Can designers take the driver's seat? A new human-centered process to design with data and machine learning

Colombo, S.; Costa, Camilla

DOI
10.1080/14606925.2023.2279835

Publication date
2024

Document Version
Final published version

Published in
The Design Journal

The Design Journal
An International Journal for All Aspects of Design

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rfdj20

Can designers take the driver's seat? A new human-centered process to design with data and machine learning

Sara Colombo & Camilla Costa

THE DESIGN JOURNAL
https://doi.org/10.1080/14606925.2023.2279835

Check for updates

Can designers take the driver's seat? A new human-centered process to design with data and machine learning

Sara Colombo[a] and Camilla Costa[b]

[a]Delft University of Technology, Delft, The Netherlands; [b]Politecnico di Milano, Milano, Italy

ABSTRACT
Machine Learning (ML) is increasingly becoming a crucial asset across diverse industries. However, designers lack human-centered processes to envision and develop innovative solutions enabled by ML. By engaging in a Research-through-Design activity, we outline a new design process to generate human-centered adaptive systems enabled by data and ML. We describe and discuss the possibilities and limits of designing with ML, the need to concurrently address user experience and ML aspects, and the implications of their mutual influence. We argue that designers can envision and design human-centered ML-enabled systems if they acquire fundamental ML knowledge, although certain tasks necessitate close collaboration with ML experts. We discuss how uncertainty and risk of failure characterize the outlined process and may limit its applicability. The proposed process serves as a foundational framework for future research in human-centered design innovation through data and ML.

KEYWORDS
Machine learning, artificial intelligence, user experience, design process, design methods, human-centered design

Introduction[1]

Machine Learning (ML), a subset of AI, is emerging as a crucial asset for a growing number of companies across diverse industries. ML can improve product and service experiences, by enabling the creation of personalized, adaptive, and learning interactive systems (Lee and Shin 2020). UX design researchers have recently started exploring ML as a new design material to generate innovative experiences (Yang et al. 2018). However, designing adaptive or personalized solutions enabled by ML poses unique challenges to designers, due to difficulties in understanding AI capabilities, tackling unpredictable outcomes of AI models, and managing AI errors (Yang et al.

CONTACT Sara Colombo sara.colombo@tudelft.nl Delft University of Technology, Delft, The Netherlands
Supplemental data for this article can be accessed online at https://doi.org/10.1080/14606925.2023.2279835.

2020). To overcome such challenges, scholars have called for the UX and interaction design communities to build a research and education agenda and to develop new human-centered design processes for the integration of data and ML (Dove et al. 2017; Yang et al. 2018, Yildirim et al. 2023).

Despite the growing demand for ad-hoc methods, the design research community lacks processes to design ML-enabled products or services from a human-centered perspective (Yang et al. 2020; Yildirim et al. 2023; Zimmerman et al. 2021). This, among other factors, is preventing designers from participating in the ideation and development of ML-enabled solutions. Scholars highlight that, within companies, designers do not contribute to the ideation of user-centered innovative systems powered by ML (Dove et al. 2017). Rather, they are usually involved after functional decisions have been made by data scientists or engineers and the main features of the solution have been defined. Yang et al. (2020) outline two possible approaches for designers to be more closely involved in ML-based innovation. Firstly, designers can draw on datasets built on users' interactions with existing solutions, to gain familiarity with the experience, identify hiccups in the interaction flow and address them by AI-enabled features. Secondly, designers can leverage existing AI libraries or models to integrate certain functionalities in the UX flow. The authors also argue that 'what AI can do for a UX problem at hand becomes clearer once a functioning AI system is built' (Yang et al. 2020). The two highlighted directions suggest that designers are not expected to envision entirely new products/ services powered by *not-yet-existing* ML models.

Through this work, we aim to explore whether it is possible for designers to envision innovative ML-enabled solutions *before* the model, or the dataset, is available, to creatively explore and fully leverage the possibilities offered by ML. In this approach, a custom ML model is built throughout the design process, to enable the intended experience. We intend to investigate the possibility for designers to lead the creation of ML-enabled solutions and to define the requirements of a ML model according to their design goals. In particular, we aim to evaluate the viability of this approach, how it affects the traditional human-centered design process in terms of new design activities and skills, and what level of cooperation with ML experts is required.

We performed a research-through-design (RtD) process, where we engaged in the design of *Procrastinate no more,* a ML-enabled solution to reduce procrastination. We carried out the process as design practitioners, adopting a human-centered design perspective. Whenever needed, we were open to collaborate with ML experts. We reflected on our design process to generate new theory, according to the research-through-design methodology (Zimmerman, Stolterman, and Forlizzi 2010; Volonté, Rampino, and Colombo 2018). As a result, we outline a new human-centered process for the design of ML-enabled solutions.

The emerging process differs in many ways from a traditional design activity. We show how, starting from a typical human-centered design process, we were at some point required to work on a parallel, but interdependent, technical route, which greatly affected our design choices. We discuss how uncertainty surrounding ML (Benjamin et al. 2021) concretely affects a human-centered design process and we reflect on the knowledge designers need to acquire to operate in this field. Finally, we identify tasks in which the collaboration with ML experts is recommended, or even necessary.

Our work contributes to the growing corpus of knowledge on designing with data and ML by outlining a new human-centered design process, by providing practice-oriented insights on the complexity of such activity and by discussing the design implications of its iterative and uncertain nature.

Existing frameworks for designing with AI

The growing use of AI technologies in human-facing applications has led the design and HCI community to develop specific knowledge and tools for designers. Previous studies discuss guidelines for practitioners to design features and applications powered by ML and AI (Amershi et al. 2019) and describe toolkits to ideate ML-enabled solutions (Jansen and Colombo 2023). A practical framework featuring ML, users, and scenarios as co-creators was developed to bridge the gap between UX and ML (Zhou et al. 2020).

Yildirim et al. (2022) investigate how designers can recommend a broad range of AI innovations. They argue that designers require a general understanding of AI capabilities and need to prepare themselves to explore datasets, envision the data pipeline, and effectively communicate impact within cross-functional teams. In their view, boundary objects (e.g. flow diagrams, system maps, and service data blueprints) help designers and AI experts to establish a shared understanding and to prototype innovative solutions.

Windl et al. (2022) outline four approaches (a priori, post-hoc, model-centric, and competence-centric) for designers to include AI in their processes. Diverging from previous research, such approaches challenge designers to be involved in technical phases of AI craft, e.g. during data collection and model testing.

Finally, Zdanowska and Taylor (2022) draw attention to the influence of design across the lifecycle of AI/ML systems, such as the role of UX practitioners in the system-feature creation, their contribution to building and testing proof-of-concept prototypes, and to envisioning how AI/ML systems should gather data to learn.

Those studies provide different perspectives on how designers and ML experts can collaborate, yet they do not outline a practical process for designers to lead the creation of innovative ML-enabled solutions. In this work, we use a design case study, *Procrastinate no more*, to outline a new human-centered process for designing with ML and to uncover, document, and reflect on the practical challenges designers face when leading the design of a ML-enabled system.

Methodology

To investigate the practical challenges of designing with ML, we engaged in the design of *Procrastinate no more,* a mobile app that helps users procrastinate less by setting personalized daily goals through ML. We describe our design activities chronologically, to best show our reasoning and choices at every stage. We outline and visualize the process and we reflect on each step, to discuss its implications.

The design team consisted of the authors of this paper and included one design student and one design researcher, both with professional experience in UX and interaction design. We had basic-to-intermediate knowledge of ML, acquired through previous research, design, and educational activities. All interactions between the authors, and with experts and users, occurred remotely *via* video conferencing platforms, due to Covid-19 restrictions. The project lasted 9 months.

To document our process, we followed a *documentation perspective* approach (Shipman and McCall 1997; Dalsgaard and Halskov 2012). This involved documenting the extensive material produced during virtual meetings (e.g. virtual ideation boards) and the material produced for internal communications and alignment, when activities were performed asynchronously (e.g. slides summarizing literature findings). After each meeting or design session, a summary was generated including the main discussion points, decisions, and future activities. Emerging issues related to the use of ML were also annotated. The whole process was summarized in a project report and was periodically updated with new material, every 2–4 wk. The report served as the main document for the subsequent analysis of the process, which was performed together by the two authors after completing the project.

In this paper, we focus on the account of the design process. Therefore, the user research methods used to inform the design process (e.g. user interviews, user tests) are mentioned, but not fully documented.

Case study: Procrastinate no more

Design brief

Our design process, based on the double-diamond process (Design Council 2005), started with investigating the issue of procrastination and reviewing existing solutions through three activities. First, we performed a literature review on procrastination causes, behavioral patterns, and effects on well-being (Ariely and Wertenbroch 2002). Second, we performed 10 semi-structured interviews and a survey with 98 subjects to explore behaviors and feelings associated to procrastination. Our findings confirmed the existence of behavioral patterns and profiles related to procrastination, e.g. procrastinating for stress, laziness, or poor management skills. The identified profiles revealed common traits, including the need to prioritize activities, break down larger goals into more manageable components, and a lack of self-encouragement and motivation. Third, we analyzed existing solutions and discussed their pros and cons. Most of the available solutions focused on helping users prioritize or perform tasks. None of them considered the behavioral side of procrastination, or its underlying causes. They had a low level of personalization and were not designed to empathize with the user's psycho-emotional sphere. We learnt that users could be encouraged and motivated to procrastinate less through gamification and positive reinforcement (Amit et al. 2021). These insights led to generate a design brief focused on creating an interactive system to reduce procrastination by:

- setting personalized, adaptive, context-aware, and achievable daily goals;
- reinforcing positive behaviors and generating enjoyable experiences;
- increasing users' awareness about their procrastination behavior.

Step 1: ML outcome and solution ideation

The first requirement was key in our decision to use ML as a design material. We realized we could use ML to set personalized challenges as to activities to complete daily. One way to personalize such challenges was to set a custom goal each day, based on the user's *predicted* tendency to procrastinate on that day. We therefore decided to train a ML model that could predict users' tendency to procrastinate activities on a certain day. We called this prediction outcome *procrastination index* (PI).

The ML ability to predict the user's daily PI was meant to be a functional component of a more complex interactive system. We therefore shifted our focus onto ideating an interactive system that could feature our personalized-goal functionality (Step 1, Figure 1).

We generated a number of ideas, e.g. virtual assistant, ambient interface, mobile app, etc. We ranked them and we selected a mobile app, where the user would input a list of activities at risk of procrastination every day (*wish-list*), and the system would suggest how many activities the user should try to accomplish, based on the predicted procrastination index for that day. The goal should positively challenge the user, by stimulating them to be slightly more productive than what the system predicted. The solution would

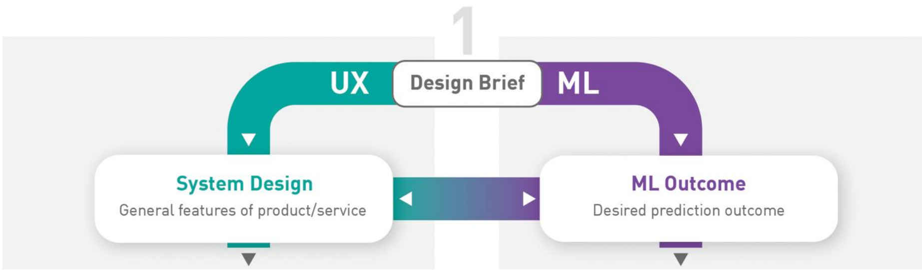

Figure 1. Step 1 of the process.

keep track of the user's progress, motivate them throughout the day, and reward them at the end of the day, if the custom goal was reached.

At this stage, both the system and ML outcome were delineated in general terms.

Step 2: UX design and data design

UX design

After outlining the overall solution, we started detailing the UX, by generating a preliminary wireframe and user interface (Figure 2). We envisioned the ideal UX in terms of main functionalities, number of screens, and interactions. However, we realized that to proceed further we needed more information on the ML model – the data it needed to operate and its specific outcome, i.e. how the PI was defined. Some UI elements were therefore temporary (e.g. Figure 3(A,F)). Moreover, during this phase, two UX elements were particularly difficult to outline: how to explain users the role of ML in setting personalized goals and its accuracy; how much we wanted users to trust our system – which would depend on the ML model accuracy and performance. The latter was a complex aspect to tackle, as we were not aware yet of the features and performance of the ML model, which did not exist yet. However, we realized that such technical aspects could impact the final user experience and should be carefully addressed by design. It became clear that the UX design could not be concluded in this stage but needed to be finalized after developing the ML model.

Dataset design

In parallel to the UX, we started the ML model design (Step 2, Figure 3). We decided to use supervised learning, structured data, and classification algorithms. A training dataset is made of one *class* (outcome of the prediction) and some *attributes*, which affect the prediction outcome. Originally, we intended to predict the daily PI (*low*, *medium*, or *high*) as our class. As we started designing our training dataset and how to populate it, we realized that predicting a daily PI could be too challenging. Therefore, we decided to

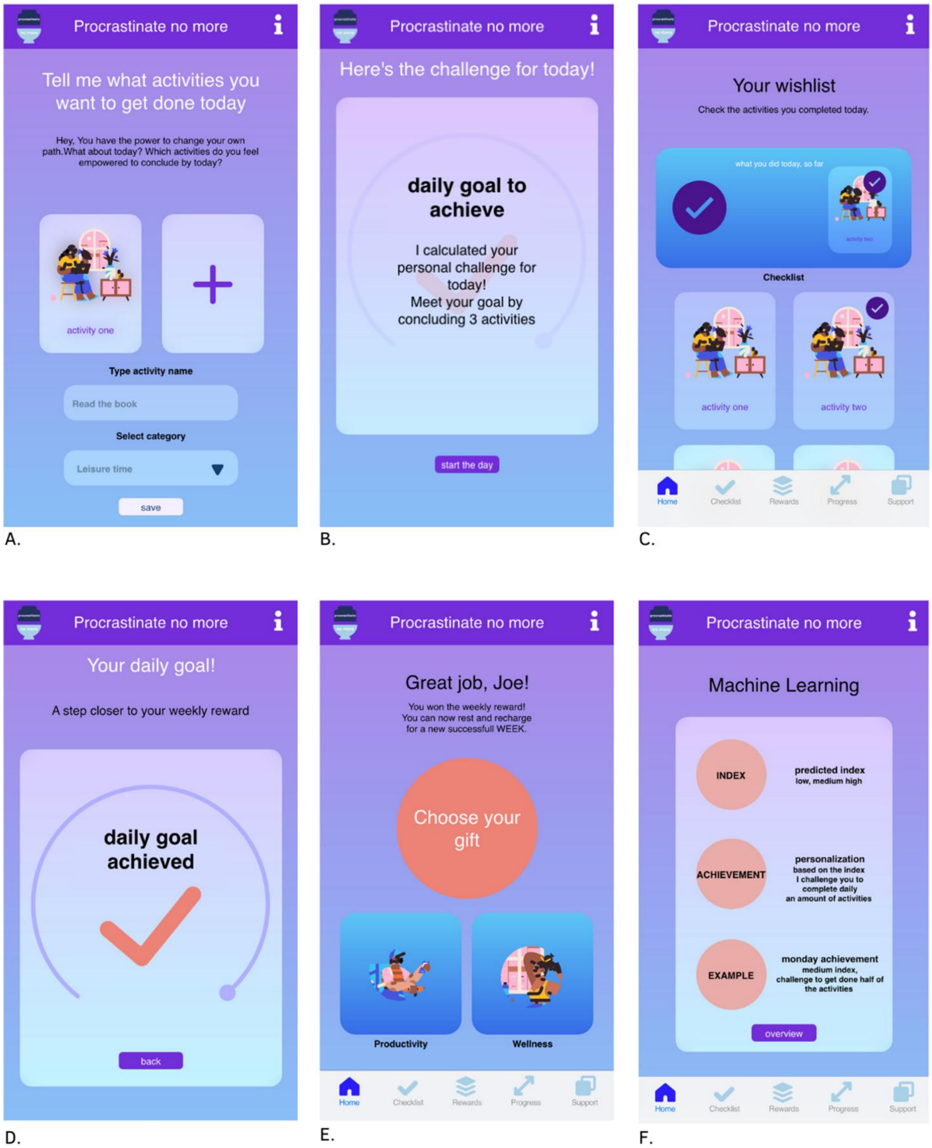

Figure 2. *Procrastinate no more*, example UI screens from the preliminary wireframe. (A) The user lists up to six activities they want to complete by the end of the day (wishlist), by adding their names and category; (B) The ML system predicts the daily PI and sets the daily goal accordingly; (C) The user marks the activities completed throughout the day; (D) When the day is over, the system rewards the user if they reached the daily goal. (E) At the end of the week, the system rewards the user with a gift to support their effort against procrastination. (F) The user can get more information on the ML operations.

predict whether each activity added to the wishlist would be procrastinated or not. The daily PI could be calculated afterwards, as a ratio between the activities predicted to be procrastinated and the total number of activities added to the wishlist. The prediction for each individual activity (procrastinated: yes/no) became the *class* in our dataset (Figure 4).

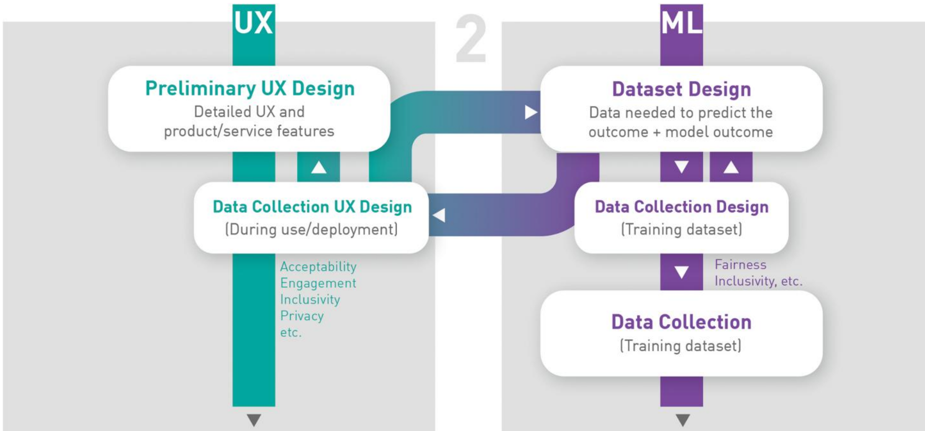

Figure 3. Step 2 of the process.

User personal details attributes					Wishlist attributes		Daily contextual data attributes				Class
User	Gender	Indolent	Stressed	Unorganized	Activity to procrastinate	Duration	Sleep hours	Vitality	Weather	Week/weekend	Procrastinated
1	male	3	2	2	dailyroutine	<15	<5	sedentary	sunny	weekday	yes
1	male	3	2	2	studentduty	>90	<5	sedentary	sunny	weekday	no
1	male	3	2	2	studentduty	>90	<5	sedentary	sunny	weekday	yes
1	male	3	2	2	leisuretime	60-90	<5	sedentary	sunny	weekday	yes
1	male	3	2	2	dailyroutine	30-60	<5	sedentary	sunny	weekday	yes
1	male	3	2	2	work	60-90	<5	sedentary	sunny	weekday	yes
2	male	3	2	2	dailyroutine	<15	<5	slightlyactive	cloudy	weekday	yes
2	male	3	2	2	homeduty	15-30	<5	slightlyactive	cloudy	weekday	yes
2	male	3	2	2	studentduty	>90	<5	slightlyactive	cloudy	weekday	yes
2	male	3	2	2	studentduty	>90	<5	slightlyactive	cloudy	weekday	no
2	male	3	2	2	leisuretime	60-90	<5	slightlyactive	cloudy	weekday	yes
2	male	3	2	2	work	60-90	<5	slightlyactive	cloudy	weekday	yes
3	male	3	2	2	dailyroutine	<15	5,6	sedentary	cloudy	weekday	yes
3	male	3	2	2	work	30-60	5,6	sedentary	cloudy	weekday	yes
3	male	3	2	2	studentduty	>90	5,6	sedentary	cloudy	weekday	yes
3	male	3	2	2	mandatoryduty	<15	5,6	sedentary	cloudy	weekday	yes
3	male	3	2	2	studentduty	15-30	5,6	sedentary	cloudy	weekday	no
3	male	3	2	2	leisuretime	30-60	5,6	sedentary	cloudy	weekday	yes
4	male	3	2	2	dailyroutine	<15	5,6	sedentary	cloudy	weekday	no
4	male	3	2	2	dailyroutine	15-30	5,6	sedentary	cloudy	weekday	yes
4	male	3	2	2	studentduty	60-90	5,6	sedentary	cloudy	weekday	yes
4	male	3	2	2	leisuretime	>90	5,6	sedentary	cloudy	weekday	yes
4	male	3	2	2	leisuretime	>90	5,6	sedentary	cloudy	weekday	yes
4	male	3	2	2	mandatoryduty	15-30	<5	sedentary	variable	weekday	yes
5	male	3	2	2	studentduty	>90	<5	sedentary	variable	weekday	yes
5	male	3	2	2	studentduty	>90	<5	sedentary	variable	weekday	yes
5	male	3	2	2	work	30-60	<5	sedentary	variable	weekday	yes
5	male	3	2	2	dailyroutine	<15	<5	active	variable	weekend	no
5	male	3	2	2	studentduty	>90	<5	active	variable	weekend	yes
5	male	3	2	2	work	60-90	<5	active	variable	weekend	yes
6	male	3	2	2	leisuretime	>90	<5	active	variable	weekend	yes
6	male	3	2	2	dailyroutine	<15	<5	slightlyactive	rainy	weekend	no
6	male	3	2	2	studentduty	>90	<5	slightlyactive	rainy	weekend	yes

Figure 4. A sample of the *Procrastinate no more* training dataset. The dataset is composed of 11 attributes and one class. Each instance corresponds to one activity in the wishlist. *Indolent*, *stressed*, *unorganized* represent procrastination profiles – participants could rate themselves from 1 to 5 on each profile. The complete dataset is available in the Supplementary material.

To proceed, we needed to define what *attributes* could influence our class – i.e. what data could help predict whether an activity would be procrastinated. We selected attributes based on our knowledge of the phenomenon under study - procrastination. The initial literature research had uncovered several elements that could affect the user's procrastination behavior, including contextual factors (e.g. day of the week, weather, activity type and duration) and personal factors (e.g. age and personality) (Steel, 2001).

We selected the following factors (attributes) for our model (Figure 4):

- User's profile data (age, gender, procrastinator profile);
- Daily contextual data (weather, weekend *vs* weekday, hours of sleep, level of physical activity of the previous day);
- Activity features (type of activity, duration).

The attributes selection was influenced by considerations on the feasibility of collecting datapoints for each attribute in building the training dataset.

Data collection for ML training

Structured training datasets for supervised learning are made of *instances* (rows in Figure 4). Instances are used to train the model on what attributes patterns determine a certain outcome. In our case, each instance corresponds to an activity and the class label indicates whether that activity has been procrastinated (yes/no).

After designing the dataset structure, we needed to collect *training* data. For each attribute of our dataset, we listed possible ways to gather the corresponding data. For instance, physical activity data could be collected from wearable fitness devices, step counter apps, or as manual inputs. Other data, such as the wishlist activities and the number of completed activities, should be collected as user inputs. However, we had multiple options there, too – chatbots, mobile app, or even simpler solutions such as online surveys. Due to time constraints, we opted for an online survey to collect all training data.

We asked 8 subjects, who self-identified as procrastinators, to fill in a survey every day, for 14 days. The survey simulated a digital diary, where users created a daily wishlist of 3 to 6 activities to complete and provided contextual and activity data. Starting from Day 2, participants had to report if the activities from the previous day's wishlist had been completed, to provide us with class labels (procrastinated: yes/no). Profile data were collected only once, on Day 1.

The resulting training dataset consisted of 306 instances (rows), each one corresponding to one activity added to the wishlist (Figure 4). This dataset initiated the training process of a ML model. Although we were aware that the dataset was limited in size, we decided to proceed with it and to add more instances later, if needed.

Data collection during use & UX design

The training dataset reflects the data needed by the ML model to predict the outcome *during use*, i.e. once deployed in the final mobile app. After designing the training dataset, we had to decide how to integrate the data collection in the overall app UX. For instance, we discussed if the wishlist activities should be collected as text or voice inputs, and in what steps of the interaction (e.g. in the homepage, or through push notifications).

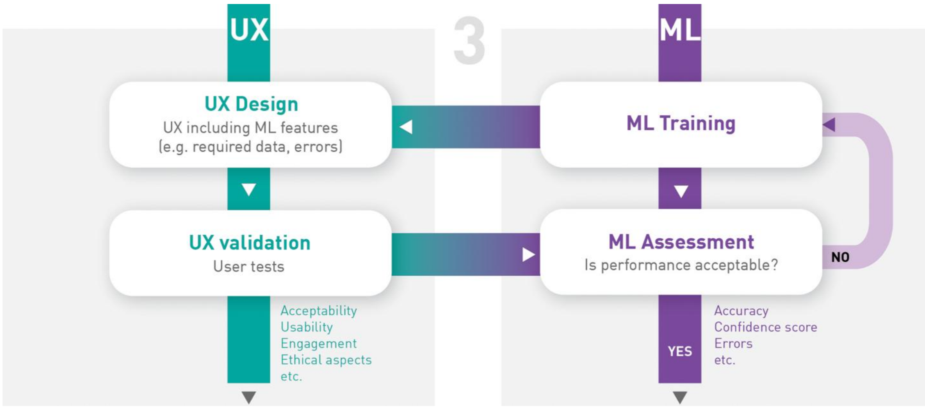

Figure 5. Step 3 of the process.

Designing the data collection experience *during use* was an essential part of the UX design process. At the same time, defining how to collect data through other sources (e.g. web APIs, connected devices, or sensors) gave a clear overview of the ecosystem that should be built for the ML model to work.

In this step, UX elements and ML features continuously influenced each other (Figure 3), which required us to repeatedly switch between a UX and ML perspective.

Step 3: ML training and user validation

ML Training

After populating and cleaning our dataset, we started the ML training phase (Figure 5). We used Weka[2] (Smith and Frank 2016), a tool for non-experts, to test different ML algorithms (i.e. One R, J48, Linear regression, and Naïve Bayes) on our dataset and to assess the performance of each resulting ML model.

ML Outcome assessment

None of the trained ML models reached a satisfying level of accuracy. Naïve Bayes performed better, but accuracy reached only 60%, possibly due to the limited number of instances in the dataset. A 60% accuracy indicates that the model can correctly predicts the outcome in 60% of the cases, therefore it is not reliable enough to be implemented in a real-life application. We agreed that such a low accuracy would negatively impact the user experience, and we decided to involve a ML expert to investigate possible ways to improve the model accuracy.

We decided to use a second tool, ILLMO[3] (Martens 2014, 2021), in collaboration with a data scientist and ML expert. We did not achieve better accuracy in the prediction, but we extrapolated some insights on i) the attributes that seemed to play a significant role in the prediction (i.e. user's gender, activity duration); ii) the fact that individual users have unique behavioral patterns, therefore it might be worth training the model on each user (which would require more training data); iii) false positives (i.e. predicting that an activity will be procrastinated, when in reality it will not) being much higher than false negatives.

By discussing the results with the ML expert, we found that they could potentially impact the UX on multiple levels. For instance, the data to collect *during use* could differ from the original training dataset, because some attributes did not seem to affect the ML outcome. Specifically, among profile data, only gender affected the prediction outcome, therefore other profile data became redundant. Likewise, we could ask users to report only the duration of the activity, since the activity type had less impact on the prediction. Moreover, different types of errors (false positives *vs* false negatives) could affect the UX and needed to be carefully considered. Indeed, false positives would result into less challenging daily goals, while false negatives could generate unreachable goals. For our scenario, we preferred false positives over false negatives, since we assumed the user would be less negatively impacted by an easier objective than a harder one. To achieve this goal, we were advised to optimize the model to minimize the number of false negatives, therefore limiting the negative impact of ML errors.

UX validation

Concurrently to the ML training, we updated the UX design to include some ML features (e.g. revised data collection and error types) and we built and tested a preliminary UX prototype with users (Figure 5). We performed 10 user tests to validate our system on the following aspects:

- *Overall design concept;*
- *UX aspects* (understandability, usability, engagement, willingness to provide feedback on ML prediction);
- *Ethical aspects* (transparency, trust);
- *Acceptability* (of data collection, use, and ML influence on users' behavior) and *pain points.*

Each user test lasted one hour and was performed remotely through a video conference platform by one author. Participants were asked to freely interact with the app prototype for 10 min, while thinking aloud. Afterwards, we performed a semi-structured interview investigating the above-mentioned aspects. The sessions were audio recorded and transcribed. Thematic analysis was performed on transcripts and notes. We report some of the insights that emerged, as examples of results that can be obtained from similar studies:

- 8 out of 10 participants were interested in explanations of the ML prediction, regardless of their ML expertise;
- 6 participants were not interested in seeing their (predicted) procrastination index every day, but wished to see a periodical report, to increase their awareness;
- 9 participants claimed they would not be negatively affected by false negatives (which would result in more challenging goals); instead, they would be motivated to be more productive.

Participants comprehended the ML-enabled goal personalization feature and found the service trustworthy due to clear data usage descriptions. Transparent communication was positively received compared to other ML applications where users may lack full awareness of data utilization. Such results provided useful insights to both improve the UX design and fine-tune the ML performance (Figure 5).

Step 4 and 5: UX refinement and ML implementation

After training the final ML model based on the user test outcomes, the UX design should be further refined considering both users' inputs and the performance of the final ML model (e.g. accuracy, error types, etc.) (Step 4, Figure 6). For instance, designers should focus on mitigating the consequences of ML errors and creating a system that can fail 'gracefully' (Google PAIR 2019). Subsequently, the ML model should be implemented in the UX prototype to perform final tests with users (Step 5, Figure 6). Based on the results, multiple iterations could occur on the UX or ML side. In our process, we did not engage in this activity, as we aimed to create a proof-of-concept and to validate it in Step 3.

Designing with data and ML: Overall process and implications

The process we followed is summarized and visualized in Figure 7. The diagram represents the iterative process that emerged from our RtD activity. It illustrates how the use of ML requires designers to engage in new activities and tasks, and it shows the relationship between UX and ML aspects. The dark green, dark purple and gradient paths show the main process. The light green and light purple paths indicate possible iterations.

The resulting process highlights some interdependences between ML and UX design tasks. We discuss the implications of such relationships.

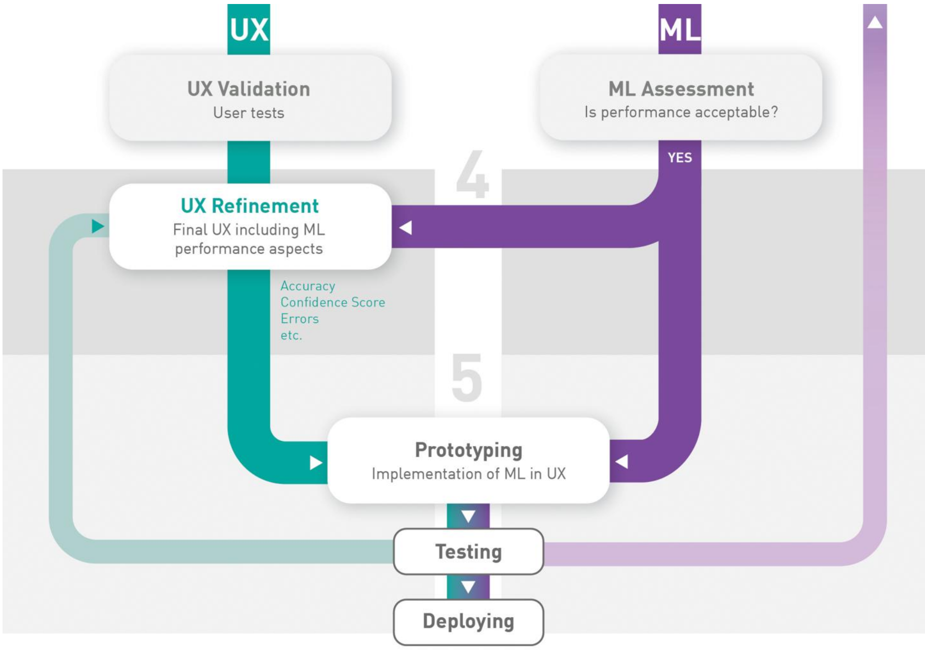

Figure 6. Step 4 and 5 of the proposed process.

Interdependence between solution ideation and ML outcome

In Step 1, we focused on two mutually influencing activities: envisioning the potential outcome of a ML model and designing the main features of an interactive system enabled by the (expected) ML outcome. These two interrelated activities can be carried out by designers, if they understand the capabilities of ML and its working mechanisms, and if they are aware of its technical limitations (Yang, Banovic, Zimmerman 2018; Yang et al. 2018). Such knowledge is necessary to avoid treating ML as 'magic' (Elish and Boyd 2017) by setting goals that are technically unachievable or too high risk (Yildirim et al. 2023). Although previous research shows that understanding the possibilities offered by ML is difficult for designers (Yang et al. 2020, Windl et al. 2022), our background knowledge helped us to envision how ML could provide unique functionalities to address our design brief. In particular, basic knowledge of ML algorithms and learning approaches, as well as examples of ML applications, were essential to imagine how to employ ML in our solution. Nevertheless, as elaborated in the Discussion section, more extensive knowledge and support could improve the quality and quantity of ideas.

Interdependence between dataset design and UX design

Designing the dataset in Step 2 turned out to be a delicate balancing act. Careful selection of attributes was crucial for building a reliable ML model; if

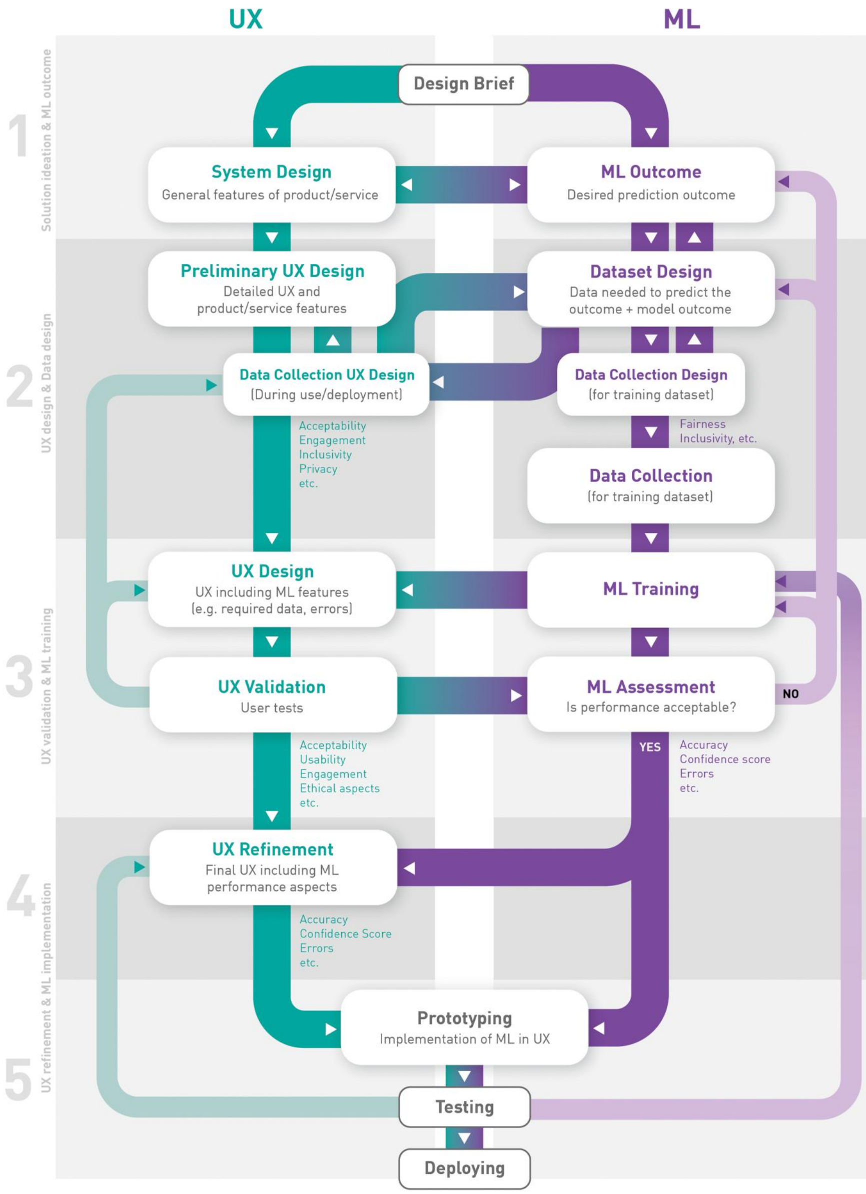

Figure 7. The iterative process for designing ML-enabled interactive systems with a human-centered perspective, as derived from our RtD activity.

the attributes lacked influence on the intended outcome, prediction accuracy would be compromised. However, we needed to assess whether and how data could be collected for each attribute, both for populating the training dataset and during use. These three elements (attributes, training data collection, and in-use collection) needed to be designed in concert, with UX considerations potentially affecting attributes selection.

Additionally, the UX design was influenced by changes in the training dataset that occurred in Step 3, after ML training. In Step 2, the UX was shaped by the initial structure of the training dataset, to ensure that vital input data (e.g. type of activity and duration) were collected through the UX. However, because the ML training revealed that only activity duration influenced the outcome, the dataset, and consequently the UX, were adjusted, adding one UX iteration in Step 3.

Due to the limited scope of our study, addressing bias, discrimination, or privacy concerns in our data collection was not feasible, although such considerations cannot be ignored (Paullada et al. 2021). Designers should envision data collection experiences ensuring the representation of different groups of users in the training dataset. Concurrently, they should design inclusive and accessible data collection experiences during use, to mitigate the risk of discrimination both in ML training and use.

Designers can uniquely shape the relation between dataset design and UX design by generating acceptable, inclusive, and engaging data collection experiences through a human-centered approach.

Interdependence between UX validation and ML training

UX validation in Step 3 may uncover issues that affect the ML training. For instance, users might feel uncomfortable towards certain data being collected, which would require re-designing the training dataset. That would, in turn, have a cascading effect on the data collection strategy (Step 2, Figure 7) and the ML training (Step 3, Figure 7). In theory, validating data collection with users at an earlier stage is possible, but our experience showed that the dataset design can be finalized only after training the model. Therefore, validations of preliminary datasets and collection strategies may be less relevant.

Users might also find the model performance unacceptable, which would require iterating back to the training phase, and likely expanding – or even re-designing, the training dataset, again affecting the UX design. Validating the UX *after* training the model also helps designers investigate the subtle relationships between ML performance, acceptability, and perceived benefit of the solution, which would be missed if the UX was validated earlier. Although validation in Step 3 is crucial for the reasons mentioned above, it demands a heavy investment on data collection and ML training, whereas the ML model can change due to the results of the UX validation. Therefore, it may be relevant to collect early user input in order to reduce the impact of late UX validations on the ML model. Preliminary assessments of some aspects of the solution may be possible, although they are not included in our process. For instance, the overall design concept, the users' reaction to potential ML errors, or the minimum accuracy required by users for certain applications could be tested in previous stages, by using tools such as Wizard of Errors (Jansen and Colombo 2022) or The Model Card Authoring Toolkit (Shen et al. 2022). Results could inform the ML model development early on. Further research on what user tests may be performed in different stages of the process could help to minimize the number of iterations.

Although iterating is common in design processes, when designing with ML such iterations have more far-reaching implications and cascade effects, which significantly increase the complexity of the process.

Discussion

ML Knowledge for UX designers

This article investigates how ML affects the human-centered design process when used as a design material. Drawing on the experience gained from our RtD project, we argue that, to execute similar processes, designers need to:

- *Acquire knowledge on ML capabilities, learning approaches, and main algorithms.* Without such knowledge, envisioning ML applications in our domain (Step 1) would not have been possible. However, our ML knowledge at the time led us to consider just a few options for ML application. In hindsight, we could have leveraged some ML capabilities other than *forecasting* procrastination, e.g. *recommending* some daily activities that could likely be concluded. However, we selected the capability and type of data (i.e. tabular) we were more familiar with.
- *Adopt toolkits or co-create to spark creativity.* To overcome the limitations due to the lack of ML knowledge, and improve the quantity and quality of ideas, designers can use ad-hoc toolkits (e.g. the Mix & Match ML toolkit by Jansen and Colombo 2023, or the resources proposed by Yildirim et al. 2023), and/or co-create with ML experts.
- *Learn to design with data.* Designing the dataset and planning data collection strategies in both ML training and deployment allowed us to stay in control of the user experience. Completely outsourcing this activity to other figures poses the risk of neglecting users' needs and expectations and designing data collection approaches that are not acceptable or engaging enough.
- *Understand ML performance aspects* (e.g. accuracy and errors). ML performance significantly influences user experience. To prevent users' frustration – or even harm, designers should learn to design UX elements that mitigate negative impacts and assess the effects of different ML performance levels.

- *Recognize ethical issues* and minimize them. By designing fair, inclusive, and unbiased data collection strategies, designer can help to minimise ethical issues. By validating the UX throughout the process, they can address transparency and accountability, to build trustworthy AI applications.

Dealing with uncertainty through Proofs of Concept

Throughout the process, we dealt with uncertainty in multiple occasions. In Step 1, we conceptualized a digital solution, not knowing if a ML model could actually provide the desired outcomes. To reduce uncertainty, we set out to train a ML model on a small dataset we designed and populated. In Step 2, we faced the entanglements of data, UX, and ML. To address uncertainty, we initially designed a preliminary UX relying on assumptions about the ML model performances and the required data. The design was refined only after the ML model was trained. In Step 3, several issues arose related to ML performance. We conducted user tests on a UX prototype, to assess how such issues could affect the UX and the overall acceptability and quality of the user experience.

The preliminary dataset (Step 2), the ML model (Step 3), and the UX wireframes and prototype (Step 2 and 3) turned out to be essential *Proofs of Concept (PoC)* (Zdanowska and Taylor 2022), which allowed us both to overcome uncertainties, and to discuss and validate our ideas with data scientists and users. Due to the potential need for substantial resource investments in these PoCs, additional risk assessment tools, like the impact-effort matrix (Yildirim et al. 2023), may be utilized for early evaluation of the quality and feasibility of concepts. However, these tools cannot fully anticipate and assess all risks comprehensively.

UX designers' and data scientists' roles

In this case study we, as designers, ideated and developed a ML-enabled solution independently in Step 1 and 2, with the support of a data scientist (DS) only in Step 3. We envision DS contribution continuing in the subsequent phases – testing, refining and deploying. Based on our experience, we reflect on what activities we believe can be led by designers, and what steps require collaborating with DSs.

In Step 1, designers can define the ML outcome and outline a solution responding to a design brief, if they have basic knowledge of ML and/or use ad-hoc toolkits. However, collaborating with DSs can improve ideation. In Step 2, designers can select a ML approach and design an inclusive and effective data collection strategy for ML training. This design activity expands the overview proposed by Windl et al. (2022) on the ways designers work with data. In this step, DSs may help to identify the best ML approach and to design the dataset accordingly. In Step 3, designers and DSs work on different goals: DSs can train and test the ML model, while designers validate the concept and UX. Although designers might train ML models with tools for non-experts, our case study shows that a technical support is highly beneficial, as it helps to build a working ML model, correctly assess the outcomes, and reflect on the impacts of technical aspects (such as error types) on the UX. In Step 4, designers finalize the UX based on technical aspects, by taking into consideration ML performance. Finally, in Step 5, they collaborate with DSs to integrate ML in the final prototype, for final user tests.

The proposed process does not aim to replace DSs in the early stages of design, yet it aims to challenge designers to take up new roles in ideating and designing ML-enabled systems. We are aware that it might not be possible for any designer to take the full lead of such a process due to the lack of vertical ML expertise. Nonetheless, we point out that designers may have much more freedom and opportunities than the ones currently presented in domain literature. We stress the importance for designers to familiarize with the challenges reported in this work, to improve the collaboration with DSs and to develop more human-centered ML-enabled solutions.

Throughout the project, we found that small datasets and preliminary ML models can be used as boundary objects (Yildirim et al. 2022) in collaborative practices. In our process, these objects were extremely helpful to interact with the DS and were effective in conveying our design concept, as well as its technical requirements. They also sparked ideas on possible uses of the dataset in different ML models, therefore providing the ground for in-depth discussions that bridged the design and technical perspectives.

Complexity and risks of designing with ML

Insights from our RtD process indicate that DSs and designers can mutually benefit from each other's expertise, collaborating effectively to mitigate the risk of failure. However, uncertainties and ambiguities characterize the process and increase the risk of failure or the need for multiple iterations, particularly in the following steps:

- *ML outcome definition*. Misjudging ML capabilities and identifying unrealistic outcomes may lead to failure;
- *Dataset design*. Unfitting, limited, or biased training datasets may yield inaccurate predictions;
- *Concept/UX validation*. Design concepts or UX may receive negative assessments from users, requiring a re-design of UX elements, data collection strategies during use, and/or training datasets;

- *ML training*. The ML model could be inaccurate, making it necessary to test other algorithms, to modify data – and consequently the UX – or even to abandon the project, if the expected performance is not achieved;
- *ML implementation and testing*. ML errors and low accuracy may render the UX unacceptable, requiring designing new ways to fail gracefully, or even re-training the ML model.

The cost of these iterations may be much higher than in traditional design processes, because any change in ML aspects can deeply impact UX elements – and vice-versa. As previously mentioned, PoCs and tools for early assessment of risks or ML performance may help to mitigate uncertainty, but they are either resource-intensive or still in their infancy (Yildirim et al. 2023).

Significance of the proposed human-centered process for designing with ML

The proposed process provides a concrete guide for designers to approach designing with ML. It does not replace, rather integrates existing design frameworks, tools and resources for designing with AI and gives them a chronological order, by positioning them within a more comprehensive human-centered design process. For instance, the 18 AI Design Guidelines (Amershi et al. 2019), the Mix&Match ML Toolkit (Jansen and Colombo 2023), and the resources to design with AI offered by Yildirim et al. (2023) concretely support designers during the ideation phase (Step 1). Google's PAIR guidelines (2019) provide examples and guidance to design for UX (Step 3). Yang et al. (2020), Elish and Boyd (2017), and Benjamin et al. (2021) address several challenges and questions related to ML limitations and uncertainty (Step 2 and 3). While some of these resources were published after our design activity took place, they can be effectively incorporated into various phases of the outlined process. Our work establishes a general framework for both current and future resources, emphasizing the procedural aspect of designing with ML. It advances the state of the art by describing in detail the steps and tasks that can be undertaken by designers – independently, or in collaborations with ML experts. Finally, it highlights the interdependencies between human-centered and technical factors, as well as specific design challenges.

Limitations

Our process was derived from the reported RtD study and from knowledge stemming from the authors' prior experiences in designing with ML, but it would benefit from further validation through more examples and cases. The dataset used to train our ML model was limited, and not inclusive; therefore, the preliminary insights reported here (e.g. relevant attributes) should be further validated. The small dataset resulted in low accuracy, which prevented us from implementing the model into a working prototype, and fully testing it.

Conclusions

Providing designers with new approaches to design with data and ML can foster human-centered design innovation and can facilitate the creation of adaptive systems that are not just optimized through ML, but fully built on its potential. We argue that such a process may give designers the ability to join from the start, or even lead, ML-based innovation processes within companies. However, it would be relevant to investigate the acceptability of the high-risk process we delineate, and to what extent companies might be inclined to adopt it, given the uncertainties it entails. Next to established companies, designer-entrepreneurs (Colombo, Cautela, and Rampino 2017) may also benefit from such an approach to generate innovative, adaptive, and personalized solutions, which could enable the creation of design start-ups hinged on data and ML.

The human-centered process for designing with ML outlined in this work is intended to provide a foundation for future research in this domain and to stimulate discourse on the role of designers in ML-based innovation. We encourage the design community to discuss, review, and build upon the proposed process, and to further explore how design can innovate through ML by leveraging its inherent potential.

Notes

1. This article presents an expanded, reworked version of a conference paper originally published in *14th International Conference of the European Academy of Design, Safe Harbours for Design Research*. See Colombo and Costa 2021 in the References list for full details.
2. https://www.cs.waikato.ac.nz/ml/weka/
3. https://researchoutreach.org/articles/illmo-a-new-platform-for-interactive-statistics/

Acknowledgments

The authors thank prof. Jean-Bernard Martens for his help with training ML models in the ILLMO environment with our dataset, and for his insights on the ML models performance and results.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Sara Colombo, PhD, Assistant Professor at the Faculty of Industrial Design Engineering, TU Delft. She holds a MS in Design & Engineering and a PhD in Design from Politecnico di Milano. Before joining TU Delft, she worked as an Assistant Professor at TU Eindhoven, a Research Scientist at Northeastern University, and a senior Research Associate and Lecturer at MIT. In recent years, she has extensively explored the intersection of AI and design, investigating AI applications for social good across diverse domains, and developing new knowledge and approaches for designing with data and AI. Her current work explores innovative methods and tools for designing human-centered and responsible AI.

Camilla Costa, Industrial and UX Designer. She received a Bachelor's degree in Industrial Design and a Master's degree in Design & Engineering from Politecnico di Milano. She is interested in how design can foster innovation and generate positive impacts on people's lives. Her current focus is design diagnostics solutions taking into consideration user needs and sustainability factors.

ORCID

Sara Colombo http://orcid.org/0000-0003-3902-1581

References

Amershi, S., D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen et al. 2019. "Guidelines for human-AI Interaction." In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems *(CHI '19)*, 1–13. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3290605.3300233

Amit, A. J., S. R. Gautam Shankararam, P. Pradeep, R. Perumalraja, and S. Kamalesh. 2021. "Framework for Preventing Procrastination and Increasing Productivity." In 2021 3rd International Conference on Signal Processing and Communication (ICPSC), Coimbatore, India, 228–232. https://doi.org/10.1109/ICSPC51351.2021.9451773

Ariely, D., and K. Wertenbroch. 2002. "Procrastination, Deadlines, and Performance: Self-Control by Precommitment." *Psychological Science* 13 (3): 219–224. https://doi.org/10.1111/1467-9280.00441

Benjamin, J. J., A. Berger, N. Merrill, and J. Pierce. 2021. "Machine Learning Uncertainty as a Design Material: A Post-Phenomenological Inquiry." In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*, Article 171, 1–14. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3411764.3445481

Colombo, S., C. Cautela, and L. Rampino. 2017. "New Design Thinking Tools for the Next Generation of Designer-Entrepreneurs." *The Design Journal* 20 (sup1): S566–S580. https://doi.org/10.1080/14606925.2017.1353004

Colombo, Sara, and Costa, Camilla. 2021. "Can Designers Take the Driver's Seat? A New User-Centered Process to Design with Data and Machine Learning." In *14th International Conference of the European Academy of Design, Safe Harbours for Design Research*. São Paulo: Blucher.

Dalsgaard, P., and K. Halskov. 2012. "Reflective Design Documentation." In Proceedings of the *Designing Interactive Systems Conference (DIS '12)*, 428–437. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/2317956.2318020

Design Council 2005. *The 'double diamond' design process model.*
Dove, G., K. Halskov, J. Forlizzi, and J. Zimmerman. 2017. "UX Design Innovation: Challenges for Working with Machine Learning as a Design Material." In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*, 278–288. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3025453.3025739
Elish, M. C., and D. Boyd. 2017. "Situating Methods in the Magic of Big Data and AI." *Communication Monographs* 85 (1): 57–80. https://doi.org/10.1080/03637751.2017.1375130
Google PAIR 2019. People + AI Guidebook. Pair.Withgoogle.Com/Guidebook.
Jansen, A., and S. Colombo. 2022. "Wizard of Errors: Introducing and Evaluating Machine Learning Errors in Wizard of Oz Studies." In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*, Article 426, 1–7. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3491101.3519684
Jansen, A., and S. Colombo. 2023. "Mix & Match Machine Learning: An Ideation Toolkit to Design Machine Learning-Enabled Solutions." In *Proceedings of the Seventeenth International Conference on Tangible, Embedded, and Embodied Interaction (T'I '23),* Association for Computing Machinery, New York, NY, USA, Article 8, 1–18. https://doi.org/10.1145/3569009.3572739
Lee, I., and Y. J. Shin. 2020. "Machine Learning for Enterprises: Applications, Algorithm Selection, and Challenges." *Business Horizons* 63 (2): 157–170. https://doi.org/10.1016/j.bushor.2019.10.005
Martens, J. B. 2014. "Interactive Statistics with Illmo." *ACM Transactions on Interactive Intelligent Systems* 4 (1): 1–28. https://doi.org/10.1145/2509108
Martens, J. B. 2021. "Comparing Experimental Conditions Using Modern Statistics." *Behavior Research Methods* 53 (3): 1240–1261. https://doi.org/10.3758/s13428-020-01471-8
Paullada, A., I. D. Raji, E. M. Bender, E. Denton, and A. Hanna. 2021. "Data and Its (Dis) Contents: A Survey of Dataset Development and Use in Machine Learning Research." *Patterns (New York, N.Y.)* 2 (11): 100336. https://doi.org/10.1016/j.patter.2021.100336
Shen, H., L. Wang, W. H. Deng, C. Brusse, R. Velgersdijk, and H. Zhu. 2022. "The Model Card Authoring Toolkit: Toward Community-Centered, Deliberation-Driven AI Design." In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*, 440–451. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3531146.3533110
Shipman, F. M., and R. J. McCall. 1997. "Integrating Different Perspectives on Design Rationale: Supporting the Emergence of Design Rationale from Design Communication." *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 11 (2): 141–154. https://doi.org/10.1017/S089006040000192X
Smith, T. C., and E. Frank. 2016. "Introducing Machine Learning Concepts with WEKA." In *Statistical Genomics: Methods and Protocols*, edited by S. Mathé and Ewy Davis, 353–378. New York: Springer. https://doi.org/10.1007/978-1-4939-3578-9_17
Steel, P., T. Brothen, and C. Wambach. 2001. "Procrastination and Personality, Performance, and Mood." *Personality and Individual Differences* 30 (1): 95–106. https://doi.org/10.1016/S0191-8869(00)00013-1
Volonté, P., L. Rampino, and S. Colombo. 2018. "The Specificity of Design Research: How Practice-Based Design Knowledge Can Enter the Great Archive of Science." In *Advancements in the Philosophy of Design*, 319–345. Cham: Springer.

Windl, M., S. S. Feger, L. Zijlstra, A. Schmidt, and P. W. Wozniak. 2022. "'It is Not Always Discovery Time': Four Pragmatic Approaches in Designing AI Systems." In *CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3491102.3501943

Yang, Q., A. Scuito, J. Zimmerman, J. Forlizzi, and A. Steinfeld. 2018. "Investigating How Experienced UX Designers Effectively Work with Machine Learning." In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*, 585–596. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3196709.3196730

Yang, Q., A. Steinfeld, C. Rosé, and J. 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 (CHI '20)*, 1–13. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3313831.3376301

Yang, Q., N. Banovic, and J. 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 (CHI '18)*, Paper 130, 1–11. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3173574.3173704

Yildirim, N., A. Kass, T. Tung, C. Upton, D. Costello, R. Giusti, S. Lacin, S. Lovic, J. M. O'Neill, R. O'Reilly Meehan, et al.. 2022. "How Experienced Designers of Enterprise Applications Engage AI as a Design Material." In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*, Article 483, 1–13. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3491102.3517491

Yildirim, N., C. Oh, D. Sayar, K. Brand, S. Challa, V. Turri, N. C. Walton, A. E. Wong, J. Forlizzi, J. McCann, et al. 2023. "Creating Design Resources to Scaffold the Ideation of AI Concepts." In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23)*, 2326–2346. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3563657.3596058

Zdanowska, S., and A. S. Taylor. 2022. "A Study of UX Practitioners Roles in Designing Real-World, Enterprise ML Systems." In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*, Article 531, 1–15. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3491102.3517607

Zhou, Z., L. Sun, Y. Zhang, X. Liu, and Q. Gong. 2020. "ML Lifecycle Canvas: Designing Machine Learning-Empowered UX with Material Lifecycle Thinking." *Human–Computer Interaction* 35 (5-6): 362–386. https://doi.org/10.1080/07370024.2020.1736075

Zimmerman, J., C. Oh, N. Yildirim, A. Kass, T. Tung, and J. Forlizzi. 2021. "UX Designers Pushing AI in the Enterprise: A Case for Adaptive UIs." *Interactions* 28 (1): 72–77. https://doi.org/10.1145/3436954

Zimmerman, J., E. Stolterman, and J. Forlizzi. 2010. "An Analysis and Critique of Research through Design: Towards a Formalization of a Research Approach." In *Proceedings of the 8th ACM Conference on Designing Interactive Systems (DIS '10)*, 310–319. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/1858171.1858228