Mix & Match Machine Learning: An Ideation Toolkit to Design Machine Learning-Enabled Solutions

Anniek Jansen*
a.jansen@tue.nl
Department of Industrial Design, Eindhoven University of Technology
Eindhoven, The Netherlands

Sara Colombo*
s.colombo@tue.nl
Department of Industrial Design, Eindhoven University of Technology
Eindhoven, The Netherlands

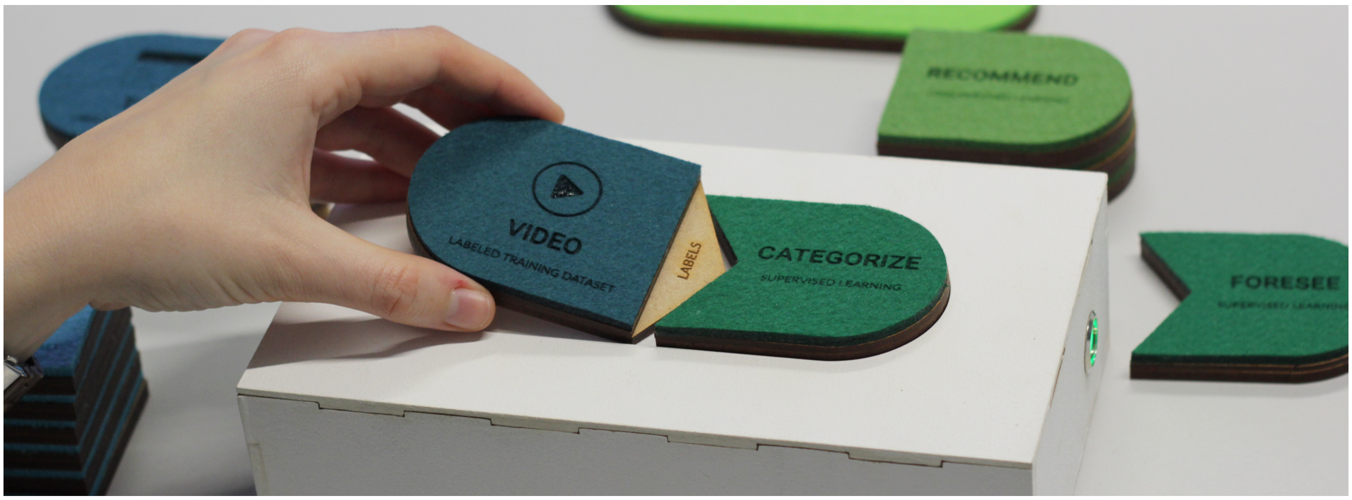


Figure 1: The sensing board with a data token and an ML capability token, which are part of the Mix & Match ML Toolkit

ABSTRACT

Machine learning (ML) provides designers with a wide range of opportunities to innovate products and services. However, the design discipline struggles to integrate ML knowledge in education and prepare designers to ideate with ML. We propose the Mix & Match Machine Learning toolkit, which provides relevant ML knowledge in the form of tangible tokens and a web interface to support designers' ideation processes. The tokens represent data types and ML capabilities. By using the toolkit, designers can explore, understand, combine, and operationalize the capabilities of ML and understand its limitations, without depending on programming or computer science knowledge. We evaluated the toolkit in two workshops with design students, and we found that it supports both learning and ideation goals. We discuss the design implications and potential impact of a hybrid toolkit for ML on design education and practice.

CCS CONCEPTS

• **Human-centered computing → Systems and tools for interaction design**; **Interactive systems and tools**; • **Computing methodologies → Machine learning**.

KEYWORDS

design ideation toolkit, machine learning, tangible user interface, ML capabilities, data types, design education

ACM Reference Format:
Anniek Jansen and Sara Colombo. 2023. Mix & Match Machine Learning: An Ideation Toolkit to Design Machine Learning-Enabled Solutions. In *TEI '23: Proceedings of the Seventeenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '23), February 26-March 1, 2023, Warsaw, Poland*. ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3569009.3572739

1 INTRODUCTION

Machine Learning (ML) is being used in an increasing number of products and services and offers many possibilities to designers to improve or innovate user experiences. ML is a core component of products and services consumers use everyday, such as recommendation systems in entertainment or online shopping platforms [30] and virtual assistants [13]. Although ML potential is wide, design education struggles to prepare the future generation of UX designers to work with ML [14]. Current professional designers often encounter ML for the first time in their job [31] and face many

*Both authors contributed equally to this research.

TEI '23, February 26-March 1, 2023, Warsaw, Poland

ACM ISBN 978-1-4503-9977-7/23/02.
https://doi.org/10.1145/3569009.3572739

challenges when working with this technology [14, 49]. Such challenges include understanding ML capabilities, envisioning novel implementations of ML to address UX problems, designing and prototyping interactions with ML systems, and collaborating with ML engineers [49].

Yang et al. [48] highlighted that the ideation of new products and services based on ML is particularly difficult for UX designers. Improving designers' technical literacy of ML can facilitate the ideation processes [49] but there is little knowledge on what aspects designers should be able to master in order to design ML-enabled solutions, and to what extent and depth they need to acquire technical knowledge. Moreover, limited research is available on how to teach ML to UX designers, as they usually do not have background knowledge in mathematics or programming [46]. A growing corpus of literature is focusing on ML education for non-majors in general [24, 26]. However, the needs of non-majors in general differ from those of UX designers, since the ability to innovate by using ML as a material requires specific skills and approaches [11, 14]. Understanding basic principles of ML, or how specific ML models work, is not sufficient for designers to ideate new solutions. They need to be able to reflect in action while working with the material [41]. For this, an understanding of the ML capabilities, i.e., understanding the breadth of ML possibilities and its limitations [49], is essential but currently missing [47].

Moreover, UX designers need knowledge that can be operationalized and situated in the design context. For instance, they need to acquire ML knowledge to a level that allows them to ideate novel solutions based on its possibilities and discuss them with data scientists [48], without necessarily diving into the more technical aspects of ML.

In the paper, we present the design of the Mix & Match Machine Learning toolkit, a hybrid physical-digital toolkit, which provides relevant ML knowledge to UX designers to support their ideation processes. The toolkit features a set of tangible tokens representing two elements, i.e., data types and ML capabilities, a sensing board, and a web interface. Data types and ML capabilities tokens are sensorized tokens, which can be individually placed on the sensing board to display detailed information in the web interface about the data or ML capability they represent. Data and ML capabilities tokens can also be combined to display examples of existing real-world applications, which leverage the selected type of data and ML capability. Such combinations are constrained by the physical shapes of the tokens, which allow only certain matches between data types (i.e., labeled or unlabeled data) and the learning approach that a certain ML capability leverages (i.e., supervised, unsupervised or reinforcement learning). The possible combinations showcase the possibilities and limitations of ML in a tangible manner. The toolkit explores physical manipulation of tangible interfaces as an approach to simplify learning and support ideation.

The toolkit was deployed in two workshops with design students, who employed it to envision and formulate a novel service or product in response to a design challenge. Our results showed that the Mix & Match ML toolkit supports designers during the ideation phase by making the ML possibility space tangible, offering clear and familiar examples that both explained ML and inspired the designers, and by providing them with a vocabulary and mental model on how to envision ML-enabled solutions. We discuss the implication of these findings for design research, practice, and education.

This work contributes to bridging the gap between UX design and ML in three ways. Firstly, it introduces the Mix & Match ML toolkit as an open source ideation toolkit for designers [1], and it describes its design rationale. Secondly, it proposes and validates an ML knowledge framework that can support design ideation with ML. Thirdly, it discusses the implications of adopting a tangible approach to develop an ideation toolkit for designing ML-enabled solutions.

2 RELATED WORK

2.1 Machine Learning and UX Design

For UX designers, incorporating ML into their design processes and into new products and services comes with challenges. Yang et al. [49] highlighted several challenges designers encounter in each phase of the design process. During the first phase of discovering and defining the idea, designers struggle with understanding the capabilities of ML, assessing the feasibility of their ideas, sketching and prototyping ML-based interactions, and foreseeing potential effects of using ML. Colombo and Costa [11] describe a new human-centered design process where ML is used as a design material. In it, the first step is to envision a solution to a user problem and identify what role ML plays in it. However, while the authors describe in detail the following steps of the design process, they point out that more research is needed to help designers envision solutions based on ML in the ideation phase. Especially in the very early phases of the ideation of new products or services, designers' difficulties seem to be due to a lack of ML literacy. Scholars argue that improving designers' technical literacy and providing them with tools that allow for easy exploration of data and/or ML models can facilitate their ideation process [49]. However, the design community is still debating what level of technical knowledge designers need to acquire in order to use ML as a design material, as well as what are the most effective forms of knowledge representation [48].

Because we are interested in how ML knowledge can be transferred to designers to facilitate their ideation process, in this section we analyze previous work on two aspects that are crucial to our investigation: how UX designers (or non-experts in general) can be educated on ML, and what tools, methods, or approaches have been developed to support the design of ML-enabled solutions.

2.2 Machine Learning Education for Designers

Little investigation has been done into what knowledge designers should possess in order to work with ML and how this knowledge can be taught. To the best of our knowledge, only Van der Vlist et al. [46] have looked into this challenge. However, their focus was on teaching two specific types of algorithms to the design students and not on ideation, which would require a broader understanding of all capabilities. Closer to the objective of this paper, is the paper by Fiebrink [16] where she designed a MOOC to teach ML to artists, musicians and creative practitioners. Based on video lectures explaining specific algorithms and using interactive machine learning in a graphical user interface (GUI), students were able to come up

[1] The open source code and instructions for the Mix & Match ML toolkit are available at https://github.com/MixMatchMLtoolkit

with and create prototypes. This study showed that teaching ML to students can support them in making creative work. However, these findings cannot be directly translated to designers since designers will often have to apply ML to a given design case, whereas students in this study were completely free in choosing what to create.

2.2.1 Machine Learning Education for Non-Experts. More work have been done to investigate how ML can be taught to non-experts in general. For example, tools have been developed to explain the process of training and evaluating models to children [24, 26, 51] or to encourage an ethical reflection of ML [27, 42]. Other studies focused on what aspects of ML were difficult to teach [43, 44] and identified that the main challenge for students lies in integrating ML into a system or new context. Yet most of the designed tools do not focus on this aspect [24, 26, 27, 42, 52].

2.3 Tangible User Interfaces for Education

One approach to improve the technical literacy of designers is by using Tangible User Interfaces (TUIs). TUIs have been highlighted in literature as having potential benefits for learning [29]. Learning with TUIs has multiple advantages. Firstly, TUIs have a playful and intuitive nature [29]. Thanks to the latter, they require minimal cognitive effort for operating and lower the threshold for participating in learning activities [34, 53], allowing the user to focus on the task instead of the tool [23]. Moreover, TUIs have been demonstrated to facilitate collaboration [29, 34, 40] as they increase visibility, communicate the current state of the work and encourage situated learning [34].

This was also shown in [40], where the TUI version outperformed a multitouch interface in terms of learning, collaboration and exploring alternative designs. TUIs can also positively impact the learning behavior as they can, among others, increase attention spans and provide a way for self-expression and communication when working in groups [29, 34]. Finally, TUIs can also affect the emotions during learning, for example increasing engagement, enjoyment, immersion, and confidence [29]. While most studies were conducted with children, several studies also showed benefits for university students and adults [29]. This motivated the authors to use a TUI to support UX designers during their ideation process for ML-enabled solutions.

2.3.1 TUIs for ML education. In the field of ML education, TUIs are also starting to emerge. Kaspersen, Billstrup and Petersen [26] introduced the Machine Learning Machine (MLM), a TUI for explaining the process of training and evaluating ML models to children. Children can iteratively create data by drawing on paper and feeding this into the *Trainer* and next carry over the model artifact over to the *Evaluator* to test how well the model works with new drawings. De Raffaele, Smith, Gemikonakli [12] compared a TUI and a GUI explaining Artificial Neural Networks to undergraduate students. This study showed that the students gained more knowledge when using the TUI and experienced the TUI to be more effective for understanding the provided information and for carrying out the tasks. Both these TUIs have focussed on one aspect of ML, either certain steps for the MLM [26] or a specific type of model in the study of De Raffaele et al. [12] but do not cover the range of capabilities of ML in general, an important learning goal for designers.

2.4 Supporting creativity

Next to educating designers to improve their technical literacy, they can also be supported in the design process with tools that support their creativity, i.e., *Creativity Support Tools (CST)* [17]. Few CSTs exist in the domain of envisioning ML-enabled solutions. The CSTs that do exist for designing with ML, are not always classified as such. An example of this is the ObjectResponder [33] where a designer can prototype ML *"in the wild"* by seeing live the results of an object identifier ML model. Other examples are the *Ideation Cards* from AIxDesign [2] which poses "What if" questions related to ML capabilities and gives example applications, the *I love Algorithms* card deck from d.school [3] explaining six common machine learning algorithms and *The Intelligence Augmentation Design Toolkit* from Futurice [4] with a card deck for ML interactions and touchpoints.

Only supporting designers to be creative is not sufficient. Toolkits also need to provide relevant ML knowledge to designers. Otherwise, designers will still run into difficulties with understanding ML and its capabilities [14, 49].

3 A FRAMEWORK FOR ML KNOWLEDGE

Existing work highlights the need to support the ideation of ML-enabled products and services, as well as the lack of approaches and toolkits that provide adequate knowledge to designers to master the complexity of data and ML in their creative processes. Current toolkits focus on specific ML models or approaches, existing ML models that can be embedded in design solutions, or narrow applications of ML; to our knowledge, no toolkit provides horizontal ML knowledge that allows an open-ended exploration of the breadth of ML capabilities, approaches, and types of data it operates with, to stimulate new design ideas. To fill this gap, we developed an ideation toolkit that supports the envisioning of novel products/services by improving ML literacy *in action*, during the ideation process.

To develop such a toolkit, we had to define two aspects: i) what (technical) knowledge could facilitate ideating with ML and ii) how to make such knowledge accessible during the ideation process, to inspire new ideas. The first aspect is addressed in the remaining of this section. The second one concerns the design of the toolkit, described in the following section.

3.1 Toolkit Content: ML Knowledge

3.1.1 ML Capabilities. Designers need to understand ML capabilities to operate with such a technology [48, 49]. ML capabilities refer to what ML "can or cannot do" [49]. Scholars argue that designers' difficulty to understand ML capabilities hinders their ability to generate new ideas from the start. Therefore, we decided to include ML capabilities in the ML knowledge featured in our toolkit. In this paper, we define ML capabilities as the ability of ML to perform certain tasks or actions, such as 'foresee' or 'recommend'. From

[2] https://www.aixdesign.co/shop/p/cards-digital
[3] https://dschool.stanford.edu/resources/i-love-algorithms
[4] https://futurice.com/ia-design-kit

a UX perspective, ML capabilities can augment products and services with specific functionalities, e.g., detecting spam emails or recommending the next movie to watch.

3.1.2 Types of Data. Datasets are the core components of ML models. Designers need to familiarize with data [48] and understand their use in both training and deployment of ML models. Knowing the breath of data types ML algorithms can work with, e.g., tabular data, images, text, etc., is an essential aspect of ML literacy. It also can inspire novel solutions. Training datasets for ML can be divided in labeled or unlabeled. That determines the learning approach that can be adopted (i.e., supervised or unsupervised), which, in turn, enables certain ML capabilities. We intended to show these relations in our toolkit.

3.1.3 Exemplars. The last element to be included in the toolkit are exemplars of ML applications. Next to abstractions, exemplars are the preferred way of designers to describe what ML can do and to communicate ideas [48]. Exemplars make ML concepts concrete and show how ML models are operationalized in existing products and services. They can sensitize designers to the possibilities provided by ML [48].

3.2 Defining ML capabilities

3.2.1 ML capabilities for design. Previous literature shows that designers need ML abstractions, or "simple insights about an ML capability" that can generate value for users [50]. According to Yang et al. [48], designers tend to describe ML capabilities as abstractions that refer to what ML is able to do, e.g., 'recognize intent'. We built on this insight to select a list of capabilities for our toolkit. The set of ML capabilities included in the toolkit are terms that represent basic functionalities of ML, which are meant to be understandable by novices with little to no background in ML or computer science. We decided not to include any technical information or jargon, rather to select a list of verbs that correspond to actions ML can perform, such as *foresee*, *cluster*, *distinguish*, or *recommend*. Such terms are meant to be general enough to be applied to different contexts and application domains, but specific enough to represent unique functionalities of ML algorithms.

3.2.2 Exploring ML capabilities through literature. We reviewed literature in the fields of computer science, HCI, and design, to look for existing classifications of ML capabilities. We collected relevant works by using different combinations of the following keywords in Google Scholar: "Machine Learning", "Artificial Intelligence", "Algorithms", "Capabilities", "Abilities", "Classification", "Framework", "Application(s)", "Design". We discarded the publications that did not present any classification or framework. The results included works featuring either *technical* or *application-oriented* classifications of ML or AI. The first cluster of publications includes works that classify ML from a technical perspective (i.e., types of algorithms) [1, 39]. The second one covers works that explore and categorize different ML applications [6, 25, 39, 50]. In most cases, the ML capabilities proposed in such classifications did not meet the level of abstraction we were aiming for. Technical frameworks categorize ML algorithms based on their technical features or the types of data they are trained on, and they require at least basic knowledge of ML to be comprehended. Application-based classifications often stem from the analysis of specific types of solutions [25] or application fields (e.g., "traffic prediction and transportation" or "e-commerce and product recommendation") [39], which are too specific and hardly generalizable. Additionally, Bawack et al. [6] provide a too high-level framework of ML capabilities, which are described as sense, comprehend, act, and learn.

3.2.3 Finalizing a set of ML capabilities for design. To generate a list of ML capabilities with the desired level of abstraction, we first clustered the ML capabilities described in the existing frameworks, to look for commonalities. We discarded the capabilities that were too generic (e.g., "analyze" or "act"), and we tried to abstract the ones that were too context-specific (e.g., "diagnose") by labelling them with one or more verbs that indicated the underlying ML functionality (i.e., "classify" an instance as associated/not associated with a disease). This initial labelling process was performed by one author and was guided by the following questions: *Is this capability specific to ML?*, *Does it suggest a clear functionality that can be used by designers in a solution?*, *Can it be understood by a non-expert in ML?*. Subsequently, both authors analyzed the resulting set of labels against the same three questions, performed additional clustering, and modified the verbs until they reached a consensus. We added a definition to each verb, to better clarify its meaning in this context.

The generated labels are umbrella terms that gather related sub-capabilities found in literature. For instance, the capability *categorize* includes sub-capabilities such as object detection and image classification (where an ML *categorizes* an object in a picture as a *face*, a *cat*, or a *pedestrian* [3, 38]). Similarly, the term *identify* includes sub-capabilities such as object recognition (where the system recognizes my face among other faces [38], or *identifies* a unique building or artwork from a picture) and speaker detection (where an ML identifies which individual is speaking by their voice). As a result of this process, we created a preliminary list of ML capabilities that could be useful for ideation. Subsequently, we collaboratively validated this list by analyzing case studies of existing ML applications (e.g., Google Lens, Netflix, Alexa). We verified if we could describe the underlying ML functionalities with the selected list of ML capabilities, and we adjusted the description or added new capabilities when we identified a mismatch.

3.2.4 Matching ML capabilities to learning approaches. Finally, we associated each ML capability to the type(s) of ML algorithm that enable it (e.g., classification, linear regression, or clustering algorithms). We classified the ML capabilities as supervised, unsupervised, or reinforcement learning, depending on the learning approach adopted by the main types of algorithms that support them. We validated this categorization with two data science experts. In this phase, we had to introduce some simplifications, to keep the knowledge approachable by non-experts. For instance, we did not consider hybrid approaches such as semi-supervised learning, and we connected each ML capability to its main learning approach, discarding the nuances that might exist. This simplification matches the goal of our toolkit, which is aimed to introduce understandable and operationable ML concepts to novices.

The final set features 12 ML capabilities. They show a consistent level of abstraction, which allows for their application to different contexts or types of solutions. The 12 selected ML capabilities,

Table 1: The twelve ML capabilities included in the toolkit with their respective learning approaches, definitions, and examples.

Machine learning capability	Learning approach	Definition	Example
Categorize	Supervised	Match an item with a category	Classify an item in a picture as a cat
Foresee	Supervised	Predict an action, event, state, behavior, intention, preferences, etc.	Forecast the gross revenue of a new movie
Identify	Supervised	Recognize the identity of a specific individual/item from a trait	Identify a specific building from a partial picture
Translate	Supervised	Transform contents from one domain to another	Translate a piece of text from one language into another
Understand	Supervised	Comprehend topics, themes, or sentiments; interpret language	Understand if a review posted on Twitter is positive or negative
Communicate	Supervised	Convey messages/content in understandable languages	A chatbot replying to your questions on a website
Cluster	Unsupervised	Group items based on their similarities	Cluster customers who buy similar items together
Distinguish	Unsupervised	Differentiate certain items from a group or average (find outliers, anomalies, etc.)	Analyze the sound of machinery and give a warning when it sounds different from usual
Recommend	Unsupervised	Provide suggestions or guidance; propose contents, activities, etc.	Recommend what song to listen next
Generate	Unsupervised	Create content (e.g. videos, images, music, text) from scratch	Generate new art based on a text prompt
Optimize	Reinforcement	Improve or perfect a certain task/route/process	Learning to play a game like chess or Go by trial and error
Navigate	Reinforcement	Steer autonomously through a physical or virtual environment	A robot finding the shortest path to the door

together with their definitions, learning approaches, and related examples, are reported in Table 1. This list does in no way aim to be exhaustive. Our goal was to create an initial set of understandable and applicable ML capabilities and test it in the context of our toolkit.

3.3 Defining data types

We focused on six data types to be included in the toolkit: *audio, image, table, text, time series, video.* The six types of data can be used both in labeled and unlabeled datasets. These data types were chosen based on an analysis of the current data types in widely-adopted online databases [15, 20, 21, 37]. They were validated by two experts in data science and machine learning. While this list is not complete, the aim was to include most of the data types that are often used by designers working with ML. For example, graph and vector data were left out, even though they can be used in e.g., clustering algorithms. However, we expected these data types to be less frequently used by designers.

3.4 Toolkit Design: TUIs for ML exploration

To facilitate designers' access to ML knowledge during their ideation process, we opted to design a hybrid physical-digital toolkit, as this allowed us to both leverage the ability of TUIs to support exploration [34] and convey more in-depth information through a web interface. To inform the design of the tangible components of the toolkit, we relied on literature and best practices in the field of TUIs.

The Mix & Match Machine Learning toolkit is a toolkit that supports designers during the ideation process by providing ML knowledge and by enabling open exploration through a tangible approach. The toolkit consists of three elements: a set of tangible tokens, a sensing board, and a web interface.

Below, we describe the design rationale used when designing the different elements of the toolkit.

3.5 Tokens

Tokens represent two categories of ML-related concepts: Data types and ML capabilities. To support the exploration of these concepts, tokens can be individually placed on the sensing board to learn more about a specific data type (e.g., labeled table, unlabeled table, labeled audio, etc.) or a certain ML capability (e.g., cluster or recommend). However, there is also a relational aspect between data and ML capabilities we wanted to represent, as ML capabilities rely on certain learning approaches (supervised, unsupervised, or reinforcement learning), which are enabled by different data types. Supervised learning algorithms require labeled training datasets, while unsupervised ones do not require labelling. Labeled datasets can still be used by unsupervised ML algorithms, by simply discarding the label. Reinforcement learning algorithms do not need

(a) One of the first concepts, using a separate token for the label

(b) The first version of the final design, which was used for an initial user test. The size and shape of the tokens were adjusted afterward

Figure 2: First versions of the tokens in cardboard and MDF

training datasets. We intended to show such relations and constraints through the tokens.

To inform the design of the tokens and evaluate different alternatives, we derived design requirements from literature. We particularly looked at how TUIs should be designed to represent the underlying concepts and to support explorative tasks and collaboration [34, 53]. As a result, the following four requirements were defined:

(1) *Matching constraints* The toolkit uses a token+constraint approach where the digital abstract information is represented as physical tokens and are constrained by their shape and the shape of the sensing board [45]. These constraints should match with the constraints of the concepts they represent - in our case, types of data and ML capabilities. Labeled data should fit supervised learning and unsupervised learning; unlabeled data should fit unsupervised learning but not supervised learning; reinforcement learning should not fit with any of the data tokens.
(2) *Intuitiveness* The valid combinations between the data type tokens and ML capability tokens should be intuitive. The same holds for the link between the token and what it represents.
(3) *Exploration* The toolkit should allow users to quickly explore, exchange, mix and match different tokens. Quick exploration is beneficial, especially during ideation, to support the creative process without interrupting it or slowing it down.
(4) *Visibility* The tokens should be visible from different perspectives and by multiple people to support collaboration.

Based on these requirements, we explored multiple designs for the tokens. The first version of data tokens used an add-on 'label' token to turn unlabeled data into labeled ones. Moreover, it did not support the combination of labeled data with unsupervised learning (see Figure 2a). In the second iteration, we used two different shapes for labeled and unlabeled data tokens. This also made it possible to combine labeled data with unsupervised learning. Since labels will not be used in unsupervised learning, the label part of the data token disappears under the unsupervised learning token when combined with it.

The first version of the final concept (see Figure 2b) was evaluated with design students (n=9) to assess what combinations they did consider possible and how these were interpreted. Based on the outcomes, the data tokens were rounded to prevent fitting two data tokens together; the total size of a combination was made equal to the size of a reinforcement token and the size of all tokens was reduced to make them easier to handle and organize. We also investigated which colors were associated to data and ML by giving the students a color wheel and letting them pick the colors they intuitively connected to data and ML. Blue colors were strongly associated with data, and cool colors in general with ML. Data tokens were therefore made in blue shades and ML capability tokens in green shades. Different shades were used to represent different categories of data and ML capabilities tokens, to improve visibility and facilitate identification.

4 THE MIX & MATCH ML TOOLKIT

4.1 Web interface

The toolkit objective is to support designers in their ideation process by providing relevant ML knowledge. This information is partly embedded in how the tokens can be combined, but most of it can be found in the connected web interface. To determine how to convey such knowledge to users, we relied on principles for learning that reduce the cognitive load of the users [22] so that they concentrate on the ideation process. The following four main principles were used:

Use of examples. Providing students with known examples makes it easier for them to connect new knowledge to existing knowledge

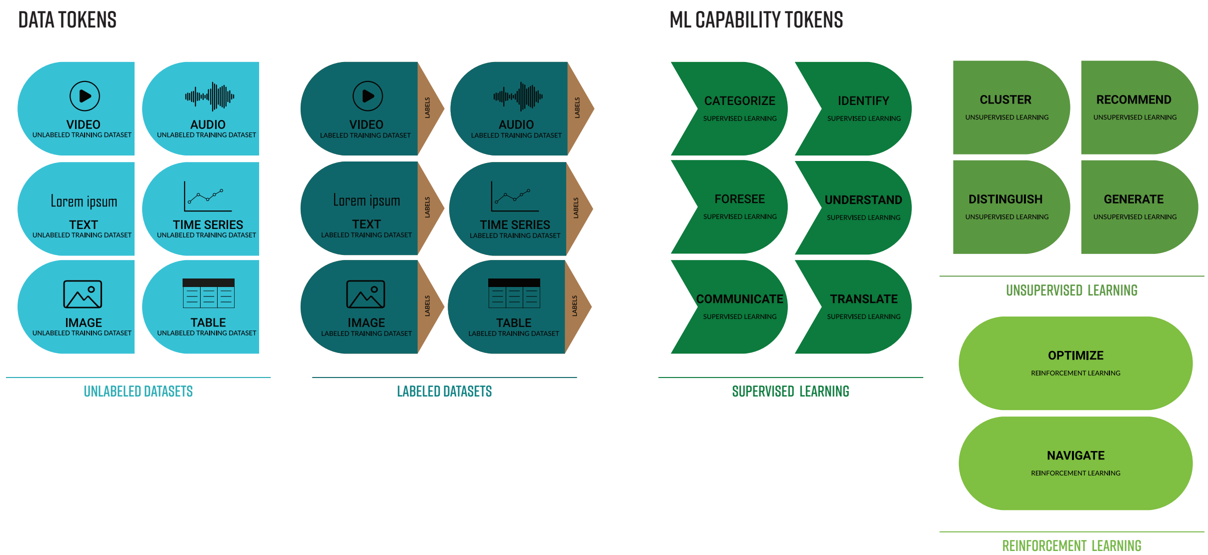


Figure 3: An overview of all the tokens included in the toolkit, with their corresponding shapes and colors.

[22]. This helps in retaining new information, as isolated facts are more easily discarded. Moreover, showing novices worked examples is more effective in showing how knowledge applies to specific cases compared to problem-solving [22]. Finally, the cognitive load can be reduced if students start with a higher order idea, and worked examples can help to achieve this idea [22]. We applied this in the toolkit by including examples of data sets, pre-trained models and existing application exemplars into the webpage.

Multimedia. When information is communicated through both text and visuals, stronger learning occurs as our brain can effectively combine these [22]. This is implemented in the toolkit by including images or schematics for the type of data and ML capability. In the example applications, the image is used to illustrate the application.

Pacing. Students learn better when they are able to pace the incoming information themselves [22]. By having no fixed order and time, the toolkit support students to control the pacing of information exploration and learning.

4.1.1 Machine Learning Simplification. For novices, clarity is more important than elaboration - while for experts it is the opposite [22]. Most design students are novices with respect to ML knowledge and often perceive ML as not accessible [43], making it even more important to have a low threshold. During the design of the toolkit, we had to find a compromise between clarity, usability, and the complexity of the subject. The toolkit is designed to be an introduction to ML for designers and to support preliminary ideation sessions. It is not intended to replace experts, instead it can help designers to get started on their own and involve experts in the subsequent stages. We are aware that not all information about ML can be captured in one toolkit, especially when the information should be understandable to non-experts. Therefore, we simplified and limited the information in the toolkit. The information provided in the toolkit was validated with a data scientist.

4.2 The final toolkit

The final design and prototyping of the Mix & Match Machine Learning toolkit components is described below.

4.2.1 The tokens. In the Mix & Match ML toolkit, two categories of tangible tokens are included: the data tokens (blue shades) and the ML capabilities tokens (green shades) (see Figure 3). All possible combinations of the tokens are shown in Figure 4.

The data tokens are split into two sets, one representing the six data types as labeled training datasets (dark blue), and one representing the same data types as unlabeled training datasets (light blue) (see Figure 3).

The 12 ML capabilities tokens represent the capabilities described in Table 1. These tokens are split into three sets, again different in shape and color (see Figure 3). Based on the learning approach that supports the ML capability, the token is either dark green and fits with the labeled data token (supervised learning); light green and does not fit with any data type (reinforcement learning); or it is middle green and fits with both labeled and unlabeled data tokens (unsupervised learning).

All tokens display some text describing the specific ML capability or type of data, and the higher category it belongs to (labeled/unlabeled, supervised/unsupervised/reinforcement). Data tokens also show a visual icon for faster recognition.

Tokens are made of a wooden base embedding an RFID label. They are covered in a layer of laser-cut felt, which improves grasping and manipulation and provides a pleasant tactile feeling (Figure 4).

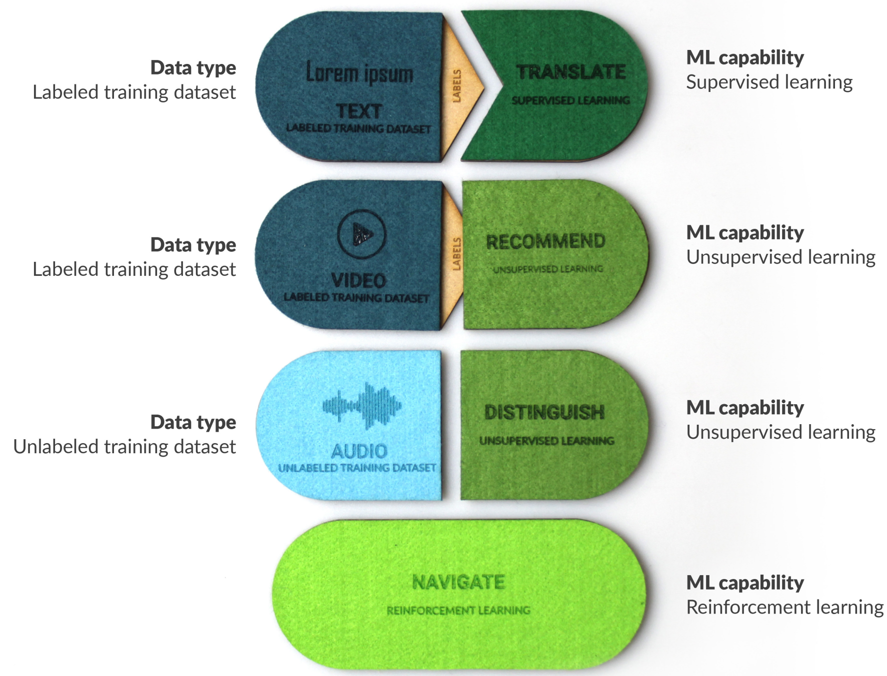


Figure 4: The five type of tokens included in the toolkit are labeled data, unlabeled data, reinforcement learning, supervised learning and unsupervised learning. The label disappears when labeled data tokens are combined with unsupervised learning tokens.

4.2.2 The sensing board. The sensing board allows the user to interact with the website by placing the tokens on top of the board (see Figure 5). Two MFRC522 RFID readers are placed directly under the surface of the sensing board and can read the IDs of Mifare classic 1K tags which are attached underneath each token. The RFID readers are connected to a LilyGO TTGO T-Energy ESP32-WROVER. The sensing board is connected to laptop via BLE, which makes it easy to connect to different laptops. The board is powered by an LG 18650 Li-ion Battery (3400mAh, 10A) to keep the setup wireless.

4.2.3 The web interface. Information about the individual tokens and their combinations can be accessed through the web interface [5]. The web interface is connected to the sensing board via BLE. A user can interact with the web interface by placing tokens on the sensing board (see Figure 5). The web interface features three types of pages (see Figure 6): (i) a *data page* if only a data token is placed on the sensing board; (ii) an *ML capability page* if only an ML capability token is placed; (iii) a *combination page* if a valid combination is placed on the sensing board. In all other cases, the web interface will give an error explaining what is wrong.

The web interface pages for individual tokens include i) short descriptions for either the selected data type or ML capability; ii) an image illustrating the data or ML capability; iii) examples, capabilities, and limitations and a list of pre-trained models for the ML capabilities or a list of example datasets for the data tokens (see Figure 6). Such datasets and models are from the public domain. In the current version, the web interface includes links to the dataset/model source page, but it does not analyze the fairness of the provided dataset or model examples. Given the increasing attention to this matter, in the near future we expect dataset/model repositories to feature a label [9], datasheet [19] or model card [36] to help users evaluate them for fairness criteria. This will also help us to select the best examples for our toolkit.

The combination page includes a short general description about the combination and a box with a real-world application example. An image, short description and links to more information

[5]The web interface of the Mix & Match ML toolkit can be accessed online at https://mixmatchmltoolkit.github.io/. The web interface is interactive and allows for the exploration of the tokens and their combinations when the tangible tokens are not available or connected.

Figure 5: The Mix & Match ML toolkit featuring the tokens, the sensing board, and the web interface

are shown, i.e., the "train it yourself" link explaining how to train similar ML models. Example applications were selected based on their clarity and ability to represent the data + ML capability combination. We evaluated and selected applications that were deemed to be non-harmful, but we cannot ensure that no biases are present in either the application, i.e., the data, model, and the design itself, or the example training data in the "train it yourself" link. For example, *audio + understand* uses smart assistants as an example application. While this application itself is not inherently harmful, it is known that some voice recognition systems are non-inclusive and have trouble recognizing female voices [5]. Other possible biases might be present, such as discriminating based on protected attributes [35] when checking for credit card fraud (*table + foresee* or in word embeddings models used for text translation or generation [7]. However, not all data and information about the trained models is publicly available, and hence an ethical evaluation of all applications is at the moment not feasible. We again emphasize that it is important to include ethical evaluations in the design process, but we decided to first validate the toolkit on its ability to support learning and ideation.

4.3 Open source toolkit

The toolkit is published as an open source project on GitHub [6]. Here, all the files for creating your own version of the toolkit can be accessed. The source code of the web interface is also made available, and the examples can be personalized if wished. A hybrid toolkit has the disadvantage of being harder to share and distribute, but by making it open source, we hope that people interested in the toolkit can build their own.

4.4 Expert validation

The design of the toolkit, i.e., the use of data and ML capability tokens and how they can be combined, and the web interface were validated with experts (n=8). These experts all worked in the field of ML and design. The majority worked at a large software company (n=6) as (senior) UX designer (n=4), product manager (n=1) or head of the human-AI interaction team (n=1). The other experts were an ethics consultant for AI and a data scientist at a company using a user-centered design approach for creating products, services, and systems with ML and AI. All except one were introduced to the toolkit remotely via a video and could access the web interface online, because they were located in different countries or regions. Only the data scientist was able to interact live with the toolkit. Overall, the experts felt the setup of the toolkit had potential for explaining ML easily to non-experts and felt it would be mainly usable for educating designers. They did anticipate designers preferring more visual information and felt it would be a good addition to include ethical aspects, although they also envisioned this being too much information for one toolkit. We decided to not directly implement these changes, as we first wanted to test if designers could grasp the basic concepts before adding extra elements like ethical considerations.

The data scientist evaluated the toolkit more in-depth. He indicated that no ML capabilities were missing, but that for completeness the data types *graph* and *vector data* should be added. This was considered before, but we decided for now to not include them to prioritize clarity over completeness, as these data types are rarely direct inputs from the real world, and hence less relevant for designers.

4.5 Interacting with the Mix & Match ML toolkit

Below, we describe one possible scenario of how the toolkit could be used in a session with a given design brief. The designers can start with exploring all types of tokens by placing them on the table and looking at which tokens can be combined and what these represent. To get more information, tokens can be placed on the sensing board. After exploring the toolkit, the designer can decide which types of data and which ML capabilities could work for their given design brief. The toolkit can be used to further define the concept by selecting specific data sets and models and by considering the capabilities and limitations of the chosen ML capability. Finally, when the concept is defined, they can get a first direction for how to realize the concept by looking for matching existing datasets and/or models. If their concept matches one of the example applications, they can also find a train-it-yourself link and more technical terms that can be used when searching for more information.

5 TOOLKIT EVALUATION

To evaluate if and how the Mix & Match ML toolkit could support designers in envisioning ML-enabled solutions, we deployed the toolkit in two different studies with design students. The aim of the first study was to validate the usability of the toolkit and evaluate how the toolkit could support the ideation process when students were already introduced to ML. In the second study, we evaluated the toolkit with design students who had little to no knowledge of ML to see if the toolkit provides sufficient relevant knowledge in an ideation process.

5.1 Study I: Method

5.1.1 Participants. Participants were recruited in a university Master elective on designing with ML. In total, eight students (n= 4 female, aged between: 18-34) participated, divided into three sessions. Session 1 (S1) had one participant, session 2 (S2) five participants and session 3 (S3) had two participants. The participants in each session were from the same group in which they worked during

[6]The open source code and instructions for the Mix & Match ML toolkit are available at https://github.com/MixMatchMLtoolkit

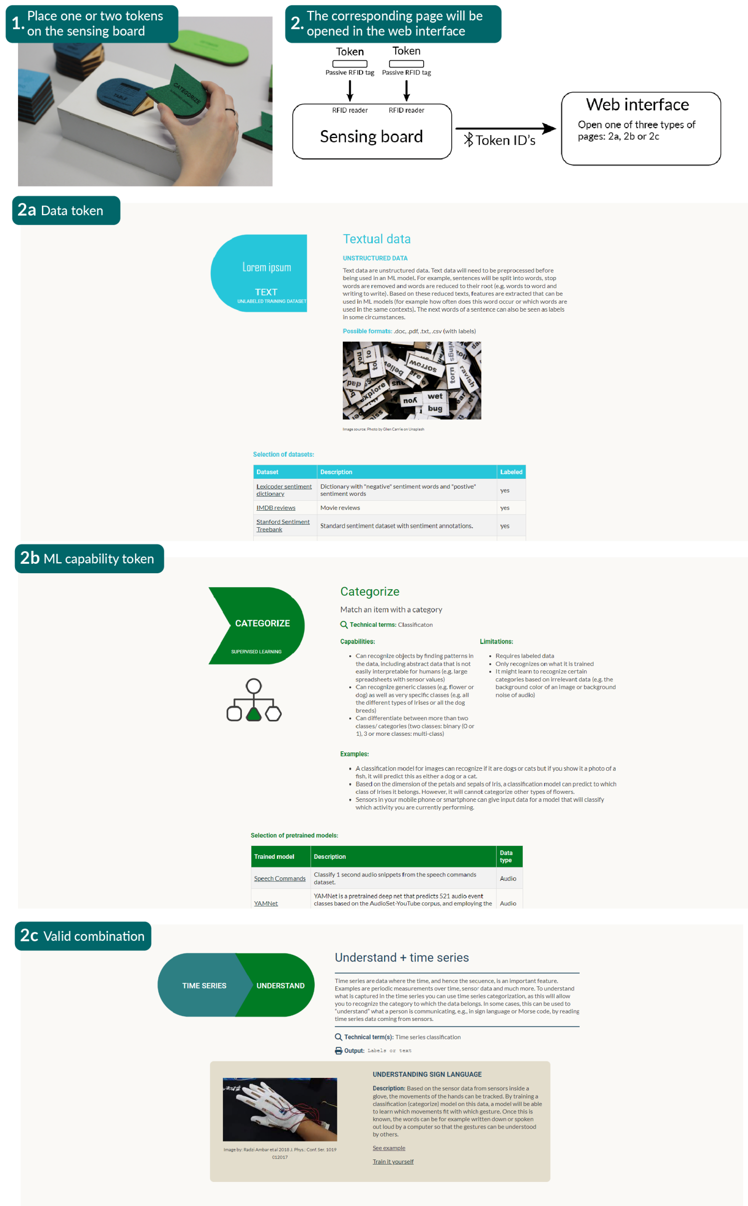


Figure 6: Based on the token(s) placed on the sensing board, one of the three types of pages is opened: the data page (image on this page by ©Glen Carrie on Unsplash), the ML capability page or the combination page (image on this page by ©Radzi Ambar et al 2018 J. Phys.: Conf. Ser. 1019 012017 [2])

the elective. In the elective, all participants had received an introduction to ML, a design case and a data set. Four participants had also followed one or more additional electives on design and ML.

5.1.2 Materials. During the study, the participants used the toolkit with all data type tokens and a selection of the ML capabilities tokens, since they had limited time. The following ML capabilities were included: categorize (sl), foresee (sl), cluster (ul), generate (ul), recommend (ul), optimize (rl).

5.1.3 Procedure. Participants were briefly instructed on how to connect their laptop to the toolkit. Next, they used the toolkit for

30-35 minutes for the design case given in the course. All groups already had one or more concepts in mind before entering the study session and they continued to work on these. At the end of the study, participants shortly presented their project and filled in a 7-point Likert scale USE questionnaire (USEQ) [32] to evaluate the toolkit usability. They were also asked open questions about their experience with the toolkit.

All sessions were audio recorded, and two sessions (S1 and S2) were video recorded as well, with a top-down view to analyze the interactions with the toolkit. Notes were taken during the sessions and these were completed with transcripts afterwards from the recordings. The study protocol was approved by the University Ethical Review Board and all data were managed in compliance with GDPR regulations.

5.1.4 Data analysis. Some studies were conducted in the local language and quotes from these studies were translated into English. We analyzed the quotes and observational notes through thematic analysis [10].

5.2 Study I: Results

5.2.1 Usability. In this first study, we evaluated the toolkit usability for participants with basic ML knowledge. Among the three sessions, we identified two main types of use. In S1, the toolkit was used as an opportunity for learning. In S2 and S3, the toolkit was used to explore existing and new concepts for the given design cases. As P2 said: *"It helped us identify our approach for the project that we're doing."*.

Regardless of the use, participants overall felt that the toolkit was easy and fun to use and made their task more efficient. For example, P7 highlighted its value in exploring alternative ideas: *"I easily and quickly explored the possibilities around our basic ideas, which I think without this tool would have taken one day of group working from us"*.

The results of the USEQ show that the toolkit scored well on all elements (usefulness M=4.91, ease of use M=5.97, ease of learning M=6.16, and satisfaction M=5.25), and especially high on ease of learning and ease of use. This result is also supported by the users' feedback. As P7 expressed: *"It is easy and fun, makes complex concepts easy to learn and apply if needed"*.

5.2.2 Use strategies. During the sessions, participants moved from many initial ideas to one concept. To achieve this, they used several steps and strategies. Some of these steps were directly supported by the toolkit, others required a workaround. It became apparent that the toolkit invited participants to start with defining the data type, as this is the first token in the reading order. The toolkit supported the selection of the data type by allowing to display all possible options on the table and to access additional information via the sensing board. Having all options in view also stimulated participants to consider alternative data types: *"What if we did not use tabular data, that could be interesting as well"* (P8). A recurring question during this step was whether data was labeled or not, and what this difference entailed. This was solved by switching between labeled and unlabeled data tokens. A direct comparison between two tokens of the same type was not supported.

Next, the discussion moved to what ML capability would fit the selected data type and their concept. Participants used the toolkit for communication and direct feedback during this step. By pointing at the tokens and making combinations on the sensing board, they would communicate to the other group members their concept while also getting feedback about whether the combination would be possible and what an example of that would be.

In the end, both groups in S2 and S3 created concepts that used two ML capabilities in sequence. For example, the concept generated in S3 first *foresees* the motivational level of a user and uses this prediction to *recommend* certain activities, exercises, and sports that fit with the motivation level of the user. The toolkit does not directly support this sequential use of ML capabilities, and in S2 they struggled with how to represent this. However, S3 tackled this problem by splitting the model into two separate combinations and placing them sequentially on the board.

5.3 Study II: method

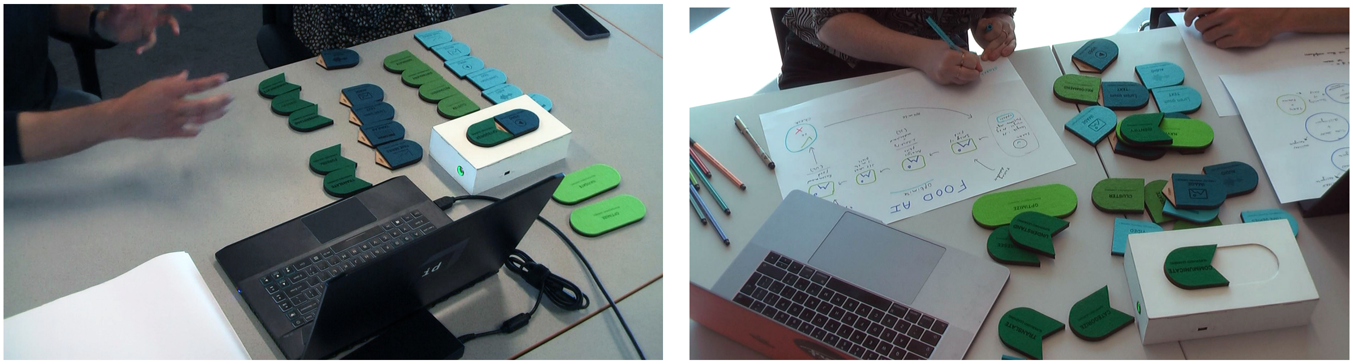
The first study showed promising results when students had already been introduced to ML beforehand. To explore if the toolkit could also be used without preliminary knowledge, we conducted a second study. The study aimed to explore the toolkit ability to support both learning and ideation. For learning, we investigated students' understanding of ML through the toolkit, as they were new to this subject. In particular, we evaluated if students would be able to explain the tokens and their underlying concepts (the "remember" and "understand" levels of Bloom's Taxonomy (BT) [28]), and if they would use these concepts to generate new ideas and describe them (the "create" and "apply" levels of BT [28]). For ideation, we investigated how the students used the toolkit to envision ML-enabled solutions to a design challenge.

5.3.1 Participants. Through convenience sampling, participants who self-identified as having little to no ML knowledge were selected within the design faculty of a technical university. In total, four students(n=2 female, age: 21-27) participated in pairs. The pairs were created randomly. The first pair consisted of a bachelor and master student, and the second pair of a master and a Ph.D. student. The students had never worked together.

5.3.2 Materials. During this study, the participants could use the complete version of the toolkit.

In the second phase - i.e. ideation, they received a design case with one central question on how to improve a situation with ML-enabled solutions. The design cases were designed to be open-ended as to not direct participants to one type of solution, and were inspired by existing projects of the researchers and recent news articles. Two different design cases were used for the two participating pairs, as one participant in the second pair was familiar with the first design case, and we wanted the design case to be novel. The first design case was: *How can we help clinicians provide personalized treatments for cardiac rehabilitation patients?* The second case was: *How to increase the joy of reading Dutch literature for young adults (aged 16-18)?* These design cases were presented to the students together with additional background information.

5.3.3 Procedure. The study was split into three phases. First, students were explained the goal of the study, and they were asked to



(a) Participants using the toolkit in the first phase, the learning phase, of study II

(b) Participants using the toolkit and sketching their ideas in the second phase, the ideation phase, of study II

Figure 7: Stills from the video recordings

freely explore the toolkit for 45 minutes without additional explanation (learning phase) (see Figure 7a). Help was offered only when connecting the sensing board to their laptop. After a short break, one of the two design cases was introduced to the participants. The participants had one hour to explore ideas and select one concept that they had to pitch in the end (ideation phase) (see Figure 7b). Finally, a semi-structured interview was conducted in which the understanding of the terminology used in the toolkit was tested and participants were asked to reflect upon their experiences. The session ended with the participants summarizing what they had learned.

All studies were audio-video recorded and transcribed afterward. The first session was conducted in the local language, and quotes from this session were translated into English. The study protocol was approved by the University Ethical Review Board and all data were managed in compliance with GDPR regulations.

5.3.4 Data analysis. All the interactions with the toolkit were logged to an online database and the data was cleaned to remove outliers. The data was aggregated for each participant pair and type of action, i.e., what type of page was opened and if one of the links on the pages were visited. The data cleaning and visualizations were done using Jupyter notebooks. The sessions were transcribed in their entirety and analyzed using thematic analysis [10].

5.4 Study II: Results

5.4.1 Learning outcomes. One of the open questions at the start of this study was what ML knowledge UX designers would need to ideate with ML and how this knowledge could be provided. As a possible solution, we proposed a framework of data types and ML capabilities.

To evaluate if participants achieved the learning goal of *remembering* and *understanding* the tokens and their underlying concepts, we asked each pair to explain both the ML capabilities and other ML concepts, i.e., labeled, unlabeled data and supervised, unsupervised and reinforcement learning, at the end of the session. Overall, these concepts were well understood. One misconception still present at the end of one session was that supervised learning meant supervised by humans, although they also explained that it needed labeled data.

Participants were able to provide definitions for all ML capability tokens, although some were incorrect or not complete. The difference between *categorize, identify* and *understand* was often unclear, as they were all understood as recognizing something. However, the analysis of the logs revealed that participants never explored all the three tokens separately, therefore they based their interpretation on the ML capability verb only. *Translate* was only defined in relation to translating text, missing the element of e.g., style transfer. *Distinguish* was considered harder to understand and sometimes seen as distinguishing between two items instead of looking at clusters or trends and finding outliers.

Finally, we were interested in assessing if participants would adopt the correct terminology when describing their ideas (i.e., *apply* level in BT). Participants indeed adopted the terminology introduced in the toolkit when describing and annotating their ideas. In Table 2, the generated ideas are reported, as they were described and/or annotated by the participants. The notes of each pair included the token combination for each idea, which are also reported in Table 2. They were not instructed to write down the tokens, but both pairs did. Pair 1 again shows some confusion between *categorize* and *understand*. Pair 2 adopted the correct terminology in their note-taking and talking, but also referred to some concepts using other names, often inspired by the example. For example, they used "chatbot" to capture the token combination of *text data* with *understand* and *communicate*.

5.4.2 Facilitating ideation. The main goal of the toolkit is to support UX designers during their ideation process. In Table 2, an overview of the ideas generated for the design cases is shown. Based on our data, we discuss three ways in which the toolkit supported the ideation process.

Firstly, having the physical tokens laid out on the table helped participants to see what was possible and to explore alternatives. This allowed them to either go broad and generate many quick ideas, as did pair 1 (see Table 2), or go deep. Pair 2 opted for depth,

Table 2: The ideas generated by the two pairs showing which tokens were used and a short description of the idea

Pair	Tokens used	Description of the idea generated
1	time series + categorize	Recognize certain patterns in heartbeat
	time series + distinguish	Recognize outliers in heartbeat
	table + foresee	Record your sleeping behavior pattern (going to sleep at what time) and predict based on that your quality of sleep
	time series + understand/foresee	Use the heartbeat to determine the type of sleep
	table + recommend	Compare the physical activity of the patient with healthy people and recommend what level of physical activity is healthy
	table + cluster	Cluster people with heart diseases and healthy people
	audio + categorize	Recognize in audio phrases like "I support you" to determine support of the family
	audio + categorize	Use audio recordings to recognize the emotions of the patient
	image + foresee	Making a photo of yourself everyday and it will recognize and predict how you are feeling
	image + understand	Recognizing what you eat based on a photo and determining if this is healthy or not
	image + understand + recommend	Recognize what you are eating and recommend a healthier alternative
2	tabular & text + recommend	Individual reading list: recommend books to read based on ratings, personal data and previous essays/reports
	text + generate	Generate genre-typical texts as a tool for understanding genre conventions (based on existing books)
	text + communicate	Reading journal chatbot: the chatbot prompts the user based on their interests (based on reading journals and book reports)
	text + communicate + recommend	Use input from reading journal chatbot to recommend new books

as they isolated four ML capabilities as potentially interesting and only used these for generating ideas.

Secondly, the toolkit knowledge contents facilitated ideation, particularly by providing examples of existing applications. These examples triggered new ideas, for instance during the ideation session of pair 2: *"But that made the communicate [ML capability token] also interesting, so we were really focused on the recommendation first and then [we] put 'communicate' there and then a whole new idea came because you saw there the chatbot."* (P12).

Finally, the toolkit enabled quick validation of initial ideas, allowing participants to rapidly check the feasibility of their concepts while generating them, without the need to interact with experts at this stage. Validation was enabled by two features: physical constraints and examples. The physical constraints of the tokens gave immediate feedback on the possible combinations of data and capabilities. Participants 'assembled' their preliminary ideas through the toolkit, by representing the building blocks of their concepts through the tokens. This gave them a sense of what combinations were possible and allowed them to test their preliminary ideas with no effort. For instance, P9 presented her idea of recommending healthy food based on images and to do so she 'assembled' the idea by combining *recommend* and *unlabeled image data* on the table. After the pair agreed on the concept being interesting, it was placed on the sensing board to be further explored.

In addition to leveraging physical constraints, ideas were validated by comparing them to the example applications included in the toolkit. For example, pair 1 generated a concept in which they wanted to recognize the food on a plate by taking a photo. They had seen the example of Google Lens before, and to verify if and how their idea was feasible, they searched for the combination of tokens that displayed that example and retrieved it for further analysis.

They later explained: *"[With the example] you know for sure, and with reading the description you weren't 100% sure if it would be something that could be done."* (P10). It is worth noting, however, that the final ML capability selected in annotating this concept was "understand" instead of "categorize", again showing a confusion between these terms.

5.4.3 Supporting multiple approaches. The two pairs adopted two different approaches when it came to learning and ideation, both being supported by the toolkit. Pair 1 used a very systematic approach of exploring almost all tokens in the learning phase and also exploring many token combinations in the ideation phase (see Tables 2, 3 and 4). To find potential interesting combinations, they adopted a data first approach and discussed how the data types offered by the toolkit could fit in their design case: *"Table, could that be useful somewhere?"* (P9). On the other hand, pair 2 worked backward, starting with the solution they envisioned. In the learning phase, this was inspired by their interests and other projects they were working on. Based on this, they chose what ML capability or data type would be interesting to look at it in detail, but they did not generate any solutions or concepts. In the ideation phase they started with the provided design case and again worked backward but this time generating new solutions. Once they envisioned a solution, they searched for the fitting ML capability and next the data: *"So the end point would be reading joy, and then we have all these things that we can do, then we can work back"* (P12). Using this

Table 3: The total number of interactions with the toolkit, the total number of tokens explored and the interaction time for both the learning phase and ideation phase of the two pairs. In the learning phase there were more interactions, more tokens explored and a longer interaction time

	Learning			Ideation		
	Actions	No. of tokens used	Interaction time (s)	Actions	No. of tokens used	Interaction time (s)
Pair 1	64	23 out of 24	35.1	46	17 out of 24	10.2
Pair 2	73	18 out of 24	27.6	25	7 out of 24	16.8

Table 4: The types of possible actions enabled by the toolkit and the frequency of each action in the second study. Overall, the participant pairs looked at combinations the most.

	Learning		Ideation	
Action	Pair 1	Pair 2	Pair 1	Pair2
Labeled data page	3	18	10	7
Unlabeled data page	14	6	6	2
Data link	0	0	0	0
Supervised learning page	11	10	6	2
Unsupervised learning page	7	6	3	2
Reinforcement learning page	2	4	1	0
Model page	0	1	0	0
Combination	27	20	20	11
Example link	0	3	0	0
DIY link	0	5	0	1
Total	64	73	46	25

approach, pair 2 explored only a small selection tokens during the ideation phase (see Tables 2, 3 and 4).

5.4.4 Benefits of a hybrid toolkit. In addition, having a hybrid toolkit resulted in the participants experiencing the task as fun. As P10 stated: *"To play around with the pieces is really nice, instead of just only a website"*. Moreover, it also made the collaboration easier: *"I do feel like, especially corporately, I did find this easier than navigating a screen interface with two people. [] we could both touch and interact with it without being like it is my laptop or what is your shortcut or your keyboard setup or like is it okay if I use your whatever. It was just very quick."* (P11).

5.4.5 Use of examples and visuals. During both phases, information in the web interface in the form of examples, images of datasets and illustrations for the ML capabilities proved to be most informative to the participants. While learning, the example applications made the concepts more concrete: *"It makes more sense if you can think of an example. Because like Google Lens we already know, then we were like oh 'Google Lens" then it uses these blocks, that makes more sense."* (P9). The participants themselves also applied this strategy by linking their own examples and experiences of ML applications to what they were reading.

The visuals and examples were also used in the ideation phase to quickly understand data and ML capabilities or to retrieve the information previously learned. During this phase, the interaction time was shorter (see Table 3) than in the learning phase and participants indicated not reading the description text. Moreover, the examples were also easier for the participants to recall concepts. During ideation, they would sometimes search for where they had seen an example that fit with their concept.

Finally, as mentioned above, the example applications were also a source of inspiration for the participants.

5.4.6 Mental framework for ML. Next to explaining the tokens and their underlying concepts, the toolkit also provided a framework for the participants' mental model about ML that could be used both for learning and ideation. As P11 expressed: *"I do find this a helpful way to think about this kind of framework, this kind of content. I definitely have some kind of takeaway that I appreciate. [...] It just provides a structured way of thinking about the possibilities that I find helpful"*.

6 DISCUSSION

In this work, we introduced the Mix & Match ML toolkit, which provides an overview of the ML possibility space to support designers in their ideation. Based on the experience of designing the Mix & Match ML toolkit and the results of the two user studies, we reflect on how the toolkit ML knowledge and design features support learning and ideation, and what elements should be improved or integrated.

6.1 ML knowledge framework for design

We evaluated the knowledge framework adopted in the toolkit through interviews with experts and two user studies with design students. The experts assessed the ML knowledge framework as a potentially effective way to represent ML to novices. Students in the two studies were able to quickly understand and operationalize the framework, both as a whole and on a more granular level (individual tokens). All data types were understood and students were able to determine the difference between labeled and unlabeled data using the toolkit. The majority of the ML capabilities (9/12) were correctly understood. Those poorly understood turned out to be interpreted based on the ML capability verb only, and not by exploring the contents of the web interface.

Further investigation is needed to determine if the ML capabilities should be revised and/or communicated differently in the toolkit, e.g., adding the definition directly to the token.

The knowledge framework also supported designers in their ideation processes. It offered a structured way to approach ML in real design challenges and made the knowledge concrete and applicable. In their ideation process and pitches, the participants naturally adopted the terminology of the data types and ML capabilities in their language.

6.2 Technical ML knowledge for ideation

An open question in the design and HCI community is what level of ML technical knowledge designers should possess [49]. With the Mix & Match ML toolkit, we provided designers with basic knowledge on the ML possibility space, instead of explaining ML from a technical perspective. At the same time, we introduced basic technical concepts and terminology such as labeled vs unlabeled data and supervised, unsupervised and reinforcement learning. This knowledge is helpful both to understand and familiarize with ML and to collaborate with ML experts in subsequent steps of the design process. The lack of a common language and the difficulty to collaborate with data scientists is another challenge identified by Yang et al. [48, 49]. The toolkit can be further tested to determine if it can be used by teams of designers and ML experts as a common ground to discuss ideas.

Overall, we argue that the toolkit provided sufficient knowledge to the participants to generate, formulate, discuss and evaluate ML-based concepts during an ideation session. It effectively addresses the lack of support in ML ideation identified by scholars [11, 48]. By using the toolkit, participants were able to outline preliminary ideas and perform an initial validation of their feasibility. The concepts generated using the toolkit describe the main solution functionalities and user interactions, together with what type of data and ML capability should be used.

6.3 Importance of visuals and examples

Design students in both studies made extensive use of the visuals and examples provided in the toolkit and commented that these were most helpful. ML can be perceived as not accessible by non-experts [43], therefore providing (familiar) examples and visuals helped students to move from thinking about ML in abstract terms to applying it concretely in their concepts. This is also in line with the findings of Yang et al. [48] that designers understand ML through designerly abstractions. Moreover, the application exemplars also appeared easier to recall, as designers referred back to those instead of specific combinations of data and capabilities, while discussing ideas. They also helped to understand the capabilities of ML, especially when the case study was familiar. Finally, students did not read all text during ideation, rather they used the visuals and examples to quickly get an understanding of the information or to recall it. Even though the participants mainly looked at tokens in combinations, to access exemplars, individual tokens helped them to gain knowledge on the lower-level concepts, e.g., the difference between labeled and unlabeled data. Therefore, the possibility to access information of individual tokens should be kept, especially to support the preliminary phases of exploration and learning.

6.4 Toolkit openness and multiple uses

One of the design rationales for the Mix & Match ML toolkit was that it should support exploration by having tokens that could be quickly exchanged and explored. Based on our experience in the design process and the findings, we believe that a toolkit for ideation with ML should not only allow exploration but should also support different types of use and be easily updatable. Our findings show that the toolkit supported different uses on multiple levels. First, it could be used for both learning and ideation. Second, within these two tasks, different strategies and approaches were supported by the toolkit, e.g., selecting data versus ML capabilities first to generate a concept, or following a breadth- versus depth-oriented ideation process. The toolkit modularity, flexibility, and open nature made it easier for designers to learn and ideate with ML following the approach they found easier or more effective.

6.5 Modularity and updatability

One functionality that the Mix & Match ML toolkit did not support, but should be included in a new design iteration of the toolkit, is the possibility to create sequential models. Existing applications make use of them and, to our surprise, the concepts created by students also featured ML capabilities in sequence, even though they can be considered more advanced. Making a modular toolkit can achieve the support of different types of uses and has the potential to facilitate sequential models. Moreover, ML is a broad and fast evolving field, making it essential that the toolkit is easily updatable and extendable to keep up with the state-of-the-art. Tokens can easily be added to the Mix & Match ML toolkit and the information in the web interface can be updated.

6.6 Potential of a tangible toolkit

We did not compare the Mix & Match ML toolkit to a digital counterpart, as our purpose was not to prove the advantages of a hybrid toolkit, but to show how UX designers could be supported by it during the ideation process. Nevertheless, our findings highlighted how the tangible aspect supported the users in learning and ideation.

6.6.1 Facilitating collaboration. The toolkit facilitated collaboration, as it made it easy for all participants to interact with the tokens and with each other. All participants could read the information on the screen and see the toolkit components being discussed. The toolkit also allowed all participants to track the current state of the work (i.e., the ideation process), in line with previous findings [34]. Having a tangible set of tokens also supported communication, as students would often point to tokens to clarify what they were saying. Furthermore, the tangible toolkit was considered by participants to make the interaction and collaboration more enjoyable.

6.6.2 Physicalizing ML knowledge. Representing ML concepts through physical tokens seems to be an effective way to lower the threshold to ML for designers. Making ML knowledge concepts visible, physical, and manipulable shortened the distance between the domains of design and ML. The use of a tangible and visual language, which designers appreciated, made ML concepts approachable and invited exploration.

6.6.3 'Assembling' preliminary ideas. Being able to grasp and manipulate ML concepts also facilitated the creation, discussion, and communication of ideas. Participants repeatedly explored their preliminary ideas by pre-assembling two or more tokens on the table before placing them on the sensing board. Concepts could therefore be represented, visually and physically, by means of the tokens. This practice of "assembling" ideas also facilitated their preliminary validation, as designers could check if two tokens fit together and hence if the right type of data was being used for the desired ML capability, and vice versa.

6.7 Implications for education

During the studies, the toolkit showed to have potential for reaching learning goals related to the "remember", "understand", and "apply" levels of the Bloom's Taxonomy [28]. While this is especially valid for the overall knowledge framework provided by the toolkit and most of the specific concepts, some limitations emerged in the ability to remember, understand, and apply certain ML capabilities. Participants also demonstrated the ability to "create" [28] new concepts based on the knowledge they acquired. However, the ideas generated in this workshop are in the form of preliminary concepts, and they are not able to demonstrate whether participants could create valid and feasible designs. More research is needed to further investigate this aspect.

This educational aspect was originally included to support the ideation process, but the learning sessions from the second study also showed the toolkit could function as a purely educational tool.

Based on these observations, we envision the possibility to use the Mix & Match ML toolkit to introduce design students to ML without relying on computer science or programming knowledge. In this instance, the toolkit will provide students with a basic understanding of the breadth of possibilities, some basic terminology and a mental framework for thinking about ML applications and their design. Using this approach, the dependency of data - one of the challenges identified by Dove et al. [14], is clear from the start. Building on this starting point, more in-depth knowledge of certain algorithms can be provided to students, as they will have a high-order idea to which they can connect specific algorithms, making it easier to learn and remember [22].

7 LIMITATIONS AND FUTURE WORK

The two studies were conducted with a small and homogeneous sample of users. Future studies with a larger and more diverse sample of potential users should be conducted to validate and expand our initial findings. It would also be valuable to see how the toolkit would be used by teams of both multidisciplinary students and practitioners. Analyzing and comparing the needs of UX designers with different experience and working in different contexts would extend the use of the Mix & Match toolkit even further. Moreover, the study did not investigate if design students would be able to further specify and design the ML components needed in their concept (e.g., specific ML models or dataset features) and/or if the toolkit could be a valuable asset for this task. Future studies should investigate this aspect as well as look into how ML experts could be involved in such a process.

While we tried to find the right balance between usability of the toolkit and complexity of ML knowledge, we are aware that the toolkit simplifies the topics and might not cover all ML possibilities. The toolkit is intended to support design ideation, but it requires interacting with ML experts to move from ideation to development. Nevertheless, although not comprehensive, the knowledge included in the toolkit and its structure were highly effective in allowing designers with no previous ML knowledge or experience to envision novel applications and ideas. More research is needed to validate, refine, and expand the ML capabilities and types of data, but our study indicates that this direction is worth exploring.

The balance between usability and complexity also resulted in the design decision of not including an explicit ethical component in the toolkit. We acknowledge the importance of this, as designers should understand, e.g., the risks of biased data/models. While we evaluated that no inherently harmful examples were included, we cannot ensure that no biases or discriminations are present. In the next iteration, we are including a disclaimer and warning on all data and ML capability pages to make users aware of these risks. In future developments, we also intend to include an additional information layer in the examples, where the generated concepts are evaluated with one of the already existing ethical toolkits [4, 8, 18].

The tangible aspect of the toolkit presents, in addition to the advantages discussed above, also some limitations in terms of scalability and adoption. Compared to a digital interface, a TUI needs to be built in multiple exemplars to be used by multiple teams. It requires time and resources to be created, which reduces its scalability. To partially address this limit, we provide detailed instructions to build the toolkit as additional material to this paper. Moreover, the use of a tangible toolkit requires teams to work in the same physical space. This might not always be possible, especially considering the recent shift towards smart and remote working. To overcome this limitation, the web interface could be modified to accommodate multiple remote users, each one operating a personal physical toolkit.

Based on the findings, we plan to improve the design of the Mix & Match ML toolkit on two main aspects. To facilitate the creation of sequential models, we envision the inclusion of connector tokens, which can act as *in-between data* and can create a link to a following ML capability. Moreover, the toolkit should facilitate the comparison of two tokens of the same type, either by modifying the sensing board, or by changing the web interface. Next to that, the interface can also be improved. Based on both the expert feedback and the findings of the studies, more visuals should be included in the toolkit or be given a more prominent role. Finally, participants preferred less text when using the toolkit for ideation, while that information was deemed useful during the learning phase. Therefore, we aim to create two use modes for the web interface, one for learning and one for ideation. In the ideation mode, parts of the text information will be hidden, and more prominence will be given to examples.

8 CONCLUSION

In this paper we introduced the Mix & Match ML toolkit, an ideation toolkit to design ML-enabled solutions. This hybrid toolkit provides designers with ML knowledge in the form of a set of tangible tokens and a web interface. The toolkit was deployed in two workshops with students to evaluate if and how it supports designers in their ideation process. Our findings show that the knowledge framework adopted in the toolkit was useful for designers to understand and learn ML basic concepts and to apply them to generate novel ideas. The tangible tokens lowered the thresholds to accessing and operationalizing ML knowledge, and supported designers' collaboration, while providing new ways to prototype preliminary concepts of ML-enabled applications. While many scholars have claimed the need to support design ideation with ML, this toolkit represents a first attempt to provide designers with an overview of the breadth of ML possibilities for ideation, by using a tangible approach that is more familiar, accessible, inviting, and enjoyable for its intended users.

Our work contributes to current research on ML in design and HCI in three ways. It provides a knowledge framework for designers to explore and learn basic concepts of ML to facilitate the ideation of ML-enabled solutions. It shows the potential of using a hybrid toolkit based on TUI principles to facilitate collaboration, learning, and ideation with ML. It provides open access to the Mix & Match ML toolkit, which can be used and customized by design researchers, educators, practitioners and students, to facilitate both learning and ideating with ML.

REFERENCES

[1] Mohammed H Alsharif, Anabi Hilary Kelechi, Khalid Yahya, and Shehzad Ashraf Chaudhry. 2020. Machine learning algorithms for smart data analysis in internet of things environment: taxonomies and research trends. *Symmetry* 12, 1 (2020), 88.

[2] Radzi Ambar, Chan Kar Fai, Mohd Helmy Abd Wahab, Muhammad Mahadi Abdul Jamil, and Ahmad Alabqari Ma'radzi. 2018. Development of a Wearable Device for Sign Language Recognition. *Journal of Physics: Conference Series* 1019, 1 (jun 2018), 012017. https://doi.org/10.1088/1742-6596/1019/1/012017

[3] Yali Amit, Pedro Felzenszwalb, and Ross Girshick. 2020. *Object Detection.* Springer International Publishing, Cham, 1–9. https://doi.org/10.1007/978-3-030-03243-2_660-1

[4] Artefact. n.d.. The Tarot Cards of Tech. Accessed on November 21, 2021 from http://tarotcardsoftech.artefactgroup.com/.

[5] Joan Palmiter Bajorek. 2019. Voice recognition still has significant race and gender biases. *Harvard Business Review* 10 (2019), 1–4.

[6] Ransome Bawack, Samuel Fosso Wamba, and Kevin Carillo. 2019. Where Information Systems Research Meets Artificial Intelligence Practice: Towards the Development of an AI Capability Framework. *DIGIT 2019 Proceedings* 2019, 5 (12 2019), 1–17. https://aisel.aisnet.org/digit2019/5

[7] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. *Proceedings of the International Conference on Advances in Neural Information Processing Systems* 29, 1 (2016), 4349–4356. https://proceedings.neurips.cc/paper/2016/file/a486cd07e4ac3d270571622f4f316ec5-Paper.pdf

[8] Ania Calderon, Dan Taber, Hong Qu, and Jeff Wen. n.d.. What are AI Blindspots? Accessed on November 21, 2021 from http://aiblindspot.media.mit.edu/.

[9] Kasia S. Chmielinski, Sarah Newman, Matt Taylor, Josh Joseph, Kemi Thomas, Jessica Yurkofsky, and Yue Chelsea Qiu. 2022. The Dataset Nutrition Label (2nd Gen): Leveraging Context to Mitigate Harms in Artificial Intelligence. https://doi.org/10.48550/ARXIV.2201.03954

[10] Victoria Clarke, Virginia Braun, and Nikki Hayfield. 2015. Thematic analysis. *Qualitative psychology: A practical guide to research methods* 222, 2015 (2015), 248.

[11] Sara Colombo and Camilla Costa. 2021. Can Designers Take the Driver's Seat? A New User-Centered Process to Design with Data and Machine Learning. *Blucher Design Proceedings* 9, 5 (2021), 435 – 446. https://doi.org/10.5151/ead2021-169

[12] Clifford De Raffaele, Serengul Smith, and Orhan Gemikonakli. 2018. An Active Tangible User Interface Framework for Teaching and Learning Artificial Intelligence. In *23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) *(IUI '18)*. Association for Computing Machinery, New York, NY, USA, 535–546. https://doi.org/10.1145/3172944.3172976

[13] Interaction design foundation. n.d.. Voice User Interfaces. https://www.interaction-design.org/literature/topics/voice-user-interfaces

[14] 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*, Vol. 2017-May. ACM, New York, NY, USA, 278–288. https://doi.org/10.1145/3025453.3025739

[15] Dheeru Dua and Casey Graff. 2017. UCI Machine Learning Repository. http://archive.ics.uci.edu/ml

[16] Rebecca Fiebrink. 2019. Machine Learning Education for Artists, Musicians, and Other Creative Practitioners. *ACM Trans. Comput. Educ.* 19, 4, Article 31 (sep 2019), 32 pages. https://doi.org/10.1145/3294008

[17] Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, Michael Mose Biskjaer, and Peter Dalsgaard. 2019. Mapping the Landscape of Creativity Support Tools in HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3290605.3300619

[18] futurice. n.d.. The Intelligence Augmentation Design Toolkit. Accessed on November 21, 2021 from https://futurice.com/ia-design-kit.

[19] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2021. Datasheets for Datasets. *Commun. ACM* 64, 12 (nov 2021), 86–92. https://doi.org/10.1145/3458723

[20] Google. n.d. Datasets. Accessed on January 11, 2022 from https://research.google/tools/datasets/.

[21] Jeff Hale. 2018. 7 Data Types: A Better Way to Think about Data Types for Machine Learning. Accessed on January 11, 2022 from https://towardsdatascience.com/7-data-types-a-better-way-to-think-about-data-types-for-machine-learning-939fae99a689.

[22] John Hattie and Gregory CR Yates. 2013. *Visible learning and the science of how we learn.* Routledge, London.

[23] Martin Heidegger. 1996. *Being and time: A translation of Sein und Zeit.* SUNY press, Albany.

[24] Tom Hitron, Yoav Orlev, Iddo Wald, Ariel Shamir, Hadas Erel, and Oren Zuckerman. 2019. Can Children Understand Machine Learning Concepts? The Effect of Uncovering Black Boxes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3290605.3300645

[25] Xiaoneng Jin, Mark Evans, Hua Dong, and Anqi Yao. 2021. Design Heuristics for Artificial Intelligence: Inspirational Design Stimuli for Supporting UX Designers in Generating AI-Powered Ideas. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 219, 8 pages. https://doi.org/10.1145/3411763.3451727

[26] Magnus Høholt Kaspersen, Karl-Emil Kjær Bilstrup, and Marianne Graves Petersen. 2021. The Machine Learning Machine: A Tangible User Interface for Teaching Machine Learning. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction* (Salzburg, Austria) *(TEI '21)*. Association for Computing Machinery, New York, NY, USA, Article 19, 12 pages. https://doi.org/10.1145/3430524.3440638

[27] Magnus Hoeholt Kaspersen, Karl-Emil Kjaer Bilstrup, Maarten Van Mechelen, Arthur Hjorth, Niels Olof Bouvin, and Marianne Graves Petersen. 2021. VotestratesML: A High School Learning Tool for Exploring Machine Learning and Its Societal Implications. In *FabLearn Europe / MakeEd 2021 - An International Conference on Computing, Design and Making in Education* (St. Gallen, Switzerland) *(FabLearn Europe / MakeEd 2021)*. Association for Computing Machinery, New York, NY, USA, Article 3, 10 pages. https://doi.org/10.1145/3466725.3466728

[28] David R. Krathwohl. 2010. A Revision of Bloom's Taxonomy: An Overview. *Theory into practice* 41 (2010), 212–218. Issue 4. https://doi.org/10.1207/S15430421TIP4104_2

[29] Yanhong Li, Meng Liang, Julian Preissing, Nadine Bachl, Michelle Melina Dutoit, Thomas Weber, Sven Mayer, and Heinrich Hussmann. 2022. A Meta-Analysis of Tangible Learning Studies from the TEI Conference. In *Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction* (Daejeon, Republic of Korea) *(TEI '22)*. Association for Computing Machinery, New York, NY, USA, Article 7, 17 pages. https://doi.org/10.1145/3490149.3501313

[30] Jie Lu, Dianshuang Wu, Mingsong Mao, Wei Wang, and Guangquan Zhang. 2015. Recommender system application developments: a survey. *Decision Support Systems* 74 (2015), 12–32.

[31] Yuwen Lu, Chengzhi Zhang, Iris Zhang, and Toby Jia-Jun Li. 2022. Bridging the Gap Between UX Practitioners' Work Practices and AI-Enabled Design Support Tools User Experience (UX), Human-AI Collaboration, design-support tools, data-driven design ACM Reference Format. In *CHI Conference on Human Factors*

in Computing Systems Extended Abstracts, Vol. 1. ACM, New York, NY, USA, 7. https://doi.org/10.1145/3491101

[32] Arnold M Lund. 2001. Measuring usability with the use questionnaire12. *Usability interface* 8, 2 (2001), 3–6.

[33] Nirav Malsattar, Tomo Kihara, and Elisa Giaccardi. 2019. Designing and Prototyping from the Perspective of AI in the Wild. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. ACM, New York, NY, USA, 1083–1088. https://doi.org/10.1145/3322276.3322351

[34] Paul Marshall. 2007. Do Tangible Interfaces Enhance Learning?. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction* (Baton Rouge, Louisiana) *(TEI '07)*. Association for Computing Machinery, New York, NY, USA, 163–170. https://doi.org/10.1145/1226969.1227004

[35] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A Survey on Bias and Fairness in Machine Learning. *ACM Comput. Surv.* 54, 6, Article 115 (jul 2021), 35 pages. https://doi.org/10.1145/3457607

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

[37] Piero Molino, Yaroslav Dudin, and Sai Sumanth Miryala. 2019. Ludwig: a type-based declarative deep learning toolbox. arXiv:arXiv:1909.07930

[38] C.P. Papageorgiou, M. Oren, and T. Poggio. 1998. A general framework for object detection. In *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*. IEEE, Bombay, India, 555–562. https://doi.org/10.1109/ICCV.1998.710772

[39] Iqbal H. Sarker. 2021. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science 2021 2:3* 2 (3 2021), 1–21. Issue 3. https://doi.org/10.1007/S42979-021-00592-X

[40] Bertrand Schneider, Patrick Jermann, Guillaume Zufferey, and Pierre Dillenbourg. 2010. Benefits of a tangible interface for collaborative learning and interaction. *IEEE Transactions on Learning Technologies* 4, 3 (2010), 222–232.

[41] Donald Schön and John Bennett. 1996. *Reflective Conversation with Materials*. Association for Computing Machinery, New York, NY, USA, 171–189. https://doi.org/10.1145/229868.230044

[42] Hong Shen, Wesley H. Deng, Aditi Chattopadhyay, Zhiwei Steven Wu, Xu Wang, and Haiyi Zhu. 2021. Value Cards: An Educational Toolkit for Teaching Social Impacts of Machine Learning through Deliberation. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 850–861. https://doi.org/10.1145/3442188.3445971

[43] Elisabeth Sulmont, Elizabeth Patitsas, and Jeremy R. Cooperstock. 2019. Can You Teach Me To Machine Learn?. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (Minneapolis, MN, USA) *(SIGCSE '19)*. Association for Computing Machinery, New York, NY, USA, 948–954. https://doi.org/10.1145/3287324.3287392

[44] Elisabeth Sulmont, Elizabeth Patitsas, and Jeremy R. Cooperstock. 2019. What Is Hard about Teaching Machine Learning to Non-Majors? Insights from Classifying Instructors' Learning Goals. *ACM Trans. Comput. Educ.* 19, 4, Article 33 (jul 2019), 16 pages. https://doi.org/10.1145/3336124

[45] Brygg Ullmer, Hiroshi Ishii, and Robert J. K. Jacob. 2005. Token+constraint Systems for Tangible Interaction with Digital Information. *ACM Trans. Comput.-Hum. Interact.* 12, 1 (mar 2005), 81–118. https://doi.org/10.1145/1057237.1057242

[46] Bram Van Der Vlist, Rick Van De Westelaken, Christoph Bartneck, Jun Hu, Rene Ahn, Emilia Barakova, Frank Delbressine, and Loe Feijs. 2008. Teaching machine learning to design students. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5093 LNCS (2008), 206–217. https://doi.org/10.1007/978-3-540-69736-7_23

[47] 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 CHI Conference on Human Factors in Computing Systems*, Vol. 2018-April. ACM, New York, NY, USA, 1–11. https://doi.org/10.1145/3173574.3173704

[48] 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*. ACM, New York, NY, USA, 585–596. https://doi.org/10.1145/3196709.3196730

[49] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376301

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

[51] Abigail Zimmermann-Niefield, Shawn Polson, Celeste Moreno, and R. Benjamin Shapiro. 2020. Youth Making Machine Learning Models for Gesture-Controlled Interactive Media. In *Proceedings of the Interaction Design and Children Conference* (London, United Kingdom) *(IDC '20)*. Association for Computing Machinery, New York, NY, USA, 63–74. https://doi.org/10.1145/3392063.3394438

[52] Abigail Zimmermann-Niefield, Shawn Polson, Celeste Moreno, and R. Benjamin Shapiro. 2020. Youth Making Machine Learning Models for Gesture-Controlled Interactive Media. In *Proceedings of the Interaction Design and Children Conference* (London, United Kingdom) *(IDC '20)*. Association for Computing Machinery, New York, NY, USA, 63–74. https://doi.org/10.1145/3392063.3394438

[53] Oren Zuckerman and Ayelet Gal-Oz. 2013. To TUI or not to TUI: Evaluating performance and preference in tangible vs. graphical user interfaces. *International Journal of Human Computer Studies* 71, 7-8 (2013), 803–820. https://doi.org/10.1016/j.ijhcs.2013.04.003