Teaching-Learning Interaction: A New Concept for Interaction Design to Support Reflective User Agency in Intelligent Systems

Hankyung Kim
Department of Industrial Design, KAIST
Daejeon, Republic of Korea
hkkim31@kaist.ac.kr

Youn-kyung Lim
Department of Industrial Design, KAIST
Daejeon, Republic of Korea
younlim@kaist.ac.kr

ABSTRACT

Intelligent systems in everyday lives learn about their users to tailor services over time. However, these systems are often designed with little consideration of user agency on their learning processes, hindering users from taking full advantage of the systems. In this paper, we propose Teaching-Learning Interaction (TLI) as a new form of interaction that affords user agency by letting users reflectively shepherd an intelligent system's manner of learning. Given such agency, users will be able to better personalize services for themselves. We first draw on Schön's notion of knowing-in-action and reflective practice to theoretically ground our concept. We then present the resulting definition of TLI and three design qualities, which are further concretized with three design examples. We end with discussion on the implications of TLI for HCI design.

CCS CONCEPTS

• **Human-centered computing** → **Interaction design theory, concepts and paradigms**.

KEYWORDS

Teaching-Learning Interaction, interaction design, intelligent systems, user agency, personalization, Human-AI interaction

ACM Reference Format:

Hankyung Kim and Youn-kyung Lim. 2021. Teaching-Learning Interaction: A New Concept for Interaction Design to Support Reflective User Agency in Intelligent Systems. In *Designing Interactive Systems Conference 2021 (DIS '21), June 28-July 2, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3461778.3462141

1 INTRODUCTION

With the advances in artificial intelligence (AI), intelligent systems have increasingly been woven into the fabric of people's everyday lives. These systems promise personalized experiences by means of their ability to learn about various facets of each user, such as usual behaviors, emotions, attitudes, preferences and goals [14]. Such input is constantly acquired and interpreted by the systems over time to shape service outcomes that are more sharply tailored to individuals.

DIS '21, June 28-July 2, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8476-6/21/06…$15.00
https://doi.org/10.1145/3461778.3462141

This capability, however, has been increasingly blamed for creating frustrating user experiences [17, 77]. Systems constantly and often inconspicuously collect a variety of input, presume user needs, and automatically personalize services. While being seemingly "smart," this enigmatic behind-the-scenes learning mechanism has largely deprived users of their sense of agency, a feeling of being a relevant actor and the capacity to exercise control over interaction [4, 69]. As a result, users are prevented from exploiting the full value of personalization and are more likely to be manipulated by the systems. In many real-world intelligent systems that people use on a daily bases, such as Youtube or Facebook for example, users are frustrated by the intricacies of learning processes for making personalized recommendations, not to mention limited, system-centric interactions that hinder users from influencing those processes. Making the most of the given level of agency, users can only take temporary measures to exert control over the system, such as repeatedly "liking" or "hiding" items at best [44].

This has been a spur to various research on user empowerment in interaction with intelligent systems, including explaining and visualizing a system's inner working [21, 42, 54], adjusting user expectation [36], and diversifying control and feedback options [3, 18, 29, 50]. However, most of these solutions have been partially empowering users, while little has been discussed about how users can be bestowed with stronger agency from the outset and exert it across the continuous process of learning and the resulting outcomes. Considering that what an intelligent system in everyday contexts learns is all about its users, for the purpose of personalizing services for them, it is reasonable that the users should have a sound understanding of the whole process of learning and exercise control over it, however dynamic and complicated.

Admittedly, considering the nature of AI, it is impossible to make every detail in learning interpretable and straightly controllable, nor is such an approach what we propose. Rather, we suggest tackling the issue of user agency by designing interactions from a new viewpoint; we suggest designers overturn the conventional perspective that an intelligent system *learns about users by itself*. We suggest that users should be able to *consciously teach the system* based on their own judgment of what the resulting outcomes should be and what should be taught to elicit them. Considering such judgement tends to emerge and be constantly reconstructed in tandem with the users' actual interactions with the systems, we theorize this schema of interaction based on Schön's notion of knowing-in-action and reflective practice [60] and propose it as a new concept that we call Teaching-Learning Interaction (TLI).

In a broader sense, TLI builds on the notion of Interactive Machine Learning (IML) [2, 16], in which users actively drive the process of building, steering and evaluating machine learning models, but they differ in motivation; the focus of IML lies more at optimizing model-building processes, whereas TLI aims for ensuring user-empowered experience. In this respect, TLI also distinguishes itself from a seemingly related notion of Interactive Machine Teaching (IMT) [55, 64]. While IMT also treats users as teachers, its chief motivation of doing so is to better leverage human teaching capabilities and domain knowledge in making model building processes more efficient, rather than to support user agency in intelligent services in everyday contexts. Unlike IMT, it is the very primary aim of TLI to reinforce end-users' agency by letting them reflectively personalize everyday intelligent systems themselves.

In the remainder of this paper, we unpack the notion of TLI in more detail. We begin with a review of previous approaches for affording user agency in interactions with intelligent systems, showing how our work resonates with the emerging trend and positioning our contribution. We then draw from Schön's theory [60] to conceptualize the new form of teaching-and-learning interactions as TLI, delineating its design qualities and three design examples that demonstrate the important aspects of the new concept. Finally, we conclude by discussing the implications of TLI for HCI design. With this paper, we aim to illustrate the potential of TLI in helping designers to step away from a currently dominant techno-centric perspective and to take a different look at how interaction between users and emergent intelligent systems should be designed.

2 AFFORDING USER AGENCY IN INTERACTION WITH INTELLIGENT SYSTEMS

There have been a variety of previous approaches in HCI to afford user agency in interactions with intelligent systems. A vast literature has been dedicated to Explainable AI [1, 27], providing insights into how to support user control by promoting their understanding of a learning system's inner work logic. These include strategies for explaining [12, 20, 25, 39, 41, 42, 45, 54, 71, 75] and visualizing [21, 30] a system's decision, as well as scaffolding of system capabilities to let users inspect and learn about different system components that gradually unveil [32, 66]. Transparency however is not always the best nor feasible solution. Too much exposure of the complicated inner logic of a system can overwhelm users and increase users' cognitive overload [57], which runs counter to the fundamental purpose of using intelligent systems [33]. Moreover, developers sometimes face situations where they should strike a balance between interpretability and performance of a model, and designers deliberately choose to sacrifice transparency for the sake of other experiential qualities.

Apart from the above approaches, researchers have opened up rooms for users to participate in interactions more actively. Indeed, people have a natural desire to exert control over their environment as the dominant motivators of behavior [31]. Especially in case of intelligent systems, users provided with control and options can better understand the underlying mechanism and hence manipulate the systems to be better personalized [38]. Accordingly, solutions have been proposed to let users influence a system with a wider range of control mechanisms. These include designing richer forms of interactions for expressing preferences [11, 59], letting users polish their own user models [3, 53], and providing multiple algorithms and filtering options to choose [18, 28, 29, 44, 50] in a more sophisticated manner. In real-world applications, we see users are allowed to explicitly turn on/off inputs used for learning, such as pins viewed on Pinterest [52] or elements of a user profile generated on Google [13], so that they can decide whether those data should continue to influence personalized recommendations. Also, many personalized news feeds have started to ask for a specific reason behind a simple "Dislike" for a post [19]. Collectively, these approaches have widened opportunities for users to play a more tangible role in interaction.

Building on this trend and a growing need for human-centered machine learning [26], intelligent systems now co-steer the learning process together with their users. The most related work has been on IML [2, 16]. Introduced by Fails and Olsen [22], the term IML is now used to encompass various approaches where end users are tightly involved in an interaction loop of iteratively adding, modifying, and deleting data and/or features to improve a machine learning model [26]. A core characteristic of IML is that it emphasizes the user role as the main driver of interaction "with control over the high-level behavior of the system" [16]. IML is distinguished from the previously described approaches whose chief focus has been capturing data, or from other human-in-the-loop machine learning techniques (e.g., supervised learning or active learning), in that the emphasis of IML lies more in a partnership between users and systems. As users play such a pivotal role, recent works have emphasized the need for more human-centered design of IML processes [2, 76].

In summary, we see that attempts to reinforce users' controllability over intelligent systems have been evolving, from increasing system transparency, to enriching forms and channels of interaction to regulate a system, to letting users exercise influence on building learning models. Now users are treated as neither a passive service recipient nor a critic, but as a *primary driver of an overall machine learning process*. Building on this recent trend, we aim to theorize how interactions that grant users a *teacher* role to play will be able to afford user agency, particularly over intelligent systems in everyday contexts. Our work extends current research in that we focus on interactions that help *users* retain agency, distinguishing itself from most prior works that focus on the interactions that help *machine learners* capitalize on human capabilities and knowledge to be trained more efficiently.

3 TEACHING-LEARNING INTERACTION

Developing our concept of TLI, we set up by noticing the recent notion of *users being the main drivers in a learning loop* as *teaching*, thereby being motivated to frame users as *teachers* and look for clues from pedagogy. In fact, several previous works have viewed interactions between users and ML systems as "teaching and being taught" and drew inspiration from educational theories as well [72]. Thomaz and Breazeal [70] have been informed by Situated Learning Theory in their work on reinforcement-learning-based teachable robots. Bengio et al. [9] have been inspired by Skinner's idea of shaping in exploring the idea of curriculum learning for ML algorithms. Most recently, how children learn their first language has inspired interactive teaching for conversational AI [51]. The notion of IMT adds to this body of work by building on Constructivism

[55]. We also leverage teachability, but unlike these prior works, we do so differently so as to propose a new form of interaction that *empowers users* rather than machine-learners, especially in interactions with more mundane everyday systems. This led us to build our concept on a theory that puts teachers at the center of discussion instead of learner-centered ones. Particularly, we envision *reflective teaching* as a form of empowerment, supporting critically reflective attitudes of users and giving them the agency to decide upon what systems should learn to personalize services. This prompted us to examine the concept of reflective practice of teachers in Schön's work [60] and reframe it in the context of human-AI interaction.

In the following, we first outline how the theory informed the concept of TLI. We then characterize the new form of interaction through three design qualities: *interactivity for surfacing tacit needs and preferences*, *purposive trial and error*, and *triggers for defamiliarization*. Next, we concretize the concept with three illustrative design examples across diverse domains of everyday intelligent systems.

3.1 Reflective practice of teachers

Within the view of framing interaction with everyday intelligent systems as a form of teaching, we initially found that such interaction resonates with Schön's notion of knowing-in-action and reflective practice of teachers [60]. In brief, the theory states that teachers' practical knowledge is tacit, becoming aware and evolving when the teachers critically think about it while teaching.

Based on Schön, teachers' knowledge is rooted and embodied in action, hence *knowing-in-action*. It is spontaneous, intuitive, tacit actions, recognitions, and judgement that are internalized in teachers themselves. It cannot be explicitly stated as words or as reasoning, but is manifested in actual performances when the teachers are involved in a teaching situation. What surfaces such knowing-in-action and makes it evolve is *reflection-in-action*. Teachers tend not to think about what they know and how they teach when their teaching is so intuitive and spontaneous that it yields expected outcomes. However, when confronted with situations where their teaching has led to "surprises, pleasing and promising or unwanted" [60], teachers reflect on their practice *in the midst of* teaching. They intuitively interpret the unique and uncertain situation, attempt to discover what contributed to the outcomes, gain new implicit knowledge, and modify their teaching behavior to test it on the spot. Schön describes this process as "a reflective conversation with the situation" [60]:

> *"The practitioner's moves also produce unintended changes which give the situation new meanings. The situation talks back, the practitioner listens, and as he appreciates what he hears, he reframes the situation once again. (...) In this reflective conversation, the practitioner's effort to solve the reframed problem yields new discoveries which call for new reflection-in-action. The process spirals through stages of appreciation, action, and reappreciation. The unique and uncertain situation comes to be understood through the attempt to change it, and changed through the attempt to understand it." [60, pp.131-132]*

We find the above concept of reflective practice sheds light on how to concretize our idea of users teaching everyday intelligent systems. What users have to teach for everyday intelligent systems is abstract, varied, and subjective information about themselves such as personal needs, preferences, perspectives, or value systems. These, unlike domain knowledge, are internalized in each person, being too tacit to be reified. Those are also highly dependent on contexts, manifested differently in different situations and evolving dramatically from time to time. In this sense, users seem to be also *knowing in their actions*. When teaching an everyday intelligent system, it is therefore critical that users are prompted to constantly surface their tacit, flexible, context-dependent needs and preferences *together with their ongoing use of a system over time* so as to convey them to the system. This characteristic aligns with Schön's description of how knowing-in-action is made aware and improved upon by reflecting on it while teaching.

Taken together, we define TLI as **a form of interaction that affords user agency by enabling users to have a teacher's mindset and reflectively guide how intelligent systems in everyday contexts learn**. It enables users to frame a system as teachable and explicitly teach it to personalize services by themselves. The users are prompted to constantly discover and elaborate on what they want from a system and how they teach it *in the midst of* interactions. When given this agency to become aware of implicit needs and make their own judgement, the users would be better able to personalize the systems. Reflection is thus a central element that TLI capitalizes on to empower users.

Before going into the detailed design qualities, we underscore that while designing interactions that support reflection, it is critical for designers to have a human-centered mindset to prevent a system from being manipulative and inducing results that are not consciously intended by users themselves. Otherwise, the design will run against the fundamental motivation behind the whole idea of TLI. For instance, an intelligent system can act proactively, but the proactivity should be leveraged only to support users' discovery of how they want to use the system, e.g., helping users reflect on what they can and should do as teachers, as well as how their teaching will potentially impact the system.

3.2 Design qualities of TLI

Being informed by the above, we propose three design qualities of TLI. They speak to the aforementioned two key characteristics of teaching that we identify to be also vital in TLI: tacit knowing and critical reflection that help users be sensitized with and improve upon that knowing. While stemming from the theory, these qualities are also substantiated by insights from our own process of developing TLI design examples.

3.2.1 Interactivity for surfacing tacit needs and preferences. Referring to the characteristics of knowing-in-action, TLI also focuses on the tacit and context-dependent nature of user needs and preferences. TLI thus prompts users to **constantly externalize their implicit needs and preferences through actual interactions with an intelligent system**. In this sense, TLI-based systems are to be highly interactive to the extent that they enable users to become better aware of and fully convey their ever-changing needs. This entails providing explicit always-open channels for teaching and learning, as well as modalities and forms of interactions that are most suitable for externalizing needs and preferences (e.g., directly sketching how a sweater looks instead of textually describing it). Understanding how users actually (want to) teach systems and accounting for such expectations will be critical to designing these interactions of self-discovery to be natural and pleasant rather than of a nuisance, echoing the importance of human-centeredness in IML emphasized in prior works [2, 16]. Designers should carefully take into account various human considerations, such as an apt timing for interruption or input types that users prefer in different contexts, so as to prevent users from having negative user experiences of being data labelers.

Considering that knowing is sensitized and improved upon through reflection-in-action, one way of affording this quality can be designing interactivity that fosters reflection in teaching. For instance, a system can ask "why" questions twice to elicit deeper reflection, adopting a strategy that Lee et al. [40] found effective in helping people reflect on their needs beyond the first answers that they came up with. At times, the system can provide users with chances to ponder on their behaviors as teachers (e.g., "*Why do you think you got this service outcome? Why did you teach that way?*") or on themselves (e.g., "*Why do you like this service? Why does it matter to you?*") to discover any unrevealed needs and teaching strategies. In either case, the users get to mull over *in situ* and think on their feet to teach differently in pursuit of the intended outcome. Further clues might be found from prior works [61, 65] that suggested principles and strategies of designing for reflection. As an example, Slovák et al. [65], also drawing from Schön's theory, have found that the process of transformative reflection should be carefully scaffolded within experience rather than being triggered by data, suggesting several strategies for such scaffolding.

3.2.2 Purposive trial and error. Schön describes what teachers do as "on-the-spot experiments" [60], since they test a new way of knowing they gained "on the spot" to see if their new view is adequate and thereby improve on their teaching. Similarly, TLI prompts users to **improve the service outcomes by themselves through iterative "road tests" while they experience the service**. Through multiple trials and errors, users continuously communicate with a system to teach what they want, test out a particular way of personalization, experience the resulting service, and modify their teaching behavior if necessary. The system then accommodates the modifications and refines outcomes which the users again experience. In short, users make ongoing and incremental changes on the manner that a system is personalized as they experience the services. While such a series of trials and errors is experimental, it is not random but *purposeful* as the users undergo the iteration with a clear intention to attain a service that works best for them.

By its name, purposive trial and error means to let users understand that services are not complete from the beginning but are able to evolve based on their own cumulative teaching interactions. This quality is thus concerned with continuously informing users how their teaching behavior has been shaping services (e.g., what kinds of user actions made use of what types of data in what context). Of course, this does not mean to expose every detail of a system's working mechanism; this is difficult because palpable links between input and output are not always created in intelligent systems, and designers must also strike a balance with business interests in reality. Rather, the key here is to make users' impact on a system's learning salient so that users can critically think about their teaching behavior and try different strategies. For example, a system might provide a "timeline" that records when and how each component in a user model has been acquired, allowing users to inspect and modify it if necessary. Prior works on human-centered approaches to explainability [41, 45, 75] will also provide valuable clues for affording this quality.

3.2.3 Triggers for defamiliarization. "Unique and unstable" [60] situations trigger reflection-in-action that enables teachers to be sensitized with their ordinary practice and improve on it. Similarly, TLI **defamiliarizes a system's routinized manner of personalization, prompting users to discover novel ways of using the system anew**. This means to design a system to surface the rigid manner of how users have been teaching it and help them consciously call that into question, even when the users seem to be satisfied with the status quo. In this way, the system helps the users avoid getting into a rut and discover new service opportunities. In this respect, affording this quality may open up the possibility to overcome problems that arise from *over-personalization*, such as the filter bubble issue [48] in many recommender systems.

An important design choice will be when, as much as how, to spark defamiliarization. It requires time for service outcomes to be well personalized so that users get used to them, so this quality is more likely to be afforded after users have interacted with a system for an enough period of time. For example, a system can spark reflection when it detects that users have not been undertaking trials and errors for a long time. This is because the idle state might imply the users' satisfaction with the current state but also the opposite; personalization is stuck in a groove and does not interest them anymore. The system then can gently provide users with previously abandoned teaching strategies and compare those with the current ones, providing a room for deliberation on whether or not the current way of teaching is the best option.

3.3 TLI Design Examples

We illustrate three design examples to show how the above design qualities can be embodied in intelligent systems in everyday contexts. In order to demonstrate that TLI can be applied across diverse instances and domains, we present examples in a range of contexts including a recommender system (RS), a private autonomous vehicle (AV), and a virtual assistant. Because TLI is a newly theorized form of interaction, existing systems or services do not fully feature its design qualities. The following examples are therefore not prior design cases but example vignettes created by ourselves, which we believe can more effectively communicate our concept.

3.3.1 Example 1. Recommender System. RS is an especially appropriate domain for applying TLI, as caring for user agency is not just beneficial but necessary in order to prevent malicious use and ethical ramifications [63]. The application areas of RS are as diverse as entertainment contents, advertisements, navigation, and news feeds of social network services. These recommendations are all widely used in everyday lives, potentially shaping and influencing people's perspectives, practices, and routines beyond supporting a single decision-making. Such potentially manipulative power notwithstanding, users are currently interacting with RS without adequate awareness and exercisable control over them. Through the lens of TLI, RS can be designed to afford much more user agency:

> Amy has signed up for a new movie recommendation service. This service provides a feature called *TeachMe* (Figure 1). Always being easily accessible, this feature lets Amy teach anything that she thinks will be useful for getting recommendations. At the beginning of the service, Amy simply thinks any "nice" movies will do fine, so she enters several keywords to specify her conception of "nice," such as "Oscar nominee" or "big budget." As she uses the system, Amy further tries teaching various things on *TeachMe*, including changing the data used for making recommendations (e.g., adding/removing keywords), controlling their weights (e.g., "*Pay attention to what I watched . Don't give me anything similar to that.*"), or adding new data types (e.g., allowing access to her web search history, explicitly imparting new information such as "I have two teenage kids."). To support this trial and error, the system highlights what part of her teaching would have caused a particular recommendation when Amy hovers over it, or it gently displays previously discarded strategies to always be tried again.
>
> Later on, the system detects that Amy spends most of the time mindlessly refreshing the feed. Having learnt that Amy has teenage kids, the system presents her several movies that are currently popular among many teens, encouraging her to compare the "nice" movies from her kids' point of view with what she likes. This calls into question Amy's usual interpretations of "nice." She mulls over how she and people of similar age as her children perceive "nice" differently, and whether she has had a blinkered perception of "nice." Also, by getting to understand better the culture shared among her children's peers, she ponders on what it means to enjoy movies together with her children as well as how she should teach the system to get support. Amy begins to personalize the system to make fresh recommendations that reflect her renewed conception of "nice."

Here, **interactivity for surfacing tacit needs and preferences** is the *TeachMe* feature itself, which allows Amy to keep elaborating on what she perceives as "nice" movies. The existence of such an explicit teaching channel can help her to effectively discover and express her own rich, sophisticated teaching intention. This also enables **purposive trial and error**, along with the system's reminder of previously discarded teaching strategies and expression of how it has been learning. These altogether facilitate Amy's hands-on experimentation with various teaching strategies for personalizing and evaluating her feed. When the service becomes somewhat predictable, the system **triggers defamiliarization** by presenting a very unfamiliar feed (i.e., a collection of the movies popular among teenagers) that leads to perspective-taking. Amy now uses the system not merely as an ordinary recommender, but as a facilitator of communication between her and her children, which differs from a typical way of recommendation that matches solely with tastes and needs of herself.

3.3.2 Example 2. Private autonomous vehicle. Private AVs are another domain suitable for applying TLI. Unlike public transportation, the fact that those vehicles are privately-owned calls for a need of user experience that is tailored to individual owners. Also, it has been constantly reported that users have different autonomous driving preferences (e.g., driving speed, inter-vehicle distance, preferred

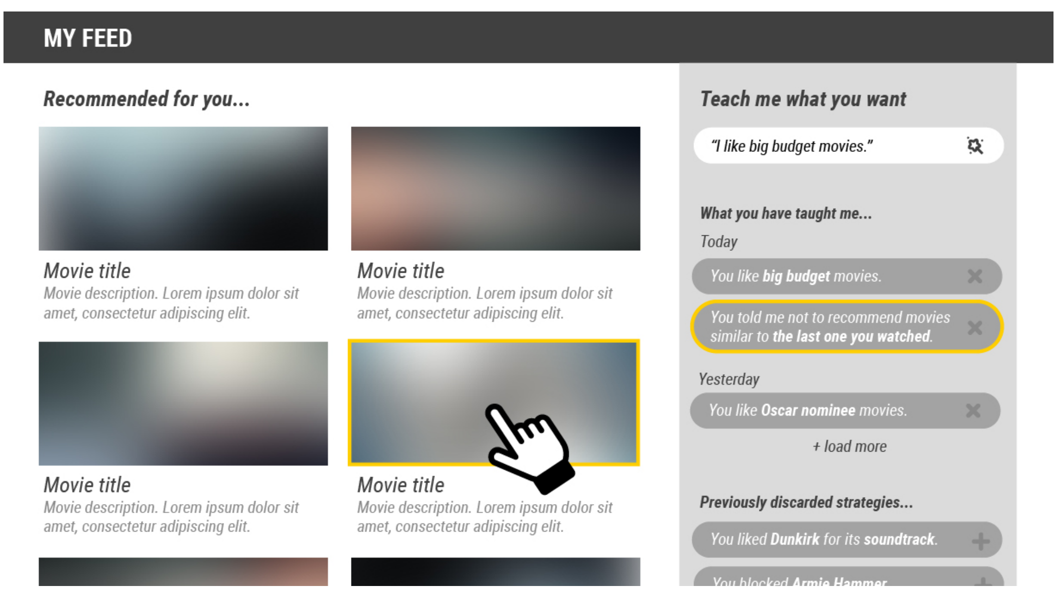

Figure 1: Interactive *TeachMe* interface design example. Amy can teach the system what she wants and simultaneously checks how the feed is being personalized in accordance with her teaching behavior.

road) [6, 15, 49]. Accordingly, learning about users and providing personalized services seem to be a promising way to enhance user experience in this domain. However, services must not be personalized based solely on preferences and needs; other factors that go beyond user control such as safety regulation and moral standards must also be considered here:

> During the first encounter with his new AV, Jack is asked to teach by demonstration how he usually drives so that it can personalize driving style based on that. After the first ride, the AV quickly finds that he is a bit short-tempered and drives recklessly. For example, Jack aggressively switches lanes to overtake cars too often. The AV tells this to Jack, saying: "*I've learnt how you drive, but I suggest a little more moderate style, for your safety. Of course, you can always teach me when to switch lanes.*" It suggests Jack compromise on several driving patterns, such as bypassing slow cars only, and Jack confirms it. As he experiences the AV's self-driving mode, he finds that he actually wants the car to drive even more cautiously. For instance, he tells his car to drive slowly and never bypass if there are any environmental risk factors (e.g., heavy rain, school zone), even if he would have ignored them while driving himself. In this way, Jack keeps elaborating on his definition of a "bypassing situation."
>
> One day, Jack gets to drive a long distance for a business trip. The AV detects his frequent interruption during its self-driving mode. When Jack is relaxing, it asks if he wants any updates in the current driving style: "*I drove as usual today, but you wanted to be at the wheel by yourself. Would you like to tell me why?*" Jack realizes he hasn't taught the system that he actually enjoys driving and the empty highway today was too nice to be wasted in self-driving mode. Jack tells the system: "*Well, I love driving. So let me take the wheel from time to time, especially when driving long distances.*" Based on this, the AV now suggests a new driving style: Jack being the main driver in collaboration with the AV while driving long distances. Jack likes the idea and starts to teach it in what ways he will need the AV's support as a second-in-command driver.

Here, **interactivity for surfacing tacit needs and preferences** is initial teaching by demonstration and on-the-go fine-tuning of driving style, which allows Jack to externalize his preferences while driving in situ. **Purposive trial and error** is supported by the AV's persuading Jack to take safety issues into consideration. The system's explicit explanation of what it does — e.g., in this example, the AV promotes safety — helps him understand why the system behaved in a certain way so that he has a clear idea of how to change or adjust the system or himself. In this example, Jack was able to polish the autonomous driving style to be safer by negotiating with the system. Later, the system **triggers defamiliarization** of the current driving service by detecting an anomaly (i.e., Jack's unusually frequent interruption) and asking Jack for its reason, which makes him reflect on why and discover a new service opportunity of collaborative driving.

3.3.3 Example 3. Virtual Assistant. This example demonstrates a case where a user teaches a system that learns from various sources of data and provides services for multiple domains. Being able to access other devices and applications, virtual assistants learn about users across different domains and selectively utilize what they learnt to shape services of various kinds. With more chances to influence such a process, users will be able to have control over service-level personalization.

> Mia has updated firmware on her AI speaker. She finds that the speaker can now provide personalized services if she teaches it well. She is asked to start by introducing herself, teaching the speaker anything that comes into her mind to describe herself: "*Well, I'm a barista. I live with two cats. I like unwinding from work by watching soccer live streams. I'm a Chelsea fan, by the way. What else… I like spicy food. I'm lactose intolerant...*" Based on what Mia has taught, the speaker suggests various services that she might like. When there is a Chelsea match, the speaker gets ready to order her favorite spicy noodles for dinner in advance. Or, it curates new dessert recipes that she might add as a new menu for her cafe.
>
> She likes most of what the speaker does for her, but sometimes she doesn't: "*Umm, no spicy noddles today. Give me something else.*" The speaker then asks: "*Can you tell me which part of the service you don't like? I acted upon your favorite soccer fixture and dinner menu.*" Mia then tells the speaker that she was going to watch the stream with her new boyfriend and he doesn't like noodles. While accepting this immediate corrective to the service, the speaker captures this as a chance for more learning. It asks: "*I'm sorry. I didn't know that. I'll look for something else except noodles. By the way, will you tell me more about him?*" Mia hesitates. She feels weird to talk about someone else like that to a machine. But at the same time, she thinks it will be interesting if her boyfriend can teach the speaker so that it can personalize services for both of them. Mia, pleased to find a solution, says: "*Actually, you can ask him directly when he comes over soon. Let's see what you can do for both of us.*"

Here, **interactivity for surfacing tacit needs and preferences** is Mia's self-disclosure through self-introduction and conversational correctives to personalized outcomes. What Mia tells the speaker in her self-introduction becomes learning materials, going beyond the conventional form of interaction in which a user can only command. Along with it, Mia goes through **purposive trial and error**, supported by the speaker's explanation for the service it provides. Especially in this example, the speaker explains what types of data are consolidated into a single service and how, so that Mia can easily spot which part needs more teaching. The system **triggers defamiliarization** by asking Mia to teach things about her boyfriend, not herself, which made her come up with a new idea of inviting her boyfriend as another user.

Throughout a series of examples, we illustrated how various design strategies can be employed to prompt the TLI design qualities. As for interactivity for surfacing tacit needs and preferences, systems were designed with triggers and channels for reflective interactions. To support purposive trial and error, cues to help people realize or newly discover their purposes of using a system were deployed. Also, systems were designed to defamiliarize the current state by intentionally escaping from the status quo or detecting unusual requests. We provide these as an initial point of consideration for designing TLI, hoping designers to further explore useful strategies for affording each quality.

4 IMPLICATIONS OF TLI FOR HCI DESIGN

So far, we have introduced the concept of TLI, its design qualities, and three examples of everyday intelligent systems designed for TLI across various domains. We hope our work would provide inspiration and research initiative for a more thought-provoking discussion among human-AI interaction designers to come up with novel design ideas and theoretical positions, rather than being considered as a solid form of prescriptive design guidelines, a method, or a process. In this light, the implications of our concept for design researchers and practitioners of future intelligent systems are as follows.

4.1 Enabling designers to re-examine the design space of human-AI interaction

TLI broadens the boundaries of a space for designing interactions in intelligent systems by sensitizing designers to reconsider what is meant by interacting with a system capable of learning. So far, the conception of interaction as a set of commands and actions has been so widely held. For example, they agonize over where to locate a "*Hide this*" button in a recommendation feed, without stepping back to speculate on whether commanding to hide an item is fundamentally the best way for a user to express disinterest. Instead, designers might let a user begin a service with setting a learning objective so that whenever s/he does not like a particular item, s/he can explain to the system why it is out of line with the objective.

We are not claiming that TLI should take over all the existing paradigms of designing interactions in intelligent systems. Instead, we believe viewing intelligent systems through the lens of TLI enables designers to focus on the open-ended, user-driven characteristic of those systems and realize how they are calling for a distinctive manner of interaction when their unique capability of learning is duly considered. We also believe our idea can produce synergy with other similar concepts for designing user-driven interactions with intelligent products and systems. For example, due to its characteristic of reciprocity, TLI might remind of the concepts of co-performance [34, 37] or co-adaptation [43]; they share the respect for the role of end-users, ensure users to remain in control [43], and shape performance together with artificial beings in everyday circumstances [34, 37]. We expect these concepts to complement and advance each other, encouraging designers to discover more creative ways to apply those into the design of everyday intelligent systems.

4.2 Inspiring the designs of new interaction and interface for everyday intelligent systems

The examples in this paper provide a glimpse of how simply reframing the conventional form of interaction into teaching and learning can already create new user experiences. Building on this potential, our idea opens up opportunities for HCI researchers and practitioners to come up with novel interaction and interface designs that transcend existing focus on system-centric data capturing, such as "*Choose what you like from here*" type of preference elicitation or "*Like/Not Interested*" type of post-hoc item feedback. Depending on service domains, every system will require its own teaching-learning methods and channels. As illustrated in our examples, in the context of AV, users might just want to "teach by doing" and let systems do the same thing for them, rather than verbally describing one by one what the system should know. In the case of virtual assistants, users might want to naturally introduce themselves and directly correct the service itself instead of selecting which apps to pair or deleting items from the activity log. For more insights, prior works on how to teach learning models [47, 67, 68, 73] and robots [35, 70] as well as principles for building an IML interface [16] might be consulted, but the strategies appropriate for guiding personalization in everyday contexts would be different from those used with the aim of building accurate models. Considering what TLI has initially drawn from, human teaching methods might be another inspirational source. For example, in the case of a system for multiple users, co-teaching strategies [23] in pedagogy might serve as a starting point.

4.3 Inspiring the designs of critically reflective interaction with intelligent systems

By reflectively teaching intelligent systems, we mean that users are invited to critical reflection, or "bringing unconscious aspects of experience to conscious awareness, thereby making them available for conscious choice" [61]. The importance of such critical reflection by users has been steadily appreciated by the design community, particularly around the works on reflective design [7, 8, 61]. Building on and extending these prior works, TLI opens a space for users to escape from the role of passive recipients of services in interactions with intelligent systems, to consciously think about values that matter to them, and to shape services to suit those qualities by themselves; in a way, users get to critically *design* their own services through TLI. Given such agency, users will become more sharply aware of many overlooked concerns residing in everyday intelligent systems. At the same time, the users will also become more sensitized with their own responsibility for making ethical choices as to input data or models steered while teaching the systems.

In this respect, we anticipate that, by designing through the perspective of TLI, designers come up with more ideas for supporting users in critical reflection while interacting with everyday intelligent systems. In this paper, we illustrated a few examples of such, including a system asking users reflective questions, "negotiating" with users, or capturing appropriate moments to make recommendations that can trigger reflective thinking. Devising more of these strategies will be the key for future TLI designs.

Many previous works on reflective design [7, 40, 61], provocative design [5, 10, 46, 56, 58], or design probes [24, 62, 74] have already suggested various ideas for promoting users' critical reflection, and designers might want to draw inspiration from those. For instance, a recent work has suggested a provocative chatbot that enables users to become critically reflective [58]. Considering that many intelligent systems adopt conversational interfaces, how this prior work has enabled provocation through conversations might be a valuable clue.

5 CONCLUSION

Building on the recent move to empower users in interactions with intelligent systems, we introduce a new perspective of viewing human-AI interaction through a concept we call Teaching-Learning Interaction (TLI). Starting with framing users as teachers, we theorized TLI as a form of interaction that affords user agency over intelligent systems in everyday contexts by enabling users to reflectively guide how a system should learn and personalize services by themselves. We developed this idea based on Schön's notion of knowing-in-action and reflective practice, and we outlined its definition, design qualities, and design examples that illustrate how the new style of interaction would support user agency. Our concept is primarily an initial suggestion of a new way of designing human-AI interaction; the main contribution of this paper is to call for a shift in perspective needed in terms of user agency in interaction with intelligent systems, while our future work will seek to prototype actual interactions and reify the proposed qualities as design guidelines. We hope this work can serve as a starting point that spurs diverse future research, including implementing new intelligent system design cases that apply TLI, as well as developing new interaction and interface designs for TLI.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No.NRF-2021R1A2C2004263).

REFERENCES

[1] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. 2018. Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3173574.3174156

[2] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the people: The role of humans in interactive machine learning. *Ai Magazine* 35, 4 (2014), 105–120.

[3] Fedor Bakalov, Marie-Jean Meurs, Birgitta König-Ries, Bahar Sateli, René Witte, Greg Butler, and Adrian Tsang. 2013. An approach to controlling user models and personalization effects in recommender systems. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. 49–56.

[4] Albert Bandura. 2001. Social cognitive theory: An agentic perspective. *Annual review of psychology* 52, 1 (2001), 1–26.

[5] Shaowen Bardzell, Jeffrey Bardzell, Jodi Forlizzi, John Zimmerman, and John Antanitis. 2012. Critical Design and Critical Theory: The Challenge of Designing for Provocation. In *Proceedings of the Designing Interactive Systems Conference* (Newcastle Upon Tyne, United Kingdom) *(DIS '12)*. Association for Computing Machinery, New York, NY, USA, 288–297. https://doi.org/10.1145/2317956.2318001

[6] Chandrayee Basu, Qian Yang, David Hungerman, Mukesh Sinahal, and Anca D Draqan. 2017. Do you want your autonomous car to drive like you?. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*. IEEE, 417–425.

[7] Eric P.S. Baumer. 2015. Reflective Informatics: Conceptual Dimensions for Designing Technologies of Reflection. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 585–594. https://doi.org/10.1145/2702123.2702234

[8] Eric P.S. Baumer, Vera Khovanskaya, Mark Matthews, Lindsay Reynolds, Victoria Schwanda Sosik, and Geri Gay. 2014. Reviewing Reflection: On the Use of Reflection in Interactive System Design. In *Proceedings of the 2014 Conference on Designing Interactive Systems* (Vancouver, BC, Canada) *(DIS '14)*. Association for Computing Machinery, New York, NY, USA, 93–102. https://doi.org/10.1145/2598510.2598598

[9] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. 2009. Curriculum Learning. In *Proceedings of the 26th Annual International Conference on Machine Learning* (Montreal, Quebec, Canada) *(ICML '09)*. Association for Computing Machinery, New York, NY, USA, 41–48. https://doi.org/10.1145/1553374.1553380

[10] Anders Bruun, Rikke Hagensby Jensen, Jesper Kjeldskov, Jeni Paay, Camilla Mejlby Hansen, Katarína Leci Sakácová, and Mette Hyllested Larsen. 2020. *Exploring the Non-Use of Mobile Devices in Families through Provocative Design*. Association for Computing Machinery, New York, NY, USA, 813–826. https://doi.org/10.1145/3357236.3395428

[11] Li Chen and Pearl Pu. 2009. Interaction design guidelines on critiquing-based recommender systems. *User Modeling and User-Adapted Interaction* 19, 3 (2009), 167.

[12] Hao-Fei Cheng, Ruotong Wang, Zheng Zhang, Fiona O'Connell, Terrance Gray, F. Maxwell Harper, and Haiyi Zhu. 2019. Explaining Decision-Making Algorithms through UI: Strategies to Help Non-Expert Stakeholders. 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, 1–12. https://doi.org/10.1145/3290605.3300789

[13] Monica Chin. 2020. How to change your Google advertising settings. https://www.theverge.com/21354028/google-advertising-settings-ads-how-to

[14] Elizabeth F. Churchill. 2013. Putting the *Person* Back into Personalization. *Interactions* 20, 5 (Sept. 2013), 12–15. https://doi.org/10.1145/2504847

[15] Nicole Dillen, Marko Ilievski, Edith Law, Lennart E Nacke, Krzysztof Czarnecki, and Oliver Schneider. 2020. Keep calm and ride along: passenger comfort and anxiety as physiological responses to autonomous driving styles. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[16] John J. Dudley and Per Ola Kristensson. 2018. A Review of User Interface Design for Interactive Machine Learning. *ACM Trans. Interact. Intell. Syst.* 8, 2, Article 8 (June 2018), 37 pages. https://doi.org/10.1145/3185517

[17] Malin Eiband, Sarah Theres Völkel, Daniel Buschek, Sophia Cook, and Heinrich Hussmann. 2019. When People and Algorithms Meet: User-Reported Problems in Intelligent Everyday Applications. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 96–106. https://doi.org/10.1145/3301275.3302262

[18] Michael D. Ekstrand, Daniel Kluver, F. Maxwell Harper, and Joseph A. Konstan. 2015. Letting Users Choose Recommender Algorithms: An Experimental Study. In *Proceedings of the 9th ACM Conference on Recommender Systems* (Vienna, Austria) *(RecSys '15)*. Association for Computing Machinery, New York, NY, USA, 11–18. https://doi.org/10.1145/2792838.2800195

[19] Essam El-Dardiry. 2019. Giving you more control over your Homepage and Up Next videos. https://blog.youtube/news-and-events/giving-you-more-control-over-homepage

[20] Motahhare Eslami, Sneha R. Krishna Kumaran, Christian Sandvig, and Karrie Karahalios. 2018. Communicating Algorithmic Process in Online Behavioral Advertising. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174006

[21] Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. "I Always Assumed That I Wasn't Really That Close to [Her]": Reasoning about Invisible Algorithms in News Feeds. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 153–162. https://doi.org/10.1145/2702123.2702556

[22] Jerry Alan Fails and Dan R. Olsen. 2003. Interactive Machine Learning. In *Proceedings of the 8th International Conference on Intelligent User Interfaces* (Miami, Florida, USA) *(IUI '03)*. Association for Computing Machinery, New York, NY, USA, 39–45. https://doi.org/10.1145/604045.604056

[23] Marilyn Friend and Lynne Cook. 1992. *Interactions: Collaboration skills for school professionals*. ERIC.

[24] William W. Gaver, Andrew Boucher, Sarah Pennington, and Brendan Walker. 2004. Cultural Probes and the Value of Uncertainty. *Interactions* 11, 5 (Sept. 2004), 53–56. https://doi.org/10.1145/1015530.1015555

[25] Fatih Gedikli, Dietmar Jannach, and Mouzhi Ge. 2014. How should I explain? A comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies* 72, 4 (2014), 367–382.

[26] Marco Gillies, Rebecca Fiebrink, Atau Tanaka, Jérémie Garcia, Frédéric Bevilacqua, Alexis Heloir, Fabrizio Nunnari, Wendy Mackay, Saleema Amershi, Bongshin Lee, Nicolas d'Alessandro, Joëlle Tilmanne, Todd Kulesza, and Baptiste Caramiaux. 2016. Human-Centred Machine Learning. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 3558–3565. https://doi.org/10.1145/2851581.2856492

[27] David Gunning. 2017. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA), nd Web* 2, 2 (2017).

[28] Jaron Harambam, Dimitrios Bountouridis, Mykola Makhortykh, and Joris van Hoboken. 2019. Designing for the Better by Taking Users into Account: A Qualitative Evaluation of User Control Mechanisms in (News) Recommender Systems. In *Proceedings of the 13th ACM Conference on Recommender Systems* (Copenhagen, Denmark) *(RecSys '19)*. Association for Computing Machinery, New York, NY, USA, 69–77. https://doi.org/10.1145/3298689.3347014

[29] F. Maxwell Harper, Funing Xu, Harmanpreet Kaur, Kyle Condiff, Shuo Chang, and Loren Terveen. 2015. Putting Users in Control of Their Recommendations. In *Proceedings of the 9th ACM Conference on Recommender Systems* (Vienna, Austria) *(RecSys '15)*. Association for Computing Machinery, New York, NY, USA, 3–10. https://doi.org/10.1145/2792838.2800179

[30] Chen He, Denis Parra, and Katrien Verbert. 2016. Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications* 56 (2016), 9–27.

[31] Jutta Heckhausen and Richard Schulz. 1995. A life-span theory of control. *Psychological review* 102, 2 (1995), 284.

[32] 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

[33] Kristina Höök. 2000. Steps to take before intelligent user interfaces become real. *Interacting with computers* 12, 4 (2000), 409–426.

[34] Da-jung Kim and Youn-kyung Lim. 2019. Co-Performing Agent: Design for Building User-Agent Partnership in Learning and Adaptive Services. 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, 1–14. https://doi.org/10.1145/3290605.3300714

[35] W Bradley Knox, Brian D Glass, Bradley C Love, W Todd Maddox, and Peter Stone. 2012. How humans teach agents. *International Journal of Social Robotics* 4, 4 (2012), 409–421.

[36] Rafal Kocielnik, Saleema Amershi, and Paul N. Bennett. 2019. Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-User Expectations of AI Systems. 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, 1–14. https://doi.org/10.1145/3290605.3300641

[37] Lenneke Kuijer and Elisa Giaccardi. 2018. Co-Performance: Conceptualizing the Role of Artificial Agency in the Design of Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173699

[38] Todd Kulesza, Simone Stumpf, Margaret Burnett, and Irwin Kwan. 2012. Tell Me More? The Effects of Mental Model Soundness on Personalizing an Intelligent Agent. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/2207676.2207678

[39] Todd Kulesza, Simone Stumpf, Margaret Burnett, Sherry Yang, Irwin Kwan, and Weng-Keen Wong. 2013. Too much, too little, or just right? Ways explanations impact end users' mental models. In *2013 IEEE Symposium on Visual Languages and Human Centric Computing*. IEEE, 3–10.

[40] Min Kyung Lee, Junsung Kim, Jodi Forlizzi, and Sara Kiesler. 2015. Personalization Revisited: A Reflective Approach Helps People Better Personalize Health Services and Motivates Them to Increase Physical Activity. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 743–754. https://doi.org/10.1145/2750858.2807552

[41] Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376590

[42] Brian Y. Lim, Anind K. Dey, and Daniel Avrahami. 2009. *Why and Why Not* Explanations Improve the Intelligibility of Context-Aware Intelligent Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 2119–2128. https://doi.org/10.1145/1518701.1519023

[43] Wendy Mackay. 2000. Responding to cognitive overload: Co-adaptation between users and technology. *Intellectica* 30, 1 (2000), 177–193.

[44] Nolwenn Maudet. 2019. Dead Angles of Personalization: Integrating Curation Algorithms in the Fabric of Design. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 1439–1448. https://doi.org/10.1145/3322276.3322322

[45] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267 (2019), 1–38.

[46] Preben Mogensen. 1992. Towards a Provotyping Approach in Systems Development. *Scand. J. Inf. Syst.* 4, 1 (1992), 5.

[47] Felicia Ng, Jina Suh, and Gonzalo Ramos. 2020. Understanding and Supporting Knowledge Decomposition for Machine Teaching. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 1183–1194. https://doi.org/10.1145/3357236.3395454

[48] Eli Pariser. 2011. *The filter bubble: What the Internet is hiding from you.* Penguin UK.

[49] So Yeon Park, Dylan James Moore, and David Sirkin. 2020. What a Driver Wants: User Preferences in Semi-Autonomous Vehicle Decision-Making. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[50] Denis Parra and Peter Brusilovsky. 2015. User-controllable personalization: A case study with SetFusion. *International Journal of Human-Computer Studies* 78 (2015), 43–67.

[51] Qing Ping, Feiyang Niu, Govind Thattai, Joel Chengottusseriyil, Qiaozi Gao, Aishwarya Reganti, Prashanth Rajagopal, Gokhan Tur, Dilek Hakkani-Tur, and Prem Nataraja. 2020. Interactive Teaching for Conversational AI. *arXiv preprint arXiv:2012.00958* (2020).

[52] Pinterest. 2019. New ways to control the ideas you see in your home feed. https://newsroom.pinterest.com/en/post/new-ways-to-control-the-ideas-you-see-in-your-home-feed

[53] Verus Pronk, Wim Verhaegh, Adolf Proidl, and Marco Tiemann. 2007. Incorporating User Control into Recommender Systems Based on Naive Bayesian Classification. In *Proceedings of the 2007 ACM Conference on Recommender Systems* (Minneapolis, MN, USA) *(RecSys '07)*. Association for Computing Machinery, New York, NY, USA, 73–80. https://doi.org/10.1145/1297231.1297244

[54] Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. *Explanations as Mechanisms for Supporting Algorithmic Transparency*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173677

[55] Gonzalo Ramos, Christopher Meek, Patrice Simard, Jina Suh, and Soroush Ghorashi. 2020. Interactive machine teaching: a human-centered approach to building machine-learned models. *Human–Computer Interaction* (2020), 1–39.

[56] Dimitrios Raptis, Rikke Hagensby Jensen, Jesper Kjeldskov, and Mikael B. Skov. 2017. Aesthetic, Functional and Conceptual Provocation in Research Through Design. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (Edinburgh, United Kingdom) *(DIS '17)*. Association for Computing Machinery, New York, NY, USA, 29–41. https://doi.org/10.1145/3064663.3064739

[57] Mitchel Resnick, Robbie Berg, and Michael Eisenberg. 2000. Beyond black boxes: Bringing transparency and aesthetics back to scientific investigation. *The Journal of the Learning Sciences* 9, 1 (2000), 7–30.

[58] Maria Roussou, Sara Perry, Akrivi Katifori, Stavros Vassos, Angeliki Tzouganatou, and Sierra McKinney. 2019. Transformation through Provocation?. 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, 1–13. https://doi.org/10.1145/3290605.3300857

[59] Tobias Schnabel, Gonzalo Ramos, and Saleema Amershi. 2020. "Who Doesn't like Dinosaurs?" Finding and Eliciting Richer Preferences for Recommendation. In *Fourteenth ACM Conference on Recommender Systems* (Virtual Event, Brazil) *(RecSys '20)*. Association for Computing Machinery, New York, NY, USA, 398–407. https://doi.org/10.1145/3383313.3412267

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

[61] Phoebe Sengers, Kirsten Boehner, Shay David, and Joseph 'Jofish' Kaye. 2005. Reflective Design. In *Proceedings of the 4th Decennial Conference on Critical Computing: Between Sense and Sensibility* (Aarhus, Denmark) *(CC '05)*. Association for Computing Machinery, New York, NY, USA, 49–58. https://doi.org/10.1145/1094562.1094569

[62] Phoebe Sengers, Rainer Liesendahi, Werner Magar, Christoph Seibert, Boris Müller, Thorston Joachims, Weidong Geng, Pia Mårtensson, and Kristina Höök. 2002. The Enigmatics of Affect. In *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (London, England) *(DIS '02)*. Association for Computing Machinery, New York, NY, USA, 87–98. https://doi.org/10.1145/778712.778728

[63] Ben Shneiderman. 2020. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction* 36, 6 (2020), 495–504.

[64] Patrice Y Simard, Saleema Amershi, David M Chickering, Alicia Edelman Pelton, Soroush Ghorashi, Christopher Meek, Gonzalo Ramos, Jina Suh, Johan Verwey, Mo Wang, et al. 2017. Machine teaching: A new paradigm for building machine learning systems. *arXiv preprint arXiv:1707.06742* (2017).

[65] Petr Slovák, Christopher Frauenberger, and Geraldine Fitzpatrick. 2017. *Reflective Practicum: A Framework of Sensitising Concepts to Design for Transformative Reflection*. Association for Computing Machinery, New York, NY, USA, 2696–2707. https://doi.org/10.1145/3025453.3025516

[66] Aaron Springer and Steve Whittaker. 2020. Progressive Disclosure: When, Why, and How Do Users Want Algorithmic Transparency Information? *ACM Trans. Interact. Intell. Syst.* 10, 4, Article 29 (Oct. 2020), 32 pages. https://doi.org/10.1145/3374218

[67] Simone Stumpf, Vidya Rajaram, Lida Li, Margaret Burnett, Thomas Dietterich, Erin Sullivan, Russell Drummond, and Jonathan Herlocker. 2007. Toward Harnessing User Feedback for Machine Learning. In *Proceedings of the 12th International Conference on Intelligent User Interfaces* (Honolulu, Hawaii, USA) *(IUI '07)*. Association for Computing Machinery, New York, NY, USA, 82–91. https://doi.org/10.1145/1216295.1216316

[68] Nicole Sultanum, Soroush Ghorashi, Christopher Meek, and Gonzalo Ramos. 2020. A Teaching Language for Building Object Detection Models. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 1223–1234. https://doi.org/10.1145/3357236.3395545

[69] S Shyam Sundar. 2008. Self as source: Agency and customization in interactive media. In *Mediated interpersonal communication*. Routledge Taylor & Francis Group, 58–74.

[70] Andrea L. Thomaz and Cynthia Breazeal. 2008. Teachable Robots: Understanding Human Teaching Behavior to Build More Effective Robot Learners. *Artif. Intell.* 172, 6–7 (April 2008), 716–737. https://doi.org/10.1016/j.artint.2007.09.009

[71] Jesse Vig, Shilad Sen, and John Riedl. 2009. Tagsplanations: Explaining Recommendations Using Tags. In *Proceedings of the 14th International Conference on Intelligent User Interfaces* (Sanibel Island, Florida, USA) *(IUI '09)*. Association for Computing Machinery, New York, NY, USA, 47–56. https://doi.org/10.1145/1502650.1502661

[72] Anna-Lisa Vollmer and Lars Schillingmann. 2018. On studying human teaching behavior with robots: a review. *Review of Philosophy and Psychology* 9, 4 (2018), 863–903.

[73] Emily Wall, Soroush Ghorashi, and Gonzalo Ramos. 2019. Using expert patterns in assisted interactive machine learning: A study in machine teaching. In *IFIP Conference on Human-Computer Interaction*. Springer, 578–599.

[74] Jayne Wallace, John McCarthy, Peter C. Wright, and Patrick Olivier. 2013. *Making Design Probes Work*. Association for Computing Machinery, New York, NY, USA, 3441–3450. https://doi.org/10.1145/2470654.2466473

[75] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing Theory-Driven User-Centric Explainable AI. 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, 1–15. https://doi.org/10.1145/3290605.3300831

[76] 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* (Hong Kong, China) *(DIS '18)*. Association for Computing Machinery, New York, NY, USA, 573–584. https://doi.org/10.1145/3196709.3196729

[77] Rayoung Yang and Mark W. Newman. 2013. Learning from a Learning Thermostat: Lessons for Intelligent Systems for the Home. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Zurich, Switzerland) *(UbiComp '13)*. Association for Computing Machinery, New York, NY, USA, 93–102. https://doi.org/10.1145/2493432.2493489