A Human-Centered Approach to Designing Teachable Systems

Christopher J. MacLellan*
Rob Sheline
Soar Technology, Inc.
Ann Arbor, MI
{chris.maclellan,rob.sheline}@soartech.com

Erik Harpstead*
Carnegie Mellon University
Pittsburgh, PA
eharpste@cs.cmu.edu

*Both authors contributed equally to this research.

ABSTRACT

Machine Learning technologies will eventually enable an unprecedented level of empowerment for end users–letting them adapt and personalize the behavior of AI systems to their own unique situations, needs, and desires through naturalistic interactions. Similar to how people can improve and extend one another through teaching, machine learning technologies promise to let users modify the behavior of systems by teaching rather than programming. However, there are many challenges faced when designing such teachable systems that have heretofore impeded their development. This paper aims to highlight two of these challenges. First, the development of learning systems has largely been restricted to technical fields, such as machine learning or cybernetics, due to the complexity of making such systems even work. Thus, there has been a lack of opportunities to provide an HCI perspective on how interactions with such systems should be designed. To address this challenge, we introduce a framework for conceptualizing teachable systems that distinguishes the interactive component from the learning component of these systems, to better enable the application of the HCI perspective on the design of the interaction between the user and system. Second, there is also a general lack of HCI methods for prototyping teachable systems, specifically when it comes to designing systems with learning capabilities that do not currently exist. To overcome this challenge, we presents a

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

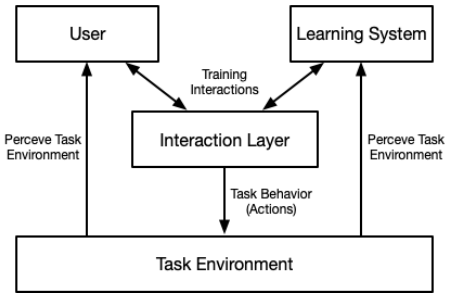

Figure 1: The context for a teachable system.

novel variation of the classic Wizard-of-Oz prototyping paradigm that specifically supports the rapid prototyping of interactive learning systems.

WHAT ARE TEACHABLE SYSTEMS?

There has been tremendous progress in the development of AI systems with many successful applications across a wide range of domains (e.g., games [8], education [3], health [1]). However, it is difficult to translate this success into technologies that make it into the hands of everyday users. Because building AI systems is often expensive, in terms of both time and expertise, they are typically built to perform a specific set of narrowly defined tasks. Users lives, on the other hand, are filled with diverse and idiosyncratic activities. While users might benefit from intelligent technologies that support these tasks, it would be difficult to develop specialized AI to support them. To address this challenge, we envision a class of systems that we refer to as *teachable systems* that enable users to personalize and adjust the behavior of AI systems through natural interaction.

Teachable systems are inherently defined by their interactions with humans, so we believe that they should be designed in a human-centered rather than technology-centered way. In particular, they should learn in ways that are better for users to express rather than ones that are better for machine-learning developers to create. Further, unlike conventional machine learning systems, which typically require large amounts of data and time to learn, teachable systems should be able to interactively learn at a scale that users can teach (e.g., tens of examples as opposed to tens of millions).

The technical approaches that will make teachable systems possible are starting to be developed (e.g., see [4]), so we argue that now is the time to start thinking about how best to design such systems. By beginning to explore the design of teachable technologies from an HCI perspective now, we hope to guide the development of these emerging technical approaches towards those that will ultimately benefit the users of teachable systems.

DESIGNING TEACHABLE SYSTEMS

To support the design of interactive teachable systems, we proposed the Natural Training Interactions framework [5] that provides a high-level structure for characterizing the relationship between a user and a teachable system. The context for this framework is the interaction of four main elements: the **task environment**, the **user**, the **learning system**, and the **interaction layer** (see Figure 1). A key feature of this decomposition is a separation of the teachable system into two parts, the learning system, which translates training interactions into new knowledge and leverages this knowledge to perform, and the interaction layer, which mediates the interactions between the other three components. Given this separation, we argue that the primary HCI challenge when building teachable systems is the design the interaction layer, rather than the technical design of the underlying learning system.

Table 1: The Natural Training Interactions Framework

Knowledge	Patterns
Goals	Passive Learning
Beliefs	Operant Conditioning
Concepts	Direct Instruction
Skills	Apprentice Learning
Experiences	After-Action Review
Dispositions	Collaborative Learning
	Programming
Types	**Modalities**
Command	Command Line
Clarify	Control Device
Acknowledge	GUI
Inform	Sketch
Spotlight	API
Annotate	Gesture
Reward	Speech
Demonstrate	Text
Direct knowledge manipulation	Multi-modal
Request <type>	

Within this structure, the user and the learning system both have the potential for prior knowledge as well as some internal state that the is not directly observable to the other. The task environment might be either physical or virtual, and while the user and the learning system are both able to independently perceive the task environment, their ability to act on the environment is mediated by the interaction layer. This mediation is crucial because it makes possible key interaction designs, such as a user verifying and approving a system's behavior before it is executed in the environment or for the learning system to observe a user as they act on the environment through the interaction layer.

It is worth contrasting this paradigm with two alternatives, one where the user and learning system interact with one another indirectly through the task environment rather than through an intermediate interaction layer and another where the interaction layer and learning system are collapsed. A key issue with these alternatives is that they confound the interaction design with the design of the learning system. Our paradigm also allows for a special training interaction channel between the user and the learning system (the interaction layer), without being bound to what can be easily expressed through the environment.

To support the design of the interaction layer, we articulated a design space for natural training interactions shown in Table 1. This design space consists of four dimension that describe how the user and learning system interact through the interaction layer. First, it assumes the user has an implicit goal of transferring some kind of knowledge to the teachable system (e.g., the user might transfer a concept or skill). This transfer process can follow one of several patterns of teaching; e.g., a didactic instruction pattern or a more learner driven pattern, such as apprentice learning. Within patterns, trainers and the learning system employ several types of interaction moves (e.g., providing feedback or asking for an example). Finally, each of these interactions can ground out into different modalities for the user (e.g., providing feedback using GUI or speech based interactions). The key insight of this design space is that decisions made in one dimension will impact the naturalness or appropriateness of the choices across the other dimensions. See MacLellan et al. [5] for more details on our framework and examples of how existing interactive learning systems fit within it.

PROTOTYPING TEACHABLE SYSTEMS

Given our framework, a key part of the design process is prototyping different designs to see which are more naturally and efficient for end users. Unfortunately, there are many challenges to prototyping interactive learning systems. Developing a machine learning systems that operates with a sufficient level of performance and fidelity to test with real users is difficult and time consuming, which often makes each prototyping iteration slow. Additionally, there are many potential interaction patterns and modalities that might not be possible with existing machine-learning approaches, making it difficult to prototype systems that rely on these capabilities. Thus, we desire a prototyping approach that

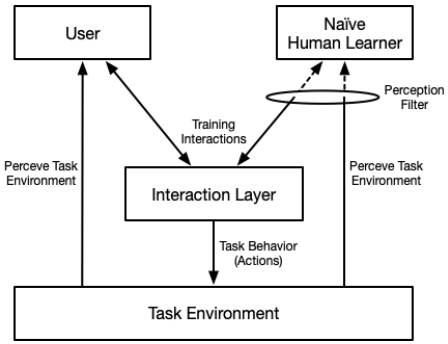

Figure 2: A Wizard-of-Oz paradigm for prototyping teachable systems.

enables designers to sidestep the need to develop a fully functional learning system in order to test out a design for the interaction layer.

To overcome this challenge we propose a new variant of the Wizard-of-Oz prototyping approach [2] wherein we replace the learning system from the framework shown in Figure 1 with a naïve experimental participant who does not know the target task and must learn it from interactions with the user through the interaction layer (see Figure 2). This substitution makes it possible for the designer to rapidly prototype the interaction layer without requiring a fully operational learning system. Like the classic Wizard-of-Oz approach, our method assumes that a human subject is a useful proxy for an arbitrary system. While this may not always be true, we believe it is reasonable for a wide variety of tasks–enabling the design of novel teachable systems.

To enable the human participant to operate in the role of the learning system, we introduce a *perception filter* between the naïve human learner and both the interaction layer and the task environment. This filter translates machine-readable representations of the task environment and training interactions into human-readable formats. Additionally, it translates the human participants interactions back into the machine readable format that might be expected by the actual machine learning system. This perception filter was inspired by the work of Sequeira et al. [6], which leveraged such a perception filter to explore different social interaction patterns for human-robot interaction. By constraining the human participant to have the same inputs and outputs as a hypothetical system, it provides an additional test that the overall system design is possible; i.e., that the system might realistically learn from the provided inputs and outputs rather than requiring some external knowledge that would not be available to the system in practice. We are currently in the process of developing a platform for conducting these Wizard-of-Oz experiments [7] and would value any feedback on this concept.

REFERENCES

[1] BG Buchanan and EH Shortliffe. 1984. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project.* Addison-Wasley.

[2] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz studies–why and how. *Knowledge-based systems* 6, 4 (1993), 258–266.

[3] Kenneth R Koedinger, John R Anderson, William H Hadley, and Mary A Mark. 1997. Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education (IJAIED)* 8 (1997), 30–43.

[4] John E Laird, Kevin Gluck, John Anderson, Kenneth D Forbus, Odest Chadwicke Jenkins, Christian Lebiere, Dario Salvucci, Matthias Scheutz, Andrea Thomaz, Greg Trafton, Robert E. Wray, Shiwali Mohan, and James R. Kirk. 2017. Interactive task learning. *IEEE Intelligent Systems* 32, 4 (2017), 6–21. https://doi.org/10.1109/MIS.2017.3121552

[5] Christopher J MacLellan, Erik Harpstead, Robert P Marinier III, and Kenneth R Koedinger. 2018. A Framework for Natural Cognitive System Training Interactions. *Advances in Cognitive Systems* 6 (2018), 1–16.

[6] Pedro Sequeira, Patrícia Alves-Oliveira, Tiago Ribeiro, Eugenio Di Tullio, Sofia Petisca, Francisco S Melo, Ginevra Castellano, and Ana Paiva. 2016. Discovering social interaction strategies for robots from restricted-perception Wizard-of-Oz studies.

In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. IEEE Press, 197–204.

[7] Rob Sheline and Christopher J MacLellan. 2018. Investigating Machine-Learning Interaction with Wizard-of-Oz Experiments. In *Proceedings of the NeurIPS 2018 Workshop on Learning by Instruction*.

[8] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. 2016. Mastering the game of Go with deep neural networks and tree search. *nature* 529, 7587 (2016), 484.