

Flatpack ML: How to support designers in creating a new generation of customizable machine learning applications

Marcus Winter ^[0000-0001-6603-325X] and Phil Jackson ^[0000-0003-4953-1112]

School of Computing, Engineering and Mathematics, University of Brighton, Brighton, UK
{marcus.winter, prj11}@brighton.ac.uk

Abstract. This paper examines how designers can be supported in creating a new generation of interactive machine learning (ML) applications that run locally on consumer-level hardware and can be trained and customized by end-users for their specific context and use case. It delineates the proposed applications against research into Interactive Machine Learning and Machine Teaching, examines the challenges designers face in these contexts and their relevance for designing the proposed new applications, and reports on the findings of a survey exploring designers' interest in ML, their understanding of ML capabilities and concepts, and their preferences for learning about ML. Based on findings from the literature and the survey, the paper identifies three overlapping research challenges in supporting designers to ideate, design and prototype the proposed ML applications.

Keywords: Machine Learning, Customization, User Experience, Design.

1 Introduction

Many current applications of Machine Learning (ML) run on cloud infrastructure and are driven by business needs, based on large data sets about users' demographics, preferences and behaviors. Our interaction with this kind of ML is usually implicit, indirect and mostly unaware: based on data that is accumulated in the background, often in a different physical, digital or temporal context than the one in which it is used to predict our needs and preferences. When these systems produce wrong or unexpected results, users often discover that they have limited options to find out how decisions are made or what information the system holds on them, and that it might be difficult or even impossible for them to correct misleading information, leading to well documented user experience problems related to transparency, control and trust.

By contrast, explicit interaction with ML that runs locally and is driven primarily by user needs to address problems in their specific context is comparatively rare and currently mostly limited to research prototypes addressing the needs of domain experts in specific contexts (e.g. [3, 4, 7, 17, 18, 24, 31]), mainly due to the complexities of designing and training ML models and their requirements for substantial computing power. Recent developments, however, are eroding these barriers, with specialized hardware emerging to run ML models locally in a performant and energy-efficient manner [8, 20, 23, 38], a host of open repositories offering free, high-quality, pre-trained

ML models for a range of problems [33-35, 42] and several ML frameworks allowing these models to be used on consumer-level hardware [37, 40, 41, 43]. While the ready availability of pre-trained models and high-level development tools shifts the focus from ML research to ML applications research, the ability to run models on edge devices such as browsers, mobile phones and single-board computers broadens deployment options and enables ML applications to take advantage of device sensors and user interaction. Combined with transfer learning approaches, which drastically reduce training times and the amount of required training data by retaining pre-trained models' previous learning derived from large general data sets [48], this enables a new generation of interactive ML applications that run locally on consumer hardware and can be trained and customized by end-users for their specific context and use case. Reflecting that fact that these applications typically rely on off-the-shelf pre-trained base models for specific problem classes and require a setup process by end-users to train and customize them in the target environment, we use the term *Flatpack ML* in reference to a popular distribution format for mass furniture that is assembled by customers at home.

This paper examines what designers need to know to ideate, design and prototype this new generation of ML applications and how they can be supported in making them relevant, usable and meaningful for users. Its main contributions are to (i) situate and delineate the proposed new generation of local, end-user customizable ML applications, (ii) provide a snapshot of designers' current understanding of ML concepts and capabilities, and (iii) identify a number of challenges that can inform the research agenda for this emerging class of ML application.

2 Related work

Situated broadly in the field of Human-Centered Machine Learning (HCML) [19], the work relates to various research efforts in the literature adopting a more user-centered perspective on ML. Regarding the user experience of ML applications and services, it draws on literature discussing the transparency, explainability and debuggability of ML systems [1, 27, 16]; regarding users' direct interaction with ML, the work draws on literature on Interactive Machine Learning (IML); and regarding the shift in focus from optimizing the accuracy of learning algorithms towards improving the efficacy of the people who "teach" algorithms how to behave [46], the research draws on literature on Machine Teaching (MT). Finally, the work draws on a range of previous efforts exploring end-users' understanding of ML concepts [16, 46] and the role of designers in creating ML applications [45, 47].

2.1 Interactive Machine Learning and Machine Teaching

The literature offers various definitions of Interactive Machine Learning (IML) in an effort to distinguish it from traditional ML. In [15], IML is defined as "an interaction paradigm in which a user or user group iteratively builds and refines a mathematical model to describe a concept" (p.4), without specifying the type of user providing the input or the context in which the machine learning happens. In [2] IML is characterized

by "rapid, focused, and incremental interaction cycles", which "facilitate end-user involvement in the machine-learning process" (p.108). As this involves a tight coupling between user and system, it "necessitates an increased focus on studying how users can effectively influence the machine-learning system and how the learning system can appropriately influence the users" (p.108). The clear focus on end-users in this definition leads the authors to call for more research into the interaction between end-user and ML system, and how the process gradually advances learning and understanding by both end-user and ML system - an aspect highly relevant to ML systems that can be trained and customized by end-users as it indicates the need to account for this development in the interface and interaction design. A third definition [36] proposes that IML "focuses on methods that empower domain experts to control and direct machine learning tools from within the deployed environment, whereas traditional machine learning does this in the development environment" (p.12). Here the focus shifts from development environment to the target environment in which the ML system is used. An important implication is that IML application design needs to support users to control and direct the IML process in-situ, shifting the focus to the support system enabling users to train and optimize ML applications.

Further emphasizing this shift in context and perspective are definitions of the field of Machine Teaching (MT). For instance, [39] see ML and MT as related but distinct research fields, where "Machine Learning research aims at making the learner better by improving ML algorithms", while "Machine teaching research aims at making the teacher more productive at building machine learning models" (p.3). An almost identical interpretation is offered in [46], which states that ML focuses on improving the accuracy of learners (i.e. learning algorithms), while MT focuses on improving the efficacy of teachers (i.e. people who interact with data and "teach" algorithms how to behave). Overall, [46] asserts that MT refers to a human-centered perspective to the process of ML, while [39] point out that MT research has "more in common with programming and human-computer interaction than with machine learning" (p.3).

Considering these definitions of IML and MT, we can delineate the problem space for Flatpack ML application design as being concerned with:

- end-users training and customizing ML applications for their specific context
- pre-trained base models and transfer learning techniques rather than custom models
- consumer-level hardware rather than high-performance professional equipment
- personal use environments rather than work environments or lab environments

Given that designers have well-developed methods to research and design for end-users and their use-contexts, this paper focuses in particular on designers' understanding of ML capabilities and concepts necessary to design the user experience and interaction of end-users training and customizing ML applications for their specific context.

2.2 Challenges for designers working with Machine Learning

According to [46] there are three categories of people who build ML systems, including experts, intermediate users and amateurs. As these categories have different perspec-

tives and operate at different levels of ML expertise, they face different sets of challenges. While ML experts are of limited interest in the context of this paper, both intermediate users and amateurs need to be considered: designers who ideate and prototype customizable ML applications arguably fall into the intermediate user category, while end-users who train and customize ML applications for their specific context can be assigned to the amateur category. Both of these perspectives are relevant when designing customizable ML applications. With regard to the former, the authors [46] identified many challenges including (c1) understanding the limits of what an ML system can learn, (c2) exploring alternative formulations of an ML problem and (c3) evaluating the performance of ML models in the context of its specific application. Regarding the latter, identified challenges include (c4) translating real-world problems into achievable learning tasks, (c5) improving the design of a ML model rather than trying to improve performance by adding more training data, and (c6) assessing model performance and accounting for bias and overfitting. There is a considerable overlap in the identified challenges for intermediate users and amateurs, with (c1, c2, c4) all relating to understanding ML capabilities and being able to map them to practical uses and (c3, c6) relating to the ability to assess how well a trained model does in a specific environment.

Focusing in particular on problems faced by UX designers, [47] identifies six distinct challenges in three case studies involving attempts to integrate UX and ML aspects in system development. They include that (c7) UX is often an afterthought rather than being thought of throughout the project, (c8) UX designs are constrained by available data, (c9) designers struggle to proactively work with data scientists, (c10) designers have limited access to competent data scientists, (c11) ML is difficult to prototype and does not integrate well with designers' *fail early fail often* approach, and (c12) designers struggle to understand ML's capabilities in the context of UX. Reflecting the context in which these challenges were identified, several relate to development methodology (c11) and communication in mixed project teams (c7, c9, c10) while others are more fundamental in nature: (c8) highlights a mismatch between ML as a design material that cannot be easily prototyped and design methodology that relies on quick iterations and successive improvement, while (c12) relates to understanding ML capabilities and applying them to problems in a specific context.

Based on a survey of 51 UX practitioners, most of whom had previous experience in collaborating with ML experts, [13] found that designers (c13) have fundamental difficulties in grasping ML as a design material due to the differences in human and machine perspectives (c14) lack suitable prototyping tools to experience ML in realistic situations, and (c15) often have misconceptions about ML capabilities and limitations. All of these are variations of previously identified challenges, with (c13, c15) relating to conceptual knowledge about ML and its possible applications, and (c14) relating to difficulties in prototyping ML in tandem with evolving design iterations.

Interestingly, [15] discuss not only problems that users face when interacting with IML systems but also problems that IML systems face when interacting with users. Among the former, they list (c16) the open-ended character of training and the related problem of assessing model accuracy and knowing when to stop, (c17) the co-adaptive character of interaction with an evolving model where both user and model influence

each other's behavior and (c18) that users associate computers with precision and struggle to interpret prediction uncertainty. Among the latter, they list that (c19) users can be imprecise and inconsistent, and that (c20) there is a degree of uncertainty when relating user input to user intent. As this aspect can be exacerbated by non-experts' lack of understanding of ML concepts and terminology (see c1, c3, c5, c6, c14, c16, c18), it suggests a need for a shared language for ML interaction not derived from the ML expert community but developed together with users from an HCML perspective.

Other challenges mentioned in IML literature cited in [2] include that (c21) users, unlike machines, easily get tired of repetitive tasks in IML workflows, (c22) users are biased towards giving more positive than negative feedback, (c23) users prefer to give guidance to ML systems rather than feedback on past decisions and (c24) users naturally provide a wide variety of input types to an ML system. Just like (c19, c20) above, all of these challenges can be seen as problems that an IML system faces when interacting with users, reinforcing the view that designers need to balance the needs of both users and ML systems to support effective interaction.

Table 1. Challenge categories and their relevance to Flatpack ML application designers.

Challenge category	Relevance to Flatpack ML
Conceptual knowledge of ML: capabilities and limitations of ML and how they can be mapped to problems in a specific context (c1, c2, c4, c12, c13, c15).	Relevant in particular during the ideation stage for new consumer-level applications for ML.
Technical knowledge of ML: modifying the design of a model or tuning its hyper-parameters to improve performance (c5).	Contingently relevant as designers typically use off-the-shelf pre-trained base models and do not engage in model design or optimization.
Operational knowledge of ML: evaluating model performance for a specific context and determining when to stop training (c3, c6, c16).	Relevant as it informs designers' decisions on how to guide users through the training and customization process.
Development culture: ensuring that UX aspects are considered throughout and communicating with data scientists and ML experts (c7, c9, c10).	Contingently relevant as projects are typically design-led, but might still require ML experts to engineer transfer learning mechanisms and custom models.
Design methodology: prototyping ML aspects of an application in tandem with evolving design iterations (c8, c11, c14)	Contingently relevant as the issue essentially becomes a design problem when training and customization is left to users.
Interaction design: supporting effective interaction between ML systems and user while accounting for the evolving nature of that interaction and for human biases and preferences (c17, c18, c19, c20, c21, c22, c23, c24).	Relevant given the challenges of guiding users through the initial training process, supporting them in selecting suitable training data and assessing model performance.

There are many overlaps between the challenges discussed in the literature, with several distinct challenge categories emerging. However, not all of the challenges discussed in the literature are equally relevant to Flatpack ML application design. The mapping in Table 1 shows conceptual and operational knowledge of ML and interaction design challenges in ML applications all being highly relevant as designers take the lead in ideating ML applications and design interactions and interfaces supporting end-users in training and customizing these applications, while challenges relating to technical knowledge of ML, development culture and design methodology are less relevant in design-led projects using off-the-shelf pre-trained base models.

2.3 Making Machine Learning more accessible to designers

Making ML more accessible to designers is not a new idea but has been subject to discussion in both popular and academic literature for some time. There are numerous online publications and books around the topic (e.g. [10, 11, 14, 22, 28, 29]) and an evolving body of academic research exploring the problem from various angles (e.g. [13, 15, 30, 45-47]). According to [45] efforts fall broadly into two categories: one taking a didactic approach based on the assumption that a better conceptual and technical understanding of ML helps designers to make better use of it, and the other taking an experiential approach based on the assumption that sensitizing designers to ML capabilities and design possibilities through hands-on engagement is more fruitful. In addition to this dichotomy of approaches, there is also a dichotomy of uses of ML: those discussing ML as a creative tool for designers that opens up new possibilities and disrupts traditional design processes, and those discussing ML as a design material that can improve, augment or allow for completely new applications and services. The work discussed in this paper focuses primarily on ML as a design material, reflecting the key idea behind Flatpack ML as a design-driven effort to develop customizable products that use ML to address user needs in a specific context.

Both the idea of sensitizing designers to the capabilities of ML through hands-on engagement [45] and the call for “research on new tools and techniques that make it easier for designers to ideate, sketch and prototype what new forms ML might take, and new contexts in which ML might be appropriate” [13, p.279] support the case for experiential learning rather than more didactic methods. Recent research exemplifying this experiential approach is described in [30], which discusses ObjectResponder as a tool for designers to quickly prototype ML applications and test them in their context of use. The tool takes the form of a smartphone application using an online computer vision service with live camera images. It provides designers with a simple interface to associate recognized objects with spoken responses using text-to-speech technology. An evaluation of the tool with design professionals found that looking at specific use contexts from a ML perspective helped designers to think more broadly about ideas, and that immediately testing them in the target environment provided valuable learning opportunities about the capabilities and limitations of ML.

2.4 Conclusions

The literature helps to distinguish the proposed Flatpack ML applications from related efforts such as IML and MT through its focus on training and customization by end-users, transfer learning techniques, consumer-level hardware and personal use environments. It also helps to inform research exploring how designers can be supported in ideating, designing and prototyping these applications. Flatpack ML application design poses a particular set of challenges with respect to communicating ML concepts to end-users and guiding them through the training and customization process. Tackling these challenges will require both conceptual and operational understanding of ML as well as suitable prototyping tools and techniques. Given that established design practice typically involves the iterative development of design artefacts, experiential approaches helping designers to develop their knowledge skills through active experimentation with ML seem a promising way forward.

3 Survey: Designers' understanding of ML

In order to complement and further qualify findings from the literature, we carried out a survey exploring designers' knowledge, understanding and interest in ML. Besides informing our work to make ML more accessible to designers, the survey also aimed to establish a baseline of current understanding and use of ML among designers, against which the outcomes of future interventions can be measured.

3.1 Instrument

The survey used an online questionnaire structured into seven sections:

1. An introduction to the research, simple definition of ML, contact details of the investigators and a question for participants to confirm their consent to take part.
2. Questions relating to participants' previous experience of designing ML applications, awareness of ML in everyday applications and ideas for novel ML applications.
3. Questions exploring participants' understanding of ML capabilities, focusing in particular on capabilities relevant to Flatpack ML due to their generality, availability of pre-trained models and suitability to run on edge devices.
4. Questions asking participants to self-rate their understanding of various conceptual and operational aspects of ML related to training, performance, uncertainty and bias.
5. Questions exploring how useful designers would find various approaches to support them in designing and prototyping ML applications.
6. Demographic questions relating to participants' age, gender, job or course title and self-rated design and programming skills.
7. A final open question asking for any other thoughts on the topic.

The questionnaire design involved a member of staff with a design background to ensure that questions made sense from a design perspective and used appropriate language for the target group, resulting in several iterations refining the introduction as well as

the wording and sequence of questions. It then was piloted and discussed post-hoc with two designers (not included in the sample), resulting in further changes to wordings and the removal of six non-essential questions. The final version of the questionnaire takes 10-15 minutes to complete.

3.2 Methodology

Participants were recruited through calls posted on popular discussion groups for various disciplines, emailed to professional networks and personal contacts of the investigators, and distributed to design students via lecturers at the University of Brighton. The questionnaire was open for a duration of three weeks and during that time was completed by 102 participants.

Answers to closed questions were analyzed quantitatively while answers to open questions were analyzed in an emergent coding process described in [32], involving first a data reduction step, where responses were coded to identify emerging themes, and then a data visualization step, where the coded data was structured into key themes for interpretation. In order to mitigate investigator bias when interpreting responses and identifying themes, both data reduction and data visualization were carried out independently by two researchers before being discussed and synthesized in a collaborative interpretation session.

Given that the survey aims to broadly inform future research across design disciplines, no attempt was made at this stage to differentiate answers between design domain or participant demographics. For the same reason, interrelations between responses to different answers were not examined at this stage, although they might be interesting to explore in future work.

3.3 Ethical considerations

Ethical considerations concerning the involvement of professionals and students in the survey are based on Anderson's guidelines for using volunteers in research projects [6]. Participants were not pressurized to participate in the survey and were informed about the context and purpose of the research. The introduction to the survey clarified that the data collection and analysis is anonymous; collected data is used only in the context of the research and not made available to third parties; and that participants can withdraw at any time without giving a reason.

With regard to student participants, calls to take part in the survey were issued by lecturers in an inclusive, non-discriminatory way, emphasizing that participation is voluntary and not connected to the curriculum or any assessment.

In order to reduce pressure on participants while filling in the survey questionnaire, all questions were optional, allowing respondents to skip questions they had no answers to or did not want to answer, without abandoning the questionnaire altogether. The survey was scrutinized and cleared by the University of Brighton's ethics approval process prior to commencement.

3.4 Sample

The survey includes 102 participants with various backgrounds, demographics, skill levels and previous experience in designing ML applications. The age distribution in the sample (Fig. 1a) is indicative of calls for participation being directed at both design students and design professionals, with the 18-25 year group being the largest segment at 34% and other segments gradually decreasing with age. However, all age groups, including over 65 years, are represented in the survey. There is a substantial gender gap (Fig. 1b) with 33% of respondents identifying as female and 64% identifying as male, while 1% identified as non-binary and 3% preferred not to answer the question. This is very similar to UK labor market statistics for design professions including traditional designers, web designers and designers in engineering, which combined have a gender split of 32% female to 68% male [44]. Only 12% of respondents report to ever having designed an application that uses ML (Fig. 1c). Answers to open questions suggest that this includes applications using ML behind the scenes rather than designing user interaction with ML components as envisaged in the customizable ML applications proposed in this paper. Most respondents rate their design skills considerably higher than their programming skills (Fig. 1d), with 21% having no programming skills at all and a further 38% having only minimal programming skills, while at the other end only 3% rate their programming skills as expert level.

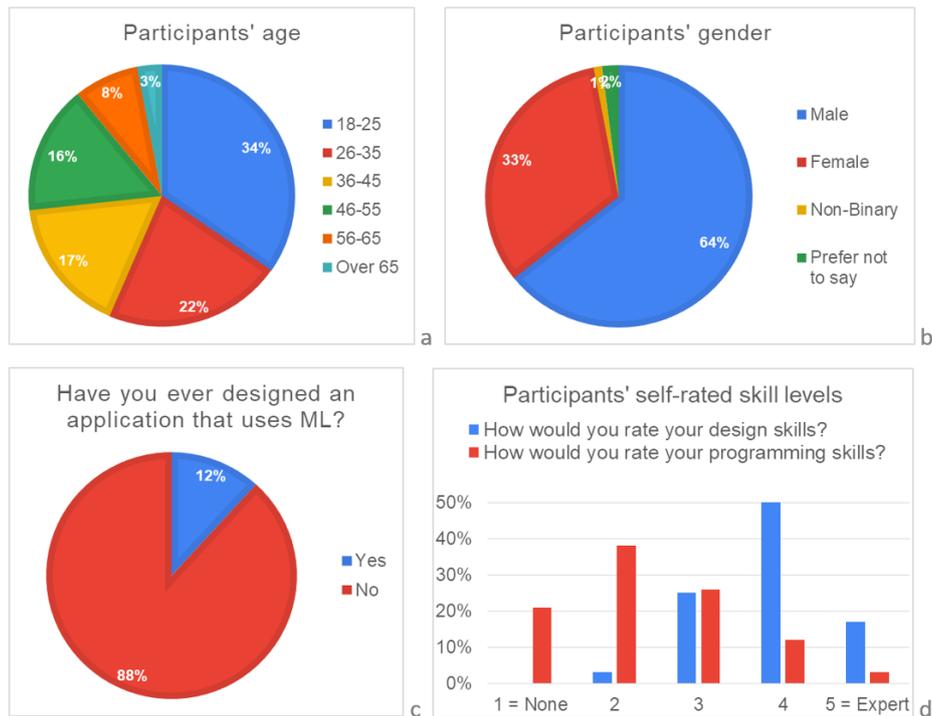


Fig. 1. Participants' (a) age, (b) gender, (c) experience in designing ML applications and (d) self-rated skill levels in design and programming.

Answers to an open question asking for participants' job or course title suggest that respondents were from a wide range of design disciplines, with the largest segments in user experience design, product design, digital media design, graphic design and building design. Answers also include more specialized disciplines such as typeface design or motion design as well as general descriptors such as studio director or creative director. The sample also includes five participants who describe themselves as artists and/or academics teaching in design disciplines.

3.5 Findings

Awareness of ML. In order to assess designers' awareness of ML in their everyday lives, the survey began with an open question asking participants to name some common applications that use ML. The fact that 72% of respondents provided at least one, often multiple, examples of common applications and services using ML suggests a high level of awareness among designers. By far the largest proportion of applications mentioned were large online services such as Google, Amazon, Facebook or Netflix, typically cited by name but in some cases referred to as "search", "social media" or "shopping recommendations", closely followed by personal assistants such as Siri, Alexa, Google Assistant used on smart speakers and mobile phones. Other answers referenced specific ML capabilities such as face recognition or object classification, or, in some cases, application areas such as predictive keyboards, financial trading and weather forecasting.

Interest in ML. Assuming that designers' actual (rather than espoused) interest in ML would manifest itself in creative engagement with it as a design material, and to see how realistic proposed applications might be as a measure of how well the capabilities and limitations of that design material are understood, the next question asked participants whether they had ever thought about possible applications that use ML, and if so, to explain in a few words what they would do. The fact that 44% of respondents answered this question with more or less concrete application ideas suggests substantial interest in ML among designers. More than half of the described application ideas reference concrete ML capabilities related to classification, typically based on image data but in some cases custom data sets, and content generation for a range of problems of varying complexity. Many other application ideas were either rather abstract (e.g. "Medical diagnostics", P53) or very complex, requiring a whole range of ML capabilities, wider AI technologies and sensors, if possible at all (e.g. "Electric Rollerblades: Braking and accelerating, reading shifts of rider's weight and moments [sic], proximity of one boot to the other when both are powered." P26).

Understanding of ML capabilities. The next set of questions asked participants how well they thought ML would perform on a range of problems, using a generic scale of 1 (Poor) to 5 (Good). The ML problems in these questions were chosen with regard to their relevance to the development of Flatpack ML applications indicated by generality, availability of pre-trained models and suitability to run on edge devices.

The overall trend in answers across capabilities (Fig. 2) suggests that most respondents assume ML to perform rather well at the specified problems, even though many are reluctant to give it the highest rating. However, there is also a substantial tail of

medium to poor ratings reflecting more critical assessments of these ML capabilities. While there are marked differences in ratings for some capabilities, with generic object recognition, body pose estimation and image style transfer attracting higher ratings, while age recognition from faces, image generation and written style transfer attracting lower ratings, average ratings for all capabilities range between 3.24 and 4.03 ($SD=0.23$) indicating that respondents overall assume better than par performance for these capabilities. One aspect particularly relevant in the context of Flatpack ML is that most respondents seem to assume that ML performs better at generic object recognition (e.g. cat, bottle, car) than at specific object or gesture recognition (e.g. your cat, thumbs up), which is typically not the case as the latter requires less generalization and can be achieved through transfer learning with comparatively small training data sets.

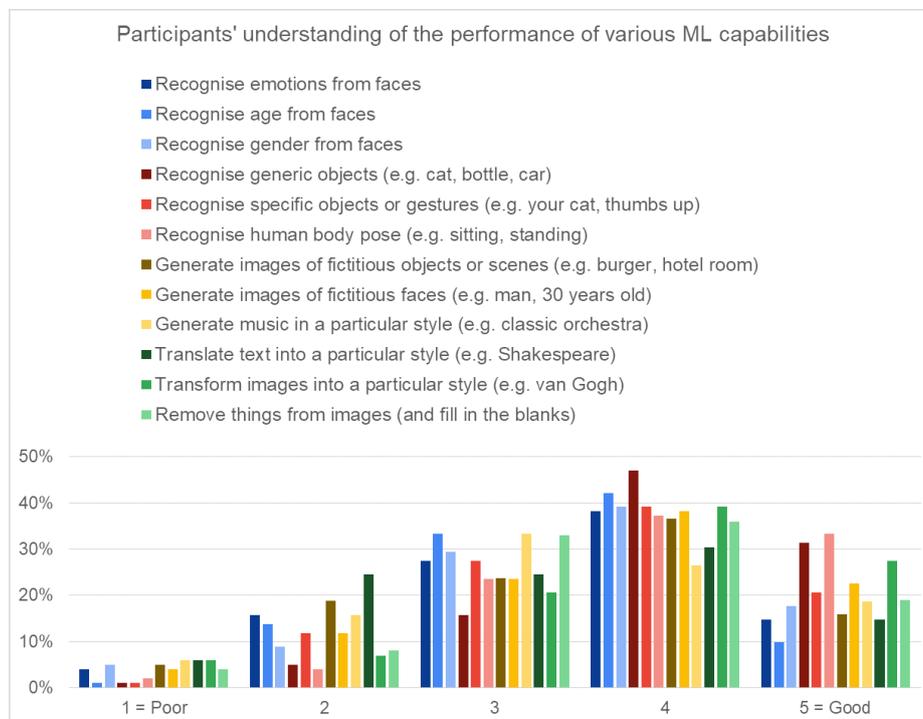


Fig. 2. Participants' understanding of the performance of various ML capabilities.

This section also included open questions asking what other ML capabilities or applications participants had heard of or would find useful, with responses describing a surprisingly wide range of ML capabilities and applications. With regard to capabilities or applications participants had heard of (50% response rate), some applications were mentioned more often than others (e.g. self-driving cars and personal assistants) and some application areas were more prominent than others (e.g. privacy and security-related applications). However, answers also included a large number of niche applications such as "Deep Fakes" (P30, P44) or "Reviewing images of crops to find disease

resistant plants" (P73). With regard to capabilities of applications participants would find useful (43% response rate), many respondents suggested applications that either already exist (e.g. "Trying glasses online and see how they fit", P11) or are actively researched (e.g. "recognising tumours or cancer from images", P29). Some suggested applications were genuinely novel, such as the idea of ML for automated usability testing of interface designs (P19) or image search based on purpose rather than content (P73). Many respondents suggested general application areas such as health or sustainability rather than specific applications and only few suggested ML capabilities not included in the provided examples, all of which either already exist (e.g. "Voice recognition" P14) or are actively researched (e.g. "precise gaze detection", P18).

Understanding of ML aspects. When asked about their awareness and understanding of various ML aspects, a large majority of respondents answered that they had never heard of, heard of but knew nothing or only knew little about these concepts (Fig. 3).

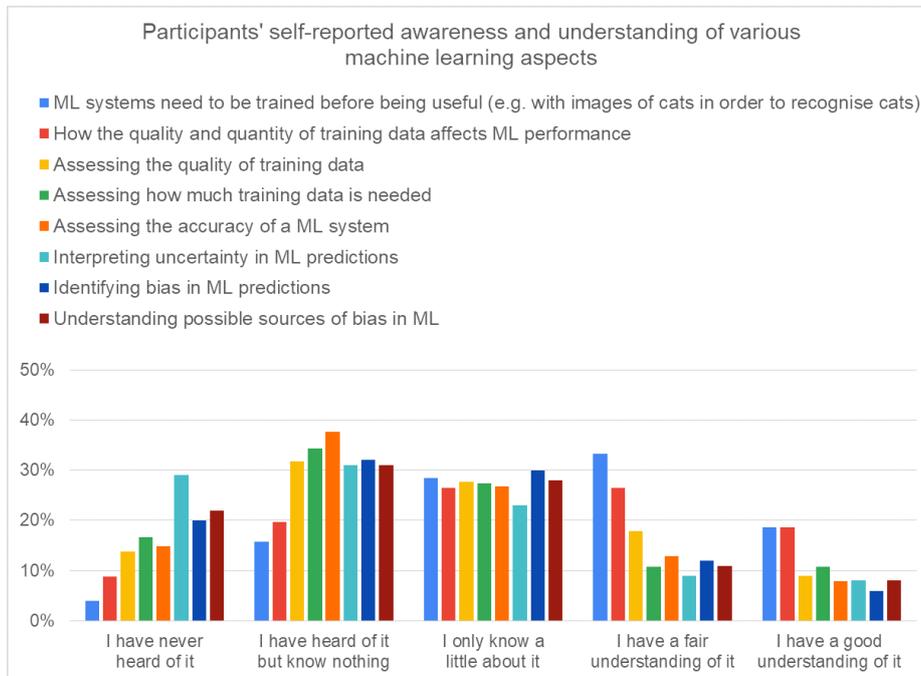


Fig. 3. Participants' self-reported awareness and understanding of various ML aspects.

In particular this included critical aspects such as assessing the quality of training data, assessing how much training data is needed and assessing the accuracy of an ML system, as well as interpreting uncertainty and identifying bias in ML predictions and understanding possible sources of bias. Notable exceptions were more fundamental aspects such as that an ML system needs to be trained before being useful and how the quality and quantity of training data affect its performance, which almost half of respondents reported to have a fair or good understanding of. A rather surprising result is that 29% of respondents answered that they had never heard of the need to interpret

uncertainty in ML predictions. The formulation of the question does not allow for inferences about whether respondents were not aware of ML systems producing confidence scores rather than definite predictions, or whether they just never heard of the need to interpret these confidence scores. However, given that a further 31% answered that they had heard of it but knew nothing about it, this is certainly a key aspect to address in any support measures as uncertainty is particularly relevant in customizable ML applications where designers might need to support end-users in setting threshold values for predictions depending on the specific use case and context. For similar reasons, it will be important to improve designers' knowledge about assessing the accuracy of an ML system, which 15% answered they had never heard of and 38% had heard of but knew nothing about, as they must find ways to meaningfully communicate this information to users.

Usefulness of support measures. When asked to rate the usefulness of various ways to support them in designing ML applications, respondents indicated that all suggested measures would be quite useful of very useful (Fig. 4). However, there is a clear preference for working examples to see and try out different ML capabilities, and for practical workshops where they build ML prototypes with guidance from an expert. Articles and tutorials, too, are seen as useful to some extent while code snippets for re-use in prototyping were rated least favorably. While the ratings for code snippets are higher for the sub-sample of participants who rated their programming skills 4 or 5 on a scale of 1 (None) to 5 (Expert), they are still lower than the ratings for working examples and practical workshops in the whole sample, underlining the importance of experiential learning methods for designers.

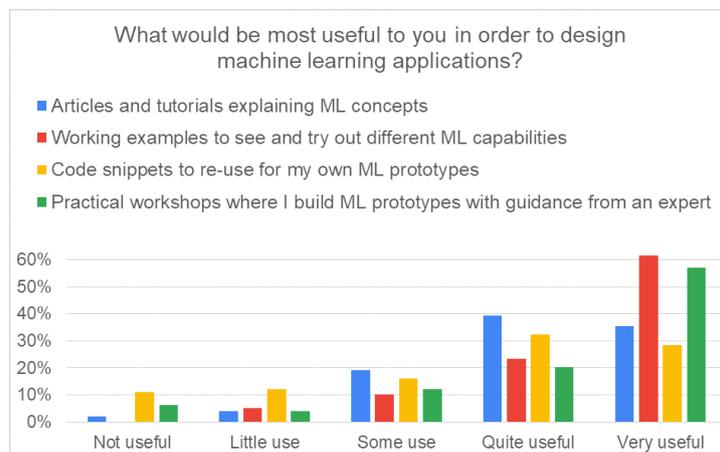


Fig. 4. Perceived usefulness of various support measures.

When asked what else could help participants to design ML applications (34% response rate), respondents expressed a strong interest in interactive, multimedia and experiential ways of learning about ML with several mentions of "step-by-step guides", "video tutorials", "workshops" and "conferences". There were also calls for more formal and

comprehensive training, in particular with regard to coding and ML-related technologies, and a distinct interest in learning about ethical aspects of ML, including bias and potential misuse, and how to address these issues.

Ethical aspects were also touched on in answers to a final open question asking for any other thoughts on ML in a design context (31% response rate). Several answers had a note of apprehensiveness (e.g. "its our future so don't mess it up", P53; "Don't take our jobs away!!!", P92), however, many respondents acknowledged that ML opens up new opportunities, both as a design material to improve the capabilities of their products and as a tool to support and speed up the design process, with some emphasizing this aspect with adjectives such as "exciting" (P2, P23, P39), "interesting" (P38, P60) and "liberating" (P37). Several respondents pointed out the general importance of ML for design professionals and expressed a need to learn more about it, with one participant stating that "it seems a bit behind the times that these tools are not more mainstream in design education" (P23).

3.6 Discussion

The findings both underline the need for this research and inform its various aspects with regard to how designers can be supported in creating ML applications that can be trained and customized by end-users.

Participants' answers suggest that designers have good awareness of ML in their everyday lives, with most participants being able to name several applications making use of the technology, and a strong interest to learn more about it. The latter is supported not only by the fact that many participants had already thought about potential ML applications, but also by statements emphasizing its importance for design professionals and calling for training and support in this area.

There is much variation in how participants rated the performance of various ML capabilities, however, above par mean ratings for all capabilities suggest that a majority of designers is rather optimistic about ML capabilities. Detailed responses indicate a misconception about ML performance in recognizing generic versus specific objects, which is of particular relevance for customizable applications using transfer learning to contextualize ML models, however, overall the results suggest that most designers have a fair understanding of what problems are more or less difficult for ML to solve. While this only applies to the particular set of ML capabilities included in the survey, it nonetheless puts into perspective claims in the literature about designers often having misconceptions about ML capabilities [13] or difficulties understanding the limits of what ML can learn [46], indicating that these difficulties might apply less to well-defined generic ML problems designers might have heard of or read about, such as object recognition or face interpretation.

Participants' self-assessed understanding of various ML aspects suggest that most designers would benefit from learning about basic concepts related to training, assessing the performance and interpreting the outputs of ML systems. While some aspects are better understood than others, there is a significant proportion of respondents who indicated that they had never heard of these aspects and a large majority who

knows little or nothing about them. This includes aspects particularly relevant to Flatpack ML, as designers need to be able to encode and operationalize them in their designs in a manner that is easy to learn and understand by end-users.

How designers can be supported to learn about and use ML as a design material is a key question in this paper, and the survey findings provide clear indications towards suitable learning approaches and materials. Participants' responses show a strong preference for experiential ways of learning about ML. Experiential learning theory [26] tells us that hands-on experimentation and reflection on the concrete experience lead to deep understanding and facilitate the transfer of learned concepts to new situations and contexts, which is particularly important when designers are asked to ideate and design novel ML applications. It also resonates with design theory, where Material Driven Design [25] provides a conceptual framework for design explorations inspired by the qualities of a design material rather than necessarily starting with a problem, while emphasizing that engagement with the design material "should also elicit meaningful user experiences in and beyond its utilitarian assessment" (p.35).

Finally, the survey findings document strong interest in ML among designers and desire to learn about it, evidenced by respondents' explicit calls for training and support. They underline the need to explore this topic further and develop suitable learning materials and programmes to support designers' engagement with ML.

3.7 Limitations

The survey involved 102 design professionals and students covering all age groups and broadly reflecting the gender distribution among UK design professionals, however, it cannot be claimed to be based on a representative sample due to the employed convenience sampling and, perhaps more importantly, the broad range of design domains and lack of a clear definition of what constitutes design or the act of designing.

Closely related to this diversity, no attempt was made to distinguish between different design domains, skill levels or demographic segments in the data analysis, although this might yield valuable insights in the future. As such, the findings can only provide a broad understanding of designers' interest, understanding and support needs, and help to inform future more targeted research efforts into making ML accessible to specific demographic groups and design domains.

4 Research challenges

Based on findings from the literature and survey, a number of overlapping research challenges emerge in the effort to support designers in ideating, designing and prototyping Flatpack ML applications.

1. Designers need conceptual knowledge about ML, in particular with regard to its capabilities, limitations and data requirements, in order to ideate realistic applications that address end-users' needs and fit a particular context. Designers also need operational knowledge of ML, in particular with regard to the quality and quantity of training data, the assessment of ML performance and the identification of bias in

trained systems. The literature suggests that designers struggle with these aspects [13, 15, 46] and survey results confirm this for the most part: while many designers seem to have a fair understanding of which classes of problems are more or less difficult for ML to tackle, they assign themselves low ratings when it comes to their understanding of operational aspects of ML. Considering that survey results show a clear preference for experiential learning methods among designers, a key research challenge in this context is to develop suitable environments and tools that enable designers to learn about ML through hands-on engagement and experimentation.

2. Experiential learning theory tells us that active reflection on concrete experience is a key mechanism to conceptualize learning, leading to deep understanding and helping learners to transfer their knowledge to new contexts [26]. A related research challenge is therefore how to encourage and support designers' reflection when engaging with these environments and tools, and how to help them relate concrete aspects of their experience to abstract ML concepts and operational knowledge. Key in this context is to recognize learning opportunities in the engagement with ML and to address wider problems relating to transparency and explainability [1, 16, 27].
3. Helping designers to encode their conceptual and operational knowledge of ML in their designs to create a good user experience for end-users training and customizing Flatpack ML systems for their specific context is another important research challenge. While the literature identifies interaction design with ML as a challenge in general [2, 15], this aspect is of particular relevance when user interaction with ML extends to training systems and assessing their performance. Key in this context is a suitable terminology and/or visual language that can describe ML aspects in ways that make sense to end-users, and design patterns and conventions for ML training and customization that increase the recognition and learnability of these features. The latter in particular can draw on recent efforts in developing design guidelines for Human-AI interaction [5, 9, 12, 15, 21].

Other more specific challenges include finding effective ways to communicate uncertainty in ML predictions, supporting end-users in identifying acceptable context-specific thresholds for that uncertainty, and helping end-users to be aware of and avoid bias when training Flatpack ML applications.

5 Summary and conclusions

Picking up on the opportunities of running ML models locally on consumer-level hardware, this paper proposes a new generation of ML applications that can be trained and customized by end-users for their specific context and use case. In reference to flat pack furniture that requires a setup process at home before it can be used, we used the term *Flatpack ML* for this type of application. While sharing many aspects with research into IML and MT, Flatpack ML design has its own problem space as it focuses on end-users training and customizing a generic application for a specific context; pre-trained base models and transfer learning techniques rather than custom models; consumer-level hardware rather than high-performance professional equipment; and personal use environments rather work- or lab environments.

The literature discusses a variety of challenges designers face when working with ML, not all of which are equally relevant in Flatpack ML design (Table 1). As this type of application uses readily available pre-trained base models, it does not require deep technical ML knowledge, is less reliant on ML experts and typically design led, mitigating many of the challenges in more traditional ML-driven projects, where UX design is often an afterthought [47]. At the same time, challenges around operational ML knowledge and interaction design become more pertinent as designers need to guide end-users through the initial training and customization process for these applications.

In order to gauge designers' interest and understanding of ML, and to find out how they can be supported in designing and prototyping ML applications, we carried out a survey of 102 design professionals and students. The results indicate a strong interest in ML among designers and a readiness to learn about it. While many designers seem to have a fair understanding of which problems are more or less difficult for ML to solve, the findings largely support claims in the literature that designers often have little understanding of basic ML concepts. There was a clear preference among participants for experiential approaches to learning. Working examples to see and experiment with ML capabilities and practical workshops to build ML applications with guidance from an expert were the most popular options, with many other suggestions involving hands-on engagement with ML.

Based on the survey results and findings from the literature, the paper identified three overlapping research challenges in supporting designers to ideate, design and prototype the proposed new generation of ML applications. They include (i) developing suitable environments and tools to support designers' exploration and hands-on engagement with ML, (ii) facilitating designers' reflection when engaging with these environments and tools to support their experiential learning, and (iii) helping designers to encode their conceptual and operational knowledge of ML in their designs to improve the end-user experience of training and customizing applications. Addressing these challenges will require substantial and multi-faceted research, to which we plan to contribute in future work.

Acknowledgements. We would like to thank the participants of our survey for volunteering their time and sharing their valuable thoughts.

References

1. Abdul, A., Vermeulen, J., Wang, D., Lim, B.Y., Kankanhalli, M.: Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda. In Proc. of 2018 CHI Conference on Human Factors in Computing Systems, paper 582. ACM (2018).
2. Amershi, S., Cakmak, M., Knox, W.B., Kulesza, T.: Power to the people: The role of humans in interactive machine learning. *AI Magazine*, 35(4), 105-120 (2014)
3. Amershi, S., Fogarty, J., Kapoor, A., Tan, D.: Examining multiple potential models in end-user interactive concept learning. In Proc. of 2010 CHI Conference on Human Factors in Computing Systems, 1357-1360. ACM (2010).
4. Amershi, S., Fogarty, J., Kapoor, A., Tan, D.: Overview based example selection in end user interactive concept learning. In Proc. of 22nd Annual ACM Symposium on User Interface Software and Technology, 247-256. ACM (2009).

5. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P.N., Inkpen, K., Teevan, J. Kikin-Gil, R., Horvitz, E.: Guidelines for human-AI interaction. In Proc. of 2019 CHI Conference on Human Factors in Computing Systems, Paper 3. ACM (2019).
6. Anderson, G.: Fundamentals of Educational Research. Falmer Press (1990).
7. Andriluka M, Uijlings JR, Ferrari V.: Fluid annotation: a human-machine collaboration interface for full image annotation. In Proc. of 2018 ACM Multimedia Conference on Multimedia Conference, 1957-1966. ACM (2018).
8. Apple A13 Bionic Chip. <https://www.macworld.com/article/3442716/inside-apples-a13-bionic-system-on-chip.html>. Last accessed 2020/01/29.
9. Argarwal, A., Regalado, M.: Lingua Franca: A Design Language for Human-Centered AI. (2020). <https://linguafranca.standardnotation.ai/>. Last accessed 2020/01/31.
10. Christina, C.: The Designer's Guide to Machine Learning. Digitalist, (2017), <https://digitalist.global/talks/the-designers-guide-to-machine-learning/>, last accessed 2020/01/17.
11. Clark, J.: Why Machine Learning and AI Matter for Design Teams. Big Medium, (2019), <https://bigmedium.com/ideas/why-machine-learning-and-ai-matter-for-design-team.html>, Last accessed 2020/01/17.
12. Corbett, E., Saul, N., Pirrung, M.: Interactive machine learning heuristics. In Proc. of Machine Learning from User Interaction for Visualization and Analytics Workshop at IEEE VIS 2018. IEEE (2018).
13. Dove, G., Halskov, K., Forlizzi, J., Zimmerman, J.: UX design innovation: Challenges for working with machine learning as a design material. In Proc. of 2017 CHI Conference on Human Factors in Computing Systems, 278-288. ACM (2017).
14. Drozdov, S.: An intro to Machine Learning for designers. (2018). <https://uxdesign.cc/an-intro-to-machine-learning-for-designers-5c74ba100257>. Last accessed 2020/01/17.
15. Dudley, J.J., Kristensson, P.O.: A review of user interface design for interactive machine learning. ACM Transactions on Interactive Intelligent Systems, 1, Article 1. ACM (2018).
16. Eiband, M., Völkel, S. T., Buschek, D., Cook, S., Hussmann, H.: When people and algorithms meet: user-reported problems in intelligent everyday applications. In Proc. of 24th International Conference on Intelligent User Interfaces, 96-106. ACM (2019).
17. Fails, J. A., Olsen Jr, D. R.: Interactive Machine Learning. In Proc. of 8th International Conference on Intelligent User Interfaces, 39-45. ACM (2003).
18. Fogarty, J., Tan, D., Kapoor, A., Winder, S.: CueFlik: Interactive Concept Learning in Image Search. In Proc. of 2008 CHI Conference on Human Factors in Computing Systems, 29-38. ACM (2008).
19. Gillies, M., Fiebrink, R., Tanaka, A., Garcia, J., Bevilacqua, F., Heloir, A., Nunnari, F., Mackay, W., Amershi, S., Lee, B., d'Alessandro, N.: Human-centred machine learning. In Proc. of 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, 3558 - 3565. ACM (2016).
20. Google Edge TPU. <https://cloud.google.com/edge-tpu/>. Last accessed 2020/01/29.
21. Google: People + AI. Designing human-centered AI products. <https://pair.withgoogle.com/>. Last accessed 2020/01/29.
22. Hebron, P.: Machine Learning for Designers. O'Reilly Media, (2016)
23. Huawei Kirin 980. <https://consumer.huawei.com/en/campaign/kirin980/>. Last accessed 2020/01/29.
24. Jo Y, Park J.: SC-FEGAN: Face Editing Generative Adversarial Network with User's Sketch and Color. arXiv preprint arXiv:1902.06838. (2019).
25. Karana, E., Barati, B., Rognoli, V., Zeeuw van der Laan, A.: Material driven design (MDD): A method to design for material experiences. Int. Journal of Design, 9(2), 35-54. (2015)

26. Kolb, D.A.: *Experiential Learning: Experience as the Source of Learning and Development*. Prentice-Hall, (1984).
27. Kulesza, T., Burnett, M., Wong, W.K., Stumpf, S.: Principles of explanatory debugging to personalize interactive machine learning. In *Proc. of 20th International Conference on Intelligent User Interfaces*, 126-137. ACM (2015).
28. Liikkanen, L.: Applications Of Machine Learning For Designers. *Smashing Magazine*, (2017), <https://www.smashingmagazine.com/2017/04/applications-machine-learning-designers/>. Last accessed 2020/01/17.
29. Lovejoy, J.: The UX of AI. *Google Design*, (2018), <https://design.google/library/ux-ai/>. Last accessed 2020/01/17.
30. Malsattar, N. Kihara, T., Giaccardi, E.: Designing and Prototyping from the Perspective of AI in the Wild. In *Proc. of 2019 Designing Interactive Systems*, 1083-1088. ACM (2019).
31. Manshaei R, Mayat U, Tarun A, DeLong S, Chiang D, Digregorio J, Khayyer S, Gupta A, Kyan M, Mazalek A. Tangible Tensors: An Interactive System for Grasping Trends in Biological Systems Modeling. In *Proc. of 2019 Conference on Creativity and Cognition*, 41-52. ACM (2019).
32. Miles, M.B., Huberman, A.M.: *Qualitative Data Analysis*. Sage (1984).
33. Modelzoo, <https://modelzoo.co>. Last accessed 2019/10/30.
34. ONNX Model Zoo, <https://github.com/onnx/models>. Last accessed 2019/10/30.
35. Open Model Zoo, https://github.com/opencv/open_model_zoo. Last accessed 2019/10/30.
36. Porter, R., Theiler, J., Hush, D.: Interactive machine learning in data exploitation. *Computing in Science & Engineering*, 15(5), 12-20.
37. PyTorch Mobile, <https://pytorch.org/mobile/home/>. Last accessed 2019/10/30.
38. Samsung Neural Processing Unit. <https://www.samsung.com/global/galaxy/what-is/npu/>. Last accessed 2020/01/17.
39. Simard, P.Y., Amershi, S., Chickering, D.M., Pelton, A.E., Ghorashi, S., Meek, C., Ramos, G., Suh, J., Verwey, J., Wang, M., Wernsing, J.: Machine Teaching: A new paradigm for building machine learning systems. *arXiv preprint arXiv:1707.06742*. (2017).
40. TensorFlow Lite, <https://www.tensorflow.org/lite>. Last accessed 2019/10/30.
41. TensorFlow.js, <https://www.tensorflow.org/js>. Last accessed 2019/10/30.
42. TF Hub, TensorFlow Hub, <https://tfhub.dev>. Last accessed 2019/10/30.
43. TorchJS, <https://github.com/torch-js/torch-js>. Last accessed 2019/10/30.
44. UK Office for National Statistics. Annual Population Survey - Employment by occupation by sex. October 2019 - September 2019. <https://www.nomisweb.co.uk/datasets/aps168/reports/employment-by-occupation?compare=K02000001>. Last accessed 2020/01/25.
45. Yang, Q., Scuito, A., Zimmerman, J., Forlizzi, J., Steinfeld, A.: Investigating How Experienced UX Designers Effectively Work with Machine Learning. In *Proc. of 2018 Designing Interactive Systems*, 585-596. ACM (2018).
46. Yang, Q., Suh, J., Chen, N.C., Ramos, G.: Grounding Interactive Machine Learning Tool Design in How Non-Experts Actually Build Models. In *Proc. of 2018 Designing Interactive Systems*, 573-584. ACM (2018).
47. Yang, Q.: Machine Learning as a UX Design Material: How Can We Imagine Beyond Automation, Recommenders, and Reminders? In *AAAI Spring Symposium Series*, (2018). <https://www.aaai.org/ocs/index.php/SSS/SSS18/paper/viewFile/17471/15475>. Last accessed 2019/10/30.
48. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems*, 27, 3320-3328. NIPS (2014).