Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design

Qian Yang   **Aaron Steinfeld**   **Carolyn Rosé**   **John Zimmerman**

Carnegie Mellon University, Pittsburgh PA, USA

{yangqian,steinfeld}@cmu.edu   cp3a@andrew.cmu.edu   johnz@cs.cmu.edu

ABSTRACT

Artificial Intelligence (AI) plays an increasingly important role in improving HCI and user experience. Yet many challenges persist in designing and innovating valuable human-AI interactions. For example, AI systems can make unpredictable errors, and these errors damage UX and even lead to undesired societal impact. However, HCI routinely grapples with complex technologies and mitigates their unintended consequences. What makes AI different? What makes human-AI interaction appear particularly difficult to design? This paper investigates these questions. We synthesize prior research, our own design and research experience, and our observations when teaching human-AI interaction. We identify two sources of AI's distinctive design challenges: 1) uncertainty surrounding AI's capabilities, 2) AI's output complexity, spanning from simple to adaptive complex. We identify four levels of AI systems. On each level, designers encounter a different subset of the design challenges. We demonstrate how these findings reveal new insights for designers, researchers, and design tool makers in productively addressing the challenges of human-AI interaction going forward.

Author Keywords

User experience, artificial intelligence, sketching, prototyping.

CCS Concepts

•**Human-centered computing** → **Human computer interaction (HCI)**; *Interaction design process and methods;*

INTRODUCTION

Advances in Artificial Intelligence (AI) have produced exciting opportunities for human-computer interaction (HCI). From mundane spam filters to autonomous driving, AI holds many promises for improved user experiences (UX), and it enables otherwise impossible forms of interaction. This trend has led to the idea of *AI as a design material* in the research community, with the hope that HCI researchers and designers can effectively envision and refine new uses for AI that have yet to be imagined [13, 26, 58].

The growing interest in how AI can improve UX implies that HCI designers have become skilled at integrating AI's capabilities into their practices. Interestingly, the research shows something else, that HCI designers largely struggle to envision and prototype AI systems [2, 3, 13, 18, 22, 26, 49]. For example, even simple AI applications can make inference errors that are difficult to anticipate. These errors impact the intended user experience, and can sometimes raise serious ethical concerns or result in societal-level consequences. However, current HCI design methods meant to mitigate unintended consequences (i.e. sketching and prototyping) can seem ill-fitted for AI. HCI professionals cannot easily *sketch* the numerous of ways an AI system might adapt to different users in different contexts [13, 58]. Nor can they easily *prototype* the types of inference errors a not-yet developed AI system might make [29, 42, 49].

Existing research frequently attributes these challenges to AI's technical complexity, demand for data, and unpredictable interactions [49, 26, 13, 42]. Less discussed is that HCI routinely grapples with complex, resource-intensive technologies using simple, inexpensive artifacts, e.g., paper prototypes and Wizard of Oz systems. What makes AI different and distinctly difficult to prototype? Equally important, designers routinely choreograph complex, dynamic, sometimes unpredictable interactions, with a focus on mitigating technologies' unintended consequences (e.g., [61]). What makes AI interactions particularly difficult to sketch? A critical first step in designing valuable human-AI interactions is to articulate the unique qualities of AI that made it so difficult to design.

The goal of this paper is to delineate *whether*, *why*, and *how* human-AI interaction is distinctly difficult to design and innovate. The paper has four parts:

1. We set the stage by cataloging the many human-AI interaction design challenges that literature has reported as well as solutions proposed.
2. We ask three provocative questions as a critique of current related work. These questions serve as a springboard for rethinking how to approach the question of why human-AI interaction appears so difficult to design.
3. We synthesize our own research, including studies on the challenges HCI practitioners faced when working with AI, our insights from making AI things via research through design, and our insights from teaching students in human-AI interaction design and innovation. This synthesis identified

Difficult to articulate what AI can/cannot do
  Technical feasibility of a design idea is highly dependent on data
  Do not know how to purposefully use AI in the design problem at hand
    Do not know how to express AI design ideas
    Difficult to sketch divergent AI interactions
      Cannot fast prototype AI system behavior
        Difficult to foresee the potential effects of AI
          Difficult to design fuzzy, open-ended interactions
            Difficult to explain AI behaviors to users
            Difficult to design shared control between the AI and the users
            Difficult to design interactions that constantly improve AI performance
              Difficult to communicate AI system evolvement over time to users
                Difficult to anticipate/mitigate unpredictable AI behaviors
                Difficult to make sure the AI is not creepy
                  Do not know whom to hold accountable for AI errors

Human-AI Interaction Design Challenges

User-Centered Design Processs

○   discover   define   ○   develop   deliver   ○

Human-AI Interaction Design Facilitators

WoZ/iML tools for exploring AI interactions
Data exploration/storytelling tools
AI technical education for designers

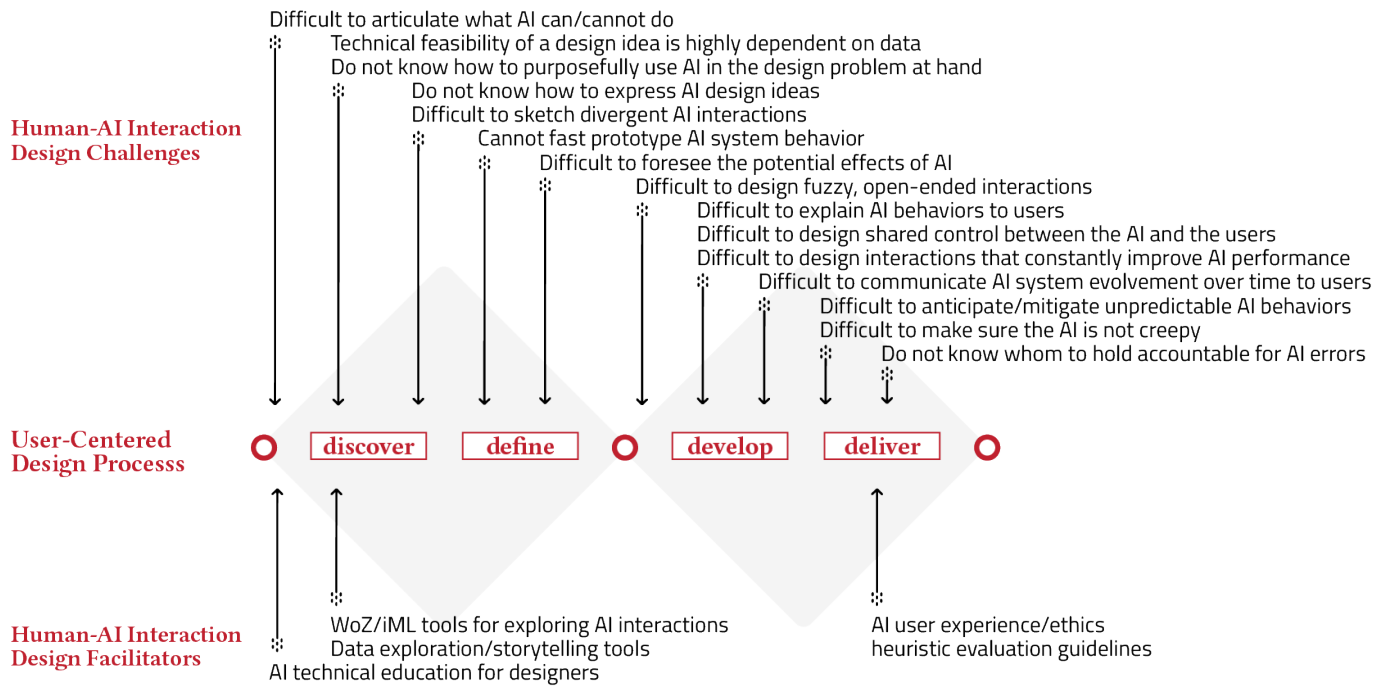AI user experience/ethics
heuristic evaluation guidelines

Figure 1: Mapping the human-AI interaction design challenges in the literature [58, 13, 26, 53]
onto a user-centered design process (Double Diamond [10])
.

two sources of AI's design complexities, and a framework that unravels their effects on design processes.

4. We demonstrate the usefulness of the framework. Specifically, its usefulness to human-AI interaction designers, to researchers of AI's HCI issues, and to AI design method innovators and tool makers.

This paper makes three contributions. First, it provides a synthesis of many human-AI interaction design challenges and emergent solutions in literature. Second, the provocation questions offer an alternative lens for approaching the human-AI interaction design challenge. We draw attention to AI's design complexity rather than technical complexity; We draw attention to how AI hinders the interaction design process rather than the end product. Finally, our framework gives structure to the currently fuzzy problem space of human-AI interaction design. This provides a first step towards systematically understanding how HCI research might best choose to approach these problems going forward.

RELATED WORK
Recent research has become increasingly interested in the opportunities and challenges AI brings to HCI and UX design. As researchers produced a wealth of valuable, novel designs, they also reported encountering many challenges in the process [2, 3, 18, 26]. Some research has investigated challenges faced by UX practitioners who do not specialize in AI but who desire to integrate it into their practice [13, 53]. Research has chosen a number of different frames for investigating these challenges including human-AI interaction

design, AI/machine learning as a design material, the design of intelligent systems, designing for/with data [6, 14, 37], and many more [33, 42, 43].

To better unpack what is known about the challenge HCI researchers and practitioners face when working with AI, we cataloged these challenges and their emergent solutions. To gain a new perspective of this work, we mapped the challenges and solutions to the familiar double diamond design process used to describe user-centered design (Figure 1) and to a diagram displaying a lean startup process with its focus on producing a minimal viable product (MVP) (Figure 2), a design approach becoming more popular with the growing use of agile software development.

UX Design Challenges of AI
Across HCI and UX communities, researchers and practitioners have reported challenges in working with AI at almost every step of a user-centered design process. From left to right on Figure 1, they reported:

- *Challenges in understanding AI capabilities (first divergent thinking stage)*: Designers frequently report that it is difficult to grasp what AI can or cannot do. This hampers designers' brainstorming and sketching processes from the start [13, 26, 51, 56].

- *Challenges in envisioning many novel, implementable AI things for a given UX problem (in both divergent thinking stages)*: AI-powered interactions can adapt to different users and use contexts, and they can evolve over time. Even when
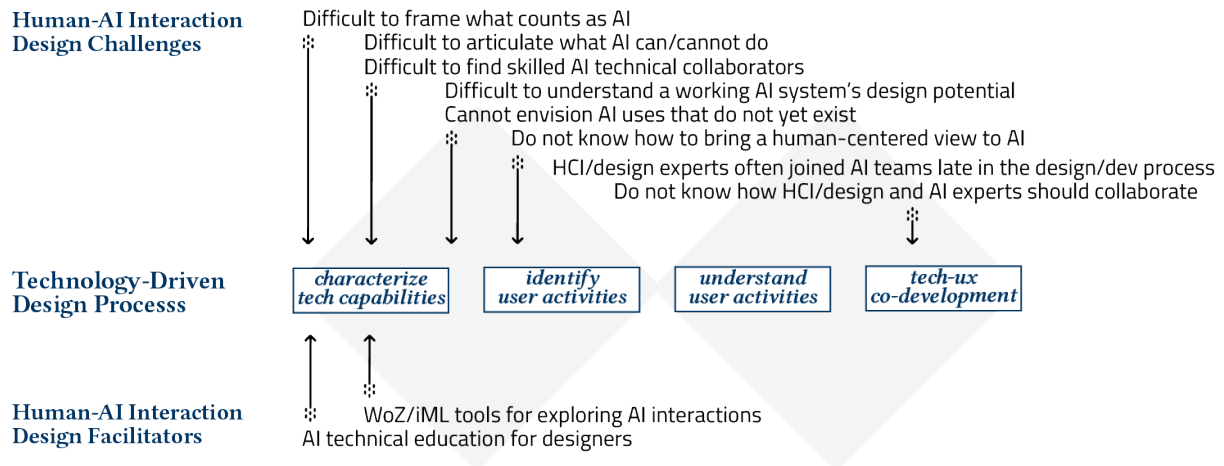
Human-AI Interaction
Design Challenges

Difficult to frame what counts as AI
  Difficult to articulate what AI can/cannot do
Difficult to find skilled AI technical collaborators
    Difficult to understand a working AI system's design potential
Cannot envision AI uses that do not yet exist
  Do not know how to bring a human-centered view to AI
    HCI/design experts often joined AI teams late in the design/dev process
Do not know how HCI/design and AI experts should collaborate

Technology-Driven
Design Processs

| *characterize tech capabilities* | *identify user activities* | *understand user activities* | *tech-ux co-development* |

Human-AI Interaction
Design Facilitators

  WoZ/iML tools for exploring AI interactions
AI technical education for designers

Figure 2: Mapping UX design challenges of AI in prior research on a technology-driven design innovation process [41, 5]

designers understand how AI works, they often found it difficult to ideate *many* possible new interactions and novel experiences with much fluidity [13, 58].

- *Challenges in iterative prototyping and testing human-AI interaction (in both convergent thinking stages):* One core practice of HCI design and innovation is rapid prototyping, assessing the human consequences of a design and iteratively improving on it. HCI practitioners cannot meaningfully do this when working with AI. As a result, AI's UX and societal consequences can seem impossible to fully anticipate. Its breakdowns can be especially harmful for under-served user populations, including people with disabilities [45].

  HCI researchers have tried two approaches to addressing this challenge. One approach is to create Wizard of Oz systems or rule-based simulators as an early-stage interactive AI prototype (e.g. as in [11, 30, 40, 44]). This approach enables HCI professionals to rapidly explore many design possibilities and probe user behaviors. However, this approach fails to address any of the UX issues that will come from AI inference errors because there is no way to simulate these errors [52]. The second approach is to create a functioning AI system, and deploy it among real users for a period of time [53]. This time-consuming, field-trial prototyping process enables designers to fully understand AI's intended and unintended consequences. However, it loses the value that comes from rapid and iterative prototyping. This approach does not protect teams from over-investing in ideas that will not work. It does not allow them to fail early and often.

- *Challenges in crafting thoughtful interactions (in the last convergent thinking stage)*: Designers struggled to set user expectations appropriately for AI's sometimes unpredictable outputs [4]. They also worried about the envisioned designs' ethics, fairness, and other societal consequences [13, 26].

- *Challenges in collaborating with AI engineers (throughout the design process*: For many UX design teams, AI

technical experts can be a scarce resource [19, 53]. Some designers also found it challenging to effectively collaborate with AI engineers, because they lacked a shared workflow, boundary objects, or a common language for scaffolding the collaboration [19, 28, 52].

Propelled by these challenges, a few researchers speculated that, when working with AI, designers should start with an elaborate matching process that pairs existing datasets or AI systems with the users and situations that are most likely to benefit from the pairing [5, 51]. This approach deviates from more traditional user-centered design in that the target user or UX problem is less fixed. It is more similar to customer discovery in an agile development process that focuses on the creation and continual evaluation of a minimal viable product (MVP) [41]. In this light, we also mapped the human-AI interaction design challenges onto an MVP innovation process. However, it seems a similar set of design challenges that curbed user-centered design also thwarted technology-driven design innovations (Figure 2, from left to right), for example:

- Challenges in understanding AI capabilities;
- Challenges in mapping out the right user stories and user cases of a "minimum viable" AI system, or envisioning how it can be applied in less obvious ways [13];
- Challenges in collaborating with AI engineers.

We found no agreed-upon set of root causes or themes around which we can easily summarize these challenges. Some researchers suggested that AI systems' technical complexity causes the interaction design problems [9]. Some considered the unpredictable system behaviors as the cause [26]. Some argued that AI appeared to be difficult to design because AI is just "*a new and difficult design material,*" suggesting that over time, known HCI methods will likely address these challenges [13]. Others argued that user-centered design needs to change in order to work for AI [19, 51]. These proposals rarely share key citations that indicate emerging agreements.

Facilitating Human-AI Interaction Design

HCI researchers have started to investigate how to make it easier to design and innovate human-AI interactions. We identify five broad themes in this body of work:

1. *Improving designers' technical literacy.* An emerging consensus holds that HCI and UX designers need *some* technical understanding of AI to productively work with it. Designer-facing AI education materials have become available to help (e.g. [8, 20, 22, 23]). However, substantial disagreement remains in what kinds of AI knowledge are relevant to UX design, and in how advanced a technical understanding is good enough for designers [9, 48, 54].

2. *Facilitating design-oriented data exploration.* This body of work encourages designers to investigate the lived-life of data and discover AI design opportunities [6, 7, 14]. For example, [37] investigated users' music app metadata as a material for designing contemplative music experience; [24] explored the design opportunities around intimate somatic data. Notably, this body of work often used terms like data-driven or smart systems; It was not always clear when the authors specifically aimed at AI.

3. *Enabling designers to more easily "play with" AI in support of design ideation, so as to gain a felt sense of what AI can do.* This work created interactive machine learning (iML) tools and rule-based simulators as AI prototyping tools, for example, Wekinator for gesture-based interactions [1] and the Delft AI Toolkit for tangible interactions [49]. Currently, almost all iML tools are application-domain-specific. In order to make the systems accessible to designers and maximally automate data prepossessing and modeling, these systems had to limit the range of possible in/outputs, therefore focused on particular application domains [38, 39].

4. *Aiding designers in evaluating AI outputs.* In recent years, technology companies have proposed more than a dozen human-AI interaction principles and guidelines (See a review here [47]). These guidelines covered a comprehensive list of design considerations such as "make clear how well the system can do, what it can do" [4] and "design graceful failure recovery" [21].

5. *Creating AI-specific design processes.* Some researchers have proposed that AI may require design processes less focused on one group of users, and instead on many user groups and stakeholders [15]; processes focused less on fast, iterative prototyping, and instead on existing datasets and functioning AI systems [51]; or processes focused less on one design as the final deliverable to engineers, and instead on closer, more frequent collaborations [19].

These themes demonstrated the remarkable heterogeneity of approaches researchers have taken to address the challenges around human-AI interaction design. Similar to most design methods published within HCI research, we found no empirical evaluations of the proposed design tools, guidelines, or workflows. It is difficult to control for and measure improvements in a design process to show that a method is producing better designs. Throwing AI into the mix only seems to increase this challenge.

THREE QUESTIONS FOR PROVOCATION

We wanted to articulate whether, why, and how AI is distinctly difficult to design. The preceding review of related work revealed a remarkable set of insights and approaches to this complex problem space. Now we step back and critically examine this rich body of research in order to more holistically understand AI's resistance to design innovation. What has research missed? Can we see gaps or emerging opportunities across this work? Our reflection of the related work led to three provocative questions. These questions served as a springboard for rethinking how we might further advance our understanding of AI's design challenges.

What is AI?

One critical question has been absent from the research discourse around human-AI interaction: *What is AI?* Or rather, what should count as AI as it relates to HCI and UX design? Prior literature has used a range of poorly-defined terms, such as machine learning systems, intelligent/smart systems, AI-infused systems, and more. The research discourse on understanding machine intelligence as a technological material is sometimes mixed with intelligence as an experiential factor.

Locating the elusive concept of AI is difficult. What is commonly referred to as AI encompasses many disconnected technologies (e.g., decision tree, Bayesian classifier, computer vision, etc.). The technical boundary of AI, even in AI research communities, is disputed and continuously evolving [46, 50]. More importantly, an actionable, designerly understanding of AI is likely to be very different from a technical definition that guides algorithmic advances.

Yet discussing AI's design challenges without *any* bounding seems problematic. What makes a challenge distinctly AI and not a part of the many challenges designers regularly face in HCI and UX work? Current research does not always make this distinction. For example, Amershi et al. systematically evaluated 20 popular AI products and proposed a set of guidelines for designing human-AI interactions [4]. These guidelines include "make clear what the system can do" and "support efficient error correction". These seem important to AI systems, but they also seem to be issues that designers must address in systems with no AI. What is less clear is if AI requires specific considerations in these areas.

What Are AI's Capabilities and Limits?

Designers need to understand the capabilities and limitations of a technology in order to know the possibilities it offers for design [17]. Engineers create new technological capabilities; designers create new, valuable products and services with existing technological capabilities and limitations [34].

Interestingly, AI's capabilities and limitations have not been the focus of current research. Instead, most work has focused on getting designers to understand how AI functions (2.1 theme 1). This is quite different from the traditional ways of moving new technology from research to design practice,

which assume designers do not need to understand the technical specifics of the technology. In addition, research has produced many rule-based and Wizard of Oz simulators to help designers better understand AI's design opportunities (themes 2 and 3). Little is known about whether these systems can sensitize designers to AI's limitations realistically. This motivates the question: Can an articulation of AI's capabilities foster a more incisive examination of its design challenge?

Why Is Prototyping AI Difficult?
AI brings challenges to almost all stages of a typical design process. However, the proposed AI design methods and tools have mostly focused on the two ends of this creative process (Figure 1 and 2); either helping designers to understand what AI is and can do generally, or enhancing the evaluation of the final design. The central activities of an interaction design process, i.e. sketching and prototyping, are under-explored. Research through Design (RtD) projects are rare when it comes to designing and innovating human-AI interaction [51].

Sketching and prototyping may constitute a fruitful lens for understanding AI's design challenges. They are cornerstones of any design process. It is through sketching and prototyping that designers understand what the technology is and can do, engage in creative thinking, and assess and improve on their designs. Interrogating why is it difficult to abstract AI-powered interactions into sketches and prototypes may shed light on how the other tangled design challenges relate to each other.

METHOD
We set out to investigate whether, why, and how human-AI interaction is uniquely difficult to design and innovate. We want to construct a framework that provides meaningful structure to the tangled challenges identified in prior research. The preceding provocation questions informed how we advance towards this goal: We first worked to identify an operational bounding of AI. Within this bounding, we curated a set of human-AI interaction sketching and prototyping processes as case studies. We synthesized these case studies, in searching for a coherent, useful framework.

One limitation of this work is that the case studies are mainly from our own research/design/teaching experiences. This is neither a representative sample nor a comprehensive one. The meta-analysis nature of our research goal calls for an extensive collection of AI design projects, ideally covering *all* kinds of AI systems for *all* kinds of design contexts. This is beyond what one paper can achieve. The synthesis of our experience and the resulting framework are intended to serve as a moderate first step in this direction.

An Operational Bounding of AI
The definitions of AI generally fall into two camps [27, 32, 50]. One describes AI as computers that perform tasks typically associated with the human mind ("*AI is whatever machines haven't done yet*" [25]). The other defines AI in relation to computational mechanisms. We chose a widely-adopted definition from the latter camp, because our focus is AI the technology, rather than what people perceive as "intelligent".

In this work, AI refers to computational systems that *interpret external data, learn from such data, and use those learnings to achieve specific goals and tasks through flexible adaptation.* [27]

Importantly, we did not intend to draw a technical boundary of what counts as AI here. We also do not consider this definition as valuable for HCI practitioners in working with AI. Instead, we used this definition only to examine whether the systems that are technically considered as AI indeed require new HCI design methods. For example, this definition describes AI as "learning" from data, yet does not specify what counts as "learning." (It remains an issue of debate in technical AI communities.) Therefore in our synthesis, we considered the challenges designers reported in working with a full range of data-driven systems, including machine learning, classic expert systems, crowd sourcing, etc. We then examined whether the challenges are different across the spectrum from systems that we all agree "learned" from data to those that we all agree did not. This way, we started to separate the design challenges that are unique to AI and those HCI routinely copes with.

UX Design Processes as Data
Within this bounding, we curated a set of AI design process from our own research, design, and teaching experience. All projects described below except teaching have been published at DIS and CHI. We re-analyzed the data collected across these projects for the purpose of this work. Below is a brief overview of these projects.

Designing the UX of AI Applications
First, we draw on our many years of experience in designing a wide range of AI systems, from simple adaptive user interfaces [58, 12], to large-scale crowd-sourced transportation systems [59]; from clinical decision supports [57, 55] to natural language productivity tools [52, 60, 16]. These experiences enabled us to give a firsthand account of the design challenges of AI, as well as a felt understanding of the solution that naturally emerged from the process.

Studying UX Practitioners
We have studied HCI/UX practitioners and their AI engineer collaborators in two projects. The first project focused on novice AI product designers [56]. We interviewed 14 product designers/managers and surveyed 98 more to understand how they incorporated, or failed to incorporate, AI in their products. We also interviewed the 10 professional AI engineers they hired to better understand where and how designers sought help. The second project focused on experienced UX practitioners [54]. We interviewed 13 designers who had designed AI applications for many, many years, in order to understand how they work with AI differently compared to working with other technologies. Synthesizing and contrasting the findings across these two studies, we were able to see how novice and expert designers approached designing AI differently.

Teaching UX Design of AI Applications
Another set of observations come from our teaching. We hosted a series of *Designing AI* workshops. Each workshop lasted for a day, with one instructor working with 2-3 students. The instructor first gave a half-hour introduction to AI, and

then provided students with a dataset and a demonstrational AI system. Students were asked to design new products/services with these materials for an enterprise client. 26 HCI Master students from two universities attended the workshop. All of them had little to no technical AI background. Throughout the series, we experimented with different ways of introducing AI. We observed how students used the AI technical knowledge in their design, where and how they struggled, and which challenges they were able to resolve with known design methods.

We also taught a one semester design studio course: *Designing AI Products and Services*. Approximately 40 undergraduate and master students took the course. About half of them had a computer science or data science background. In comparison to the workshops, the course allowed us to observe students working with a more diverse set of AI systems and design tasks, e.g. designing crowd as a proxy for AI, designing simple UI adaptions, designing natural language interactions.

Data Analysis
With this diverse set of design processes and observations, we synthesized a framework meant to give structure to the many challenges around human-AI interaction design. We started by proposing many themes that might summarize these challenges. We then analyzed the emergent themes via affinity diagramming, with a focus on the characteristics of AI that may scaffold a full range of design challenges. Specifically, we critiqued these frameworks based on three criteria:

- *Analytical leverage*: The framework should effectively scaffold a wide range of AI's design opportunities and challenges. It should help separate design challenges unique to AI from others;

- *Explanatory power*: The framework should help researchers articulate how a proposed design method/tool/ workflow contributes to the challenges of human-AI interaction design, and the limits of its generalizability.

- *Constructive potential*: The framework should not only serve as a holder of AI's known challenges and solutions; It should also provide new insights for future research.

We proposed and discussed more than 50 thematic constructs and frameworks. The three authors, an external faculty, and an industry researcher participated in this process. All have spent at least 5 years researching AI and HCI. We also presented this work to two research groups. One included about 40 HCI researchers. The other included 12 machine learning researchers. They provided additional valuable critiques and helped us refine the framework.

THE FRAMEWORK
Our synthesis identified two attributes of AI that are central to the struggles of human-AI interaction design: *capability uncertainty* (uncertainties surrounding what the system can do and how well it performs) and *output complexity* (complexity of the outputs that the system might generate). Both dimensions function along a continuum. Together they form a valuable framework for articulating the challenges of human-AI interaction. This section describes the framework. In the next section, we demonstrate its usefulness.

Two Sources of AI Design Complexity

Capability Uncertainty
When speaking of the capabilities of AI, we broadly refer to the functionality AI systems can afford (e.g. detect spam emails, rank news feeds, find optimal driving routes), how well the system performs, and the kinds of errors it produces. The capabilities of AI is highly uncertain. We illustrate this by walking through the lifetime of an AI system, moving from an emergent algorithmic capability in AI research labs to situated user experience in the wild (Figure 3, left to right).

AI's capability uncertainty is at its peak in the early design ideation stage, when designers work to understand what design possibilities AI can offer generally. This is not easy because there exists no catalog of available AI capabilities. What might seem like a blue-sky AI design idea may suddenly become possible because of a newly available dataset. The performance of a deployed AI system can constantly fluctuate and diverge when it gains new data to improve its learning. This great uncertainty in AI's capabilities makes it difficult for designers to evaluate the feasibility of their emergent ideas, thereby hindering their creative processes.

The range of AI's available capabilities includes more than the capabilities of existing AI systems. It includes any AI things that are technically feasible. When envisioning AI systems that do not yet exist, designers face additional capability uncertainties. For example, designers may choose to harvest their own dataset from users' traces of interaction. This approach gives designers a relatively high degree of control over the data they will eventually work with. However, it is often very difficult to estimate how long it might take to collect enough high-quality data and to achieve the intended functionality. Designers frequently worked with user-generated data in order to understand available AI capabilities. To understand AI's capabilities, to a great extent, is to understand this gap between what the data appear to promise and the AI system built from that data can concretely achieve. As one expert designer we interviewed describes: To understand what AI can do is to conceptualize "*a funnel of what (data and/or system) exists and what is possible.*" [53]

Alternatively, designers may choose to leverage existing AI libraries or pre-built models to address their design problem at hand. These systems free designers from the data troubles and allow them to get a felt experience of the AI interactions. Unfortunately, these toolkits represent a very narrow subset of the whole landscape of AI capabilities.

What AI can do for a UX problem at hand becomes clearer once a functioning AI system is built. For most systems trained on self-contained datasets, designers can measure their performance and error modes, and then make design choices accordingly. However, this performance is limited by any biases present in a dataset and should only be viewed as an initial estimate (*system performance* in Figure 3).

Some AI systems continue to learn from new data after deployment (labeled as "deployed system performance over time" in Figure 3). In the ideal case, the system will "grow," integrating new insights from new data and adapting flexibly to more va-

rieties of users and use contexts. Unfortunately, the new data might also drive system performance in the wrong direction. Tay, the Twitter bot, provides an extreme example [36]. More typically, the system's performance improves for users and use contexts that have produced rich data. It performs worse for less frequent users and less typical situations. That the system capability can constantly evolve, fluctuate, and diversify is another part of AI's capability uncertainty.

Finally, user profiles and use contexts could also impact an AI system's capability. Many context-aware and personalization systems fall into this category. Consider the social media news feed ranker, Amazon shopping recommendations, and ride-hailing app's driver-rider matching as examples. It is not difficult to conceptualize what these systems can do in general (e.g. ranking news, recommending items); however, it is no trivial task to envision, for a particular user in a particular use context, what error the AI system might make, and how the user might perceive that error *in-situ*. Anticipating the situated, *user-encountered capability* of AI is difficult, yet it is fundamental to *user experience* design.

Output Complexity
The second source of human-AI interaction challenges concerns what an AI system produces as a possible output. While *capability uncertainty* is responsible for the HCI design challenges around understanding what AI can do, AI's *output complexity* affects how designers conceptualize the system's behaviors in order to choreograph its interactions.

Many valuable AI systems generate a small set of possible outputs. Designing interactions for these systems is similar to designing for non-AI systems that generate probabilistic outputs. A face detection tool, for example, outputs either "face" or "not face." To design its interactions, the designer considers four scenarios: when a face is correctly detected (true positive), when no face is detected (true negative), when there is no face and a face is mistakenly detected (false positive), and when the image contains a face but the system fails to detect it (false negative). Designers consider each condition and design accordingly.

When designing systems that produce many possible outputs, sketching and prototyping become more complex and cognitively demanding. Imagine designing the interactions of a driving route recommender. How many types of errors could the recommender possibly produce? How might a user encounter, experience, and interpret each kind of error, in various use contexts? How can interaction design helps the user to recover from each error elegantly? Some simulation-based methods or iML tools can seem necessary for prototyping and accounting for the route recommender's virtually infinite variability of outputs. The route recommender exemplifies the many AI systems that produce open-ended, adaptive outputs. The traditional, manual sketching and prototyping methods struggle to fully capture the UX ramifications of such systems.

The system outputs that entail most design complexities are those that cannot be simulated. Consider Siri as an example. Similar to route recommenders, Siri can generate infinite possibilities of outputs. Yet unlike route recommenders, the relationship between Siri's in- and outputs follow complex patterns that cannot be concisely described. As a result, rule-based simulators cannot meaningfully simulate Siri's utterances; nor can a human wizard. We refer to such AI system outputs as "complex."

Notably, output *complexity* is not output *unpredictability*. While prior research often viewed AI systems' unpredictable errors as causing UX troubles, we argue that AI's output complexity is the root cause. Let us illustrate this by considering how designers might account for AI errors when designing two different conversational systems. One is Siri. The other is a system that always replies to user requests with a random word picked from a dictionary. While highly unpredictable, the interactions of the latter system can be easily simulated by a random word generator. Then following a traditional prototyping process, designers can start to identify and mitigate the AI' costly errors. In contrast, Siri's outputs are only quasi-random, therefore resist abstraction or simulation. To date, it remains unclear how to systematically prototype the UX of such systems, in order to account for its breakdowns.
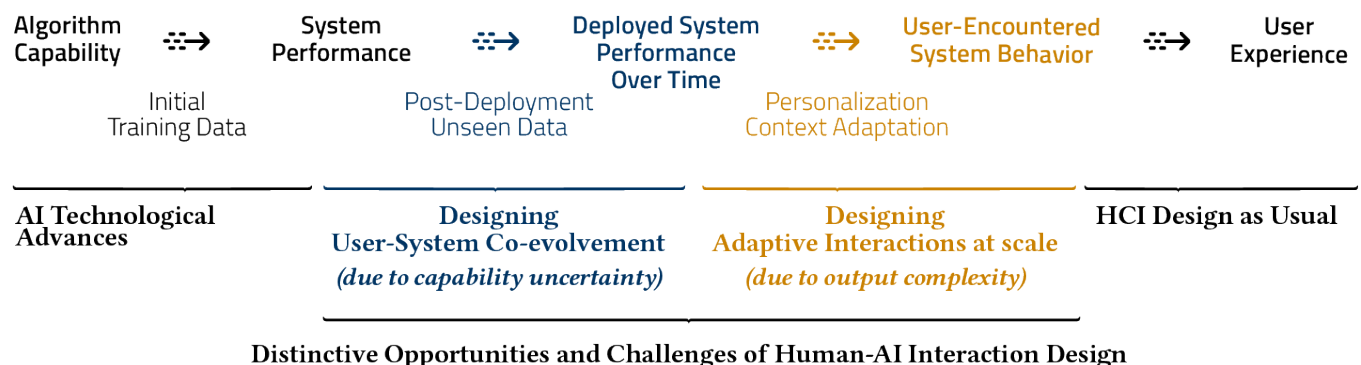
Algorithm Capability ⇢ System Performance ⇢ Deployed System Performance Over Time ⇢ User-Encountered System Behavior ⇢ User Experience

Initial Training Data   Post-Deployment Unseen Data   Personalization Context Adaptation

AI Technological Advances	Designing User-System Co-evolvement *(due to capability uncertainty)*	Designing Adaptive Interactions at scale *(due to output complexity)*	HCI Design as Usual

Distinctive Opportunities and Challenges of Human-AI Interaction Design

Figure 3: The conceptual pathway translating between AI's capabilities and thoughtful designs of human-AI interaction. AI's capability uncertainty and output complexity add additional steps (the colored segments) to a typical HCI pathway, make some systems distinctly difficult to design. Designers encounter these challenges from left to right when taking a technology-driven innovation approach; right to left when following a user-centered design process.

Two Complexity Sources Taken Together

Prior research has identified a wealth of human-AI interaction design challenges. These challenges stem from AI's *capability uncertainty* and *output complexity*. For instance, designers struggled to understand what AI can and cannot do even when they understood how AI works [13]; This is because the capabilities of an AI system can be inherently uncertain and constantly evolving. Designers struggled to rapidly prototype human-AI interaction [52] because the interactions of two mutually adaptive agents resist easy abstraction or simulation. Designers struggled to follow a typical user-centered design workflow when designing human-AI interactions [19, 51]. This is because the central point of a double diamond process is to identify a preferred future, a defined design goal that existing technologies can achieve. However, AI systems have capabilities that do not fully take shape until after deployment, so the preferred future can seem like "*a funnel of what's possible*", rather than what is concretely achievable.

Figure 3 maps the challenges onto the translation process between technological capabilities and user experience. When taking a user-centered design approach, designers will encounter the challenges from the right to left. Taking a technology-driven design innovation approach, from left to right. This diagram explains why a similar set of design challenges appeared to have thwarted both technology-driven and user-centered AI design processes.

AI's evolving capabilities and adaptive behaviors have made it a particularly powerful material for HCI and UX design. The same qualities also bring distinctive design challenges. Human-AI interaction design and research, therefore, should not simplistically reject AI's capability uncertainty or output complexity/unpredictability. Rather, it is important to understand how to leverage these distinctive qualities of AI for desirable human ends, while minimizing their unintended consequences.
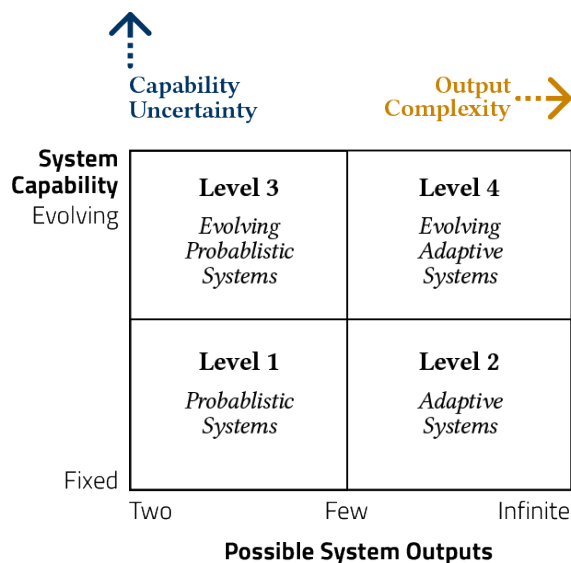

Figure 4: The AI design complexity map.

USING THE FRAMEWORK

In this section, we demonstrate the usefulness of the framework. Specifically, its usefulness to human-AI interaction designers (section 6.1), to researchers of AI's HCI issues (6.2), and to AI design method innovators and tool makers (6.3).

Four Levels of AI Systems

The framework can help expose *whether* and *how* a given AI system is difficult to design with traditional HCI design processes and methods. Existing HCI sketching and prototyping methods can sufficiently cover level one systems, systems with known capability with few possible outputs. New challenges emerge when designers work with systems that produce a broad set of possible outputs, and when the deployed system continues to learn from new user data. Therefore, for practitioners, the framework can help identify the low-hanging fruit in integrating AI into their practice. For HCI researchers, the framework can help identify the unique challenges of human-AI interaction design and make a targeted contribution.

To make the framework easier to use as an analytical tool, we summarized four levels of AI systems according to their design complexity (Figure 5). We demonstrate its usefulness using Levels 1 and 4 systems as examples since they represent the two extremes of AI's design complexity. The design challenges of Level 4 are also a superset of issues encountered in Levels 2 and 3.

Level One: Designing Probabilistic Systems

Level one systems learn from a self-contained dataset. They produce a small, fixed set of outputs. For example, face detection in camera apps, adaptive menus that ranks which option the user is more likely to choose, text toxicity detectors that classify a sentence as profane or not. Designers can design the UX of these systems in similar ways as designing non-AI, probabilistic systems. They are unlikely to encounter the distinctive challenges of human-AI interaction design.

Consider this design situation: a design team wants to help online community moderators to more easily promote civil discourses by using a text classifier that flags toxic comments.

☐ *No particular challenges in understanding AI capabilities:* By playing with the system, the designers can develop a felt understanding of what the classifier can and cannot do. Because the system will not learn from new data, this understanding will remain valid post-deployment.

☐ *No particular challenges in envisioning novel and technically feasible designs of the technology*: Designers can easily imagine many use scenarios in which the flagging-profane-text functionality can provide value.

☐ *No particular challenges in iterative prototyping and testing:* Because the outputs of the system are limited (profane, not profane), designers can enumerate all the ways in which the interactions may unfold (false positive, false negative, etc.) and making interactive prototypes accordingly.

☐ *No particular challenges in collaborating with engineers*: Once the designers understand the functionality and the likely performance and errors of the classifier, they can
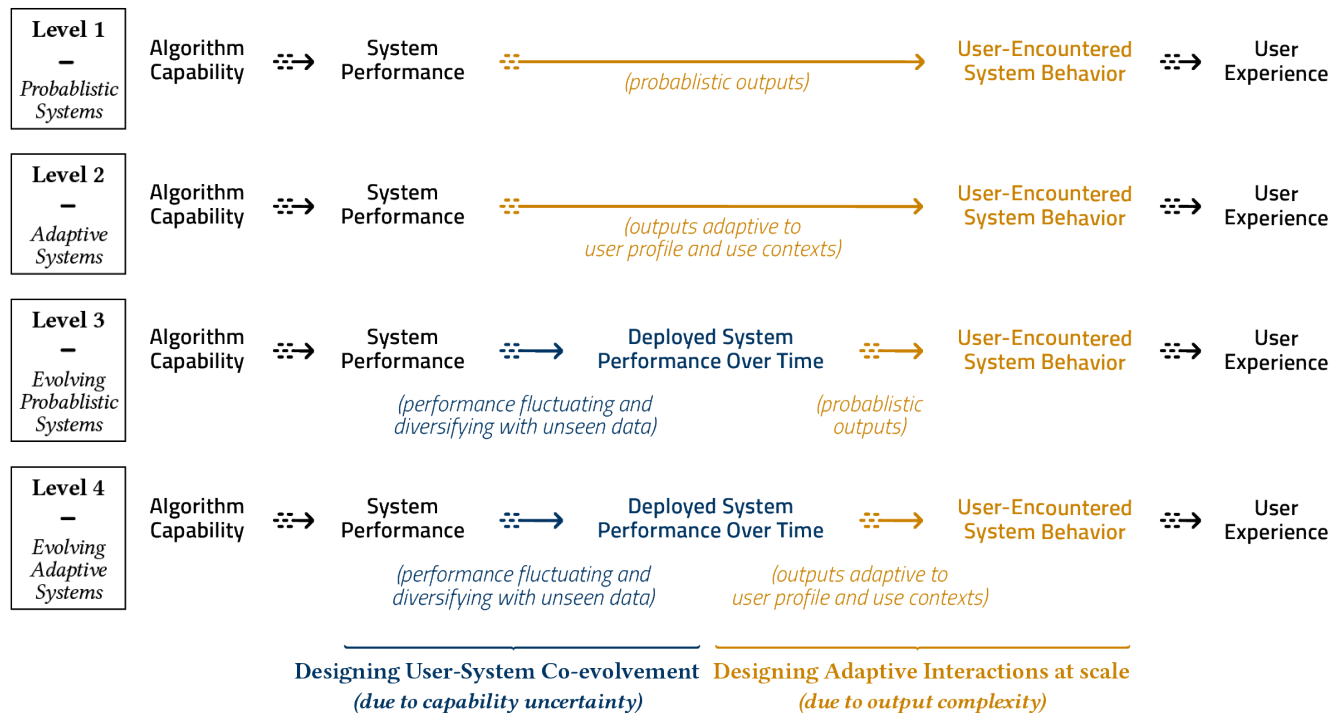
Figure 5: Four levels of AI systems according to design complexity.

design as usual and provide wireframes as a deliverable to engineers at the end of their design process.

Language toxicity detection is a complex technical problem at the frontier of AI research. However, because the system's capabilities are bounded and the outputs are simple, existing HCI design methods are sufficient in supporting designers in sketching, prototyping, and assessing its interactions. Language toxicity exemplifies level one systems; They are valuable, low-hanging fruits for HCI practitioners to integrate into today's products and services.

Level Four: Designing Evolving, Adaptive Systems
Level four systems learn from new data even after deployment. They also produce adaptive, open-ended outputs that resist abstraction. Search engines, newsfeed rankers, automated email replies, a recommender system that suggests "items you might like," would all fit in this category. In designing such systems, designers can encounter a full range of human-AI interaction design challenges. Consider the face recognition system within a photos app. It learns from the photos the user uploaded, clusters similar faces across photos, and automatically tags the face with the name inferred from the user's previous manual tags.

■ *Challenges in understanding AI capabilities*: The system's performance and error modes are likely to change as it learns from new images and tags. Therefore it is difficult to anticipate what the system can reliably do, when and how it is likely to fail. This, in turn, makes it difficult to design appropriate interactions for these scenarios.

■ *Challenges in envisioning novel and technically feasible designs of the technology*: Re-imagining many new uses of a face-recognition-and-tagging tool – beyond tagging people on photos – can be difficult. This is because its capabilities are highly evolved and specialized for its intended functionality and interactions.

■ *Challenges in iterative prototyping and testing:* The system's capabilities evolve over time as users contribute more images and manually tags, challenging the very idea of rapid prototyping.

■ *Challenges in collaborating with engineers*. The system requires a closer and more complex HCI-AI collaboration than as in a traditional double-diamond process. Engineers and designers need to collaborate on understanding how the face-recognition performance will evolve *with* users' newly uploaded photos and tags, how to mitigate the AI's potential biases and errors, as well as how to detect AI errors from user interactions so as to improve system learning.

Face recognition and tagging are a relatively mature technology that many people use every day. However, because its capabilities are constantly evolving and the outputs are diverse, systematically sketching, prototyping, and assessing the UX of face tagging remains challenging. It exemplifies level four systems; These are opportune areas for HCI and RtD researchers to study human-AI interaction and design, without getting deeply involved in technological complexities.
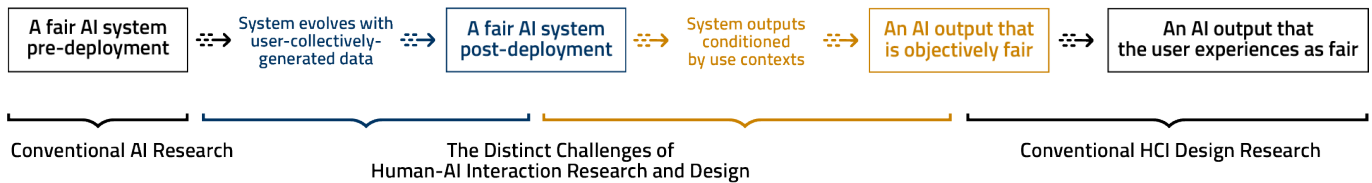
Figure 6: An example of the framework in use. Using the framework, researchers can easily outline the problem space of a human-AI interaction issue of their interest, for example, the issue of AI fairness.

The Anatomy of AI's HCI Issue

For researchers who study specific human-AI interaction design issues (e.g. fairness, intelligibility, users' senses of control, etc.), the proposed framework gives a preliminary structure to these vast issues. Take as an example the challenges surrounding *accounting for AI biases*, a challenge that many critical AI systems face across application domains such as healthcare and criminal justice. Building a "fair" AI application is widely considered as difficult, due to the complexity both in defining fairness goals and in algorithmically achieving the defined goals. Prior research has been addressing these challenges by promoting interaction design guidelines [4, 35].

Our framework provides a more holistic structure to the problem space of "AI fairness" (Figure 6). It illustrates that the current work has mostly focused on building "a fair AI system pre-deployment"; that algorithmic fairness is only part of the whole "AI fairness" problem space. There is a real need for HCI and AI research in collaboratively translating fairness as an optimization problem into a feature of AI the socio-technical system (Figure 6, blue segment), and into a situated, user experience of fairness (yellow segment). The framework suggests a tentative agenda for these important future research topics.

Implications for Design Methods and Tools

Finally, the proposed framework intends to allow for a more principled discussion on how to support human-AI interaction design practice. It can help identify the core challenges AI brings to HCI practice across application domains. It can help researchers to articulate the contribution of their emergent AI design methods/tools/workflows as well as their scope of generalizability. Finally, it can provide new insights into how to address the remaining challenges.

We consider UX prototyping methods of AI as an example.

1. Identifying root challenges.   Current research typically attributes the difficulty of prototyping AI to AI's technical complexity or reliance on big data. However, HCI routinely grapples with complex, resource-intensive technologies using simple prototypes. What makes AI unique? Our framework suggests that the root challenges are that AI's capabilities are adaptive and its outputs can autonomously diverge at a massive scale. Such systems problematize the conventional HCI prototyping methods that treat technology's affordance as bounded and interactions prescriptive. These methods can work when prototyping AI as an optimization system in the lab (level one). They could fail in fully addressing AI's ramifications over time as a real-world, sociotechnical system.

2. Articulating the contributions and limits of emergent design methods/tools/processes.   To make prototyping human-AI interaction easier, researchers have created simple-rule-based simulators [49, 7]) as AI prototyping tools. Mapping the characteristics of rule-based interactions onto the AI design complexity map (Figure 5), it becomes evident that rule-based simulators are most effective in prototyping level 1-2 systems. They can be particularly valuable for systems that generate a broad set of outputs (level 2) where traditional, manual prototyping methods struggle. However, rule-based simulators cannot easily prototype systems that autonomously learn from user-generated data (level 3-4). These are living, sociotechnical systems; the rules that map their inputs to outputs evolve in complex ways over time.

3. Providing new insights for future research.   Framing level 3 and 4 AI systems as living, sociotechnical systems reveal new insights into how we might more effectively prototype their interactions. For example, CSCW research has investigated how to prototype workplace knowledge sharing systems whose affordance co-evolves with its users' behaviors, the interactions among its users, and the organizational contexts at large [31]. These are too living, sociotechnical systems with uncertain capabilities and complex outputs. This body of work, though not typically considered as related to AI, could offer a valuable starting place for considering how we might design prototype human-AI interactions in the wild, over time. In this light, the proposed conceptual framework offers actionable insights for addressing the challenges of prototyping AI methodologically.

CONCLUSION AND ACKNOWLEDGEMENT

AI plays an increasingly important role in HCI and UX design. Today, designers face many challenges in working with AI. Prior research often attributed these challenges to AI's algorithmic complexity and unpredictable system behaviors. Our synthesis of these challenges provided an alternative view to this common assumption. We encourage fellow researchers to critique, evaluate, and improve on this proposed framework based on their respective design and research experiences.

REFERENCES

[1] 2009. wekinator: Software for real-time, interactive machine learning. `http://www.wekinator.org/`. (2009).

[2] 2017. *Designing the User Experience of Machine Learning Systems, Papers from the 2017 AAAI Spring Symposium, Technical Report SS-17-04, Palo Alto, California, USA, March 27-29, 2017*. AAAI. `https://mikek-parc.github.io/AAAI-UX-ML/`

[3] 2018. *The Design of the User Experience for Artificial Intelligence (The UX of AI), Papers from the 2018 AAAI Spring Symposium, Palo Alto, California, USA, March 26-28, 2018*. AAAI. `https://mikek-parc.github.io/AAAI-UX-AI/`

[4] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. ACM. `https://www.microsoft.com/en-us/research/publication/guidelines-for-human-ai-interaction/`

[5] Sara Bly and Elizabeth F Churchill. 1999. Design through matchmaking: technology in search of users. *interactions* 6, 2 (1999), 23–31.

[6] Kirsten Boehner and Carl DiSalvo. 2016. Data, Design and Civics: An Exploratory Study of Civic Tech. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2970–2981. DOI: `http://dx.doi.org/10.1145/2858036.2858326`

[7] Sander Bogers, Joep Frens, Janne van Kollenburg, Eva Deckers, and Caroline Hummels. 2016. Connected Baby Bottle: A Design Case Study Towards a Framework for Data-Enabled Design. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. ACM, New York, NY, USA, 301–311. DOI: `http://dx.doi.org/10.1145/2901790.2901855`

[8] Shan Carter and Michael Nielsen. 2017. Using Artificial Intelligence to Augment Human Intelligence. *Distill* (2017). DOI:`http://dx.doi.org/10.23915/distill.00009` https://distill.pub/2017/aia.

[9] Amber Cartwright. 2016. Invisible Design: Co-Designing with Machines. (2016). `http://airbnb.design/invisible-design/`

[10] Design Council. 2005. The 'double diamond' design process model. *Design Council* (2005).

[11] Justin Cranshaw, Emad Elwany, Todd Newman, Rafal Kocielnik, Bowen Yu, Sandeep Soni, Jaime Teevan, and Andrés Monroy-Hernández. 2017. Calendar. help: Designing a workflow-based scheduling agent with humans in the loop. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2382–2393.

[12] Scott Davidoff, Brian D. Ziebart, John Zimmerman, and Anind K. Dey. 2011. Learning Patterns of Pick-Ups and Drop-Offs to Support Busy Family Coordination. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 1175–1184. DOI: `http://dx.doi.org/10.1145/1978942.1979119`

[13] Graham Dove, Kim Halskov, Jodi Forlizzi, and John Zimmerman. 2017. UX Design Innovation: Challenges for Working with Machine Learning as a Design Material. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. ACM Press, New York, New York, USA, 278–288. DOI: `http://dx.doi.org/10.1145/3025453.3025739`

[14] Melanie Feinberg. 2017. A Design Perspective on Data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 2952–2963. DOI: `http://dx.doi.org/10.1145/3025453.3025837`

[15] Jodi Forlizzi. 2018. Moving Beyond User-centered Design. *Interactions* 25, 5 (Aug. 2018), 22–23. DOI: `http://dx.doi.org/10.1145/3239558`

[16] Michael Freed, Jaime G Carbonell, Geoffrey J Gordon, Jordan Hayes, Brad A Myers, Daniel P Siewiorek, Stephen F Smith, Aaron Steinfeld, and Anthony Tomasic. 2008. RADAR: A Personal Assistant that Learns to Reduce Email Overload.. In *AAAI*. 1287–1293.

[17] William W. Gaver. 1991. Technology Affordances. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '91)*. ACM, New York, NY, USA, 79–84. DOI: `http://dx.doi.org/10.1145/108844.108856`

[18] Marco Gillies, Bongshin Lee, Nicolas D'Alessandro, Joëlle Tilmanne, Todd Kulesza, Baptiste Caramiaux, Rebecca Fiebrink, Atau Tanaka, Jérémie Garcia, Frédéric Bevilacqua, Alexis Heloir, Fabrizio Nunnari, Wendy Mackay, and Saleema Amershi. 2016. Human-Centred Machine Learning. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*. ACM Press, New York, New York, USA, 3558–3565. DOI: `http://dx.doi.org/10.1145/2851581.2856492`

[19] Fabien Girardin and Neal Lathia. 2017. When User Experience Designers Partner with Data Scientists. In *The AAAI Spring Symposium Series Technical Report: Designing the User Experience of Machine Learning Systems*. The AAAI Press, Palo Alto, California. `https://www.aaai.org/ocs/index.php/SSS/SSS17/paper/view/15364`

[20] Mayank Goel, Nils Hammerla, Thomas Ploetz, and Anind K. Dey. 2015. Bridging the Gap: Machine Learning for Ubicomp - Tutorial @UbiComp 2015. `https://openlab.ncl.ac.uk/bridging-the-gap/`. (2015).

[21] Google. 2019. People + AI Guidebook: Designing human-centered AI products. `pair.withgoogle.com/`. (2019).

[22] Patrick Hebron. 2016a. *Machine learning for designers*. O'Reilly Media.

[23] Patrick Hebron. 2016b. New York University Tisch School of the Arts Course: Learning Machines. (2016). `http://www.patrickhebron.com/learning-machines/`

[24] Karey Helms. 2019. Do You Have to Pee?: A Design Space for Intimate and Somatic Data. In *Proceedings of the 2019 on Designing Interactive Systems Conference (DIS '19)*. ACM, New York, NY, USA, 1209–1222. DOI: `http://dx.doi.org/10.1145/3322276.3322290`

[25] Douglas R Hofstadter and others. 1979. *Gödel, Escher, Bach: an eternal golden braid*. Vol. 20. Basic books New York.

[26] Lars Erik Holmquist. 2017. Intelligence on tap: artificial intelligence as a new design material. *interactions* 24, 4 (2017), 28–33.

[27] Andreas Kaplan and Michael Haenlein. 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 (2019), 15–25.

[28] Claire Kayacik, Sherol Chen, Signe Noerly, Jess Holbrook, Adam Roberts, and Douglas Eck. 2019. Identifying the Intersections: User Experience + Research Scientist Collaboration in a Generative Machine Learning Interface. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, New York, NY, USA, Article CS09, 8 pages. DOI: `http://dx.doi.org/10.1145/3290607.3299059`

[29] Scott R. Klemmer, Anoop K. Sinha, Jack Chen, James A. Landay, Nadeem Aboobaker, and Annie Wang. 2000a. Suede: A Wizard of Oz Prototyping Tool for Speech User Interfaces. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology (UIST '00)*. ACM, New York, NY, USA, 1–10. DOI:`http://dx.doi.org/10.1145/354401.354406`

[30] Scott R. Klemmer, Anoop K. Sinha, Jack Chen, James A. Landay, Nadeem Aboobaker, and Annie Wang. 2000b. Suede: A Wizard of Oz Prototyping Tool for Speech User Interfaces. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology (UIST '00)*. ACM, New York, NY, USA, 1–10. DOI:`http://dx.doi.org/10.1145/354401.354406`

[31] Esko Kurvinen, Ilpo Koskinen, and Katja Battarbee. 2008. Prototyping social interaction. *Design Issues* 24, 3 (2008), 46–57.

[32] Shane Legg, Marcus Hutter, and others. 2007. A collection of definitions of intelligence. (2007).

[33] Brian Y Lim, Anind K Dey, and Daniel Avrahami. 2009. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2119–2128.

[34] Panagiotis Louridas. 1999. Design as bricolage: anthropology meets design thinking. *Design Studies* 20, 6 (1999), 517–535.

[35] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 220–229. DOI: `http://dx.doi.org/10.1145/3287560.3287596`

[36] GINA NEFF and PETER NAGY. 2016. Talking to Bots: Symbiotic Agency and the Case of Tay. *International Journal of Communication (19328036)* 10 (2016).

[37] William Odom and Tijs Duel. 2018. On the Design of OLO Radio: Investigating Metadata As a Design Material. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 104, 9 pages. DOI: `http://dx.doi.org/10.1145/3173574.3173678`

[38] Kayur Patel, James Fogarty, James A. Landay, and Beverly Harrison. 2008. Examining difficulties software developers encounter in the adoption of statistical machine learning. In *23rd AAAI Conference on Artificial Intelligence and the 20th Innovative Applications of Artificial Intelligence Conference*. Chicago, IL, United States, 1563–1566.

[39] Kayur Dushyant Patel. 2012. *Lowering the Barrier to Applying Machine Learning*. Ph.D. Dissertation. University of Washington.

[40] Laurel D. Riek. 2012. Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines. *J. Hum.-Robot Interact.* 1, 1 (July 2012), 119–136. DOI: `http://dx.doi.org/10.5898/JHRI.1.1.Riek`

[41] Eric Ries. 2011. *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. Crown Books.

[42] Antonio Rizzo, Francesco Montefoschi, Maurizio Caporali, Antonio Gisondi, Giovanni Burresi, and Roberto Giorgi. 2017. Rapid Prototyping IoT Solutions Based on Machine Learning. In *Proceedings of the European Conference on Cognitive Ergonomics 2017 (ECCE 2017)*. ACM, New York, NY, USA, 184–187. DOI:`http://dx.doi.org/10.1145/3121283.3121291`

[43] Albrecht Schmidt. 2000. Implicit human computer interaction through context. *Personal technologies* 4, 2-3 (2000), 191–199.

[44] Lisa Stifelman, Adam Elman, and Anne Sullivan. 2013. Designing Natural Speech Interactions for the Living Room. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. ACM, New York, NY, USA, 1215–1220. DOI: `http://dx.doi.org/10.1145/2468356.2468574`

[45] Maria Stone, Frank Bentley, Brooke White, and Mike Shebanek. 2016a. Embedding User Understanding in the Corporate Culture: UX Research and Accessibility at Yahoo. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 823–832.

[46] Peter Stone, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyanakrishnan, Ece Kamar, Sarit Kraus, and others. 2016b. Artificial intelligence and life in 2030. *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel* (2016).

[47] Jennifer Sukis. 2019. AI Design & Practices Guidelines (A Review). `https://medium.com/design-ibm/ai-design-guidelines-e06f7e92d864`. (2019).

[48] Mary Treseler. 2017. *Designing with Data: Improving the User Experience with A/B Testing*. O'Reilly Media, Chapter Designers as data scientists. `http://radar.oreilly.com/2015/05/designers-as-data-scientists.html`

[49] Philip van Allen. 2018. Prototyping Ways of Prototyping AI. *Interactions* 25, 6 (Oct. 2018), 46–51. `DOI: http://dx.doi.org/10.1145/3274566`

[50] Wikipedia contributors. 2019. Artificial intelligence — Wikipedia, The Free Encyclopedia. `https://en.wikipedia.org/wiki/Artificial_intelligence`. (2019).

[51] Qian Yang, Nikola Banovic, and John Zimmerman. 2018. Mapping Machine Learning Advances from HCI Research to Reveal Starting Places for Design Research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '18 (CHI '18)*. ACM.

[52] Qian Yang, Justin Cranshaw, Saleema Amershi, Shamsi T. Iqbal, and Jaime Teevan. 2019. Sketching NLP: A Case Study of Exploring the Right Things To Design with Language Intelligence. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 185, 12 pages. `DOI: http://dx.doi.org/10.1145/3290605.3300415`

[53] Qian Yang, Alex Scuito, John Zimmerman, Jodi Forlizzi, and Aaron Steinfeld. 2018a. Investigating How Experienced UX Designers Effectively Work with Machine Learning. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA, 585–596. `DOI: http://dx.doi.org/10.1145/3196709.3196730`

[54] Qian Yang, Alex Scuito, John Zimmerman, Jodi Forlizzi, and Aaron Steinfeld. 2018b. Investigating How Experienced UX Designers Effectively Work with Machine Learning. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA, 585–596. `DOI: http://dx.doi.org/10.1145/3196709.3196730`

[55] Qian Yang, Aaron Steinfeld, and John Zimmerman. 2019. Unremarkable AI: Fitting Intelligent Decision Support into Critical, Clinical Decision-Making Processes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 238, 11 pages. `DOI: http://dx.doi.org/10.1145/3290605.3300468`

[56] Qian Yang, Jina Suh, Nan-Chen Chen, and Gonzalo Ramos. 2018. Grounding Interactive Machine Learning Tool Design in How Non-Experts Actually Build Models. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA, 573–584. `DOI: http://dx.doi.org/10.1145/3196709.3196729`

[57] Qian Yang, John Zimmerman, Aaron Steinfeld, Lisa Carey, and James F Antaki. 2016b. Investigating the Heart Pump Implant Decision Process: Opportunities for Decision Support Tools to Help. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 4477–4488.

[58] Qian Yang, John Zimmerman, Aaron Steinfeld, and Anthony Tomasic. 2016a. Planning Adaptive Mobile Experiences When Wireframing. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems - DIS '16*. ACM Press, Brisbane, QLD, Australia, 565–576. `DOI: http://dx.doi.org/10.1145/2901790.2901858`

[59] John Zimmerman, Anthony Tomasic, Charles Garrod, Daisy Yoo, Chaya Hiruncharoenvate, Rafae Aziz, Nikhil Ravi Thiruvengadam, Yun Huang, and Aaron Steinfeld. 2011. Field Trial of Tiramisu: Crowd-Sourcing Bus Arrival Times to Spur Co-Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 1677–1686. `DOI: http://dx.doi.org/10.1145/1978942.1979187`

[60] John Zimmerman, Anthony Tomasic, Isaac Simmons, Ian Hargraves, Ken Mohnkern, Jason Cornwell, and Robert Martin McGuire. 2007. Vio: A Mixed-Initiative Approach to Learning and Automating Procedural Update Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. Association for Computing Machinery, New York, NY, USA, 1445–1454. `DOI: http://dx.doi.org/10.1145/1240624.1240843`

[61] Lamia Zouhaier, Yousra Bendaly Hlaoui, and Leila Jemni Ben Ayed. 2013. Building adaptive accessible context-aware for user interface tailored to disable users. In *2013 IEEE 37th Annual Computer Software and Applications Conference Workshops*. IEEE, 157–162.