Investigating How Experienced UX Designers Effectively Work with Machine Learning

Qian Yang[1] **Alex Scuito**[1] **John Zimmerman**[1] **Jodi Forlizzi** [1] **Aaron Steinfeld**[2]
HCI Institute[1] Robotics Institute[2]
Carnegie Mellon University, Pittsburgh PA, USA
{yangqian, steinfeld}@cmu.edu {scuitoalex, johnz, forlizzi}@cs.cmu.edu

ABSTRACT
Machine learning (ML) plays an increasingly important role in improving a user's experience. However, most UX practitioners face challenges in understanding ML's capabilities or envisioning what it might be. We interviewed 13 designers who had many years of experience designing the UX of ML-enhanced products and services. We probed them to characterize their practices. They shared they do not view themselves as ML experts, nor do they think learning more about ML would make them better designers. Instead, our participants appeared to be the most successful when they engaged in ongoing collaboration with data scientists to help envision what to make and when they embraced a data-centric culture. We discuss the implications of these findings in terms of UX education and as opportunities for additional design research in support of UX designers working with ML.

Author Keywords
User Experience Design; UX Practice; Machine Learning; Design Material; Interaction Design.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
Machine learning (ML) plays an increasingly important role in improving a user's experience. From mundane spam filters to personalized newsfeeds to conversational agents like Alexa to the promise of driverless cars, many products and services now improve user experience (UX) with algorithms that learn from an underlying data source. This growing reliance on ML somewhat implies that UX designers have become quite skilled at envisioning new products and services that leverage ML's capabilities. Interestingly, recent research indicates that many UX designers are unprepared to effectively leverage ML capabilities [9, 32]. For example, a recent survey showed that many UX designers struggle to understand the capabilities and limitations of ML. Also, they typically joined projects towards the end, after the functional decisions had been made. "*Design teams are simply putting lipstick on the pig*" [9]. Other work showed designers often fail to notice obvious places where ML could improve UX [32].

Recently, design researchers and educators began taking actions to address this problem. A few have developed designer focused education materials, meant to teach the technical concepts of ML [13, 14, 15]. This work implies that designers should understand the mechanics of algorithms in order to work effectively with ML. Other researchers created design patterns to help practitioners recognize common situations where ML can improve UX [32]. Others organized workshops, bringing together groups of artists, designers, and technologists to collectively explore how ML might function as a creative material [11, 16].

The work to make ML more accessible to designers has led some to discuss *ML as a design material* [31]. Our research adds to this growing area of inquiry. Instead of investigating problems designers face when working with ML or working on tools meant to make ML more accessible to designers, we chose to investigate the design practices of some of the few UX designers who regularly create new products and services that use ML to enhance UX. We hoped that their approach and reflections would reveal new insights around UX design education and insights on the kinds of tools needed for enhancing UX with ML.

We interviewed 13 designers who all had at least four years of experience designing ML enhanced UX. The interviews produced several interesting findings: 1) Designers shared that they knew very little about how ML works, and this was not a priority for them. They instead used designerly abstractions and popular exemplars to explain what ML is and to communicate design ideas with each other. 2) ML projects are longer in preparation and scope than other design projects. During the preparation stage, designers evolved their ideas in close collaboration with data scientists; They did not deliver fully formed designs to a technical team. 3) Designers "play" with quantitative data during all phases of a design project. Our findings suggest an alternative to the common assumption that teaching designers how ML works as the most effective way of helping them engage with it

DIS '18, June 9–13, 2018, , Hong Kong

ISBN 978-1-4503-5198-0/18/06…$15.00
DOI: 10.1145/3196709.3196730

as a design material. Instead, design researchers hoping to aid practitioners might focus on providing abstractions, exemplars, and new tools and methods that support designers collaborating with data scientists.

This paper makes two contributions. First, it presents a rare description of how experienced designers effectively worked with ML. It offers a point of reference for UX education; equipping novice designers with knowledge and skills needed for the best practice in industry. Second, this work illuminates many open research questions in understanding and supporting UX design of ML. It suggests an agenda for important future research topics.

RELATED WORK

Making Machine Learning Available to Designers

Design researchers investigating ML noted 1) designers overlook simple opportunities to utilize ML and that 2) design-led innovation of ML is rare [32]. In a recent survey, practitioners stated they could not easily envision novel and interesting uses of ML. While they understood how ML worked broadly, they did not understand what was needed to design with it [9]. Based on observations of these challenges, design researchers and practitioners have begun making efforts to help UX designers work with ML as a design material. These efforts advanced in a variety of directions, but they typically fall into two broad characterizations.

One characterization presumes that designers cannot work with a material they do not fully understand, redirecting designers' focus away from the interactive artifact towards the statistical processes of ML products [6, 27]. Several books, courses and online collections of materials help designers learn about ML (i.e. [5, 13, 14, 15]). The book *Machine Learning for Designers*, for example, reviews concepts of supervised and unsupervised learning, as well as common analogies of how ML works [14]. 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 materials they use to design.

A second characterization, less popular than the first, works to demonstrate divergent ways of designing ML, sensitizing designers to its rich design possibilities. Yang et al. developed a set of design patterns and a boundary object to help designers better communicate their designs to data scientists [32]. Other researchers have explored various themes related to the interaction design challenges ML imposes, such as intelligibility [18], animistic design [19], and design for the peace of mind [12]. An university design course asked students to create a "useless artificial intelligence" product, encouraging students to engage with ML through a making process that probes affordances, capabilities, and limitations of this technology [29]. Instead of teaching designers what ML is, this approach challenges designers to envision applications technologists would likely never imagine.

Design Practice and Collaboration with Data Scientists

Designing a partially known material for uncertain situations is a common task for any interaction designer [24, 17]. To envision things that have never before existed, designers engage in reflective conversations with design materials [10, 23]. Understanding how designers engage in this conversation and how the materials "talk back to the designer" is an important component in understanding design practice [23].

Research about design practice has shown that designers can handle complex design materials by using simple methods, such as sketching, brainstorming, body-storming, and diagramming with sticky notes [24]. HCI research on UX practitioners discussed software as a design material, noting that designers often need to use developers as a proxy in conversations around the possibilities of software design [21].

With the increasing use of ML to enhance UX, there has been speculation that data scientists and ML will decrease the need for UX design [5, 8]. Girardin and Lathia's investigation of the designer-data scientists collaboration within their own organization seems to confirm this speculation [12]. Additionally, some HCI researchers observed that when developers use statistics to make design decisions, the role of designers can diminish [2]. As AI becoming the new UX [4, 32], some even claimed that "data scientists are the next UX designers" [8]. In spite of this speculation, we found no work investigating how design practitioners collaborate with data scientists in industry.

We see a need for more studies UX practice as it relates to ML. The HCI and IxD communities have learned over time that the effectiveness of design research depends on its fit with the nature of design practice [1, 24]. In the midst of diverse orientations and approaches to moving ML to practice, we need to understand how UX practitioners actually work with ML. We need to understand not only the challenges, but also the successful attempts in designing with ML that have produced the many ML products people now use every day. This paper takes a step towards addressing this need.

METHOD

We wanted to understand the current practices used by UX designers who regularly design ML products and services. Previous work focused on the struggles that less experienced designers often face. We were interested in how designers with lots of experience designing with ML work compared to traditional UX design. We hoped to identify opportunities for new design research that could make it easier for all UX designers to work more effectively with ML as a design material.

We conducted retrospective interviews with UX design practitioners who have played an active role in creating widely adopted ML-driven products. We recruited participants by leveraging our network of alumni who work in practice; mostly former master's students now working in industry. We then used snowball sampling, asking them to recommend other designers with experience working with ML, to grow our set of participants.

Education	Professional Role	Example Design Project	Org Size	Exp.
HCI	Design Manager	Intelligent features in multiple messaging or conversation apps	>10,000	10+ yrs
Design	Design Manager	Intelligent features in a media consumption app to better match content providers and users	>10,000	10+ yrs
Design	UX Designer	Decision support app for physicians	1,000 - 10,000	10+ yrs
Design	Designer-turned PM	Language learning app; patient coaching app.	100 - 1,000	10+ yrs
Design and HCI	Designer	Intelligent tutoring feature for education apps	<100	10+ yrs
CS and HCI	Designer-turned entrepreneur	New messaging and efficiency apps	<100	10+ yrs
Design and HCI	Designer	Recommender in a shared service app	>10,000	7-10 yrs
CogPsy	UX researcher	Intelligent reminder in a social media app	>10,000	7-10 yrs
EE and Design	UX designer	New wearable health product	>10,000	5-7 yrs
HCI	Designer	Recommender in a text processing software	>10,000	
Design	Service designer	New health coaching app	>10,000	5-7 yrs
Psychology	UX researcher	Intelligent feature in a social media app	>10,000	4 yrs
HCI	UX researcher	Intelligent content recommendation for a media consumption product	>10,000	4 yrs

Table 1. Interview Participants. We interviewed UX designers who had more than four years of experience in designing ML-enhanced products. Many had more than 10 years of related experience.

We chose retrospective interviews because observations of current practice seemed impractical. First, actual cases unfold very slowly, over many months. Second, the practitioners we encountered all worked on new projects they could not discuss for reasons of industrial competition.

We interviewed 13 participants. All had designed products and services that enhance UX with ML for more than 4 years. Most had worked on these types of products and services for more than 10 years. Nine participants designed products used by more than one billion users. Two worked on successful special-purpose ML applications: a clinical decision support systems and a wearable health coaching system. The others worked to extend existing products with new ML features. Table 1 provides a summary of the participants' background and the type of projects they described in our study.

We asked all participants to complete a pre-interview survey where they described their education and professional background, as well as their familiarity with concepts from UX design, statistics, machine learning, and data storage. We then interviewed participants, asking them to walk us through a recent design case where they used ML to improve the UX. Throughout the interviews, we probed them to get details on their process and to surface the triggers that drove specific decisions for what to do and how to work. We paid particular attention to two previously expressed challenges from the literature [9]. First, UX designers often struggle to understand the capabilities of ML, so we probed on what they did to understand these. Second, designers working on ML projects are often only included at the end of a project, so we wanted to probe on when they joined projects or if they ever initiated projects on their own.

To better understand how they perceived ML, we asked them to draw an illustration of how the systems they designed worked. We wanted to see how they represented things like data collecting, labeling, and training. At the end of the interview, we asked participants to reflect on what they viewed as the major differences between how they design when working with ML and when working on products and services that don't use ML. We asked them to share, "*the things you wish you had known about before you started your career in designing ML systems*".

We recorded and transcribed the interviews. We then reviewed the transcripts, pulled out important insights, and used affinity diagrams to synthesize across the interviews in order to identify thematic patterns. We created and consolidated process models detailing how the different projects unfolded and how the designers collaborated with data scientists. We chose affinity diagrams and models instead of grounded theory for several reasons. These methods are typically used in HCI and interaction design practice. Our intention was not to build up a detailed theory of current design practices performed by only a tiny number of designers. Instead, we chose to follow a more practice-based approach because our focus was on generating ideas that could help practitioners other than the practitioners we were studying. We wanted to inform the design of new resources and tools for less experienced practitioners by better understanding what these few experienced practitioners do.

FINDINGS

We organized our findings around three themes: work participants did to understand ML capabilities, changes to the

design process that seem directly related to working with ML, and participants' acknowledgement of a data culture. These themes are not exclusive, and some aspects spill across two or all three.

We noticed two distinct work cultures that had a large impact on how participants worked. A few participants worked in large, AI-focused organizations, where designers had access to many data scientists and the companies actively promoted innovation via ML. The others worked in environments with much more limited access to data scientists. In the detailed findings below, we note where this difference impacted the work they described.

Understanding Machine Learning Capabilities

Participants almost all stated that they "*know very little*" about how ML works. They characterized their ML literacy as "*understanding at a very high level... [at the level of] knowing what a classifier is and what a label is.*" Only one participant (P6) had taken any ML course; an online course taken well after graduation. P6 was the only participant with a CS degree, and had transitioned from CS to UX while earning an MHCI degree. All other participants claimed they gained familiarity with ML mostly by reading online articles found on Medium and Reddit.

Interestingly, participants did not feel their lack of technical knowledge hindered their ability to design or to collaborate with data scientists. They shared that they worked on the design issues, not the technical issues, and that working with ML required "*... more design savvy*". Several participants claimed designing the user interaction was "*the actual challenge*."

> *I don't think of ML as affecting my interaction designs or not. I think about it more like impacting certain algorithms that are inputs into users' experience.* (P14)

This issue often came up when they made illustrations of their ML systems. Many drew a line in the middle, creating a distinct break between the users and the technology. P10's illustration, for example, had a dense collection of design factors on the left: usage scenario, user's frustration, and "*a very simple thing*" the design team thought ML could do to aid the user. They then drew a line followed by several boxes representing the back-end mechanisms. The ML was one of these boxes, and they noted it was really not so simple.

Most claimed that they learned to work with ML similarly to how they learned to work with other interactive technologies. "*ML is just like JavaScript*", several participants claimed. They did not seem to view themselves as technology experts, but instead as UX design experts who had great comfort working with a variety of technologies, and ML was one of these.

Designerly Abstractions and Examplars

We found almost nothing in common with the way designers spoke about ML and the way education materials meant to teach ML for engineers. Participants rarely spoke of ML in technical terms. For example, they never talked about supervised or unsupervised learning (common starting point for teaching ML). Instead, they appeared to think with, and work with, abstractions; simple insights about an ML capability had had implicitly linked with generating value for users. These were much more abstract than design patterns. More similar to design patterns, they often used exemplars to communicate these abstractions.

> Q: What does machine learning do?
> P7: *Some try to recognize intent, a bit like auto-correct. Some are intent prediction like Clippy. Anyone who is working on assistive technology, is working on some class of that problem.*

In the excerpt above, P7 describes ML using abstractions of its capabilities: recognize intents and predict intents. He then ground these capabilities through the use of exemplars: auto-correct and Clippy. We found this manner of describing ML across most participants.

Participants most often described the capabilities of ML as it related to the user's utility. Their abstractions narrowly oriented towards both the users and scenarios related to their designs. They never spoke of general taxonomies of ML functionality or specific algorithms. Some examples:

> *We use machine learning so that we can build something that can personalize for a lot of people.* (P3)
>
> *In consumer tech, we try to raise the level of abstraction [of user commands] rather than doing everything manually.* (P7)
>
> *ML gets users directly into the task they really need to do.* (P1)
>
> *We are doing a recommendation system of sorts. As a product designer [not a technologist], I think about that as how can we show an evolving relationship between a user and our service... [I want users] foreseeing our relationship improve, where the relationship is the recommendations we are giving them.* (P14)

The abstractions almost always appeared with exemplars. The abstractions served as a general insight about an ML capability and provided an understanding of how it worked. The design exemplars provided specific interaction possibilities and a glimpse of a possible felt experience. In our interviews, participants frequently referenced widely-known exemplars including Clippy, autocorrect, email spam filtering, and Tay the Twitter bot. They used these to help describe the capabilities of ML; the form, function, and user experience; and the potential breakdowns that might occur.

Participants actively grew their working set of abstractions and exemplars. They looked for these in both online articles and in competitors' products. One participant recalled that she carried around many conversational virtual assistants so that she could learn the feedback different products gave in response to the same instruction, in different contexts.

The number and variety of exemplars participants used varied wildly, and those with the largest working sets seemed to be the most successful and comfortable at using ML to enhance UX. Participants working at AI-focused organizations had

a significant advantage in building their working set of exemplars and abstractions. These organizations had data scientists frequently giving demos as one way to sensitize their design teams to emerging ML capabilities. "*So many people demo for me. I don't even know whom to call if I have an idea and want to consult a data scientist.*" Participants at smaller or less AI-focused organizations had much more limited access to data scientists. They appeared to consider and propose fewer design alternatives. They were also more likely to use only the most familiar interaction forms, such as recommender or reminder, when describing their process.

Design Process and Collaboration

Participants shared that working with ML took much longer than when designing other UX products and services. We wanted to synthesize the actual time span ML projects required. However, we found that none of our participants had worked on an ML product from its initial ideation until its final release. We could see that some ML projects had lasted for more than four years.

> *ML has a different time-frame for design iteration. Longer initial development, but then ongoing iteration. It felt like building a feature versus building a framework. When you ship it, it's not the end of it. They cycle, and data drives the next step.* (P3)

A consolidation of participants' design process narratives revealed an ML design process. It starts with a long preparatory stage (stage 1), during which designers and data scientists identify a design goal that is both technically viable and that appears to have a high likelihood of improving the user's experience. Once a team had settled on the design goal, scientists and developers developed the system and designers crafted its interaction design (stage 2). In stage 3, they together invested in frequent iterative releases and assessments in order to improve user adoption.

All participants described stage 2 in their ML projects. Stage 2 is a central stage that all projects that participants described went through. Stage 1 happened only at the large, AI-focused organizations. Few projects had made it to stage 3. Below we describe three distinct design stages.

Stage 1: Concept Development

New ML design seemed to only happen within the AI-focused organizations. It started with a long, preparatory stage involving two activities. First. participants collected log data form current services their organization provided. Participants examined the log with sensitivity to specific patterns. They often seemed to search for patterns they thought might be there more than they would data mine to discover unexpected patterns. They imagined what user behavior might be worth learning and what learned interactions might be valuable for users. One designer described this process as "*just me being really excited about the product*" (P14).

Second, participants would share these inchoate ideas with data scientists. This happened informally, outside of any specific development projects, and they had iterative interactions, where they might return to the same data scientists with new or refined ideas. They noted that the data scientists had a very different view of scenario-based design. They used scenarios to validate what an idea might be instead of using scenarios to generate and refine new ideas. "*Data scientists use scenarios to validate designers' assumptions about how the product should work, [and then] toss back ideas*".

The conversation between designers and data scientists focused on identifying a design goal worth pursuing. Often this would lead the discussion away from working with ML. The discussions focused on coming up with a "good enough" idea, and it did not address either its exact technical feasibility or experiential quality. Participants particularly pushed the data scientists to understand what might be technically possible.

> *(I) framed the questions not as do you know what would work, but in your gut, do you think this would be possible. Possible on a scale of 1-10.* (P8)
>
> *The level of detail I'll need to discuss with [the data scientists] is understanding the capability of what could be possible. I didn't need to get into specific detail about it.* (P3)
>
> *Those [design ideas] are a series of aspirations. Rather than saying what data do we have [...] Could we challenge the data team to figure out how to get close to that?* (P9)

The collaboration at this stage focused on co-evolving a shared vision between the two areas of expertise, an ideal user experience that was worth pursing and ML could potentially help achieve. This shared vision often took form of a unique abstraction of ML capabilities that emerged out of both UX and ML. Instead of saying "machine learning", designers and data scientists might select an expression better suited to the context of their product. For example, some used "affinity" to describe the match between a user and a piece of content. Others used "personalization" to describe the intended, evolution of the relationship between the product and the user. Participants characterized these discussions as a chance for both sides to learn from the others' expertise.

> *There is no such a framework or something, but I think later there is a kind of an acknowledgment when we talk about "personalization"... An acknowledgment that a more personalized experience is a better experience that one is less likely to walk away from.* (P14)
>
> *I gained a better understanding of the capabilities of the algorithms. The data scientist gained a better understanding of what was worth pursuing.* (P3)

Stage 2: Interaction Design and Assessment

Stage 1 produced preliminary understanding of the data, and it established a shared vision between the designers and the data scientists that met the company's goals. Stage 2 focused on refining this vision. In many organizations, this was the stage when UX designers were invited to join a project. During stage 2, participants described developing "*a funnel of visions, a funnel of what exists and what is possible in the company*" (P8). This would advance towards defining

a single, valuable ML feature that was both experientially valuable and technically feasible.

> *It's definitely an ongoing process. We had a lot of basic things to get working. [...] Once we get that improved, then we will probably be adding more ML stack onto the system.* (P14)

> *Design is always about a new product and a complete, larger vision. But that doesn't make it easy to build. We chunk. We talk about stages of development.*" (P8)

Once a product idea was clearly formulated and agreed upon, participants shared that they would next put the required resources in place. Participants would start to design the interactions in parallel with the technical development. Participants explored many possibilities. They also spoke about assessing their interaction designs based on following criteria:

- *Could users produce effective labels needed to train the ML system?*
- *Could users make sense of the ML inference [or adaptation] and did they view it as valuable?*
- *Could users easily recover from ML [inference] errors?*

Participants used a combination of traditional UX methods such as user studies, sketching, and usability testing. They also engaged in continued negotiations with technologists. Collectively, these activities allowed them to craft the user experience. This iterative process continued until the interaction design had matured to a point it could be handed over to a front-end development team.

Today, most new features for online products and services only see broad release after passing a series of A/B tests, and the same process was used for ML features. These tests are meant to reduce the risk of user abandonment and/or reduction in new user conversion rates. Almost all participants spoke of A/B testing as a critical part of improving their design, and this related to both improving the algorithms' performance and improving the interaction. In most cases, the participants shared that they offered users the option to turn a new ML feature off. This appeared to help them view passing of the A/B tests as less challenging.

Stage 3 - Release and Continued Refinement

Few of the projects described by participants made it to stage 3. One notable exception was P7's worked on a message classifier. At the beginning of stage 3, they worked to improve the poor adoption rate. This became "*a really key moment in thinking about design of systems like this*". Stage 3 had to do with refinement that could make a project successful.

> *We had a classifier that predicts whether or not an incoming message is important . . . and the performance is scary good. . . . However, a vast majority of users didn't turn the feature on or soon opted out. When you dug into why that was, people would say "I don't even know what's important to me, how am I gonna trust the ML system to know it's important to me?"*

> *The core insight my research team had was that people are trying to figure out whether they are gonna trust that system, and the way they figure it out to think of it as a person. If I hand a stack of messages to a person, someone I don't know, is it possible that person could figure out with some accuracy if that message is important to me. Most people's take was "no, I barely know it myself".*

Stage 3 sent the UX designers into the field to investigate how people used their system in order to improve the design. The team improved the design by moving away from a binary classification (important or not important) to a small set of categories. The new design made it clear why something might not be important by classifying it as "promotional materials," something the team felt any human would be able to do. The new design used the same ML technology and proved to be a huge success.

The few participants who had projects reach stage 3 stated that such a re-design was inevitable because some mental model issues are difficult to catch before products launch. They felt they "*have to do this level of work on the design side*" (P2,7,13) so that users cannot only recover from the errors, but can understand the errors, and at the same time preserve some level of trust in the system.

> P1 describing why design problems for a chatbot auto-reply feature were unforeseeable: *We help you [users] say that you already gonna say. We do it a bit faster. We are actually influencing what you are saying, but not predicting what you are saying.*

Embracing a Data-Driven Culture

We probed participants on what they needed to be effective and what UX students should know in order to effectively envision ML innovations. Their responses collectively revealed an acknowledgement that they all worked in data centric environments, and that designers needed to embrace this data centricity in order to have impact. They spoke of the importance of learning to speak the language of quantitative data and data science (e.g., telemetry, analytics, A/B testing, covariance, correlation). In the pre-interview surveys, participants responded that they worked with these concepts constantly (Figure 1). "*This is how engineers measure and businesses do things. This can influence how you design.*"

To help embrace a more data-centric culture, participants shared that they engaged in new design activities. They used a combination of qualitative and quantitative methods to advocate for their design. Specifically, they developed new skills around collecting telemetry data (data remotely collected from current products and services), and they generated data visualizations as one way of making sense of their data. In addition to working with data, participants shared that their collaborating data scientists share their role as a user advocate and as facilitator of incorporating domain experts' insights; something that did not happen with other, technical development efforts.

Designing Telemetry Data

Participants frequently asked ML specialists about how people use the product, how they feel about it, and what

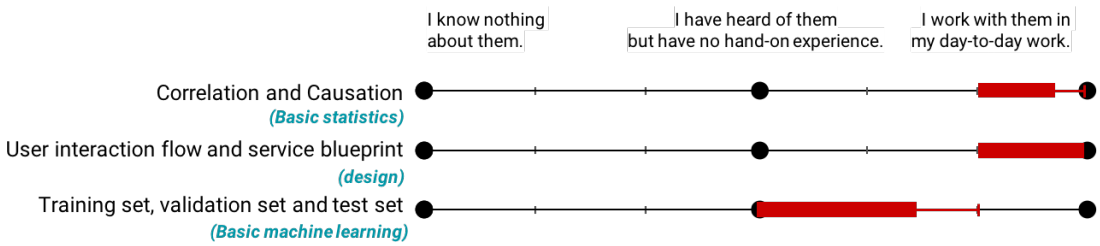

Figure 1. Participants' familiarity with concepts from statistics, UX design and machine learning.

problems they had that could be rectified by design. To help answer these questions, participants shared that they designed telemetry systems so their products can better track users' interactions. They used the data to capture meaningful and accurate snapshots of user behaviors.

Designers began to define their design goals in relation to the user behavioral data collected telemetrically, in the form of a "*matrix for a good experience*" (P2,8,12). This matrix would later be used to measure the success of their designs through a suite of A/B tests. Designers use these matrices to "*influence the engineers to think a different way, not in terms of single A/B test, but a suite of tests*." The telemetry data thus became a space of exchange between designers and data scientists. The data helped to expand the narrow scope of A/B tests and to more holistically address user experience over a larger course of interactions.

Designing Data Visualizations

Participants worked to interpret user behavior data. Most utilized customized dashboards that translated mundane log data snapshots into user stories and insights. Several taught themselves to create data visualization tools, creating visualizations that used by data scientists, engineers, and fellow designers in their organizations. They designed dashboards and visualizations to combine an immersive and an analytical way of understanding, so that the quantitative analysis of user behavior "*do not privilege data scientists*".

> *Data scientists have their methods, and I have my bag of tricks. They have kernels and clusters, and we are good at telling rich, compelling user stories. How can we look at hundreds and thousands of attempts, and also reach out to inquire more about what happened, what the breakdown was?* (P14)

A few designers further utilized these data visualizations as vehicles for conducting user studies. One participant recalled that the first thing he did when joining the project was to list "*a series of questions about user behavior that data analytics can answer*" (P8). They then bookmarked the corresponding data matrix on the telemetry analytics tool the team used. Every morning they checked the new incoming log data against this matrix. When they noticed an intriguing user behavior, they emailed the user to inquire the back story, details, and sometimes set up follow-up interviews. They labeled this method as "*qualitative study in a quantitative scale*."

Advocating for Users Along with Domain Experts

Besides data scientists, designers also worked closely with domain experts during all phases of a ML project, especially projects that design special-purpose applications. For example, P12 designed a reminder feature of a social network product in collaboration with psychologists who specialized in human perception and memory; P15 designed an intelligent tutoring system on biology in collaboration with middle school teachers and biologists.

While cross-disciplinary collaboration is not new nor unique to ML projects, participants suggested that when working with ML, data scientists and domain experts overlapped with the responsibility in their work to make sense of and predict user behavior. They were no longer the sole user advocates. Most participants repeatedly stressed the unique skills designers brought to the table in relation to other human-centered ML roles; some designers also adapted new methods to manage and communicate their designs:

> *I embedded an information-motivation-behavior framework (in user data analysis). It showed everyone that just (user) information is not enough.* (P8)

> *Designers have domain knowledge about how people typically respond to certain things, how easy is it to change a certain behavior. Designers helping with shaping the experience, with the missing part, the things you have not thought about.* (P9)

DISCUSSION

ML plays an increasingly important role in how new products and services deliver an improved user experience. Today, UX designers face challenges in understanding ML capabilities, in envisioning new products and services, and in collaborating effectively with data scientists. Our study of UX designers who have had years of experience designing ML enhanced products and services adds to this growing area of research. We found that these UX designers do not view themselves as ML experts, nor do they think that learning more about ML would make them better ML designers. Instead, participants appeared to be the most successful when they engaged in ongoing collaboration with data scientists to help envision what to make and when they embraced a data-centric culture and became proactive at using data in their design practice.

Below, we reflect on how these findings point to needs and opportunities related to the education of new UX designers

who wish to enter a field where they can innovate with the possibilities of ML. We also reflect on opportunities for additional design research that can better inform the HCI community about the current disconnect between UX practice and innovation via ML and that can create new resources and tools that allow all UX designers to more effectively engage with ML as a design material.

Preparing UX Design Students

Like others in the HCI community [4, 9, 32], we promote the idea of ML as a design material. We hope to enable designers to leverage ML in their designs, and to conceive of entirely new ways to use ML. The practice of experienced designers provides a point of reference for this effort. We see two main ways HCI and UX educators can better prepare students to more successfully enter industry: training them to work and think with data and providing opportunities for them to collaborate with data scientists.

Designing in a Data-Driven Culture

Many education efforts aiming to improve designers' ability to work with ML has focused on teaching statistics and ML algorithms (i.e. [5, 14, 32]). Our findings suggest an additional skillset — basic interpretation and manipulation of quantitative data. Use of telemetry data and A/B tests to assess and measure user behavior is not new or unique to ML. However, we observed that ML design teams used these skills with great frequency.

Basic data literacy and the ability to draw quantitative user insights from telemetry data would benefit designers in the increasingly data-driven cooperate culture. Student designers should learn to speak the language of data, analytics, and A/B tests as an extension of their user-centered training. Existing cross-disciplinary courses such as "Data Storytelling" and "Data Visualization" could be leveraged in support of this. Learning these skill has the added benefit of help designers work on non-ML projects that also rely on analytics for insights.

Additional experience in integrating data analytics and qualitative user study methods is also likely to benefit student designers. UX education can provide a disciplined discussion on the advantages and disadvantages of each approach, and how they might best be used together. In our study, all participants used both qualitative and quantitative methods. They invented several design techniques toward this goal, for example, using telemetry data system as a way of understanding user behavior quantitatively, recruiting user study participants based on log data, and identifying design goals through a matrix of telemetry data so as to evaluate their designs through a suite of A/B tests. These methods could serve a starting point for UX courses and programs that aim to cultivate UX designers who can work with both quantitative and qualitative methods.

Collaboration with Data Scientists

Effective collaboration with data scientists was critical for all of our participants. In working with ML, participants started by establishing a shared understanding with collaborating data scientists in order to discover a goal or innovation to pursue. Participants then experimented with various design ideas, pushing the boundaries of technical constraint. Data scientists tested design ideas and fed technical limitations back to designers. Students would benefit from having some exposure to this new kind of ideation and iterative design process. This pattern of collaboration enabled our participants to play a central role in shaping both the function and form of ML applications. Our participants used data scientists as a proxy to have a conversation with the material of ML. New designs arose as a part of their design intention, rather than from the availability of data or algorithm.

This emerging form of collaboration should not yet be considered a best practice, but it was the best practice we observed. Student designers entering industry now and in the near future will need to work with data scientists in this unstructured manner, with very few boundary objects to help scaffold their collaborations. We recommend that UX and HCI educators create more ML related courses designed to attract cross-disciplinary students; for example, courses that require user researchers, designers, data scientists, and software engineers to work together. Students would benefit from experiencing this collaborative, creative approach.

Research Opportunities in Supporting UX Practice and Education

This study also surfaced many immediate research opportunities in supporting UX practice: 1) developing designerly abstractions of ML's technical capabilities, 2) developing examplars and sensitizing concepts of ML-enhanced designs, and 3) creating boundary objects that bridge UX and ML expertise.

Designerly Abstractions of ML

We advocate for research to develop resources that make the mechanisms of ML more available to designers. Previous research reported that some UX designers could fail to understand ML "specifically," even when they understand how ML systems generally work [9]. Our study echoes this. The designers comprehend ML in notably different ways than its textbook definitions; they understand ML largely through abstractions and examplars.

The abstractions served as a frame through which designers reflected on the design challenges at hand and made new assertions about how ML might provide value for users. They freed designers from grappling with technical limitations when sketching, empowering them to envision ML applications that moved beyond current archetypical forms. They served as boundary objects, allowing designers to discuss what users value with data scientists and to address issues of context. They also fostered new design ideas, serving as bridges between technical capabilities and design possibilities.

We suspect that many of the abstractions participants shared would generalize well beyond the specific applications they were working on. For example, participants stated that ML enables "*an experience personalized for everyone*" (P3), "*an evolving relationship with the users*" (P14), and "*handling more abstract user instructions*" (P7), and these

typically matched with other UX values that HCI research has raised over time [12, 31]. Extending, evaluating, and documenting these abstractions offers a clear space for design research. We encourage HCI researchers aid UX practitioners to develop a large suite of these abstractions, possibly by deconstructing current products and services that employ ML. Our participants developed such abstractions over months and even years of collaboration with data scientists. A ready repository of ML abstractions could potentially expedite this process, enabling novice designers to grasp the design possibilities of ML and build a rapport with technologists much faster.

A robust set of these abstractions would help to evolve the understanding of ML as a design material. It could function as a kind of taxonomy that is likely to be radically different from ones used by data scientists; a taxonomy focused on the match of contextual capability and user value.

Design researchers could advance the abstractions by creating many instantiations of the same abstraction as a way of investigating the opportunity space. For example, P14 described how ML enables users to engage in an unfolding relationship with the service, which evolves with user interaction. Design researchers can investigate many different kinds of evolving relationships between users and ML-enhanced products, looking for the best opportunities and unintended consequences linked to different forms. These design inquiries could then function as sensitizing concepts, allowing practitioners better operationalize the suite of abstractions.

Design Examplars and Sensitizing Concepts

In addition to abstractions, our participants used examplars to comprehend and communicate ML's more nuanced design qualities. A set of widely known examplars – Clippy, Tay, auto-correct and email spam filters – codified a rudimentary set of algorithm capabilities and interactions, as well as limitations and frustrations they have personally experienced.

Interestingly, designers in our study only referenced a few classic, and some failed, designs despite the fact that many commercial applications employ ML today. They seemed to only notice successful designs when they studied competitor products, comparing multiple ML interactions across different products, users, and contexts. We suspect that designers in our study failed to recognize many ML designs that they use every day, especially the ones that have so successfully adapted to their interactions that they have become invisible or unremarkable [26].

The frequent use of exemplars suggests that they would be a great way to communicate ML capabilities and limitations to UX practitioners. To help expand designers' perceived application of ML, design researchers could assemble a collection of well-documented examplars. A wide variety of HCI technical research that uses ML could also provide a source of exemplars and this could help to transfer these ideas to practice.

We speculate that this reliance on explicitly documented exemplars points reveals a huge opportunity for sensitizing concepts; designs created to open up the space for design innovation by sensitizing researchers and practitioners to the breadth of possibility [3, 33]. Design researchers could create sensitizing concepts to communicate ML design opportunities that are not instantly recognizable. Demonstrating a functional ML system to designers might not be enough to sensitize them, as many of these systems have weak connections to divergent user experiences after repeated use. Some work is needed to explore the diverse forms of research artifacts and knowledge representations, and to deliberately choose the ones that most effectively sensitize practitioners.

Boundary Objects for UX and Machine Learning Expertise

Our interviews captured an intimate, constant, cross-disciplinary collaboration when creating ML products. This is somewhat different from the design and technical collaborations found in a traditional UX design [7], where designers typically deliver a fully formed design to a technical team to implement. We see opportunities for new collaboration tools that help designers better work with data scientists. Previous work has proposed the use of boundary objects that scaffold the conversation between UX and ML expertise in creating adaptive UIs [32]. Our findings suggest similar work to be potentially valuable for many other application domains of ML.

We also see value in data visualizations and tools that serve the particular needs of UX designers. In our study, participants taught themselves to capture "rich and compelling user stories" from telemetry data that were different from the data scientists' insights. Designers frequently ask: How do users use the system? What is the matrix for a good user experience? What are the situations where design can make things better? Data tools for UX designers could surface some answers to these frequently asked questions. The visualizations could also serve as a boundary object for discussing user behavior patterns.

Our interviews showed that designers who lacked effective access to data scientists explored fewer design ideas and more often quickly resorted to familiar designs of ML. Previous design research has focused on enabling and improving designer and data scientist collaboration, assuming that capable data scientists are readily available [12, 32]. However, our study indicates that this assumption in not always true. Many designers in our study lacked access to dedicated or even proficient data scientists, especially the ones working at startups, small technology companies, and non-IT-focused companies. Despite efforts to make ML available to everyone [28, 25], the fact that proficient data scientists were scarce might be a reality that most designers will face.

There is a real need for design tools and methodologies that support designers who lack constant access to capable data scientists. For example, ML tools for designers could simulate the role of the data scientists, enabling designers to quickly evaluate the feasibility of their ideas when sketching. We also see opportunities for constructive design research to demonstrate creative designs that use off-the-shelf, ML

plug-ins; designs that do not need intensive ML development effort to implement.

Open Research Questions

We have discussed research opportunities around disseminating the current best practices in industry for UX designers working with ML. Parallel to these efforts should be work to critically examine if these are indeed the "best" practices. We highlight two key issues on which prior research in designing ML and current practice are in conflict, and therefore merit further study.

Do Designers Need Technical Machine Learning Knowledge?

Designers in this study understood ML at a high level, treating it largely as a black box. When sketching, they asked the data scientists "not whether this would work, but in your gut, whether this would be possible on a scale 1-10", acquiring only enough understanding to evoke a response to the design questions they had at that point in time. This behavior somewhat conflicts with the efforts being made to teach designers the statistical taxonomies and descriptive mechanisms of ML [13, 14, 15]. Designers instead used data scientists as a proxy for their design material.

This study showed that designers were *capable* of engaging in a reflective conversation with ML without a formal ML education. Yet it remains an open question whether a more advanced understanding of ML would inspire new design perspectives and new designs. On the one hand, design research has long promoted a "designerly way of knowing and doing" — the idea that designers can work with technologies and situations that they do not fully understand [1, 20, 23]. More investigation is needed to better define what is a "good enough" understanding of ML, and to understand when a more detailed knowledge of ML might help.

On the other hand, previous investigations of technology materials are often based on in-depth descriptions of their special properties, and then progressed to envisage a novel interaction design [30]. This approach allows designers to redirect their focus from the artifacts toward the materials from which the artifacts are built. This kind of discussion has not yet happened in designing ML. What are ML's special properties in improving UX? How do its experiential qualities differ across algorithms or datasets? Answers to these questions would allow for a more reflective and disciplined discussion of how to make ML available to design practitioners and enable design-led ML innovation. This would also inform educators whether and how ML education for designers should be different from the one for general non-ML experts.

Should the User-Centered Design Process Change?

Previous work shows some designers do not know how to bring their UX expertise to bear on ML [9]. Our study offered one possible solution, a divide-and-conquer approach. Designers first identified the right problem for ML to solve, validated its technical feasibility, and then iterated on the diverging-converging ideation process to craft the right manifestation of ML functionality. We found their overall approach to be similar to descriptions of the lean startup with a focus on making a minimal viable product [22]. They were similar in the way teams would assess the viability of an idea and pivot when it seemed to fail.

Future research should reflect on and improve the process we observed. One advantage of this process is that it did not require designers to acquire extensive ML skills, use many new tools, or radically change their familiar design activities. After a working ML application was launched, designers would go through a second design and evaluation process, fixing the design problems that the initial iteration failed to capture. However, a complete ML design process seemed to take way more time than a conventional double-diamond UX process [7]. Can the function and form of a ML design be addressed at once? Are there new ways to prototype and evaluate the technical viability and design quality before product launch? Answers to these questions have the potential to radically lower the cost of developing ML systems, and to enable many more resource-sensitive organizations to insert ML to their products and services.

At a higher level, the procedural knowledge of designing ML marks a clear space for HCI/design research. Existing work has offered valuable declarative knowledge and conceptual understandings of ML from a design perspective (i.e. [18, 19]). Embedding this growing body of new knowledge into organizational and procedural contexts opens up new research opportunities and promises real impact on UX practice.

CONCLUSION AND LIMITATION

In this paper, we report on an interview study to understand how expert designers who regularly design ML products do their work. We expanded previous work by illustrating the best practice of incorporating ML in the design process in industry. Our findings offer points of reference for design education, research and tool development with the goal of making ML available to UX practice. Through this work, we hope to start a reflective discussion around how HCI research can better support actual UX-ML design practice.

One limitation of this study is that it is exclusively from the perspective of a small and very specific sample of UX practitioners. We do not know the design process from the perspective of their data scientist or other collaborators. We do not know if or how other organizations have experienced resistance to ML approaches. We encourage the IxD and HCI research community to join us, and understand the UX practice more holistically.

ACKNOWLEDGEMENTS

This material is based in part upon work supported by the National Science Foundation under Grant Numbers SES 1734456. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This material is also in part based upon work supported by the National Institute on Disability, Independent Living, and Rehabilitation Research under NIDILRR 90RE5011. The first author is also supported by the Center for Machine Learning and Health (CMLH) Fellowship in Digital Health.

REFERENCES

1. Christoph Bartneck. 2007. Design methodology is not design science. In *the CHI 2007 workshop: Converging on a" science of design" through the synthesis of design methodologies, San Jose, CA*.
2. Jonathan Bean and Daniela Rosner. 2014. Big data, diminished design? *interactions* 21, 3 (2014), 18–19.
3. Glenn A Bowen. 2006. Grounded theory and sensitizing concepts. *International journal of qualitative methods* 5, 3 (2006), 12–23.
4. John Brownlee. 2015. Apple Finally Learns AI Is The New UI. (2015). https://www.fastcodesign.com/3047199/apple-finally-learns-ai-is-the-new-ui
5. 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.
6. Amber Cartwright. 2016. Invisible Design: Co-Designing with Machines. (2016). http://airbnb.design/invisible-design/
7. Design Council. 2005. The 'double diamond' design process model. *Design Council* (2005).
8. Laura Denham. 2017. Data Scientists Are The Next UX Designers. (2017). https://channels.theinnovationenterprise.com/articles/data-scientists-are-the-next-ux-designers
9. 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
10. Daniel Fallman. 2003. Design-oriented human-computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 225–232.
11. 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
12. 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
13. 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).
14. Patrick Hebron. 2016a. *Machine learning for designers*. O'Reilly Media.
15. Patrick Hebron. 2016b. New York University Tisch School of the Arts Course: Learning Machines. (2016). http://www.patrickhebron.com/learning-machines/
16. Mike Kuniavsky, Elizabeth Churchill, and Molly Wright Steenson. 2017. *The AAAI Spring Symposium Series Technical Report: Designing the User Experience of Machine Learning Systems*. Technical Report SS-17-04. Palo Alto, California.
17. Bryan Lawson. 2012. *What designers know*. Routledge.
18. 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.
19. Betti Marenko and Philip Van Allen. 2016. Animistic design: how to reimagine digital interaction between the human and the nonhuman. *Digital Creativity* 27, 1 (2016), 52–70.
20. Wanda J Orlikowski and Debra C Gash. 1994. Technological frames: making sense of information technology in organizations. *ACM Transactions on Information Systems (TOIS)* 12, 2 (1994), 174–207.
21. Fatih Kursat Ozenc, Miso Kim, John Zimmerman, Stephen Oney, and Brad Myers. 2010. How to support designers in getting hold of the immaterial material of software. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2513–2522.
22. Eric Ries. 2011. *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. Crown Books.
23. Donald A Schon. 1984. *The reflective practitioner: How professionals think in action*. Vol. 5126. Basic books.
24. Erik Stolterman. 2008. The nature of design practice and implications for interaction design research. *International Journal of Design* 2, 1 (2008).
25. TensorFlow: Smarter machine learning, for everyone 2017. (2017). https://www.google.com/intl/en/about/main/tensorflow/
26. Peter Tolmie, James Pycock, Tim Diggins, Allan MacLean, and Alain Karsenty. 2002. Unremarkable Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*. ACM, New York, NY, USA, 399–406. DOI: http://dx.doi.org/10.1145/503376.503448

27. 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
28. Udemy Online Course: Applied machine learning for Everyone 2018. (2018). https://www.udemy.com/applied-machine-learning-for-everyone/
29. Philip van Allen and Ben Hooker. ArtCenter College of Design Course: Internet of Enlightened Things. (????). https://canvas.instructure.com/courses/1111888
30. Mikael Wiberg. 2014. Methodology for materiality: interaction design research through a material lens. *Personal and ubiquitous computing* 18, 3 (2014), 625–636.
31. 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.
32. Qian Yang, John Zimmerman, Aaron Steinfeld, and Anthony Tomasic. 2016. 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
33. John Zimmerman, Erik Stolterman, and Jodi Forlizzi. 2010. An analysis and critique of Research through Design: towards a formalization of a research approach. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*. ACM, 310–319.