derive an output. From the findings, P4 mentions how designers could be more involved in designing for data collection while both P5 and P6 mentions how designers should be included to help design and validate AI models. In this sense, designers may need to also be closely involved with the data and model, much like the way interaction designers would emphasize the use qualities of the screen as a digital material (Löwgren, 2002).

What are the prospects of regulating the design of AI models and outputs?

From the findings, a lot of discussions revolve around the data and, to an extension, the concerns of privacy (P4, P7) and fairness (P4, P6, P7, P8) which reflects the principle of *non-maleficence* and *justice* respectively (Floridi & Cowls, 2019). In striving to uphold the value of privacy, the account of P7 on how their organization has a dedicated data governance team and has devised a standard of operation when it comes to the

processing of customer data along with the example of how P6 implements techniques such as data augmentation to minimize bias can be an example of how applying values and considering ethics can be implemented at the level of a technical component.

Regulations, specifically the European GDPR[22], were a factor implied by P7 in establishing the standard of operations that their organization has for data. While regulations such as the GDPR helps to regulate the data of AI, perhaps this regulatory approach can also be considered towards the model and the output of AI systems. Perhaps an organizational example of this is how P5 and their organization enforces explainability efforts to be included in the scope when developing AI for clients, ensuring that AI outputs are transparent to an extent. What are then the prospects of extending that approach to regulate AI models and outputs?

5.2. AI as a design material

The materials that designers work with largely define design work (Holmquist, 2017). For example, graphic designers working with the printing medium would be particularly concerned about properties of the paper, its coating types, and so on (Holmquist, 2017). On the other hand, interaction designers would typically concern themselves with properties of the digital material such as pliability or responsiveness (Löwgren, 2002). However, the unique properties that AI brings, such as its capability to learn and evolve after its release into the wild, suggests a new kind of design material that is distinct in its challenge (Holmquist, 2017; Höök & Löwgren, 2021).

While the technical properties of AI have been laid out in the previous section, by framing *AI as a design material* potentially gives insights regarding the way AI is regarded within the process of designing and issues pertaining to its unique materiality are synthesized. The issues can then be seen from the design process and how designers engage with its distinct materiality.

Focusing on the unique capability of evolving over time that AI is capable of, Yang et al. (2020) suggests a framework which, in essence, encapsulates a conceptual pathway illustrating the areas concerning the design and development of AI systems. The diagram below shows the way in which a highly complex AI system could be approached, where a user-centered design approach can be seen as beginning from the right side, initially

[22] https://gdpr-info.eu/

defining the desired user experience and consequently moving to the left towards the development of algorithmic capability (Yang et al., 2020).

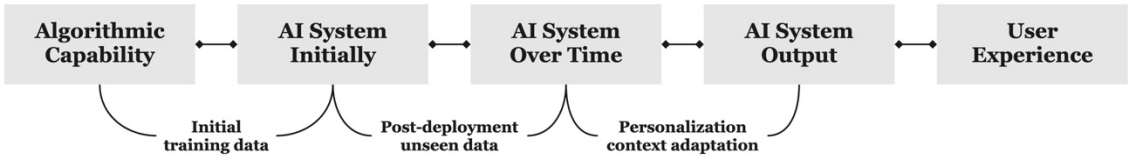


Figure 20. AI as a design material (reproduced from Yang et al., 2020)

How should design processes account for the evolving nature of AI?

With reference to the framework conceptualized by Yang et al. (2020), the findings shows that designers' contributions lay mainly in the conceptualization of the AI solution, whether it is through facilitating ideation sessions (P1, P4, P5), designing the interactions and the interface (P2), or making AI approachable so as to involve users in participatory activities (P1). In this regard, designers are designing the initial AI system based on a desired user experience. However, as the framework above shows, there are usually phases that are seemingly not accounted for in which the AI system evolves over time before producing an output that ultimately influences the actual user experience.

From the findings, there is also not much indication as to how the designers account for the evolving nature of AI systems in their initial design or throughout the engagement with its users after it is deployed. While guidelines on designing for AI exist and seem to suggest accounting for its evolution over time (Amershi et al., 2019), it seems to be primarily focused on designing adjustments in the interactions in accordance to changes rather than giving guidance on designing designs that change. This is interesting to note, as this is arguably one of the distinct challenges in designing AI (Holmquist, 2017; Höök & Löwgren, 2021; Yang et al., 2020), which perhaps can lead to causing unintended consequences even in cases where it seems to have been sufficiently tested such as the case with Microsoft's Tay chatbot[23] as conveyed by P3.

This may relate to the discrepancy in operationalizing the process of designing AI described in preceding literature, namely the challenge of prototyping a system capable

[23] https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/

of adaptation (Bratteteig & Verne, 2018; Yang et al., 2020), although indeed there are emerging studies on solving the issue as found in the literature (Subramonyam et al., 2021; van Allen, 2018). Instead, challenges described in the findings seem to revolve mostly around the lack of understanding towards AI (P1, P2, P4, P6, P8) and finding better ways to communicate and present it (P1, P2, P5, P6). This may also suggest that designing AI is an over-exaggerated challenge hyped by its recent popular appeal while in reality it may not be so different to designing other systems. However, I find the former argument supported by the existing body of work in this topic to be more convincing. What are the gaps between designing the initial AI system and accounting for the evolving nature of AI? How might we reduce these gaps? Why are there gaps? How then should design processes account for the evolving nature of AI after it is deployed? What kind of shifts are required in the process to operationalize designing AI?

The design emphasis on human-centric participatory approaches alone may not be enough

The findings show the many experiences and emphasises on 'human-centricity' and participatory approaches (P1, P2, P4, P5, P6, P8). From the findings, this can either mean involving the users in the processes (P1, P4, P5) or putting human interests at the forefront of design objectives (P2, P6, P8). In addition, human-centered AI can also mean *'building the intelligent systems to understand the (often culturally specific) expectations and needs of humans and to help humans understand them in return'* (Riedl, 2019, p. 36). However, if the approach towards this is by simply encouraging users' participation in the design process then there are critical arguments on why this alone may not be enough.

As conveyed by P3, participatory approaches where users and stakeholders alike are involved in the process does not seem to be enough for a few reasons. First, there could be a conceptual *gulf of execution* (Hutchins et al., 1985) between what the AI is capable of and the actual reality of its ability due to the minimal AI literacy of the users (P1, P4, P6). While this lack of comprehension can easily be bridged through educational means (Long & Magerko, 2020), or as demonstrated by P1, another challenge against purely participatory approaches is in predicting the potential outputs of the AI system, as neither relying on human minds to imagine possibilities (P3) and creating dummy prototypes seem sufficient (Bratteteig & Verne, 2018). Indeed this may lead to some credibility in the

arguments to shift the sole focus from human-centered design towards other paradigms in design (Norman, 2005; van Allen, 2017).

Despite its important role and contribution in designing AI as suggested by most of the findings, some accounts suggest that design does not seem to be able to contribute much in situations where technical rigor is required and within these situations, developers take the center stage (P2, P7, P8). Perhaps this is expected, as different parts of the development process require different sets of expertise. However, as crucial parts of developing an AI system lies within technical decisions (Morley et al., 2019), developers are implied to have greater power and responsibility over the design of AI compared to designers (P2, P8). Perhaps this may suggest that future AI designers are to be more attuned to the technical rigors inherent to the development of AI systems.

Contrary to the suggestion however, Holmquist (2017) foresees AI to become more common in the future and that *"designers will no longer have to be experts in neural networking to use AI, just as they do not need to know the ins and outs of TCP/IP or even HTML to design Web pages."* Designers would need to discover how to work with AI as a new material, indicating a radical departure in design practice in much the same way as the early days of designing for the digital (Holmquist, 2017).

What kind of designerly abstractions are needed in the design process?

Rather than obtaining in-depth technical expertise, a reflective study by Yang et al. (2018) also suggests that forming designerly abstractions of AI and building streamlined means of collaborating with developers would be of greater benefit for future AI designers. This may be in line with the experiences of P1 in creating design artifacts such as AI cards that make understanding AI easier and more approachable for users participating in the process.

While these cards focus on illustrating the vast range of AI capabilities, perhaps another avenue in which design can contribute is by creating an abstraction which could sufficiently communicate why and how an AI system evolves. An avenue that is recently explored is the idea of compiling a list of 'design heuristics' unique to AI which can serve to support the early ideation phase for in providing an overview of AI capabilities useful for both designers and users (Jin et al., 2021). This might play an important role in the efforts as it deals with challenges to participatory approaches in designing AI (Bratteteig & Verne, 2018). How else then might designers effectively communicate and present AI

during participatory processes? What kind of designerly abstractions are needed in the design process?

5.3. AI as a sociotechnical system

From both the findings and the literature there exists a substantial amount of insights that seem to prescribe in seeing AI from a broader point of view. For this, a frame to see AI further beyond both technical artifacts and design material is needed, extending the scope of AI to account its social underpinnings (Crawford & Joler, 2018; van de Poel, 2020).

Taking both the perspectives of P6 on how AI reflects and amplifies behaviours of humans and W1 in seeing it as a technological extension of society as examples, many of the findings speak of the complexity beyond technicality that is inherent in AI (P3, P5, P6, P8, W1, W3), but it may perhaps be even more than it seems. For example, in the anatomy of AI, the true scale of building artificial intelligence, an Amazon Echo in this illustration, is presented as a mind-bogglingly complex map of logistical relationships dependent on exhaustive extraction of human labour, data, and planetary resources (Crawford & Joler, 2018).

The societal entanglement for AI lies also in its core capability to learn and adapt from data which must be actively acquired throughout its lifetime (Holmquist, 2017). This establishes an active relationship in which AI systems require human engagement to be able to continuously improve its function (Crawford & Joler, 2018; Holmquist, 2017). In this sense, the workings of AI can be seen to depend not only on technical means but also on societal aspects to enable its intended function (van de Poel, 2020). In this regard, van de Poel's (2020) perspective of seeing *AI as a sociotechnical system* is an interesting proposition that can be further discussed in tandem with the findings. By adopting the framing of AI as a sociotechnical system, the issues can then be seen from a broader lens extending beyond the design of the AI technology itself, encompassing the active role of components such as the people and the institutions interacting with the system. This implies that we can also consider a wider perspective and take into account blocks that are otherwise seemingly out of scope in tackling the issues pertaining to AI.

In van de Poel (2020), a sociotechnical system is understood as systems that depend on technical hardware, human behavior, and social institutions for their proper

functioning, and AI as a sociotechnical system consists of 5 distinct blocks: (1) technical artifacts, (2) human agents, and (3) institutions. The blocks (4) artificial agents and (5) technical norms are the distinct differentiator between AI and other sociotechnical systems (van de Poel, 2020).

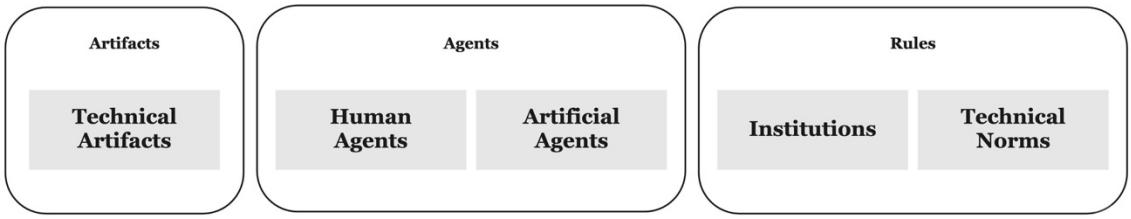


Figure 21. AI as a sociotechnical system (reproduced from van de Poel, 2020)

A prime example of human agents are the users of AI systems can then be seen as an integral as they continuously shape the system throughout their engagement with it (van de Poel, 2020). In this regard, the findings show a resounding call to extend the users' involvement into the design process as well through a participatory approach (P1, P4, P5, P6, P8). With relation to applying values and considering ethics, this can be beneficial as P5 suggests that the users should be more involved in actually considering what matters most to them. P5 observes that there is typically a focus on testing the usability of products, but there is little focus on actually testing for the ethical considerations from the perspectives of the users. This can perhaps be facilitated through the usage of AI ethic cards as designed by P1.

However, as elaborated in the previous section, the involvement of users in the process is difficult due to the common challenge of fully understanding AI and all of its complexities (P1, P2, P3, P4, P5, P6, P8). If left without proper understanding of AI, then there are potential risks of misjudgments especially if the interpretations of AI are based on works of fiction as commonly encountered by P6. This supports the case for the necessity for striving towards explainability in AI systems, designing AI literacy as an essential competency in a world where AI becomes increasingly common, and creating designerly abstractions to better present AI capabilities (Jin et al., 2021; Liao et al., 2020; Long & Magerko, 2020).

How might we design environments to foster a culture of applying values and considering ethics in designing AI?

When it comes to applying values, *institutions* play an important role for AI systems (Umbrello, 2019; van de Poel, 2020). In the case of institutions, findings have shown examples of developers consciously embedding or enacting certain values as governed by their institutions (P7). Having a shared organizational stance for AI ethics has also been beneficial for P5 in enforcing certain standards when dealing with clients. And perhaps institutions can potentially establish social norms to also prevent and disincentivize certain behaviours from users that may affect the values of the AI system as reflected by W3's idea of an AI culture in harmony with the community. While values can be distilled from top-to-bottom through the role of institutions, Umbrello (2019) argues that values can also emerge from a bottom-up approach. Thus, if the environments that an AI is situated in plays an important role, how might we design environments to foster a mindful culture of applying values and considering ethics in designing AI?

How might design efforts account for the complexities in values and ethics?

While van de Poel (2020) makes a compelling argument that all the components of the AI sociotechnical system have values embedded in them and thus play a crucial role in the efforts to apply values and consider ethics, P3 reminds us that the discussions of values and ethics are rather situated and diverse in its forms. And it's not too far in its example, as P4 suggests the seemingly common lack of awareness or even apathy towards data safety and privacy in Indonesia may be attributed to the cultural differences when compared to Europe while P7 wonders whether they would even consider AI ethics had they not been educated in the United States where the issues of fairness and diversity seem to be a prominent point of discussion in their field. Not to mention that, as once P3 reminds us, what humans define as good and bad does change over time. In this regard, Whittlestone et al., (2019, p. 197) argues to emphasize on the *tensions* instead when it comes to AI ethics, either as discussions revolving around 'a strict moral tradeoff' or as a result of current technological and societal constraints, as the role of principles can be limited. Nonetheless, how then might designers of AI systems consider the differences in values and ethics? And how might we design what is good and bad for AI systems capable of change when the standards of good and bad itself changes over time?

6. Results

To summarize the results of the synthesis from both the literature and the findings, the explorative inquiry towards how design(ers) apply values and consider ethics in designing AI has yielded some hypotheses and further questions outlined in Table 11 below. These points of departure in thoughts will hopefully be beneficial for future research in the topical interaction between AI, design, and ethics. Aside from that, the framing conceptualized in this project can be useful as each frame offers a different perspective to AI and therefore, a different focus to the issues of applying values and considering ethics (see Figure 22).

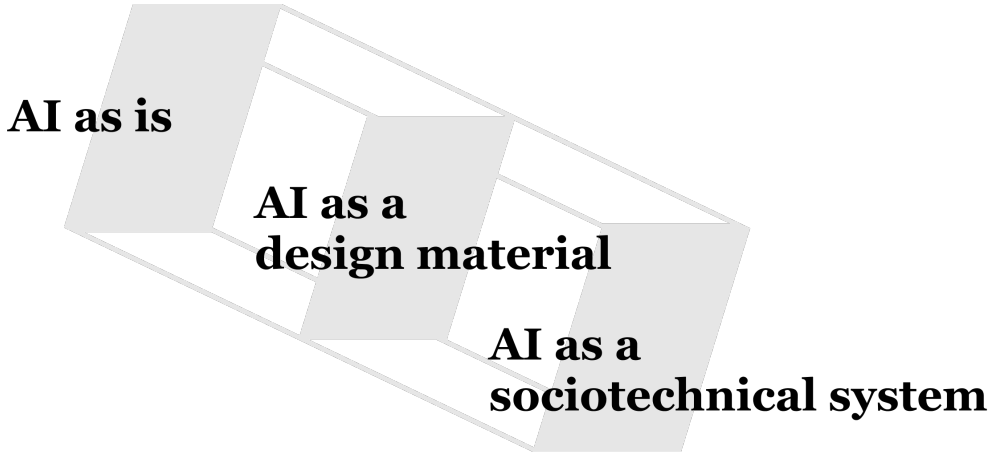


Figure 22. The three framings of AI

Based on these results, there are future avenues that can be worth pursuing as future research directions. In using the framing of AI as is, it can be seen that designers are generally focused on designing for the output of the AI system. However, the data and the AI models are subsequently important parts of the system as well. In this regard, it might be beneficial to discover what kind of role that design can contribute to the activities related to the data and the models of an AI.

Another avenue filled with research opportunities is on how the designers can account for the challenges inherent to AI as a design material with emerging topics such as prototyping AI systems (Koch et al., 2019; Subramonyam et al., 2021; van Allen, 2018), investigations into how explainability and better AI literacy can overcome the challenges

(Liao et al., 2020; Long & Magerko, 2020), and in creating designerly abstractions useful for the process of designing AI systems (Jin et al., 2021).

Table 11. Summary of the results

Framing	Hypotheses and further questions
AI as is	What are the opportunities for designers to help with the data and model?
	What are the prospects of regulating the design of AI models and outputs?
AI as a design material	How should design processes account for the evolving nature of AI? What are the gaps between designing the initial AI system and accounting for the evolving nature of AI? How might we reduce these gaps? What kind of shifts are required in the design process to fully operationalize designing AI?
	What kind of designerly abstractions are needed in the design process? How then might designers effectively communicate and present AI during participatory processes?
	To account for risks and negative consequences in the process, participatory approaches alone may not be enough due to the challenges in understanding AI and the difficulty in predicting the evolution of systems.
AI as a sociotechnical system	How might we design environments to foster a culture of applying values and considering ethics in designing AI?
	How might the designing of AI systems consider the differences in values and ethics? How might we design what is good and bad for AI systems capable of change when the standards of good and bad itself changes over time?

In applying values and considering ethics in designing AI, the framing of AI as a sociotechnical system by van de Poel (2020) seems like a good approach to take as it gives a holistic flow of how values can be intended, embodied, then realized through the various components beyond just the AI itself. This gives a holistic starting point on how evaluating the implementation of values from AI ethics documents into practice can potentially be done. However, future research must also consider the complexities of values and ethics, such as its diversity, its divergence in interpretation, and its dynamic nature, and how that can be designed into a system capable of learning and evolving.

Despite the challenges in accounting for values and ethics, the potential benefits of AI cannot be denied and its impact on design and the broader societal dimensions must be considered. Perhaps this emphasizes the importance of future work in this intersection as AI is here to stay.

7. Conclusions

7.1. Research Summary

With the increasing concern over the risks of AI (Cave & ÓhÉigeartaigh, 2018; Floridi et al., 2018; Neff, 2016; Turchin & Denkenberger, 2020) and the proliferation of ethical principles (Fjeld et al., 2020; Floridi & Cowls, 2019; Hagendorff, 2020; Jobin et al., 2019), design plays a potentially important role as it has operationalized applying values and considering ethics throughout history (Devon & van de Poel, 2004; Friedman, 1996; Monteiro, 2019; Shilton, 2013). However designers face unique challenges due to the distinct materiality of AI (Bratteteig & Verne, 2018; Holmquist, 2017; Stoimenova & Price, 2020; Yang et al., 2020) and thus, the emerging studies on novel ways of designing AI (Amershi et al., 2019; Koch et al., 2019; Liao et al., 2020; Subramonyam et al., 2021; van Allen, 2017, 2017).

While some investigations have been conducted into how designers design AI (Dove et al., 2017; Girardin & Lathia, 2017; Liao et al., 2020; Yang et al., 2018), not many inquiries have seemingly been made on how designers apply values and consider ethics in designing AI despite the emerging studies that argue on how design could theoretically embed values into the design of AI systems (Dignum, 2017; Liao & Muller, 2019; Umbrello, 2019; van de Poel, 2020). Based on this research gap, this project asks the question: *how do design(ers) apply values and consider ethics in designing AI?*

To explore this inquiry through a design approach, 8 interviews with both AI designers and developers were conducted to give insights on how their AI design processes are, what role design played in the process, the challenges they faced in designing AI, their thoughts on the implication of AI for societies, how they see AI ethical values and their attempts in applying them to their actual practice. Subsequently, a workshop session with 5 designers were held to generate as many ideas as possible to how they might apply values in designing AI then following it up with a reflection on whether, how, and why these ideas should be applied in practice.

The findings show emphasis on human-centered and participatory approaches to apply values and consider ethics in designing AI. However, these efforts are hindered with inherent challenges such as the lack of comprehension of AI systems and designing for the evolving nature of its outputs post-deployment. Moreover, there are other factors

outside of the design process that complicates how values and ethics can be translated into practice. Nonetheless, participants both imply and explicitly state the essential role and contribution of designers in the development of AI although these findings lead to the notion that a paradigm shift in design practice within the context of AI may be required. To further showcase some of the findings:

- Designers mainly contribute at the earlier stages of the process and not much is gathered on accounting for the evolving nature of AI. On top of bringing human-centered approaches, design can contribute in other ways such as designing concepts to present AI in an easily understandable manner. A common challenge in the process and for designers is on understanding AI and accounting for its capability to evolve.
- Most participants see the importance of ethical AI with some emphasizing on values that are reflected in some AI ethical principles. Moreover, striving to apply values and consider ethics in designing AI is not only as an answer to concerns but can also be beneficial for businesses.
- Many ways and examples of applying values and considering ethics emphasize a human-centered participatory approach although some argue that that alone is not enough.
- There are challenges hard to overcome such as vague ethical statements that are difficult to translate into practice and there are further complications to consider such as the diverse and situated nature of both AI solutions and human values.

In further synthesis of the findings from both the interviews and the workshop combined with the literature highlights, a framework was proposed to reframe AI in different perspectives: (1) AI as is, (2) AI as a design material, and (3) AI as a sociotechnical system. From these reframings, further questions and recommendations for future research directions were generated that can serve as basis for further endeavours into the inquiries on the intersection between AI, design, and ethics:

- What are the opportunities for designers to help with the data and model?
- What are the prospects of regulating the design of AI models and outputs?
- How should design processes account for the evolving nature of AI? What are the gaps between designing the initial AI system and accounting for the evolving nature of AI? How might we reduce these gaps? What kind of shifts are required in the design process to fully operationalize designing AI?

- What kind of designerly abstractions are needed in the design process? How then might designers effectively communicate and present AI during participatory processes?
- To account for risks and negative consequences in the process, participatory approaches alone may not be enough due to the challenges in understanding AI and the difficulty in predicting the evolution of systems.
- How might we design environments to foster a culture of applying values and considering ethics in designing AI?
- How might the designing of AI systems consider the differences in values and ethics? How might we design what is good and bad for AI systems capable of change when the standards of good and bad itself changes over time?

As a result in exploring how design(ers) apply values and consider ethics, the insights from the findings of both interviews and the workshop were synthesized along with the literature highlights which produced further ideas and questions seemingly worth considering for future research directions.

7.2. Limitations

It goes without saying that this research project has its limitations. Despite the efforts to explore the landscape of how designers apply values and consider ethics in designing AI, the research faced both limitations in terms of operationalizing the methodology aside from the limited time.

First, there are limitations in terms of sampling for the participants of both the interviews and the workshop session as it was quite the challenge to connect with designers that have relevant experience in designing AI despite the increasing adoption of AI technologies. As one participant has indicated, it seems like positions on AI projects are still very much exclusive and restrictive to relatively highly experienced individuals. Therefore, the small sample that this research may not be a thorough representation of the landscape in how designers apply values and consider ethics.

Second, there are limitations in conducting the workshop session. In retrospect, conducting the workshop session may not be the best idea to gauge on how designers apply values and consider ethics as the session was too focused on generating potential ideas rather than reflecting the reality of their daily practice. In hindsight, this was also

an adjustment made to account for the more diverse background of the participants that have attended the session compared to those that have attended the interviews. Due to both limitations, this research project could benefit from further empirical investigations as the inquiry to how designers apply value and consider ethics in designing AI is a crucial part of the intersection between design, AI, and ethics.

Bibliography

AI HLEG. (2019). *Ethics guidelines for trustworthy AI*. European Commission.

Amershi, S., Inkpen, K., Teevan, J., Kikin-Gil, R., Horvitz, E., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., & Bennett, P. N. (2019). Guidelines for Human-AI Interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–13. https://doi.org/10.1145/3290605.3300233

Ballard, S., Chappell, K. M., & Kennedy, K. (2019). Judgment call the game: Using value sensitive design and design fiction to surface ethical concerns related to technology. *Proceedings of the 2019 on Designing Interactive Systems Conference*, 421–433.

Benton, S., Miller, S., & Reid, S. (2018). *The Design Economy 2018*. Design Council (UK).

Bleecker, J. (2009). Design Fiction: A short essay on design, science, fact and fiction. *Near Future Laboratory*, *29*.

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-Volume 2*, 1–5.

Brown, T. (2008). *Design thinking*.

Bryman, A. (2016). *Social research methods*. Oxford university press.

Buchanan, R. (1992). Wicked problems in design thinking. *Design Issues*, *8*(2), 5–21.

Buchanan, R. (1998). Branzi's dilemma: Design in contemporary culture. *Design Issues*, *14*(1), 3–20.

Buchanan, R. (2001). Design research and the new learning. *Design Issues*, *17*(4), 3–23.

Buley, L., Avore, C., Gates, S., Stephen, S., Goodman, R., & Walter, A. (2019). *The New Design Frontier*. DesignBetter by InVision.

Cam, A., Chui, M., & Hall, B. (2019). *Global AI Survey: AI proves its worth, but few scale impact*. McKinsey & Company.

Castro, D., & McLaughlin, M. (2021). *Who Is Winning the AI Race: China, the EU, or the United States? —2021 Update*. Center for Data Innovation.

Cave, S., & ÓhÉigeartaigh, S. S. (2018). An AI race for strategic advantage: Rhetoric and risks. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 36–40.

Compton, M., & Barrett, S. (2016). A Brush with Research: Teaching Grounded Theory in the Art and Design Classroom. *Universal Journal of Educational Research*, *4*(2), 335–348.

Cramer, H., & Kim, J. (2019). Confronting the tensions where UX meets AI. *Interactions*, *26*(6), 69–71.

Crawford, K., & Joler, V. (2018). Anatomy of an AI System: The Amazon Echo as an anatomical map of human labor, data and planetary resources. *AI Now Institute and Share Lab*, *7*.

Cross, N. (1982). Designerly ways of knowing. *Design Studies*, *3*(4), 221–227.

Cross, N. (2001). Designerly ways of knowing: Design discipline versus design science. *Design Issues*, *17*(3), 49–55.

Cummings, M. L. (2006). Integrating ethics in design through the value-sensitive design approach. *Science and Engineering Ethics*, *12*(4), 701–715.

Dam, R. F., & Siang, T. Y. (n.d.). *Affinity Diagrams – Learn How to Cluster and Bundle Ideas and Facts*. The Interaction Design Foundation. Retrieved May 17, 2021, from https://www.interaction-design.org/literature/article/affinity-diagrams-learn-how-to-cluster-and-bundle-ideas-and-facts

Danish Design Centre. (2018). *Design Delivers 2018: How Design Accelerates Your Business*. Danish Design Centre.

Dervin, B. (1998). Sense-making theory and practice: An overview of user interests in knowledge seeking and use. *Journal of Knowledge Management*.

Devon, R., & van de Poel, I. (2004). Design ethics: The social ethics paradigm. *International Journal of Engineering Education*, *20*(3), 461–469.

Dignum, V. (2017). *Responsible artificial intelligence: Designing AI for human values*.

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

European Commission. (2018). *The European Artificial Intelligence landscape*.

Fischer, H. R. (2001). Abductive reasoning as a way of worldmaking. *Foundations of Science*, *6*(4), 361–383.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center Research Publication*, *2020–1*.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, *1*(1).

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., & Rossi, F. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707.

Friedman, B. (1996). Value-sensitive design. *Interactions*, *3*(6), 16–23.

Friedman, B., & Kahn Jr, P. H. (2003). Human values, ethics, and design. *The Human-Computer Interaction Handbook*, 1177–1201.

Friedman, B., Kahn, P., & Borning, A. (2002). Value sensitive design: Theory and methods. *University of Washington Technical Report*, *2–12*.

Gillies, M., Fiebrink, R., Tanaka, A., Garcia, J., Bevilacqua, F., Heloir, A., Nunnari, F., Mackay, W., Amershi, S., & Lee, B. (2016). Human-centred machine learning. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 3558–3565.

Girardin, F., & Lathia, N. (2017). When User Experience Designers Partner with Data Scientists. *AAAI Spring Symposia*.

Golsby-Smith, T. (1996). Fourth Order Design: A Practical Perspective Tony Golsby-Smith. *Design Issues*, *12*(1), 5–25.

Google. (2018, June 7). *AI at Google: Our principles*. https://blog.google/technology/ai/ai-principles/

Grudin, J. (2008). A moving target: The evolution of HCI. *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*, 1–24.

Haenlein, M., & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*, *61*(4), 5–14. https://doi.org/10.1177/0008125619864925

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 1–22.

Harper, R. H. R. (2019). The Role of HCI in the Age of AI. *International Journal of Human–Computer Interaction*, *35*(15), 1331–1344. https://doi.org/10.1080/10447318.2019.1631527

Holmquist, L. (2017). Intelligence on tap: AI as a new design material. *Intelligence*.

Holtzblatt, K., & Beyer, H. (2014). Contextual design: Evolved. *Synthesis Lectures on Human-Centered Informatics*, *7*(4), 1–91.

Höök, K. (2000). Steps to take before intelligent user interfaces become real. *Interacting with Computers*, *12*(4), 409–426.

Höök, K., & Löwgren, J. (2021). Characterizing Interaction Design by Its Ideals: A Discipline in Transition. *She Ji: The Journal of Design, Economics, and Innovation*, *7*(1), 24–40. https://doi.org/10.1016/j.sheji.2020.12.001

Hutchins, E. L., Hollan, J. D., & Norman, D. A. (1985). Direct manipulation interfaces. *Human–Computer Interaction*, *1*(4), 311–338.

IBM. (2021, July 15). *AI Ethics*. https://www.ibm.com/artificial-intelligence/ethics

IDEO.org. (n.d.). *Bundle Ideas*. Design Kit by IDEO.Org. Retrieved May 17, 2021, from https://www.designkit.org/methods/30

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2019). *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems* (First Edition).

Information Commissioner's Office. (2021, June 24). *The basics of explaining AI: Definitions*. Information Commissioner's Office; ICO. https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence/part-1-the-basics-of-explaining-ai/definitions/

Iversen, O. S., Halskov, K., & Leong, T. W. (2012). Values-led participatory design. *CoDesign*, *8*(2–3), 87–103.

Jin, X., Evans, M., Dong, H., & Yao, A. (2021). Design heuristics for artificial intelligence: Inspirational design stimuli for supporting UX designers in generating AI-powered ideas. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–8.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399.

Johnson, J., Roberts, T. L., Verplank, W., Smith, D. C., Irby, C. H., Beard, M., & Mackey, K. (1989). The Xerox Star: A retrospective. *Computer*, *22*(9), 11–26. https://doi.org/10.1109/2.35211

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*(1), 15–25.

Koch, J., Lucero, A., Hegemann, L., & Oulasvirta, A. (2019). May AI? Design ideation with cooperative contextual bandits. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12.

Kolko, J. (2010). Abductive thinking and sensemaking: The drivers of design synthesis. *Design Issues*, *26*(1), 15–28.

Kolko, J. (2011). Craftsmanship. *Interactions*, *18*(6), 78–81. https://doi.org/10.1145/2029976.2029996

Kolko, J. (2018). The divisiveness of design thinking. In *Interactions* (Vol. 25, Issue 3, pp. 28–34). Association for Computing Machinery.

Krippendorff, K. (2005). *The semantic turn: A new foundation for design*. crc Press.

Krippendorff, K. (2011). Principles of design and a trajectory of artificiality. *Journal of Product Innovation Management*, *28*(3), 411–418.

Kvale, S. (2008). *Doing interviews*. Sage.

Lee, P. (2016, March 25). *Learning from Tay's introduction*. The Official Microsoft Blog. https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/

Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: informing design practices for explainable AI user experiences. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15.

Liao, Q. V., & Muller, M. (2019). Enabling Value Sensitive AI Systems through Participatory Design Fictions. *ArXiv Preprint ArXiv:1912.07381*.

Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16.

Lovejoy, J. (2021, January 22). *When are we going to start designing AI with purpose?* Medium. https://uxdesign.cc/when-are-we-going-to-start-designing-ai-with-purpose-e196f986974b

Löwgren, J. (2002). *The use qualities of digital designs*.

Löwgren, J., & Stolterman, E. (2004). *Thoughtful interaction design: A design perspective on information technology*. Mit Press.

Maeda, J. (2017). *Design In Tech Report 2017*.

Maeda, J. (2018). *Design in Tech Report 2018*.

Maeda, J. (2019). *Design in Tech Report 2019*.

Microsoft. (n.d.). *Responsible AI principles from Microsoft*. Retrieved August 16, 2020, from https://www.microsoft.com/en-us/ai/responsible-ai

Microsoft. (2018, August 9). *What is Artificial Intelligence?* Microsoft. https://azure.microsoft.com/en-us/blog/what-is-artificial-intelligence/

MIT Technology Review Insights. (2020). *The global AI agenda: Promise, reality, and a future of data sharing*. MIT Technology Review.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, *1*(11), 501–507.

Monteiro, M. (2017, July 7). *A Designer's Code of Ethics*. Medium. https://deardesignstudent.com/a-designers-code-of-ethics-f4a88aca9e95

Monteiro, M. (2019). *Ruined by design: How designers destroyed the world, and what we can do to fix it*. Mule Design.

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2019). From what to how. An overview of AI ethics tools, methods and research to translate principles into practices. *ArXiv Preprint ArXiv:1905.06876*.

Muller, M. J., & Kuhn, S. (1993). Participatory design. *Communications of the ACM*, *36*(6), 24–28.

Naumer, C., Fisher, K., & Dervin, B. (2008). Sense-Making: A methodological perspective. *Sensemaking Workshop, CHI*, *8*.

Neff, G. (2016). Talking to bots: Symbiotic agency and the case of Tay. *International Journal of Communication*.

Norman, D. A. (1994). How might people interact with agents. *Communications of the ACM*, *37*(7), 68–71.

Norman, D. A. (2005). Human-centered design considered harmful. *Interactions*, *12*(4), 14. https://doi.org/10.1145/1070960.1070976

Partadiredja, R. A. (2020). Datafication and Design: Are we designing the right thing? *Kommunikation.Medien*, *Vol. 2020*(Issue 12: Themenschwerpunkt "Not with us! Formen und Dynamiken der Protestkommunikation"), 1–10. https://doi.org/10.25598/JKM/2020-12.9

Partadiredja, R. A., Serrano, C. E., & Ljubenkov, D. (2020). AI or Human: The Socio-ethical Implications of AI-Generated Media Content. *2020 13th CMI Conference on Cybersecurity and Privacy (CMI)-Digital Transformation-Potentials and Challenges (51275)*, 1–6.

Plain, C. (2007). Build an affinity for KJ method. *Quality Progress*, *40*(3), 88.

Ramos, G., Suh, J., Ghorashi, S., Meek, C., Banks, R., Amershi, S., Fiebrink, R., Smith-Renner, A., & Bansal, G. (2019). Emerging perspectives in human-centered machine learning. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–8.

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

Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*(2), 155–169.

Rogers, Y. (2012). HCI theory: Classical, modern, and contemporary. *Synthesis Lectures on Human-Centered Informatics*, *5*(2), 1–129.

Schön, D. A. (1984). *The reflective practitioner: How professionals think in action* (Vol. 5126). Basic books.

Sheppard, B., Kouyoumjian, G., Sarrazin, H., & Dore, F. (2018). *The Business Value of Design* (McKinsey Quarterly). McKinsey & Company.

Shilton, K. (2013). Values levers: Building ethics into design. *Science, Technology, & Human Values*, *38*(3), 374–397.

Simon, H. A. (1988). The science of design: Creating the artificial. *Design Issues*, 67–82.

Spinuzzi, C. (2005). The methodology of participatory design. *Technical Communication*, *52*(2), 163–174.

Spool, J. M. (2013, December 30). Design is the Rendering of Intent. *UX Articles by UIE*. https://articles.uie.com/design_rendering_intent/

Stephanidis, C., Salvendy, G., Antona, M., Chen, J. Y., Dong, J., Duffy, V. G., Fang, X., Fidopiastis, C., Fragomeni, G., & Fu, L. P. (2019). Seven HCI grand challenges. *International Journal of Human–Computer Interaction*, *35*(14), 1229–1269.

Stoimenova, N., & Kleinsmann, M. (2020). *Identifying and addressing unintended values when designing (with) Artificial Intelligence*.

Stoimenova, N., & Price, R. (2020). Exploring the Nuances of Designing (with/for) Artificial Intelligence. *Design Issues*, *36*(4), 45–55. https://doi.org/10.1162/desi_a_00613

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

Subramonyam, H., Seifert, C., & Adar, E. (2021). ProtoAI: Model-Informed Prototyping for AI-Powered Interfaces. *26th International Conference on Intelligent User Interfaces*, 48–58.

Turchin, A., & Denkenberger, D. (2020). Classification of global catastrophic risks connected with artificial intelligence. *Ai & Society*, *35*(1), 147–163.

Umbrello, S. (2019). Beneficial artificial intelligence coordination by means of a value sensitive design approach. *Big Data and Cognitive Computing*, *3*(1), 5.

van Allen, P. (2017). Reimagining the Goals and Methods of UX for ML/AI. *2017 AAAI Spring Symposium Series*.

van Allen, P. (2018). Prototyping ways of prototyping AI. *Interactions*, *25*(6), 46–51.

van de Poel, I. (2020). Embedding Values in Artificial Intelligence (AI) Systems. *Minds and Machines*, *30*(3), 385–409.

Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019). The role and limits of principles in AI ethics: Towards a focus on tensions. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 195–200.

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, *109*(1), 121–136. JSTOR.

Winograd, T. (2006). Shifting viewpoints: Artificial intelligence and human–computer interaction. *Artificial Intelligence*, *170*(18), 1256–1258.

Yang, Q., Scuito, A., Zimmerman, J., Forlizzi, J., & Steinfeld, A. (2018). Investigating how experienced UX designers effectively work with machine learning. *Proceedings of the 2018 Designing Interactive Systems Conference*, 585–596.

Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376301

Zimmerman, J., Oh, C., Yildirim, N., Kass, A., Tung, T., & Forlizzi, J. (2020). UX designers pushing AI in the enterprise: A case for adaptive UIs. *Interactions*, *28*(1), 72–77.

Appendix

Table 12. Codebook

1A — What is AI	
Definition	The meanings and definitions of artificial intelligence.
Coding Rules	Statements and workshop notes that elaborates on the definitions of AI and how it means relative to the participants.
Example	*"AI is an intelligent way of automation 'dumb' or repetitive tasks"* – W2 from workshop

1B — Why AI	
Definition	The reason and goal behind the utilization of AI.
Coding Rules	Statements and workshop notes implying the goal of why AI is utilized in certain situations.
Example	*"I was asking like, why, why you're doing this? What, what's your main goal? And they say like, I know this is a very, very yeah… A shallow reason. They say that we basically just want to do something with our user data. We have all this phone number and we want to do something. At least we want to blast like a promotion."* – P4 from interviews

1C — The data	
Definition	Data as a technical part of AI.
Coding Rules	Statements and workshop notes with focusing and elaborating on the data aspect of AI.
Example	*"Data accounts for 80% of the success. A good data highly leads to a good model. But what happens if you don't have good data? The key is in synthetic data and the data augmentation."* – P6 from interviews

1D — The model	
Definition	Models as a technical part of AI.
Coding Rules	Statements and workshop notes with focusing and elaborating on the model aspect of AI.

Example	*"Um, I mean, I guess like one problem I was facing early on was just that. That I didn't have like the computing power to really like train these like models. I mean, it is one like it, yeah, it takes a lot of like computing time to like properly train a model. Um, and yeah, just yeah, to if, um, you know, it requires a lot of like GPU resources and um, that costs money."* – P8 from interviews

1E — The output	
Definition	The outputs of AI systems.
Coding Rules	Statements and workshop notes indicating the nature of AI outputs.
Example	*"I think one of the interesting things about AI or some of the issues is you can't really understand exactly or immediately see why the results or predictions are like they are."* – P2 from interviews

1F — The challenges of AI	
Definition	The general challenges revolving around the implementation of AI systems.
Coding Rules	Statements and workshop notes expressing the challenges experienced in implementing AI.
Example	*"And after, oh yeah, after the, after we let's say we gather the data input there, I think there will be like a long and painful process to actually acquire the data. Because based on my experience, talking to some AI vendors it's always been their problem. Like… To, to even to get an access to something takes what like one month for, for like bureaucracy"* – P4 from interviews

1G — The outcomes of AI	
Definition	The perceived outcomes (potentially) brought by the implementation of AI technologies.
Coding Rules	Statements and workshop notes expressing the outcomes, both negative and positive, towards the potential changes brought by AI.
Example	*"AI facial recognition can excessively be used for surveillance leading to nightmares such as police state"* – W3 from workshop

2A — The process	
Definition	The processes pertaining to the approaches and the processes in designing AI.
Coding Rules	Statements on the approach designing AI alongside elaborations on its processes.
Example	*"So basically we could start with an opportunity mapping and usually either it's on, um, a company level where we look at the entire company or whether we look at one department for example..."* – P1 from interviews

2B — Collaborations throughout the process	
Definition	Identifying collaborators and ways of collaborations between differing roles in the process of creating AI.
Coding Rules	Statements regarding the roles involved and the ways of involvement in the creation process of AI.
Example	*"And um, yeah, then we had, uh, uh, yeah, computer scientists on our team as well, that came from actually, another department, but worked on those specific and AI related, uh, coding. And, uh, and then, yeah, and some old engineering people we used to work with, uh, yeah. Clients, people, uh, yeah. So..."* – P2 from interviews

2C — Challenges during the process	
Definition	General difficulties and challenges encountered during the process of designing and developing AI.
Coding Rules	Statements and workshop notes on general challenges either experienced or perceived for the process of designing and developing AI.
Example	*"Because most of the times it's always very, uh, I think at part, at times, it's very difficult to make informed decisions. Because simply, because of the level of complexity of the things we are dealing with."* – P3 from interviews

3A — The goal and role of design(ers)	
Definition	The goal and role of designers in the process of creating AI systems.
Coding Rules	● Statements explicitly noting design goals to

	achieve during the process. ● Statements indicating the roles that designers take during the process of creating AI systems.
Example	*"Um, and we're also collaborating with, uh, with a few design companies helping us to facilitate these workshops. So, so there's some, some really, really good companies out there that, that does that way better than we do."* – P5 from interviews

3B — Design(ers) contributions	
Definition	Ways in which design have or can potentially contribute to the development of AI systems.
Coding Rules	Statements crediting design for the potential contributions that it may bring to the development of AI.
Example	*"And what I think one of the things that design can offer and designers are good at, well, I, that's a very big statement, um, but I think design, we were very, we're very good at what is called abductive reasoning and the notion of synthesizing information and dealing with ill-defined problems."* – P3 from interviews

3C — Challenges for design(ers)	
Definition	Design challenges and difficulties faced by designers specific to designing with and for AI technologies.
Coding Rules	Statements by designers reflecting on their challenges or by developers giving assessment on potential challenges for design.
Example	*"I mean, having worked for like deeply in like the sort of like mathematics behind it and the, you know, like sort of, um, like, um, like computational, like theory behind like machine learning. I think that's like a really difficult space to kind of like bring, um, design into, um, this, because like that sort of like research field is like evolving like really fast. And prior to just like a lot of… Kind of like background knowledge and like understanding of like how, how these systems work on like yeah, like a mathematical and computational level. Um, that's kind of like, yeah, that that'd be like a difficult area for designers to come in."* – P8 from interviews

4A — Awareness of ethical considerations and its effects	
Definition	The awareness towards ethical considerations and its subsequent effects.

Coding Rules	Statements and workshop notes indicating awareness towards different ethical considerations alongside some of its subsequent effects.
Example	*"At least right now, there's also a lot of ethical discussions about the, and just in the public. So, so they are, um, they nearly, it's also really, when they come into an AI Design Sprint session, they are already, um, some of the questions or like, um, an AI Design Sprint sessions are very interactive, so sort of the, um, sometimes like the first thing people say it's already, um, uh, uh, ethical questions."* – P1 from interviews

4B — Reasons of applying values and considering ethics	
Definition	The reasons as to why values should be applied and ethics should be considered for AI systems.
Coding Rules	Statements and workshop notes elaborating reasons as to why values are applied and ethical concerns are considered.
Example	*"Um, and in the US you have all kinds of composition. You have all the demographics, you have all the races, you have all of that stuff. And we have to be very very conscious on that. So when I work in, um, in that cybersecurity... I'm sorry, civil engineering firm. And when I do a lot with, um, urban projects, I have to be very conscious on that, uh, for sure."* – P7 from interviews

4C — Applications of values and considering ethics in practice	
Definition	Existing and potential applications of how values are applied and ethics are considered in the day-to-day practice of creating AI systems.
Coding Rules	● Statements on how practitioners have applied values and ethics into their day-to-day practice ● Workshop notes on ideas how values and ethics can be put into practice
Example	*"There are quick fixes like having multidisciplinary teams, and inclusive workshops when designing an AI system."* – W1 from workshop

4D — Barriers in applying values and considering ethics in practice	
Definition	The challenges and barriers that creates difficulty in applying values and considering ethics in the practice of designing AI.

Coding Rules	Statements on the implied challenges, barriers, and difficulties in applying values and ethics in practice.
Example	*"Yeah. I mean, it's just not a, um, part of like the Silicon Valley sort of like culture and like mindset. I think that really kind of like stems from that, um, you know, I think a lot of the kind of work that happens, um, in Silicon Valley, isn't really like, you know, focused towards like human needs or like community needs. It's more focused on like, I dunno, what's like this new cool, like technology and like, you know, how can we use it to like, make a lot of money? And, you know, that was like, certainly like my mindset, like, you know, a few years ago."* – P8 from interviews

4E — Complexities in applying values and considering ethics	
Definition	The complexities rooted in the diversity and situated nature of values and ethics.
Coding Rules	Statements and workshop notes illustrating the complex issues of values and ethics.
Example	*"But this is, I think this is primarily going to be decided that this country based on country basis. And of course, so we are going now into very difficult debate on whether Western values are the best values. For me, European, not so much Western, but European values because American fellows are very different than European values. Are European values are the best thing? I think. But I'm a European I've... I've lived with these values. I like my privacy. I think this is the best thing for me. Um, again, with the… But I also, we have to be very conscious about the fact that this could be like a new version of colonialism right. So we are exporting our values. I really don't like that. Yeah."* – P3 from interviews