



DEGREE PROJECT IN COMPUTER SCIENCE AND ENGINEERING,
SECOND CYCLE, 30 CREDITS
STOCKHOLM, SWEDEN 2019

Probing User Perceptions on Machine Learning

THEOFRONIA ANDROULAKAKI

**KTH ROYAL INSTITUTE OF TECHNOLOGY
SCHOOL OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE**

English title

Probing user perceptions on Machine Learning

Swedish title

Att Sondera Användares Förståelse av Maskininlärning

Author

Theofronia Androulakaki, androu@kth.se

Submitted for the completion of the KTH program;
Human Computer Interaction, Master of Science in Computer Science and
Engineering

Supervisor: Marie Louise Juul Sondergaard, KTH, School of Electrical
Engineering and Computer Science, Department of Media Technology and
Interaction Design.

Examiner: Kristina Höök, KTH, School of Electrical Engineering and Computer
Science, Department of Media Technology and Interaction Design.

Date of submission: 2019-12-10

ABSTRACT

Machine Learning is a technology that has risen in popularity in the last decade. Designers face difficulties in working with Machine Learning as a design material. In order to help designers to cope with this material, many different approaches have been suggested, from books to insights of experienced designers with Machine Learning. In this research, the focus is on the users' perceptions on Machine Learning and how these could contribute to better design. For this purpose, 10 participants deployed probes to investigate the term Machine Learning. Probes consisted of simple tasks that provoked participants to recognize Machine Learning elements in applications they already use and were deployed with the use of their smart phones. Participants formed personalized perceptions on Machine Learning which varied from creativity in Machine Learning to preoccupations about data use. Based on these findings, suggestions to designers were proposed. Moreover, a secondary research question that emerged was the difficulties the researcher faced while working with probing on Machine Learning user experiences for the specific research.

SAMMANFATTNING

Maskininlärning är ett teknologi som har blivit populär det senaste decenniet. Som designer kan det vara svårt att jobba med maskininlärning som ett "designmaterial". Olika tillvägagångssätt har föreslagits för att hjälpa designers att hantera det här material. I studien som presenteras här läggs fokus på användarens uppfattningar om maskininlärning och hur deras förståelse skulle kunna bidra till bättre design. Tio deltagare använde så kallade "probes" i syfte att undersöka hur vi möter maskininlärning i vardagen. Dessa "probes" bestod av enkla uppgifter som uppmuntrade deltagare att notera och utforska hur maskininlärning ingår som element i tillämpningar som de använder i t ex smartphones. Deltagarna uttryckte sin personliga förståelse och funderingar om maskininlärning, vilket omfattade allt från kreativitet till oro kring hur personliga data används i dessa system. Baserat på en analys av resultaten formulerar vi råd till hur en designer ska utforma interaktion med maskininlärningssystem. Slutligen adderar vi en reflektion om svårigheterna med att använda probes för att studera maskininlärning.

Probing User Perceptions on Machine Learning

Theofronia Androulakaki

KTH Royal Institute of

Technology

Stockholm, Sweden

androu@kth.se

ABSTRACT

Machine Learning has risen in popularity over the last decade. Designers face difficulties in working with machine learning as a design material. Whole books have been written about the difficulties in designing with machine learning, and even experienced designers find it hard to cope with this material. This research will focus on the users' perceptions on machine learning and how these could contribute to better design. For this purpose, 10 participants deployed probes to investigate the term machine learning. Probes consisted of simple tasks that provoked participants to recognize machine learning elements in applications they already use and were deployed with the use of their smart phones. Participants formed personalized perceptions on machine learning which varied from questions on creativity in machine learning to data use concerns. Based on these findings, suggestions to designers were proposed. Moreover, a secondary research question emerged regarding the difficulties the researcher faced while working with probing on machine learning user experiences for the specific research.

Author Keywords

Probes; Machine Learning; User Experience.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI):

INTRODUCTION

Since 2011 McKinsey Global Institute pointed out that machine learning was going to take a leading role in the upcoming technological innovations [19]. Tech companies such as Facebook, Google and Apple have focused on machine learning and “reoriented themselves around” this technology [7]. Machine Learning is on the rise, which can be attributed to the increase of computational power and the extension of data sets [18]. Applications of machine learning can take various forms such as recommender systems, spam filters, ad placement, search result ranking, driverless cars, translation of speech to text, conversational agents and typing autocorrect [6, 8].

The rise of machine learning in the last decade has also brought to the surface various aspects of this technology that are controversial. Artificial Intelligent Agents that are developed with reinforcement learning beat humans in very demanding video games such as StarCraft II [23]. In the future, will such agents become stronger? Another aspect that raises concerns is the collection and use of data on which machine learning technology is based. One

prominent issue with data is how they are used to train a system and the fact that a system is susceptible to the training set data, which can depict the views and the prejudices of the people that provide the data. An example where machine learning was used as a vehicle of prejudice was the training of Microsoft's Artificial Intelligence chatbot which replied to questions with racist messages [22]. Racism, sexism and discrimination may be an actual part of machine learning algorithms that can classify or make recommendations [5].

Machine Learning imposes new challenges and struggles upon designers. According to Dove [8] focus on common practices such as utility and usability will not be sufficient for machine learning technologies. Yang and Banovic [24] argue that design research should “*open up the space for design innovation that uses machine learning*” but in the case of machine learning the design space seems to be much wider in comparison to previous technologies. According to previous research [24, 8], User Experience (UX) designers do not seem ready to handle what a design material such as machine learning can offer. For this reason, several efforts towards addressing the issue have been made. Material and books have been published to help designers approach machine learning technology [13]. Yang, Scuito, Zimmerman, Forlizzi and Steinfeld [26] took a different approach, interviewing UX designers that had experience in machine learning projects and considered that these insights could contribute to the improvement of the education of UX designers.

Although there is a plethora of efforts towards machine learning as design material, users' perception is not taken into consideration in most of them. The end-user is a crucial link to how machine learning uses training data and evolves. According to Holmquist's [14] suggestions for Artificial Intelligence, the control of these systems “*should be designed to allow the sharing of control with the user*”. Since the user is the final recipient of every technological artifact it would be useful for the user's aspect to be taken into account. The difficulty lies not only in designing experiences with machine learning but also in understanding how users respond to these experiences. To that end this research attempts to investigate how users characterize machine learning. This could contribute to better use of machine learning and enhance UX. Machine learning offers personalized experiences, thus users'

perceptions on machine learning can lead to a better design of these personalized experiences.

A way to understand how users characterize these experiences is the use of probes. Cultural Probes is a Research through Design method from which researchers may gain insights on users' experiences [1]. Although previous use of probes in experiences with machine learning concerned mainly Big Data [12], cultural probes as a method disposes of certain characteristics that can match with machine learning such as their capacity to provide insights into users' experience in an engaging and creative way in the field [11]. Moreover, a characteristic of Cultural Probes that is appropriate for machine learning is that they investigate the user's perceptions and personal context [20]. Another element that I consider important is the adaptability and the variations of the Cultural Probes method [1] that allows the researcher to use it with a difficult design material [8] such as machine learning.

In order to examine how users characterize these experiences on machine learning, a task-oriented probing inspired by Gaver's, Dunne's and Pacenti's [11] initial Cultural Probes was used as a method. Probes were deployed by 10 participants. Participants reflected on already existing uses of machine learning using their smartphones in their own space.

The contribution of this paper is twofold. First it focuses on the characterization of machine learning by users which can lead to a tighter engagement of the user in the designing process with machine learning. Second, this work gives prominence to the challenges in designing probes for machine learning. The emergence of the difficulties I encountered during design may contribute to the use of probes for more extensive research on machine learning and its applications.

BACKGROUND

UX and Machine Learning

Machine Learning can be considered as a tool to enhance UX through services and technical advances such as conversational agents, personal assistants and recommender systems. A common practice in technical advances is that soon after they appear, they are followed by designers' innovations which lead to new products and artifacts [24]. However, this is still not the case for machine learning technology [24]. In most cases, the initiation of a new machine learning product comes from the machine learning team and not from the UX team [24]. The role of designers is limited to improving the product or service but Dove attributes this lack of initiation from designers to the gap of proper education in machine learning technology [8]. An additional factor that makes machine learning a difficult material to work with is the fact that there are no prototyping tools to work with machine learning [8]. The difficulties in machine learning also lie to the unpredictability that characterizes it and the difficulty to understand it [14,26]. On the other hand, the fact that

machine learning is more difficult to understand may actually challenge designers and lead to innovation [26].

In an effort to focus on the UX of machine learning [25] suggests 7 clusters of machine learning technical capabilities within HCI based on the bibliography on machine learning and UX. These clusters in parallel with 4 different channels where machine learning adds value to users' lives and may motivate designers in a quest for innovation. The 4 channels include inferences about yourself, the world, optimization and one more general category. Based on these 4 channels [4] designers would be able to use machine learning in order to find unusual user demands that might also apply to larger user groups.

Another factor that could contribute to better UX with machine learning is focusing on the user who is the final receiver of machine learning advances. Moving the conversation from technology to people may provide insights and inspiration for innovation [8]. HCI research is based on a user- approach [25] that takes into consideration the context and tries to understand people [8]. Although interactive machine learning may not only concern end-users but also practitioners, it is a step towards the direction of involving users in machine learning. In Interactive machine learning "*a user or user group iteratively trains a model by selecting, labeling, and/or generating training examples to deliver a desired function*" [9]. Interactive machine learning is an example of how people can contribute to the machine learning process. This paper will focus on people and how their experiences with current machine learning technology may lead to new insights.

Cultural Probes

Cultural Probes is a tool from which researchers may gain interpretations of users' experiences [1]. Gaver *et al.* [11] were the first to introduce the term in order to explore and understand technology in terms of experiences and Cultural Probes in a more unconventional way. The main purpose of Gaver's *et al.* [11] research was to help researchers understand the daily lives of elderly people in 3 different countries. Cultural Probes can include activity packs with material such as postcards, cameras, notes that are based on open, ambiguous questions [17]. Participants of the research are introduced with the probe package and then they are left with it for a specific period of time to process and reflect on the material on their own [11]. This way, Cultural Probes "*provide a way of gathering information about people and their activities*" [10].

Cultural Probes have the ability to transmit the intentions of the researchers in a more relaxed manner and they develop a relationship of engagement between the researchers and participants [2]. The process of completing the questions and the tasks themselves motivate participants to express themselves and give their own interpretations. Another advantage of the use of this research method is that it helps participants think from a different perspective and reshape

preexistent notions [1]. According to Graham, Rouncefield, Gibbs, Vetere and Cheverst [13] probes engage users and elicit deeper responses from them.

Moreover, Cultural Probes differentiate from the questionnaire or interview methods because participants can complete the tasks at their own pace, in the real world and not in the restricted conditions of an experiment in a lab. This fact allows researchers to have more unguarded insights into users' lives and users to show a more realistic aspect of their lives [2].

Using Probes for Machine Learning research

Adaptation of Cultural Probes

Cultural Probes have been used by researchers in various adaptations. The way probes are adopted vary in both scope and diversity [27]. Some of their adaptations include technology, urban, cognitive and other probe variants [1]. Many researchers have chosen the characteristics of probes that matched their research and ignored parameters that were a fundamental part of [11]'s initial probes. As Boehner, Vertesi, Sengers and Dourish [1] mention, the original Cultural Probes should not be considered the "one best way" to inspire design. I have used Gaver's *et al.* Cultural Probes [11] as an inspiration to deploy probes that matched the needs of the particular research.

Compliances and Differentiations to Gaver's Cultural Probes

The main characteristics of the original probes that matched the purpose of the specific research were the playfulness and the element of freedom that they provide to the participants. Playfulness was chosen to elicit participants' interest and motivate them to perform the tasks. As already mentioned, two prominent features about machine learning are its unpredictability and the difficulty to understand. The unpredictability lies in the fact that even if a machine learning system is well trained, it is still, to some extent, drawing its own conclusions from given data [14]. Cultural Probes is a method that allows an extended level of freedom to the participant. It is not limited to the environment of a lab and could enable participants to experiment with a technology that might not always be predictable. It provides them the time and space to perceive and experience machine learning. Cultural Probes may enable users of machine learning to express themselves and be creative. This may lead to deeper insights about user experiences with machine learning.

Machine Learning as technology has certain characteristics that match the nature of Cultural Probes. Machine learning is a ubiquitous technology because it can be found as part of smart phone apps, in devices, in cars, in credit cards, at airports or public spaces. Lately, there is a tendency in HCI to use methods that take advantage of ubiquitous technologies. According to Boehner *et al.* [1] the most prominent one is Cultural Probes. The main part of Artificial Intelligence where probes have been used is games and Big Data [20, 12]. I believe that Cultural Probes

can make a significant contribution to research on machine learning. When probes were introduced as a research method by [11], their main purpose was to help the researchers understand the daily lives of elderly people in three different countries. Machine Learning is present in every aspect of everyday life, from working to sleeping or running. Probes could help users discover machine learning in various applications.

On the other hand, there were characteristics of the original Cultural Probes that were unsuitable for the specific adaptation of the probes. One of them is the materiality that characterizes them. Besides the fact that in this research Cultural Probes' tasks are executed in various places, they are also mainly deployed in the digital rather than the physical space. Machine Learning is a technology that is developed as software. The artifacts that are based on machine learning exist in the digital world. For this reason, the probes were not physical objects but tasks that participants had to perform. All the probes were task-based due to the nature of machine learning but also due to the restriction that the research was for the most part conducted remotely.

Another element that differentiates the probes of the specific research from [11]'s is that they resemble more to technology probes. They do not apply to one specific prototype or artifact and their main purpose is to "track how users respond to and engage with a technology over time" which is the main goal of technology probes as well [1]. Moreover, another differentiation in the current probes that makes them more relevant to technology probes is the fact that I have chosen to analyze them. The original probes are not analyzed, they are used as a means to provoke design inspiration, while technology probes can be analyzed as they are a variation to the original probes [3].

Another divergence to Cultural Probes is the lack of the aesthetics part. In the initial Cultural Probes but also in later adoptions, aesthetics is a dominant aspect of the method [11,21]. The focus on digital space was the main reason aesthetics was not a prominent aspect of these probes. I considered that giving a direction through a specific aesthetic approach would limit participants' perception on the field of machine learning

METHOD

Research through Design

This particular body of work fits into Research through Design. In Research through Design, design is used "to explore the problem and the solution" and thus "we gain knowledge via the act of making" [16]. Probes is an approach to design research because it is mainly built through experimentation and imagination. In this research Gaver's *et al.* Cultural Probes [11] was adapted after a process of reflection and learning. The researcher had initially an approach closer to Gaver's *et al.* method [11]

which was adapted iteratively to explore the different elements of machine learning.

Participants

The main criterion for the prescreening was the availability of the participants for at least 10 to 15 minutes per day for a week. Participants were recruited from the researcher’s social circle.

During the research 10 participants deployed the probes. 5 participants participated from distance in Greece and 5 participants lived in the same city as the researcher. Participants’ demographics are shown in Table 1.

Age	25 to 43, $\mu= 33.7, \sigma = 4.98$
Nationality	Greek (5), Swedish (1), Albanian (2), Belarusian (1) and Turkish (1)
Background	computer science, human-computer interaction, law (2), education, engineering (2), logistics, medicine
Technology literacy	60% of the participants graded themselves at a medium level
	30% graded themselves at a high level
	10% graded themselves at a high level
Education	70% Master’s Degree 20% Doctoral degree 10% Bachelor’s Degree
Familiarity with machine learning	60% were not familiar with the term machine learning
	20% had only heard the term machine learning but it was not clear to them
	20% were familiar with the term

Table 1: Demographics

A pilot study was conducted before the main study to define the final tasks of the probes. The purpose of the pilot study was to examine if the tasks of the probes were appropriate for different levels of technological literacy. In the pilot study, 3 participants deployed the probes for a shorter period of time which lasted 3 days and they devoted around 1-2 hours per day. In the end of the pilot study, participants provided feedback to the researcher in a semi-formal interview in order to improve the design of the probes. The focus was on the difficulties they faced during their experience.

Probes’ toolkit

The probes toolkit included limited physical materials, such as a notepad and 2 colored paper frames, red and green. The green and red frames were only used for task number 7. The frames are depicted in Figure 1. Even for the tasks where physical equipment was used, participants had to take a photo of the artifact and then send it via a messaging app.

All probes consisted of tasks that participants had to perform. Participants that participated remotely had a session with the researcher via video call. During this call, they were informed about the research and they received a document file that included all the tasks. The researcher presented the tasks to each participant and discussed the pace, the means of communication and the duration of the research with them. Participants that participated in person had a session with the participant as well. The only difference was that the aforementioned physical toolkit was delivered to them during the meeting. In the case of the distance participants, the red and green frames were replaced by graphical elements of the application that participants used as shown in Figure 1.



Figure 1. Frames that were used for the 7th task of the probes

Tasks of the probes

The probes were task-oriented and consisted of 10 tasks. Tasks were the main format of the probes in order to be consistent for the participants and also easy to deliver and deployed via a messaging app. The tasks were reformed and adapted after the pilot study. Two of the tasks were based on the diary format to depict the ubiquitous nature of ML. These tasks are summarized in Table 2.

Day	Task	Description
1	Machine Learning Diary	You should write down in which of your interactions with technology within one weekday you consider machine learning.
2a	Phone check	Check your phone apps and take screenshots of the apps that use machine learning technology. Send an emoji that represents the way you feel about the machine learning element in each app. With each screenshot you add a sentence to justify their choice on why it is machine learning
2b	Favorite/worse machine learning app	You should take a photo or a screenshot of an app/program/device that uses machine learning that you really like. Add a sentence to the

		picture to justify your choice. Do the same for the one that you consider the worst machine learning app/program/device.
3	Spot machine learning at home	You are at home. Can you spot machine learning applications/devices? You should send screenshots or photos of these apps. Moreover, attach a gif that represents the way you feel about each specific app. In case you cannot spot such an app you should try to imagine one.
4	Spot machine learning at work	Repeat the same task but this time you are at work.
5	5a Spot machine learning at a park	Repeat the same task but this time you are at a park
6a	Most influential aspect of machine learning	You should take a picture of, draw or describe something/an aspect in our lives that you consider that machine learning would be more influential in the future.
6b	Machine Learning Diary	Machine Learning Diary but this time for a weekend day
7	Ethical or unethical machine learning	You have been given a red frame and a green frame. Use these frames to evaluate machine learning apps. You will take a pic of the app or the use of machine learning in your environment and use the green frame in your photo if you consider it ethical or the red for unethical respectively. When you send the photo, you should use tweet hashtags to elaborate on your choice of color frame.

Table 2: Tasks of the probes

Design Decisions

Throughout this research, my main effort while designing the tasks was to avoid static deployment of the probes. By the term static I mean that the probes are deployed in one and the same physical space. There are many examples of research where the deployment of the probes included a specific physical space or were meant to be deployed at home [21,20]. A main feature of machine learning is its ubiquitous nature. Many uses of machine learning take place in the exterior environment in settings such as

transportation or entertainment. By placing tasks inside and outside of the house, participants are motivated to think not only the apps on their smartphones but also more uses of machine learning and with that in mind, the design of tasks included various physical spaces where the user has to reflect on machine learning applications/devices that are available there. The same task had to be executed in different spaces such as at work, at a park so that the participants have the opportunity to explore the concept of ubiquity of machine learning.

On the other hand, the deployment of the probes in different spaces raised the level of effort and commitment for the participants. This entailed participants spending time on the probes during work hours or during entertainment or at home. In order to maintain the ubiquity level in the design of the tasks and be able to find committed participants, the number and the complexity of the tasks had to be reduced. The tasks mainly included photos and screenshots that participants could take with their smart phones and certain tasks were repeated in different contexts. In this way, participants do not focus on understanding a new task or creating an artifact but spend more time reflecting on the actual task.

A decision that had to be taken at the beginning of the process is whether the probes should be based on machine learning or on one artifact or app on machine learning. The approach that was followed is that the probes would be designed on machine learning as a concept and not be limited to one specific app. I wanted to examine how users perceive machine learning based on its broader scope. This decision led to new challenges because I had to come up with tasks that would focus on how to provoke users to reflect on the different aspects of this technology. For this purpose, the tasks were not limited to smart phone apps, but I also tried to make participants think of machine learning in various situations and contexts.

Another challenge in the design process was creating tasks for a diverse target group. Participants would come from different backgrounds, with different levels of technological literacy. The main challenge lays in the fact that certain participants would be informed about the term machine learning, might even have developed machine learning applications and other participants would never have even heard of the term before. Both categories would have to perform the same tasks and be motivated to do so. Should the ones that were not aware of the term be informed by the moderator? But then are the participants free to establish their own perception on the concept of machine learning? They might be biased by the moderator's approach to the concept. Eventually, the design decision was not to provide any information to the participants on the term. They were going to be free to establish their own approach to it. They were also free to search for the term on the web in case they did not know it or were not sure about it. The design approach on the tasks for this challenge was to keep them

simple and, in a way, playful so that they motivate the users that were familiar with the term and might not feel willing to redefine it. The playful touch in the probes was enhanced by asking participants to add emoji, gifs and colored frames to their answers for the tasks.

A great challenge was to design the probes for a technology that does not have a strong physical aspect, at least for the time being. Participants would have to search in the digital space to recognize and find elements of machine learning in applications. A difficulty that I was faced with was not to distract participants from their pursuit of machine learning in the digital space. My goal was to retain participants' focus on the digital world. The decision that was taken in order to be consistent with this direction, was that all tasks had to be completed by the participants in the digital space and not in a physical one. The digital space that was chosen was a messaging application of the participant's preference. Although a limited physical toolkit was provided to the participants, all the completed tasks had to be sent via the messaging app. The reasoning behind the choice of a messaging app was that participants would not spend time to learn a new app. This amount of time would be more useful to be spent on reflecting on the actual tasks.

Participants had to spend time to reflect on and approach the term machine learning. This is the reason that the period of deployment was scheduled for one week. Of course, one week requires a high level of commitment from the participant's side. To face this challenge and since there was no other type of incentive, I decided to limit the daily time engagement for the participants. This time was defined from 10' to 15' per day and participants were informed about the time demands of the research before agreeing to participate. The reasoning behind the 15'time was that users would be more motivated if they were aware that they did not have to spend too much time. Another design approach to enhance the motivation of participants was to keep the tasks simple and specify the exact day each task had to be executed. Each day consisted of one or two tasks and participants had to send these tasks by the end of the day. In this way, participants kept a daily communication with the moderator and did not lose their focus during the specific week. Moreover, in order to keep the participants focused daily messages were sent to them via the messaging app of their choice. These messages are a more personalized kind of notification to remind the participants of their commitment to the research.

Finally, the use of only one means of response of the participants was also a design decision. All tasks should be designed for messaging apps. The features of these apps such as gifs and emojis had to be exploited. They add the playful tone that was desired but also, they are the main means of expression that users of these applications use daily. Moreover, they allow users a way to articulate how they feel about the perceived experience with machine learning. It should be mentioned that the participants were

already familiar with messaging apps, so an advantage was that they did not have to spend time in learning one.

FINDINGS

The data collected via probing consisted of gifs, pictures, emojis and text responses. They were analyzed by the researcher under the lens of 4 axes. Since the probes were task-based, these axes derived from the tasks and depict the way machine learning perceptions were investigated in this research. The purpose of the probes is to examine user perceptions and not impose a specific design idea. The first axis is whether participants were able to recognize applications that included machine learning elements and how machine learning is characterized based on that. The second axis is reactions towards machine learning applications. The third axis is machine learning in different contexts. Finally, the fourth axis is users' concerns about machine learning and which aspects of it they consider ethical and unethical.

First Axis: Characterization of Machine Learning

The tasks that mainly formed the characterization of machine learning were 2a where participants had to check their smart phones for apps and justify their choice and the 1 which was the machine learning diary. After the deployment of the probes, participants were asked if they had searched for the term on the web and they all answered positively. As an outcome of the web search, the characterization of machine learning was formed on its most popular applications and uses. Among these were recommender systems, natural language processing and automated driving. Moreover, another prominent element in participants' answers on these tasks was the personalized element of machine learning. The main characterizations are presented in this section.

Machine Learning Perceived Based on its Main Applications
Participants managed to recognize various aspects of machine learning across different applications. The most prominent perspective after the analysis of the data was the use of machine learning for recommender systems. Except for one participant, all participants managed to recognize this aspect of machine learning in platforms like YouTube, Facebook, Netflix and Instagram. Although recommender systems on Facebook and YouTube platforms are the most obvious answers when you search the term machine learning on the web, this does not mean that participants did not understand this aspect of machine learning and just used the existing knowledge about it from the web. A quote from a participant who had no previous knowledge of machine learning is indicative of the presence of machine learning in different apps as part of recommender systems *"the most interactive parts seem to me, although I might be wrong, the parts that involve prediction of my items of interest, suggestions of songs videos that might be interesting for me by Spotify, YouTube, Netflix or contextual ads in Google Chrome"*. Other machine learning uses that participants were able to recognize within different applications included speech recognition, autocorrect,

automated and assisted driving, games, natural language processing and spam e-mails. The majority of participants were able to recognize which applications on their phones included machine learning and in particular to name the specific part of the application that used it. One example is that a participant considers voice instructions in his phone assistant in order to make calls while driving or to dictate emails as machine learning.

Machine Learning a Personalized Technology

Participants managed to recognize several applications that offer personalized services and are based on machine learning. Except from personalized ads and recommender systems, participants mentioned applications for walking and fitness. Some screenshots of these applications can be seen in Figure 2. According to the justification of the specific participant the health data application uses machine learning technology because *“it provides personalized advice for my health”*. Another participant suggested a different context for the recommendation part of machine learning where the personalized element is highlighted. The participant mentions *“Imagine: I’m at a new restaurant and don’t know what to order but somehow they know my preferences, what I like and dislike, patterns that even I am not aware of and gives me a list of suggestions according to that”*. Another participant stressed the personalized element in machine learning by providing an example of personalized medicine. The participant mentions *“In fact, once I came across an article where scientists used machine learning to determine specific brain areas that are affected by depression and to predict the response to the treatment based on that.”*



Figure 2. Fitness applications screenshots from the probes.

Misleading perceptions on Machine Learning

Part of the characterization was some misleading perceptions on machine learning. There were some participants that formed a characterization on machine learning which was not close to the actual technology. Moreover, although some participants were able to approach certain characteristics of machine learning, there

were still some misunderstandings regarding the concept. These issues are presented below.

Can a device be considered Machine Learning?

An issue that emerges from the analysis of the data is the oscillation between machine learning as a program or software component and machine learning as a device or part of a device. This was an issue that troubled mostly the participants that did not have previous background in programming but considered themselves technologically literate. The activities did not limit participants’ interpretation of the term. They were free to consider machine learning a device, a program, an app or part of the above. Some participants mentioned devices that use machine learning such as Siri by Apple or self-driving cars and some other mentioned their smartphone or laptop device as a machine learning application.

Machine Learning versus programming.

One of the characterizations of machine learning which was not close to the actual meaning of machine learning is the perception that it is related to everything that includes programming and logic in a device or a program. From this perspective, participants could not distinguish the self-learning part and autonomous element of machine learning in applications. The examples of machine learning they provided included everything that was based on programming and automation. Both participants that formed this account on machine learning had not heard the term before and considered themselves not particularly or averagely technologically literate.

Machine Learning about the use of data but without the learning part.

Another interesting user perspective was one that recognized only the data but not the learning part of machine learning. Participants that belonged to this category were not aware of the term machine learning and according to their answer after the deployment of the probes, they searched for the term on the web. They realized that a prominent part of machine learning is about data but did not get the essence of machine learning which is learning through datasets. The most repeated element in their examples was the personalization feature of these apps. The apps did use their personal data and preferences to provide them a service but did not always use machine learning for this personalization. There was confusion over personalization and machine learning which are not always connected with a prerequisite relation. Participants were able to recognize some popular machine learning applications but they also provided some false positives examples. These examples included services that store their credentials such as chrome auto sign-in or their public transportation card but on their elaborations, it was obvious that they disregarded the learning part of the machine learning process. These participants managed to reflect on only a limited part of the definition which led them to a broader interpretation of machine learning but failed to recognize the artificial intelligence element.

Second Axis: Reactions Towards Machine Learning

Part of tasks 3,4 and 5 was to make participants express how they felt about machine learning apps with the use of gifs. The reactions on the apps varied greatly from excitement and wonder to disappointment and disapproval. Participants used emojis and gifs in order to demonstrate their reactions. Overall the most common reactions were positive such as satisfaction, liking, joy, relief. Nevertheless, negative or neutral reactions also existed. These reactions included skepticism, doubt, disappointment, indifference, even creepiness.

The positive reactions towards machine learning formed around three different categories. These categories are entertainment, usefulness and the fact that it facilitates their work life. Participants expressed their satisfaction, appreciation and awe by using emojis such as thumbs up, winking face, smiling face, smiling face with smiling eyes, cool and gifs in the same context. A participant found Google Ads particularly useful for his job and expressed that with the use of a winking face emoji (see Figure 3 (a)). Another participant expressed his satisfaction and awe for the usefulness of Google Home by sending the gif shown in Figure 3 (b). The other category that gathered many positive reactions by participants was the use of Machine learning for entertainment in popular applications and devices such as Spotify, YouTube and smart TVs. In Figure 4 (a) a participant expressed excitement for the smart TV device with the use of the particular gif. Another participant attached a smiling face with heart-eyes emoji to Spotify and elaborated on that. *“The random online station function provided by Spotify, creates clever playlists based on my interests and previous playbacks”* as seen in Figure 4 (b).

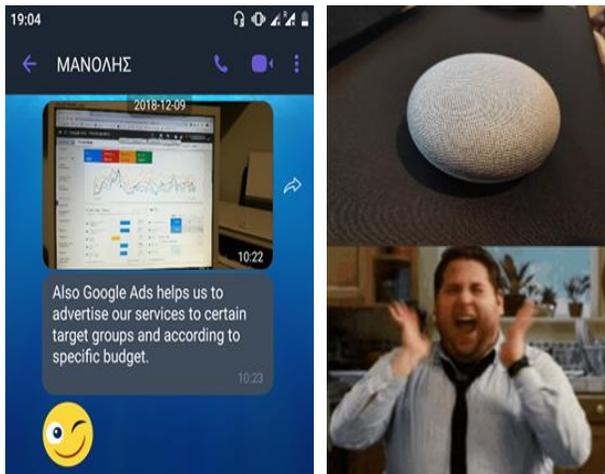


Figure 3 (a) and (b). Screenshots from the probes that were interpreted to convey an experience of astonishment.

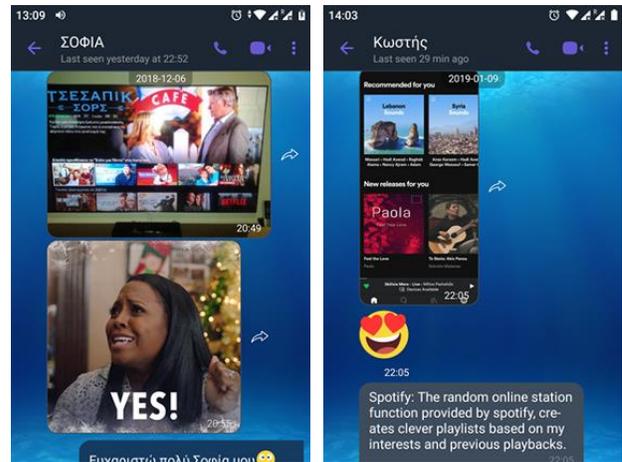


Figure 4 (a) and (b). Screenshots from the probes that were interpreted to convey an experience of awe.

The negative reactions mainly revolved around safety and use of personal data, difficulty of use and the level of intrusiveness of this technology in participants' lives. For the use of data, many participants named Facebook and Instagram applications and the way they might use personal data for commercial purposes. This was prominent in the last task where participants formed their judgment on ethical and unethical aspects of machine learning by providing hashtags. One participant used the hashtags #privacyviolation and #biasedjudgement to show disapproval towards the policy of data used by Facebook. In the same context of reactions for Facebook, another participant used the hashtags #spy, #gdpr, #dataAbuse, #ads, #aggressiveAds and #privacyAbuse.

As already mentioned, the other category of negative reactions included difficulty of use. Some participants had greater expectations for certain applications based on machine learning. These applications included Google Drive or public transportation apps and devices that used Natural Language Processing such as Siri and Google Translate. In Figure 5 two screenshots from the same participant are depicted. In a) the participant uses a thinking face emoji to show skepticism over Google Drive app and its ease of use and in b) disapproval for Google Translate is expressed by using the thumbs down emoji. The participant provides the elaboration that Google Translate is an application that uses machine learning but without much success. Another participant uses the "happy cry" emoji to express displeasure for the Google Translate app and mentions *“It is machine learning, but for me it is not so correct in translating”*. Another participant mentions Siri as an app that definitely *“uses machine learning because it learns and improves by searching based on given commands, but it is not efficient”*.



Figure 5 (a) and (b). Screenshots from the probes that were interpreted to convey negative reactions.

The third notion around which participants expressed negative feelings towards machine learning was the level of intrusiveness in their lives. Among the participants, there were few that expressed frustration and disagreement on the way machine learning can be used. They feel machine learning can be an intrusion in their lives on certain occasions. For two participants this included the recommendation part of popular applications. One participant commented on the recommendation part of messenger messaging app when using it because “*it gives the feeling of being watched which isn’t nice at all*”. Another participant mentioned personalized ads in Messenger “*you have a conversation about holidays and you mention booking.com, the next day an advertisement about hotels appears*”. The marketing emails sent by Amazon based on previous searches were considered invasive by one more participant.

Third Axis: Machine Learning Under Different Contexts

Tasks 3,4 and 5 prompted participants to reflect on machine learning under different contexts. Despite the fact that for the tasks of the first 2 days they focused on the most popular uses of ML, participants reflected more in different aspects of ML for the tasks where the context varied ; more importantly, they made suggestions that could be useful for ML designing in applications concerning their field of expertise.

Technology that is Plastic and Allows Creativity

In this approach, the users were able to view machine learning as a technology that allowed them to be creative, improve their professional life or solve everyday problems. Participants consider machine learning as a kind of “plastic material” technology that can be used in various contexts and take different forms and shapes. Participants were able to spot machine learning in various spaces and devices and recognize the ubiquitousness and polymorphism of this technology. Moreover, the plasticity of the technology allowed them to imagine future uses of it. In the task where participants were asked to provide examples of machine learning related to their job, a participant who is a biology expert suggested a specific image recognition software that

would incorporate functions desired for biology lab results and today does not exist. For the task where participants had to provide examples of machine learning used in a park, the majority of them used their creativity and suggested new uses. These examples included apps about interactive maps providing information on the actual state of the plants and ratings from the visitors or even drones that would use image recognition at a park.

Fourth Axis: Users’ concerns regarding Machine Learning

The seventh task of the probes wanted participants to express their opinions on ethical or unethical aspects of machine learning. Based on the responses from this task and participants’ comments on task 2b where they had to justify their choice on their least liked machine learning applications, two main areas of concern emerged.

Do It for me or Do It Myself? Where do I want it to stop?

A theme that emerged was the level that technology takes action on behalf of the user. Some participants mentioned the fact that machine learning has taken over certain tasks that they would prefer to perform themselves. For example, a participant mentioned “*I prefer to use my brain for reminders and orientation demands instead of using a GPS destinator*”. Another participant mentioned Siri Apple device and the implications of giving orders to a system inside your house. Some participants consider that this kind of intelligence sometimes does not help the user and prevents them from using their own abilities. On the other hand, other participants mentioned Siri and personal assistants as a means that helps them in their everyday life and saves them time. This issue belongs to the known dilemma in intelligent technologies which is summarized as “do it for me” or “do it myself” [8]. A representative example from a participant is on the recommender system of Netflix and the lack of freedom that this might cause. “*However, I was happier about the machine learning element before when I didn’t really know what I wanted to watch. Now I feel like it is too good somehow and doesn’t give me enough freedom to explore and that the same things keep popping up all the time*”. Participants recognize the benefits of machine learning such as ease of use, reminders, recommendations but are skeptical about the level of automation in their lives and the restrictions this may imply.

Visibility of Machine Learning

Another interesting perspective from a participant who was familiar with the term was that the users do not realize that machine learning uses the data they provide. The process of machine learning takes place in the background and the user has as little sense as possible. The participant questions whether machine learning should be visible to the user or not. How informed should we be on how a machine learning service is provided to us and how it uses our data? Is it really necessary for this technology to be explained?

On the other hand, other participants expressed the opinion that the way machine learning uses their preferences and makes suggestions was too obvious and too static. They considered that machine learning establishes a profile based on the user's preference and this does not mean that it matches the actual preferences of the user. The participant that shared this concern mentions "*I think that once they are based on a certain choice that I made at one certain moment my part here is done and the rest is decided for me and I'm not influencing it anymore very difficult to explain but I think this could be the reason why this kind of suggestions from the apps do not always match my interests*".

DISCUSSION

Suggestions for designers

Machine learning is a new design material with which designers are already working and probably will be working extensively in the future. Machine Learning is based on intelligence and designers need to figure out how to work with it. According to Holmquist [15], this will cause a level of difficulty because similarly to the transition from paper to screen, designers need to develop new design practices.

The perceptions users form on machine learning could potentially lead designers to better future uses of this technology and more insightful designs. In what way can the characterization of machine learning by users be useful to the design process? Based on the findings of this research, they can be summarized into the 3 following design suggestions.

User Control over Machine Learning

Participants were worried about the level of control they have over machine learning applications. The element of intelligence in the apps and devices makes them skeptical regarding the limits of their freedom and choice. Users feel threatened by the possibility of an app taking control over them and making decisions on their behalf. Designers need to take this concern into consideration and design machine learning in apps and devices in a way that makes the user feel secure. This means combining the fact that machine learning has the potential to self-act and at the same time reassure users' need for control. More flexible designs that distribute the tasks between the user and machine learning could allow the sense of freedom to users. Holmquist [15] suggests interface elements that inform the user when a system makes a decision, on what the decision was based and revert the decision in case of disagreement.

Machine Learning for non-experts

One interesting finding of this research is that the user is willing and able to have an active role in shaping machine learning services and technologies. The majority of the participants that were non-experts managed to understand the main aspects of machine learning and suggest applications that could be useful in their field of expertise. This outcome stresses the role of non-experts in machine learning technology and comes in agreement with

Holmquist's opinion [15]. Holmquist [15] believes that in the future, Artificial Intelligence technologies such as neural networks, will be used and trained by non-experts with the use of well-defined interfaces. Designers can focus on how to create interfaces that will use complex Artificial Intelligence technology but will be used by non-experts and these systems will support their training. This suggests that they could use elements of abstraction in the design of interfaces in order to hide the complexity of such technologies and make it more usable.

Ethical and privacy issues to take under consideration

A prominent theme in the findings is the ethical and privacy restrictions related to machine learning. Participants were concerned about the safety personal data and the implications this may have in the future. These concerns were mainly focused on applications such as Facebook, Instagram and Amazon. In particular, there was distress about the way their data are used for advertisement purposes. Some participants were annoyed at the way their habits and personal information are used to create a profile of them. They do not consider that this profile necessarily expresses them. Another factor that caused some participants' concern was the face and voice recognition application elements in Facebook and devices like Siri. They are not sure how this type of personal data can be used and if they are safe to provide it.

Although participants are informed about cookies and data according to the recent General Data Protection Regulation that came in practice by the European Union in May 2018 this does not seem to reassure them. This concern can be valid since many issues with data in machine learning have not been clarified. For example, what happens with the material that machine learning produces based on the users' data? Who owns it? At this point, designers have to deal with the uncertainty and at the same time earn users' trust. An approach could be a design that is open about data issues that might arise and also informs the user about how their data is used and possible implications.

Design Challenges for Probes When Investigating User Experiences with Machine Learning

Ubiquity of Machine Learning

Although I believe that the nature of machine learning is appropriate for research with probes at the same time probing machine learning poses some significant challenges. One of the main characteristics of machine learning is its ubiquity. Machine Learning can be part of other applications. It can be visible, or not, lying in the background. This means that in my probes design I should take into consideration all these different forms of machine learning and try to provoke participants to think of these different aspects and uses. In this particular research, this was attempted by placing the tasks in various spaces and provoke participants to imagine uses of machine learning other than those apps of their smart phones. Participants were able to recognize different aspects of machine learning to a certain degree. The fact that due to the nature of the

probes they did not get any feedback from the researcher, led some of them to repeat their answers and not think various uses of machine learning. Although the context for the specific task changed, certain participants stuck to the same applications of machine learning such as recommendation systems. This characteristic of machine learning should be taken into consideration when designing probes. It is of great difficulty to think of tasks that help the participants think in this direction; tasks that would help them conceive new aspects of ML unknown to them so far or not thought of, and at the same time tasks that would not limit the concept to applications in their smart phones or in a specific physical space.

Machine Learning Is not Only Recommendation Systems

A challenge that became more obvious to me after the analysis of the results is that the design of the probes should provoke insights not only on the recommendation systems use of machine learning but also on other aspects of the concept. I should have focused more on the fact that machine learning has various uses and applications. The most well-known use of machine learning is recommendation systems. For a participant that has no previous knowledge of machine learning and will search for the term on the web, this is the first use of machine learning that will appear. This was not taken into consideration in the design of the probes and the result was that some participants mainly focused on recommendation systems in various platforms and did not seek less known uses of machine learning. The tasks of the probes could have included more of the variety of machine learning apps so that participants could characterize it in a broader range.

Focus on a Specific Application of Machine Learning

A similar challenge was the choice not to focus on a specific machine learning artifact, device or application. The purpose of this research was to approach the general concept of machine learning. This demanded a lot of effort from the part of the participants in order to think about different applications and different uses of machine learning. Focusing on one artifact can be easier during the design of the probes. Moreover, it is more convenient for the participants. They can focus more on a specific app and not try to spot different uses of machine learning in various spaces and contexts.

Level of Detail in the Directions

According to Moser, Fuchsberger and Tscheligi [20] an important element for the successful deployment of the probes is not the open-endedness of material but how much detail exists in the instructions of the probes. The degree of detail makes the task clearer and easier to understand [20]. In this research, an effort towards this direction was attempted. The tasks were broken down into sub-tasks in order to provide clear directions to the participants. General directions were provided for the messaging app as a tool of communication with the moderator and specific instructions for each task. Participants did not face any problems following the specific instructions for each task, but they

faced some regarding the general directions. Two of them ignored the direction about using an app of their choice and started sending the material via e-mail. This can be attributed to the fact that the specific participants were not particularly fond of messaging apps or that the way that instructions were provided was not clear. The general instructions were placed in the file after the specific tasks and maybe this was a factor that led the reader to ignore them.

CONCLUSION

This research used probing to examine user perceptions on machine learning. The deployment of the probes was conducted entirely through a messaging application. 10 participants had to reflect on the meaning of machine learning by performing various tasks for a period of one week.

I think that this research was able to showcase personalized perceptions on the experiences of the participants with machine learning. Through a positive and creative approach, participants expressed their opinion on machine learning, were critical about it and managed to realize different aspects of a technology they were already using. Personalized views on machine learning varied from plasticity and creativity elicited by machine learning to worries about the level of control and security. The majority of participants were able to approach the most common uses of machine learning. Some of them imagined new applications based on machine learning that could be useful to them or to their profession. This element of cooperation with the users could lead to better and more usable applications of machine learning. According to Holmquist [15] apps on Artificial Intelligence in the future will be co-controlled by users and machine learning technology. Knowing the individualized perception of the users and the negative preconceptions they form on machine learning could contribute towards [15]'s direction and provide a better co-control in the future. Future research could shed more light to the role of the user in the advancement of UX based on machine learning.

A sub-question that emerged during the design and deployment of the probes is the design challenges that the researcher faced during this process. These design challenges point out difficulties that designers could take into consideration in future research with probing on machine learning.

REFERENCES

1. Kirsten Boehner, Janet Vertesi, Phoebe Sengers, and Paul Dourish. 2007. How HCI interprets the probes. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems (CHI '07)*, 1077-1086.
2. Andy Boucher, Dean Brown, Liliana Ovalle, Andy Sheen, Mike Vanis, William Odom, Doenja Oogjes and William Gaver, W. 2018. TaskCam: Designing and Testing an Open Tool for Cultural

- Probes Studies. In *Proceedings of the 2018 Conference on Human Factors in Computing Systems* (CHI '18), 71-83.
3. Barry Brown, Moira McGregor, and Eric Laurier. 2013. iPhone in vivo: video analysis of mobile device use. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, (CHI '13), 1031-1040.
 4. Kim Carmona, Erin Finley, and Meng Li. 2018. *The Relationship Between User Experience and Machine Learning*. Available at SSRN 3173932.
 5. Kate Crawford. 2016. Artificial intelligence's white guy problem. Retrieved October 12, 2018 from <https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html>
 6. Pedro Domingos. 2012. A few useful things to know about machine learning. *Communications of the ACM* 55, 10: 78-87.
 7. Catherine Dong. 2017. The evolution of machine learning. Retrieved September 30, 2018 from <https://techcrunch.com/2017/08/08/the-evolution-of-machine-learning/>
 8. 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), 278-288.
 9. John J. Dudley and Per Ola Kristensson. 2018. A Review of User Interface Design for Interactive Machine Learning. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 8, 2: 1-37.
 10. Gerry Gaffney. 2006. What is a Cultural Probe?. *Information & Design. Info Design [Em.*
 11. Bill Gaver, Tony Dunne, and Elena Pacenti. 1999. Design: cultural probes. *Interactions* 6, 1: 21-29.
 12. Gaver, William, Andy Boucher, Nadine Jarvis, David Cameron, Mark Hauenstein, Sarah Pennington, John Bowers, James Pike, Robin Beitra, and Liliana Ovalle. 2016. The Datacatcher: batch deployment and documentation of 130 location-aware, mobile devices that put sociopolitically-relevant big data in people's hands: polyphonic interpretation at scale. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 1597-1607.
 13. Connor Graham, Mark Rouncefield, Martin Gibbs, Frank Vetere, and Keith Cheverst. 2007. How probes work. In *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces* (OzCHI '07), 29-37.
 14. Patrick Hebron. 2016. *Machine learning for designers*. O'Reilly Media.
 15. Lars Erik Holmquist. 2017. Intelligence on tap: artificial intelligence as a new design material. *Interactions* 24, 4: 28-33.
 16. Kristina Höök, Martin Jonsson, Anna Ståhl, and Johanna Mercurio. 2016. Somaesthetic appreciation design. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 3131-3142.
 17. Filip Lange-Nielsen, Xavier Vijay Lafont, Benjamin Cassar, and Rilla Khaled. 2012. Involving players earlier in the game design process using cultural probes. In *Proceedings of the 4th International Conference on Fun and Games* (FnG '12), 45-54.
 18. Panos Louridas and Christof Ebert. 2016. Machine Learning. *IEEE Software* 33, 5: 110-115.
 19. James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela H. Byers. 2011. *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.
 20. Christiane Moser, Verena Fuchsberger, and Manfred Tscheligi. 2011. Using probes to create child personas for games. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology* (ACE '11), 8 pages.
 21. Doenja Oogjes, William Odom, and Pete Fung. 2018. Designing for another Home: Expanding and Speculating on Different Forms of Domestic Life. In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018* (DIS '18), 313-326.
 22. Rob Price. 2016. Microsoft is deleting its AI chatbot's incredibly racist tweets. Retrieved October 12, 2018 from <https://img.sauf.ca/pictures/2016-03-24/d360716e3199095063ebd4749b78fc4c.pdf>
 23. Oriol Vinyals, Timo Ewalds, Sergey Bartunov, Petko Georgiev, Alexander S. Vezhnevets, Michelle Yeo, ... and John Quan. 2017. *StarCraft II: A New Challenge for Reinforcement Learning*. arXiv:1708.04782.
 24. 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), 565-576.

25. Qian Yang, Nikola Banovic, and John Zimmerman. 2018. Mapping Machine Learning Advances from HCI Research to Reveal Starting Places for Design Innovation. In *Proceedings of the 2018 Conference on Human Factors in Computing Systems (CHI '18)*, 130.
26. Qian Yang, Alex Scuito, John Zimmerman, Jodi Forlizzi, and Aaron Steinfeld. 2018. Investigating How Experienced UX Designers Effectively Work with Machine Learning. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*, 585-596.
27. Jayne Wallace, John McCarthy, Peter C.Wright, and Patrick Olivier. 2013. Making design probes work. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*, 3441-3450.

TRITA -EECS-EX-2019:821