Mapping Machine Learning Advances From HCI Research to Reveal Starting Places for Design Innovation

Qian Yang
Carnegie Mellon University
HCI Institute
yangqian@cmu.edu

Nikola Banovic
Carnegie Mellon University
HCI Institute
nbanovic@cs.cmu.edu

John Zimmerman
Carnegie Mellon University
HCI Institute
johnz@cs.cmu.edu

ABSTRACT
HCI has become particularly interested in using machine learning (ML) to improve user experience (UX). However, some design researchers claim that there is a lack of design innovation in envisioning how ML might improve UX. We investigate this claim by analyzing 2,494 related HCI research publications. Our review confirmed a lack of research integrating UX and ML. To help span this gap, we mined our corpus to generate a topic landscape, mapping out 7 clusters of ML technical capabilities within HCI. Among them, we identified 3 under-explored clusters that design researchers can dig in and create sensitizing concepts for. To help operationalize these technical design materials, our analysis then identified value channels through which the technical capabilities can provide value for users: self, context, optimal, and utility-capability. The clusters and the value channels collectively mark starting places for envisioning new ways for ML technology to improve people's lives.

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

Author Keywords
User Experience; Machine Learning; Sensitizing Concept; Data Mining; Bibliometric; Research Transfer.

INTRODUCTION
Machine learning (ML) is increasingly being used to improve the quality of user experience. From mundane spam filters that save people time to conversational agents that offer a more human mode of interaction, to personalized media recommendations, it can seem like ML is in almost every new technology product and service. Both UX practitioners and researchers note this trend and become especially interested in design opportunities surrounding ML [9, 18, 14]. Some even speculate "ML is the new UX" [4, 36].

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

ISBN 978-1-4503-5620-6/18/04. . . $15.00

DOI: https://doi.org/10.1145/3173574.3173704

Interestingly, one study recently noted a lack of design innovation with ML, suggesting a potential gap preventing HCI's ML research advances from moving into commercial products [6].This claim is based on the assumption that technical advances are followed by design innovation, where designers envision many new product forms that fit the technical advance into different aspects of peoples' lives. For example, the 1962 cassette recorder represented a technical advance, providing an easier way to make audio recordings. This technical advance was followed by a wave of design innovations. Designers created many new forms for audio devices that employed the cassette tape, including: home stereo systems, boom boxes, personal cassette players like the Walkman, automotive tape players, duo-cassette player-recorders that better supported making mix tapes, dictation devices, and phone answering machines. It is difficult to see a similar type of design innovation taking place with ML. It is hard to trace a technical advance being followed by many new product forms that operationalize the tech for a new target user, a new activity, and/or a new context.

This raises the question: How can we help HCI design researchers and practitioners better work with ML's technical advances?

This challenge of making technical advances more accessible to design practitioners is not new or unique to ML. In previous cases, design researchers have made sensitizing concepts as one way to better expose the capabilities and possibilities of new technology. Design researchers made intriguing sensitizing concepts for interactive textiles [28]; they held workshops to expose haptics' design possibilities beyond a buzzing phone to design practitioners [24]; and they created many different social robots to show a variety of ways robots might fit into the various aspects of peoples' lives; robot forms not likely to emerge from a robotics engineering lab [8].

We wanted to build on the success of these previous design research endeavors. We wanted to help design research make contributions that open up the space for design innovation that uses ML. But where to start? The design space, where ML might improve UX, feels significantly larger than the spaces of these previous design inquiries. As a first step, we chose to investigate HCI research that used or mentioned ML. We chose this bounding because while there have been many ML advances outside of HCI, HCI research generally has a focus on human needs and desires.

We searched for HCI publications that mention ML. This produced a corpus of 2,494 publications. Our first goal was to reveal the topics HCI has worked on with respect to ML. Second, we wanted to identify opportunities where design researchers might dig in and create sensitizing concepts that aid HCI-ML advances in transferring to design practice. We carried out our analytic literature review in three phases:

1. We generated an overview of the corpus using simple statistics. We noticed a lack of HCI research that addresses both UX and ML. To date, only 9 DIS papers have mentioned "machine learning," and only 3 described or reflected on the design of a ML-enhanced system.
2. We generated 7 clusters of work where HCI researchers have repeatedly employed ML techniques to make an advance. Among them, we identified clusters where systems have frequently use similar interactive forms to create value for users and clusters where new ML capabilities seem much less bound to any specific interactive form.
3. To help operationalize these genres and make them more useful as starting places for creating sensitizing concepts, we created a conceptual model containing 4 value channels through which ML advances provide experiential value for users. We used these to demonstrate how design researchers can ideate concepts using the technology within a cluster.

This paper makes two contributions in helping spanning the gap between HCI-ML advances and UX design. First, it provides a set of labeled topic clusters of ML technology. The clusters form a map of HCI advances that employ ML technology and are ready for design researchers to work on. Second, it provides a conceptual framework of ML's user value. The value channels, when used with the topic clusters, can support ideation of possible sensitizing concepts that make the ML technical advances in HCI more accessible to practitioners.

BACKGROUND ON DESIGN INNOVATION WITH NEW TECHNOLOGIES

We first provide a brief description of how designers design with new or less-understood technologies, and link this to what sensitizing concepts in HCI are and how they help move new technologies into practice.

Challenges of Designing New Technology

Designers integrate known technologies into novel and valuable new products and services. Louridas notes this when describing the difference between designers and engineers. He notes that engineers create the "means"; they create new technology that allows new capabilities. Designers, he claims, work like bricoleurs. They do not invent new things, but instead they create novel and valuable assemblies of known things [23].

To envision things that have never before existed, designers innovate by engaging in reflective conversations with design materials [32]. Schön describes how designers reflect in action, how they conceive of what they want to make while in the act of making it [31]. Reflection in action works well when designers have a tacit understanding of the materials they are using, of the materials' capabilities. This creative process happens less readily with new technology, because designers lack a ready-at-hand knowledge of what it is and what it can do. Neither Louridas or Schön described how designers become familiar with new and emerging technology in order to innovate with it.

From Technology to Design

In addressing the challenge of starting with technology and searching for valuable new things to make, HCI researchers have proposed *matchmaking* [3]. Matchmaking starts by asking designers to detail the technical *capabilities* of the tech they are working on. Next, they systematically work to discover *activities* related to these capabilities, *domains* related to the activities, and finally target *users* connected to the revealed domains. Unfortunately, this approach is both under-investigated as a design method and underutilized by both design researchers and practitioners.

Design researchers have claimed that they can help with the challenge of transferring technical advances to practice. When proposing that HCI researchers accept research through design as a research contribution, design researchers claimed their work would embedded the latest HCI technical advances into design exemplars that would be easier for practitioners to understand, and that would aid in the transfer to practice [38].

Sensitizing Concepts

The term "sensitizing concept" comes from the social sciences and research on grounded theory [2]. Designers have appropriated this idea, using sensitizing concept to mean the creation an artifact meant to open up space for future design innovation; the artifact is meant to sensitize other designers to many new possibilities beyond the specific artifact that was created. Design researchers using research through design often make sensitizing concepts as a way of producing design knowledge [38].

Sensitizing concepts carry and manifest all the combined knowledge about the material that has influenced the design [33]. They help designers grasp and feel the new design materials. For example, one design researcher had a desire to work with haptics; however, he had no easy way of playing with haptics as a design material in order to develop tacit knowledge of what could be [24]. He started an ambitious project to sketch with haptics, setting a goal for himself to make a new haptic device each day for several weeks. His work produced new language for talking about the aesthetics of haptics as well as several simple devices that could produce very different kinds of haptic feedback. He worked to transfer this knowledge through workshops with designers where they together build several of these devices in order to experience the range of possible tactile sensations.

Sensitizing concepts push the boundary of a design space. For example, researchers were motivated by the fact roboticists seemed to be making robots with little concern for the people who might be living with them and interacting with them. They produced many novel designs that recast the role of a robot. For example, the hug robot recast a robotic, plush device as a new way for grandparents and grandchildren to share hugs both remotely and asynchronously [5]. The work showed a radical new form as well as a radical new purpose for a social robot; one not likely to emerge from a robotics engineering lab. Similarly, HCI researchers have developed sensitizing concepts for interactive textiles [28], producing new opportunities for the textile industry based on IT innovation.

The use of sensitizing concepts is not new to design. For many years designers have made and shared new concepts with the public to get feedback on what might be good [33],to create a dialog between industry, design practice, and end users.

Previous research demonstrated that by understanding and engaging with new technology in a designerly way, design researchers can envision new forms and new purposes for the technology through the creation of sensitizing concepts, and help initiate a wave of design innovations. Recently design researchers stated that today's ML systems are as creative and interesting as the data scientists that make them [6]. Machine learning technology, as another new and less-understood design material, seems a perfect place for design researchers to step in and make a difference.

MAPPING THE SPACE OF HCI-ML INNOVATION

We aimed to map the HCI-ML innovation space and to identify opportunities for design researchers to dig in and create sensitizing concepts as a research contribution. We began our process by assembling a corpus of HCI publications that used or mentioned ML. We chose not to look at all research on machine learning as we suspect much of it has no clear relationship to user experience. All work in HCI, including the technical advances has some relationship to people/users, and thus some connection to user experience. We thought the HCI research would provide a good enough set to reveal a rich set of starting places for ideating possible sensitizing concepts. We extracted papers and their metadata from the ACM SIGCHI database and its mirror (`dl.acm.org` and `hcibib.org`), which contains conference proceedings and journal publications published in all ACM HCI venues.

We searched for articles using a set of ML-related terms. We searched for "machine learning", as well as the common variations of this term used in HCI (artificial intelligence, ambient intelligence, intelligent interface, and pattern recognition). Our search returned 2,494 HCI publications that contain at least one ML-related terms in the meta data (title, abstract, author keywords, general terms, ACM category). This formed our corpus.

PHASE 1: GROUNDING THE GAP

Method

For the first phase of our analysis, we wanted to ground the claim previous design researchers made about a lack of design innovation with ML technology. We wanted to gain some insight as to whether the ML advances made by HCI researchers had transferred to practice. However, this was not directly possible with our corpus, as the papers do not discuss transfer to commercial products, nor the use of HCI advances in commercial products. Therefore, as a proxy for HCI-ML technology transfer to practice, we chose to look for evidence that these technical advances had been picked up by UX design researchers publishing in HCI venues. Papers discussing research through design in HCI have claimed that design researchers would integrate the latest HCI technical advances in their exemplars as one path for transferring this knowledge to practitioners [37]. We wanted to see if this was happening. If the technology was not being used by design researchers, then this could provide preliminary evidence that it might not be effectively being picked up by practitioners.

To search for a gap between HCI technical research and design research, we computed descriptive statistics of the corpus and then plotted trends over years. As part of this we paid specific attention to different publishing venues as they focus on different sub-communities within HCI. Throughout this process, we selected individual papers to read, to help us make better sense of the trends.

Findings

Our plot of ML terms in HCI shows a clear trend that ML is an increasingly important topic (Figure 1). A total of 1,939 HCI publications mentioned 'machine learning' since the publication of the first mention in 1969. More than half of the papers addressing ML were published in the past five years. 176 ML-related CHI papers have been published since year 2000 and 235 ML-related UbiComp papers were published since 2007. The overwhelming majority of papers appeared in HCI venues with a technical focus. Publications from four technically focused conferences plus the CHI conference – ICMI, CHI, UbiComp, IUI, and RecSys. – comprised more than half of the corpus (Table 1).

We looked specifically at DIS, as this is a conference most devoted to design research within ACM publications. Only nine DIS papers mentioned "machine learning." Of these, only five described or reflected on the design of an ML-enhanced system.

Discussion

Our preliminary analysis of the corpus suggests that the explosive development of ML within HCI community has had a strong technical focus. Design researchers working within HCI do not seem to be participating in this trend, and this may indicate a breakdown in the transfer of ML advances

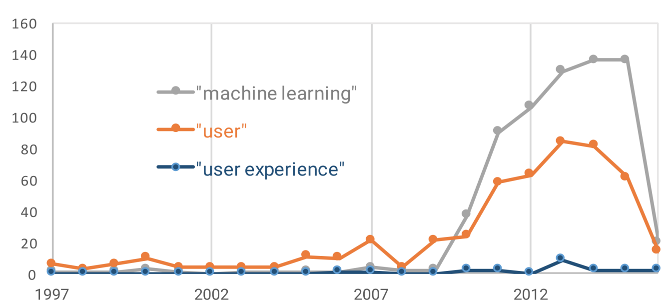

Figure 1. The numbers of HCI publications that mentioned "machine learning", "machine learning" and "user", and "machine learning" and "user experience" over the years.

Number of papers	CHI	DIS	CSCW	UbiComp	UIST	IUI
Mentioned *ML*	266	9	29	268	68	194
Mentioned *ML* and *user*	158	5	11	266	68	194
Mentioned *ML* and *UX*	1	0	0	0	0	0

Table 1. The number of ML-related HCI publications in different venues as of September 1st, 2017.

to UX design practitioners, thus inhibiting design innovation with this technology.

Drawing on previous HCI research on designing for new technologies, constructive design researchers can potentially help overcome this gap and spur more designerly innovation with ML technology. They can do this by making sensitizing concepts that demonstrate how ML capabilities can generate new kinds of value in people's lives. With regard to ML, however, a recent paper showed that UX practitioners struggle to understand ML's capabilities and limitations [6]. Another paper noted that even UX researchers with lots of experience in ML still miss simple opportunities to augment their products with well-known ML techniques [36]. To help designers and researchers better grasp the capabilities of ML, we continue to mine our corpus with the goals of revealing the ML topics HCI has extensively investigated, and forming a map of HCI-ML technical capabilities that are ready for design researchers to work on.

PHASE 2: TECHNOLOGY MATERIALS

Method

We wanted to reveal the topics HCI has worked on with respect to ML. We wanted to identify well-established topics, where the ML technology might have formed some UX design conventions as well as those that suggested more emergent technical advances. Given the size of the corpus (2,494 conference and journal publications) and the heterogeneity of its intellectual content, we felt a manual analysis process would be ineffective. Therefore, we chose to investigate if datamining might reveal interesting clusters; interesting starting places for design research. Design researchers do not typically use this datamining to find topics for new investigations. We have had some experience working on designs that employ ML; however, none used natural language processing. Fortunately, in recent years, dozens of walk-up-and-use text mining tools have become available (see [34] for a comprehensive list).

We used document clustering and topic modeling to gain some preliminary indications as to the structures and topics within the corpus. Both are commonly used and well-established techniques in text analytics. Both have been used in other sub-fields of HCI to reveal and overview of research topics and trends [22, 26]. We first prepared our corpus for the text mining procedures. We then followed a publically-available, step-by-step tutorial [30] to mine the clusters and topics of the corpus.

Construct the corpus. From previous experience, we knew that datamining algorithms are generally sensitive to noise in the data. In our case, noise might include words that do not represent the true topic of the papers. To reduce this noise, we chose to not include the body of the papers. We used only the metadata for clustering and modeling.

Tokenization and preparation. We parsed texts in our corpus into words and phrases (tokens), and removed punctuation and stop words – words that don't convey significant meaning such as "a" or "the" – from the tokens. We broke the words down into their roots. This is typically referred to as stemming. For example, "crowd-sourcing" and "crowd-sourced" were both transformed into "crowd-source," making them both the same token.

Calculating publication similarity. For each paper, we computed the weighted frequencies of the different tokens. Tokens that occurred frequently within a paper, but not frequently within the whole corpus, receive a higher weight. These words are assumed to contain more meaning in relation to the topic of the paper. Next, we computed the difference in weighted frequencies between any two papers in the corpus. This is typically referred to as the semantic distance.

Clustering. Using the resulting matrix of between-paper similarities, we ran clustering algorithms to reveal the hidden structure among the papers. We first experimented with the most commonly used and simplest algorithm, K-means. We manually input the number of clusters we wanted and the algorithm iteratively assigned papers to clusters in a way that minimizes the within-cluster semantic differences. We experimented by generating three to nine clusters, searching for the most fitting and informative choice. However, we could not easily determine an optimal number of clusters.

Next, we experimented with hierarchical clustering. This approach does not require a pre-determined number of clusters. Instead, it generates a hierarchy tree of paper clusters, automatically determining which clusters to combine together and which clusters to split apart. This provided us additional insights into the relationships among the papers and into what might be an appropriate number of clusters.

Cluster evaluation. We examined the clusters through a number of approaches. Because tokenized documents in the clusters are in a hyper-dimensional space, they cannot be visualized. Therefore, we tried both *multidimensional scaling* and *principal component analysis* as two different ways to reduce the dimensionality of the clusters in order to create visualizations. These are two of the most common dimension-reduction methods. In addition, we visualized the hierarchical clusters as a *dendrogram*; a tree diagram.

Topic modeling. Another approach to understanding clusters is topic modeling. We applied Latent Dirichlet Allocation (LDA) to each cluster, creating a model of their topics. LDA assumes papers are a mixture of topics and that each token influences the paper's topics. LDA generates a set of topics for each cluster. Each topic has a set of words that defines it, along with a certain probability.

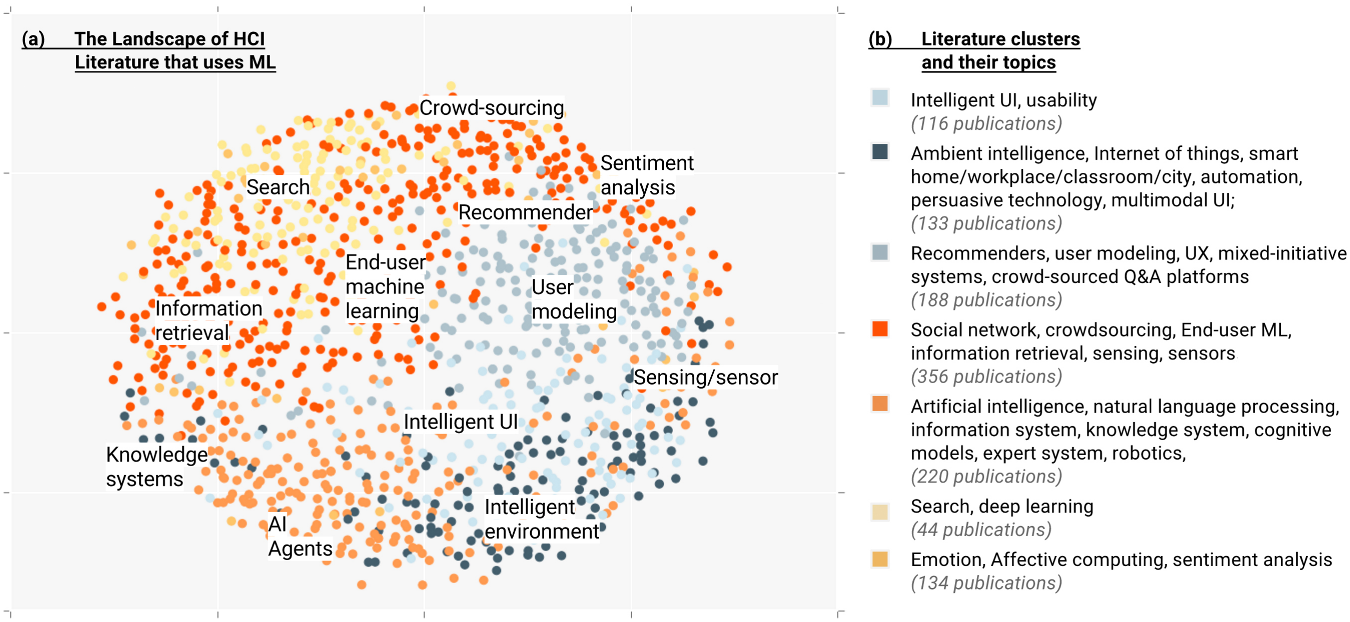

Figure 2. An overview of HCI research that uses machine learning. (a) An illustration of literature landscape based on the semantic distances among each publication. Each dot represents a publication, color-coded by cluster. (b) major topics of each cluster. The illustration and topic keywords are algorithmically generated; we then manually interpret and annotate the topics.

The above procedure generated multiple sets of paper clusters, and a set of algorithmically-generated topics for each cluster. These results were not particularly informative or insightful in terms of revealing staring places for design research. For example, LDA represented cluster topics as word lists, but the lists contain quite a few generic words (e.g., model, predict, data,). It also produced combinations of words that are not instantly understandable. A more designerly analysis was necessary.

To make better sense of the clusters, we sampled papers from each. When reading them, we interpreted the topics LDA generated, mapped the papers onto affinity diagrams, compared them to the algorithmic mapping, formed some conceptualization of the clusters, and then validated or adjusted the cluster through additional reading and descriptive statistics (i.e. frequent keywords, venues, and authors). Through this iterative process, we read hundreds of papers in the corpus and gradually formed our understanding of the clusters. We finally selected a clustering method whose results were most informative and useable to designers, and we produced a set of meaningful labels for each cluster.

Findings: Clusters of ML-HCI Research

We examined these clusters through the lens of UX innovation. First, we wanted to identify well-established ML topics as they indicate opportunities for researchers to create design patterns for UX practitioners to use. Second, we wanted to identify technical topics whose design space is under-explored, as they indicate open places for design innovation and constructive design research.

Our analysis revealed seven clusters of HCI literatures in relation to ML. Based on our extensive reading, affinity diagramming and algorithmic topic modeling, we labeled the clusters as: 1) intelligent UI and usability; 2) intelligent environment; 3) recommenders and user modeling; 4) social network and sensor framework; 5) AI and knowledge system; 6) search and deep learning) and 7) sentiment analysis and affective computing. These genres cover most but not all papers in our corpus. Some of the contributions were novel enough to not fit anywhere. The clusters were not necessarily mutually exclusive. Some papers fit in more than one clusters as these publications may speak to multiple literatures or multiple application domains. Figure 2 details the topics of and the semantic distances among each of these clusters.

In four of the clusters, there was a noticeable funnel effect, with researchers frequently devising vastly differing ML techniques and coming to similar, convenient interaction forms. A typical example of this can be seen in the Intelligent Environment cluster. Here a wide variety of technical topics – from internet of things to smart homes, classrooms, workplaces and cities – were tied to only a few interaction forms: automation, persuasion or multimodal interfaces. These common pairings illustrate that the application domain has frequently produce these same interaction forms. Among the papers we sampled, we found no papers that talked about sketching to produce ML enhanced designs or any work describing attempts to generate a divergent set of possible design solutions. The vast majority of papers seemed to use a “convenient" interaction choice. The different application domains often used different “convenient" design choices.

cluster #	**Technology Topics**	**UX Topics**
1	Intelligent UI	Usability
2	Ambient Intelligence; Internet of things; Smart Home/workplace/classroom/city;	Automation; Multi-modal UI; Persuasive Technology.
3	User Modeling; Mixed-initiative Systems; Crowd-sourced Q&A platforms	User Experience; Recommenders
4	Social Network; Crowd-sourcing; End-user ML; Information Retrieval; Sensing & Sensors	N/A
5	Expert System & Knowledge System; Robotics; Natural Language Processing; Cognitive Models	AI Agent
6	Search; Deep Learning	N/A
7	Emotion & Affective Computing; Sentiment Analysis	N/A

Table 2. Topic Clusters in ML-related HCI research. Topic co-occurrences in these clusters surfaced some common combinations of ML techniques and interaction forms.

Topic co-occurrences in these clusters surfaced some common combinations of ML techniques and interaction forms. For example, user modeling appeared often with recommender, suggesting that the latter is a typical use of user-modeling techniques. Likewise, in the sensor network cluster, mobile sensing appeared often with notification, visualizations, and persuasive technologies. Deep learning appeared often with search and not with other interaction-related topics.

The other three clusters center in ML technologies and appear less bound to any particular interaction form. Relatively few papers in relation to these ML technologies have attempted to address the interactive forms they might take. For example, one cluster shows that deep learning falls close to search, and not close to any other topics. This suggests that research on deep learning within HCI might predominantly have been used to improve the search user experience.

Sentiment analysis is another topic where we found many technical advances but few interaction design discussions. In detecting sentiments at-large on online social networks, for example, the remarkable heterogeneity of techniques, perspectives and orientations have contributed to a new theme of HCI research on understanding sociocultural dynamitic and structure (i.e. [16, 17]). There was little work investigating if the detection of sentiment could inform the design of better online communities or produce better experiences for users.

There were valuable exceptions to the general lack of design deliberation in HCI-ML. In one particular example [20], a researcher proposed a radically new approach to build and use a particular deep learning algorithm: neural network (NN). They investigated a set of NNs that were built for recognizing entities in images and discovered that one particular network could also trace relations between the entities. Inspired by this capability, they proposed that NNs might be a good way to learn relationships, rather than to trace the entities themselves. They then designed a new way of building NNs that trace relations between interactive systems and users; a radically new way to craft human-machine relationships.

Discussion

Our goal was to identify opportunities where design researchers can dig in and make sensitizing concepts as research contribution. We produced seven clusters of ML topics within HCI research. We also documented pairings between certain interaction forms and ML advances. Design researchers might apply their envisioning skills to the most under-investigated ML technologies; creating sensitizing concepts that offer new forms for ML to take and many new ways it might deliver value to people.

Our clustering showed that in each application domain, researchers frequently resort to a small, fixed set of convenient interaction forms. On one hand, identifying these common pairs of ML techniques and interactions is useful for researchers who are looking for how to make a well-defined intelligent design contribution, or to seek new ground within the space of UX-ML by avoiding design space already substantially covered. On the other hand, such convenient designs choices do not advance us as a field towards the objectives of design innovation, innovations that radically re-imagine new forms and purposes for ML. To avoid reinventing the wheel, our topic landscape can help design researchers identify major under-explored design spaces, and create sensitizing concepts to release the UX potential of these technologies as a research contribution.

We see opportunities for UX research in expanding design possibilities beyond the conventional forms their domain frequently resorts to. We encourage designers not only to actively re-imagine what they could do differently with the ML technology but also to report on the divergent set of possible designs that emerge from in their sketching process.

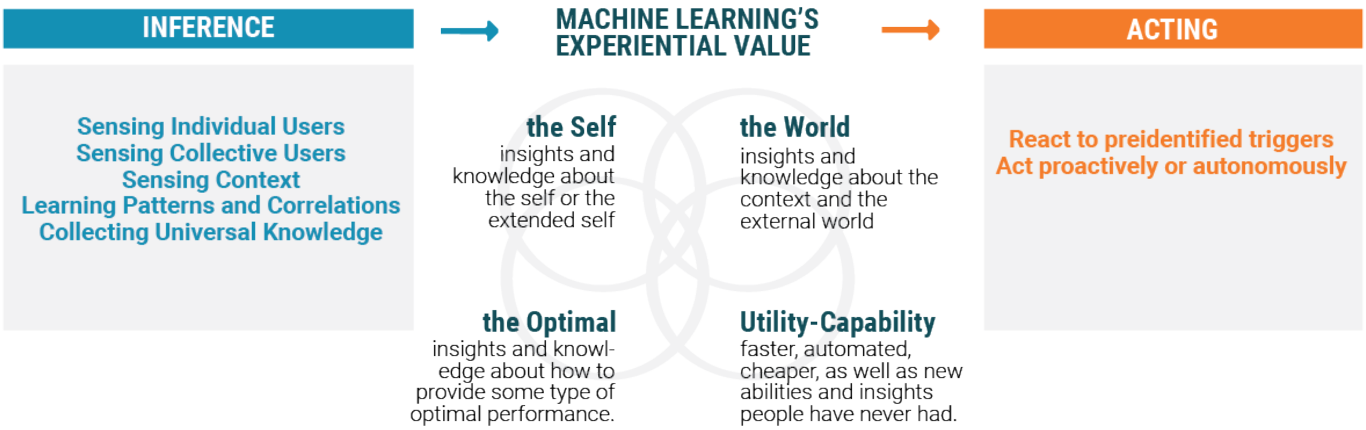

Figure 3. Value channels of Machine Learning. We identified four channels through which ML advances in HCI provide value to users.

Lastly, our analysis identified three clusters of ML technical advances that have not yet been bound to particular utilities, interactions or user experiences. For example, sentiment analysis - What can we design when one is able to capture societal happiness [1] or the mood of a city [27]? Or social network mining - How would we design better online communities given the capability to predict its collective attitude towards an upcoming social event? We believe operationalizing these techniques mark exciting new themes in the design space of ML-driven product and services.

PHASE 3: UX VALUE FROM MACHINE LEARNING

The seven clusters provide a set of technically defined starting places to develop sensitizing concepts. However, HCI research and practice employs a user-centered orientation that discourages starting design work from the perspective of a specific technology and then searching for users. To aid in making these seven clusters more approachable from a user-centered process, we focused on how ML has previously been used to create new value for end users. We assumed a mapping of the value space would provide several user-centered perspectives that could help when ideating concepts within each of the clusters.

To help reveal how ML has generated UX value, we revisited the seven clusters and searched for recurrent themes within each and across the entire set. This process helped us consider creation of value beyond the simple rationales used in many of the technical papers where their goal was to demonstrate the technology could work, and not to show every possible way their technology might help people. For example, recommender papers would often hand-wave around who needed or benefited from recommendations and would instead focus on improving the performance of the recommender. We wanted to abstract from this to move closer to revealing the many stakeholders that might benefit from a recommendation.

We sampled and read papers from each cluster's center; papers written by leading authors within a cluster, or papers addressing the most popular sub-topic within a cluster. We also sampled "boundary" papers; papers with a high likelihood of being in more than one cluster. For each paper, we made an inference describing how a user might derive value from the advance. We mapped our resulting set of inferred user value using affinity diagrams. This produced a new set of themes. In addition to the affinity diagram, we also experimented with mapping these advances offered by the papers onto different UX representations. These included: 1) a double diamond model of the UX design process, 2) a set of design challenges UX designers face when working with ML [6], 3) a Venn diagram that showed how ML creates UX value for commercial products and services [34], and 4) an inference-to-action interaction flow (Figure 4). While mapping, we continually discussed inspiring ML applications and trends; we discussed what might be missing; and we discussed what might be wrong with our affinity themes. Through an iterative process involving many rounds of remaking the affinity clusters and mapping the papers to the different models, themes gradually emerged. Finally, we consolidated our insights from the models and the affinities to create a conceptual model describing how HCI-ML research innovations create UX value (Figure 3).

Findings: User Value of ML

The HCI advances using ML offered inferences and actions that increase a user's perception of value. They also provided new capabilities simply not available prior to the use of ML. We created a schema to illustrate ML capabilities HCI researchers have developed (Figure 4). As we map and analyze these capabilities, the user value ML offers gradually become clear.

Our conceptual model shows four channels where ML creates or augments value for users:

- inferences about self (e.g., automatically inferring and logging sleep stage and sleep quality [11])
- inferences about the world (e.g. prediction of public health crisis outbreak [12])
- inferences about what might be optimal (e.g. recognizing the patterns of engaging conversations to design engagement-aware conversational agents [35];)
- inferences that provide utility and/or a new capability that is not directly related to self-understanding, contextual

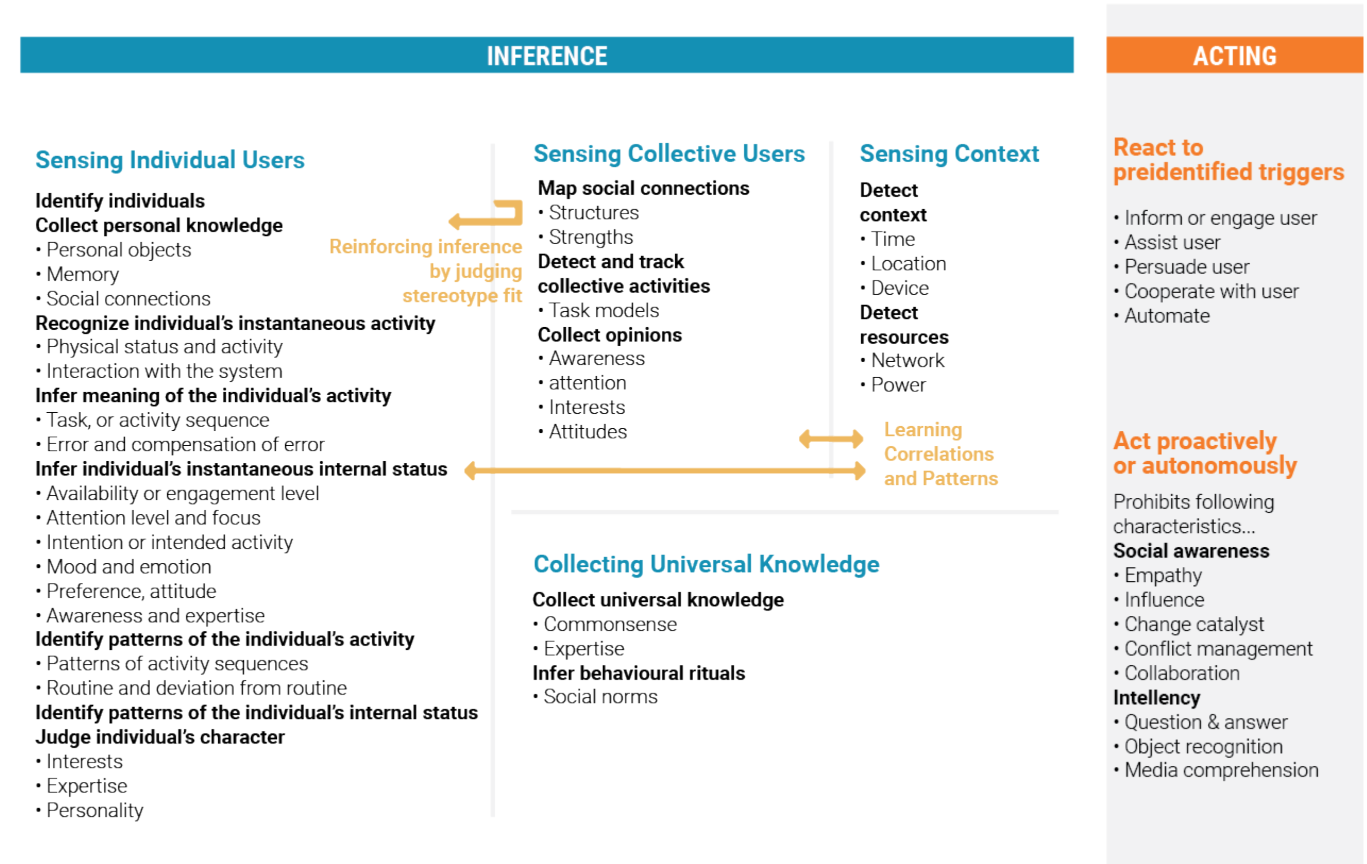

Figure 4. A schema of machine learning capabilities in terms of enhancing inference and actuation. These capabilities increase a user's perception of the aforementioned experiential value.

awareness, or optimization. Sort of a catch-all for everything else. (e.g. the ability to turn people's skin into a touch interface because devices might shrink to be too small to effectively interact with [13]).

Inferences about self: One important way ML has been used to create value for users is by monitoring and logging the user's actions and then using this data to produce inferences about the user or about a group the user shares behavioral characteristics with. This value channel deals with the insights and knowledge that allow users to understand themselves. For example, HCI research explored simple reminders about a user. These drew on the user's past personal experiences, the objects they use, and even the names of their friends. These reminders could enhance cognition, trigger the recall of memories [21] and generate sentimental value [7]. We found instances where ML significantly enhances the capabilities to make sense of such information by inferring stereotypes and social dynamics of groups one belongs to, based on the similarity of attitudes and actions. ML can detect meaning in a user's subtle social actions, such as detecting sarcasm in outgoing tweets [10]. It can infer a user's internal state, such as their emotion, attention level, knowledge growth over time, or intention at the moment. These techniques can be applied to groups of users too, groups that ML can dynamically and implicitly form by identifying some shared behaviors [29].

Inferences about the world refers to the value of knowing about user's current context, a distant context, or knowing information relevant to a currently unfolding interaction. Popular contextual cues that researchers used include:

- Context of use: time, location, motion, device, etc.;
- Computing capabilities: devices accessible for user input and display, available computing resources, connectivity, costs of computing, etc.;
- Physical environment: lighting and noise level, etc.;

In addition to simple facts, ML also makes it possible for machines to make sense of, and even autonomously act on human knowledge. For example, clinical data mining can surface implicit in electronic medical records. This can help clinicians make better medical decisions. This makes them better collaborators with people. ML can transform facts about the external world into machine intelligence. Robots can collect signals with regard to human-interactions, learn social norms and common-sense knowledge from them, and perform new tasks in a socially fluid manner [15].

Inferences about the optimal refers to insights and information about an arbitrarily defined “optimal" or “better" status. This sense of optimal can often be found in persuasive technologies, personal coaching applications, and intelligent tutoring systems that work to increase learning efficiency. For example, ML can monitor a students’ level of participation and predict students that are likely to dropout within massive open online courses (MOOCs). Early detection of a likely dropout might trigger an appropriate intervention. In the long term, this type of prediction can provide valuable insights into more optimal course design. Similarly, commercial media applications often rely on ML to deliver recommendations that increase dwell time and the long-term engagement of users. In both examples, ML infers a predictive pattern of the optimal behaviors, such as active participation in a class, and enables the system to intervene to increase the likelihood of an optimal outcome.

Utility and/or new capability. In general, all designs that employ ML to improve UX provide some type of utility to the user, often through a new and desirable capability. This bounds the entire solution space when the ML benefits the user as opposed to benefiting some other stakeholder. Within this bounded space, we have identified three specific categories: inferences about the self, the optimal, and the world. We classify the rest of the solution space as *utility*.

For utility, we mean things such as saving the user time, saving them effort, or providing them with an increased sense of control. One common example is the automation of a formerly manual task. This channel covers user value that does not directly come from increased self-understanding or an increased understanding of context. This is also not the same as optimal. Giving people new capabilities or automating mundane tasks to save them time and effort differs from optimal because the value is pre-set by the designer and not discovered by the ML system. differs from value of the optimal in that whilst most utilities represent universally shared values (i.e. efficiency in performing mundane tasks), the definitions of the optimal (i.e. long product dwell time) can be subjective and fluid. Utility value includes interaction efficiency, availability, reduced cognitive and interaction efforts, etc. Example ML applications that deliver utility value includes: adaptive mobile user interfaces that minimize the users’ navigation efforts (i.e. [19, 25]), conversational agents that allow users to interact in a more convenient or natural way, and system transform any surface into a touch screen [13].

Discussion: Ideating Sensitizing Concepts

The previous phase generated seven clusters that describe where HCI research has employed ML. In this phase we created a conceptual model showing how ML has been used to generate or augment value for users. We documented four value channels that function as different perspectives or lenses for connecting the technology within a cluster to potential users. Collectively, use of the cluster and model should help design researchers ideate many possible sensitizing concepts. We recommend a process of selecting a technology and then systematically generating ideas from each of the four value channels: user, context, optimal, and utility-capability. This is one form of match-making [3].

To show how this might work, here we provide a quick demonstration using the sentiment analysis cluster as our technology focus. We start by first browsing HCI literature on sentiment analysis for inspiration. Based on this, we selected a specific ML technique named *identity sentiment analysis* [16]. This technique can automatically extract “who did what to whom" (social event actors’ behaviors and identities) from text. It can also reveal the sentiment expressed by the author towards the event, towards a specific behavior or action, and any of the individual people discussed in the text. To ideate novel design concepts, we first put this technical capability against the backdrop of the whole HCI-ML topic map. We noted that sentiment analysis falls close to the topics of recommenders and user models (Figure 2a). This led us to think about recommending news stories that share the individual user’s sentiments of a given day.

We then view this technology through the lens of user value channels; This helped us generate many concepts from the perspectives of different possible stakeholders.

- *From the perspective of inferring individual users*, we can design to replicate a lived experience of social events. For example, we can design an immersive environment that simulates the actions and emotions in public religious practice so that users can experience a sacred, meditating moment at home after a busy day;
- *From the perspective of contextual value*, we can design this technique to translate social events in an exotic culture into analogous events in users’ native culture. This mapping can be based on mirroring actions and shared public sentiments between two events;
- *From the perspective of inferring the optimal*, we consider “the optimal" being an unbiased view towards different cultural groups or opinion groups. We can create de-biased narratives of social events and groups, inducing users’ rational engagement with the events;
- *From a utility point of view*, we imagine this technique can hugely augment users’ text comprehension. Given that this technical can automatically extract event sequence, actor identity and sentimental interpretations, we can design a text visualization tool that mass produce timeline summaries for textual narratives, i.e. legal documents, historical stories, news narratives, etc.

As such, the two constructs (technical topics and user value channels) can be used together to aid design researchers in ideating what to make with innovating UX of ML. These ideas are not the end, but instead they are meant as a starting point for an iterative user-centered design process through which the designs will both evolve and be reframed. We encourage fellow design researchers to apply this method to their respective topic of interest, realizing the design potential of ML’s various techniques and forms.

REFLECTION

The goal of this paper was to provide an analysis of the current landscape of HCI research in relation to ML, and organize the diverse design opportunities that comprise this field. To do so, we categorized the HCI-ML work into two frameworks. We created seven topic clusters of ML technologies and a conceptual model showing how users experience value from ML enhanced systems. The technical topics and value channels can serve as descriptors for what has been done and provide starting points for future UX innovations. We encourage fellow researchers to use, evaluate, discuss and improve these frameworks, and join us in releasing the potential for better, more creative, more sophisticated design with ML.

We also want to step back for a moment to more broadly consider the idea of mining ML advances in research to inform and inspire design researchers. One limitation of the presented frameworks is a lack of evaluation. The intended outcome of this approach — inspiring design researchers and making them desire to create sensitizing concepts or undertaking research on ML — is a very difficult thing to measure because it is nearly impossible to control for. Instead, if HCI design researchers accept the argument that ML as a design material adds value to UX, what our frameworks provide is a concrete way for them to begin to investigate ML in their own research and design. In addition, the ML research topic landscape and the inference-acting ML capability schema offer the practical value for their design, taking stock of what is known and to identify major unknown topics as a basis for their future research endeavor. We strongly encourage the UX and HCI research community to join us and start a serious discussion around the innovation issues related to the idea that *ML is the new UX*.

ACKNOWLEDGEMENTS

This work was developed under a grant from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR grant number 90RE5011-01-00). NIDILRR is a Center within the Administration for Community Living (ACL), Department of Health and Human Services (HHS). This work was also supported by grants from NIH, National Heart, Lung, and Blood Institute (NHLBI) # 1R01HL122639-01A1.

The first author is also supported by the Center for Machine Learning and Health (CMLH) Fellowship in Digital Health.

REFERENCES

1. Saeed Abdullah, Elizabeth L Murnane, Jean MR Costa, and Tanzeem Choudhury. 2015. Collective smile: Measuring societal happiness from geolocated images. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 361–374.
2. Herbert Blumer. 1954. What is wrong with social theory? *American sociological review* 19, 1 (1954), 3–10.
3. Sara Bly and Elizabeth F Churchill. 1999. Design through matchmaking: technology in search of users. *interactions* 6, 2 (1999), 23–31.
4. John Brownlee. 1999. Apple Finally Learns AI Is The New UI. (1999). https://www.fastcodesign.com/3047199/apple-finally-learns-ai-is-the-new-ui
5. Carl DiSalvo, Francine Gemperle, Jodi Forlizzi, and Elliott Montgomery. 2003. The hug: an exploration of robotic form for intimate communication. In *Robot and human interactive communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE international workshop on*. IEEE, 403–408.
6. 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*. ACM, 278–288.
7. Chris Elsden and David S Kirk. 2014. A quantified past: remembering with personal informatics. In *Proceedings of the 2014 companion publication on Designing interactive systems*. ACM, 45–48.
8. Jodi Forlizzi. 2007. How robotic products become social products: an ethnographic study of cleaning in the home. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*. ACM, 129–136.
9. 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
10. Roberto González-Ibánez, Smaranda Muresan, and Nina Wacholder. 2011. Identifying sarcasm in Twitter: a closer look. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2*. Association for Computational Linguistics, 581–586.
11. Weixi Gu, Zheng Yang, Longfei Shangguan, Wei Sun, Kun Jin, and Yunhao Liu. 2014. Intelligent sleep stage mining service with smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 649–660.
12. Xinning Gui, Yubo Kou, Kathleen H Pine, and Yunan Chen. 2017. Managing Uncertainty: Using Social Media for Risk Assessment during a Public Health Crisis. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 4520–4533.
13. Chris Harrison, Desney Tan, and Dan Morris. 2010. Skinput: appropriating the body as an input surface. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 453–462.
14. Patrick Hebron. 2016. *Machine learning for designers*. O'Reilly Media.

15. Chien-Ming Huang and Bilge Mutlu. 2012. Robot behavior toolkit: generating effective social behaviors for robots. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. ACM, 25–32.

16. Kenneth Joseph and Kathleen M Carley. 2015. Culture, Networks, Twitter and foursquare: Testing a Model of Cultural Conversion with Social Media Data.. In *ICWSM*. 626–629.

17. Kenneth Joseph, Wei Wei, Matthew Benigni, and Kathleen M Carley. 2016. A social-event based approach to sentiment analysis of identities and behaviors in text. *The Journal of Mathematical Sociology* 40, 3 (2016), 137–166.

18. Mike Kuniavsky, Elizabeth Churchill, and Molly Wright Steenson. 2017. *The 2017 AAAI Spring Symposium Series Technical Reports: Designing the User Experience of Machine Learning Systems*. Technical Report SS-17-04. Palo Alto, California.

19. Pat Langley. 1999. User modeling in adaptive interface. In *UM99 User Modeling*. Springer, 357–370.

20. Lucian Leahu. 2016. Ontological Surprises: A Relational Perspective on Machine Learning. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. ACM, 182–186.

21. Hosub Lee, Cameron Upright, Steven Eliuk, and Alfred Kobsa. 2016. Personalized object recognition for augmenting human memory. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM, 1054–1061.

22. Yong Liu, Jorge Goncalves, Denzil Ferreira, Simo Hosio, and Vassilis Kostakos. 2014. Identity crisis of ubicomp?: mapping 15 years of the field's development and paradigm change. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 75–86.

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

24. Camille Moussette. 2012. *Simple haptics: Sketching perspectives for the design of haptic interactions*. Ph.D. Dissertation. Umeå Universitet.

25. Syavash Nobarany, Mona Haraty, Brian Fisher, and Steve DiPaola. Epistemically Active Adaptive User Interfaces. (????).

26. Stefano Padilla, Thomas S Methven, David W Corne, and Mike J Chantler. 2014. Hot topics in CHI: trend maps for visualising research. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. ACM, 815–824.

27. Viktoria Pammer. 2015. Mood in the city-data-driven reflection on mood in relation to public spaces.. In *ARTEL@ EC-TEL*. 75–81.

28. Johan Redström, Maria Redström, and Ramia Mazé. 2005. *IT+ textiles*. Edita Publishing Oy.

29. Matteo Roffilli and Alessandro Lomi. 2006. Identifying and classifying social groups: A machine learning approach. *Data Science and Classification* (2006), 149–157.

30. Brandon Rose. Document Clustering with Python. `brandonrose.org/clustering`. (????). `brandonrose.org/clustering`

31. Donald Schön and John Bennett. 1996. Reflective conversation with materials. In *Bringing design to software*. ACM, 171–189.

32. Donald A Schön. 1984. *The reflective practitioner: How professionals think in action*. Vol. 5126. Basic books.

33. Erik Stolterman and Mikael Wiberg. 2010. Concept-driven interaction design research. *Human–Computer Interaction* 25, 2 (2010), 95–118.

34. Wikipedia. List of text mining software. `https://en.wikipedia.org/wiki/List_of_text_mining_software`. (????). `https://en.wikipedia.org/wiki/List_of_text_mining_software`

35. Qianli Xu, Liyuan Li, and Gang Wang. 2013. Designing engagement-aware agents for multiparty conversations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2233–2242.

36. 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*. ACM, 565–576.

37. John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 493–502.

38. 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.