Designing Teachable Systems for Intelligent Tutor Authoring

Adit Gupta, Christopher J. MacLellan
College of Computing and Informatics
Drexel University
3675 Market Street,
Philadelphia, Pennsylvania 19104
{adit.gupta, christopher.maclellan}@drexel.edu

Abstract

Intelligent tutoring systems (ITS) consistently improve students' educational outcomes when used alone or in combination with traditional instruction (MacLellan et al. 2018). One major barrier to the wider use of AI tutoring systems is that they are non-trivial to build, requiring both time and expertise. Typically, authoring a tutor takes 200-300 hours of developer time to produce an hour of instruction time (Aleven et al. 2009; Weitekamp, Harpstead, and Koedinger 2020). Existing authoring methods, including Cognitive Tutor Authoring Tool's (CTAT) Example Tracing and SimStudent's Authoring by Tutoring approaches, let authors create ITSs quicker than traditional approaches, such as hand programming. While Example Tracing and Authoring by Tutoring reduce the time and expertise required to create an ITS, such techniques do not allow humans to teach AI technologies in ways that are natural to humans. In this paper, we propose a research plan based on Natural Training Interactions (NTI) framework (MacLellan et al. 2018) that aims to create more human-centered and efficient tutor authoring tools. We propose dual-sided, restricted-perception Wizard-of-Oz (WoZ) experiments, a novel variant of commonly used WoZ experiments, to prototype teachable AI technologies for tutor authoring. We engineered the NTI testbed to allow novel tasks to be studied in WoZ experiments without having to start from scratch for each experiment. Lastly, we propose three research questions that we believe will help us understand how to create teachable AI technology to power tutoring systems. The NTI framework aims to produce teachable agents that can be used by teachers and other non-programmers to naturally and efficiently author ITSs.

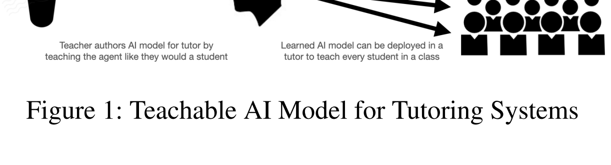
Teacher authors AI model for tutor by teaching the agent like they would a student
Learned AI model can be deployed in a tutor to teach every student in a class

Figure 1: Teachable AI Model for Tutoring Systems

Introduction

Intelligent tutoring systems (ITS) are a type of computerized educational technology that tutors students on problems by providing them with additional interactions such as feedback messages, next-step hints, and demonstration. A number of studies show consistent improvement of students' learning outcomes through the use of ITSs (Pane et al. 2014). While effective and beneficial to learners, ITSs are burdensome to create. Authoring an ITS may take up to 300 hours of development time per hour of instruction time (Murray 1999).

Existing efforts to support tutor authoring, such as Cognitive Tutor Authoring Tool's (CTAT) Example Tracing and SimStudent's Authoring by Tutoring, have been shown to reduce authoring time as much as four times (Aleven et al. 2009; Weitekamp, Harpstead, and Koedinger 2020). Example Tracing lets authors build tutors by demonstrating correct solutions to problems, so they can create and deploy tutors without any programming expertise (Aleven et al. 2009). SimStudent is a more recent machine-learning approach that enables authors to build tutor system by ”teaching” an agent via worked examples and feedback (Li et al. 2012; Matsuda 2010). SimStudent generalizes from these examples, so it requires less demonstrations to learn a correct and complete model than example tracing (MacLellan, Koedinger, and Matsuda 2014). While these existing tutor authoring tools reduce the time and expertise needed to author a tutor, they still struggle to support non-programmers to build tutors for complex domains (Maclellan et al. 2015).

The Apprentice Learner (AL) Architecture builds on the prior SimStudent work with the aim of providing a general platform for investigating and comparing simulated student models. A recent study with the AL architecture (MacLellan and Koedinger 2020) showed that AL agents were able to support tutor authoring across 8 different tutoring systems across a wide range of domains (math, language, chemistry, and engineering). Additionally, this work showed that for domains with complex solution spaces, authoring tutors with AL agents takes substantially less time than authoring with example tracing (authoring an experimental design tutor took 27 minutes with example tracing, but only 9.19 minutes with an AL agent).

While tutor authoring approaches, such as AL, promise superior performance to predecessor models, they have not been tested with real humans. Recent work by (Weitekamp, Harpstead, and Koedinger 2020) propose interaction loops where the student is substituted with a simulated learner. The AI agents that mimic the inductive learning progress undergone by human students are known as Simulated Students (Weitekamp, Harpstead, and Koedinger 2020). Failures with this approach include limitations in techniques to effectively train Simulated Students (Weitekamp, Harpstead, and Koedinger 2020). While this is a step in the right direction, more work is needed to create teachable systems for tutor authoring. Reducing the time it takes to author a tutor can certainly promote quicker iterations of the tutor model, however, one must still posses programming knowledge to achieve this feat. Moreover, options accessible to non-programmers often only support a few domains (e.g., Math), increasing both number of examples and time needed to complete tutor authoring.

The goal of this paper is to outline a research plan to design teachable systems for tutor authoring, utilizing a human-centered approach. To do so, we build on the Natural Training Interactions (NTI) Framework (MacLellan et al. 2018). We will describe how to create tutor models that can be authored through interaction that are natural for humans. Lastly, we will describe broader implications of this work that could facilitate in empowering non-programmers such as K-12 educators to author tutors.

Background

As we try to understand how to enable naturally teachable AI systems to support tutor authoring, we must first describe what is a natural interaction. Based on prior work by (MacLellan et al. 2018) we see that naturalness can be identified using four characteristics: they (1) support the goal of the user, (2) do what the user expects, (3) allow the user to work the way they want, and (4) leverage users' experience to minimize training. As we design teaching interaction patterns, we must examine the interactions that occur between a teacher and a learner. We must use HCI methodology which enables us to learn from real human and tutor interactions.

Natural Training Interactions Framework

The Natural Training Interactions (NTI) Framework (MacLellan et al. 2018) is a framework for cognitive systems training interaction that aligns with the four characteristics of naturalness described above. The framework characterizes four dimensions of interaction between a human and an AI agent. These dimensions include: (1) **knowledge**, (2) **patterns**, (3) **types**, (4) **modalities** (see Table 1). The goal of an interaction between a human and AI agent is to change an aspect of the agent's **knowledge**. The trainer and agent follow an instructional **pattern**. Trainers engage in several **types** of interactions within patterns. And lastly, these interactions are done through different **modalities**. We apply this framework to see which instruction patterns might be most natural to humans.

Wizard of Oz Experiments

Designing interactive AI systems like the teachable systems that we envision is recognized as a difficult HCI problem (Yang et al. 2020). One of the only known approaches to prototyping such systems are Wizard-of-Oz (WoZ) experiments (Dahlbäck, Jönsson, and Ahrenberg 1993). WoZ studies circumvent the need for a fully developed AI system by substituting the intelligent agent with a human "behind the curtain" who simulates the desired behavior—letting us quickly test interactions, even if we do not yet have the AI/ML models to support them. In classic Wizard-of-Oz experiments, a researcher rapidly tests a hypothetical Artificial Intelligence system with real users by manually simulating its capabilities. (Yang et al. 2020) recognizes WoZ as one of the only known AI prototyping approaches.

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	

Dual-Sided Wizard-of-Oz Experiment

In order to build human-centered teachable agents for tutor authoring, we must observe learning interactions between a teacher and a learner. However, WoZ experiments historically are not meant for prototyping learning systems. Traditionally, WoZ systems have a fully trained confederate ”behind the curtain”. To combat this dilemma, we propose a novel variant of the WoZ experiment, known as dual-sided, restricted-perception Wizard-of-Oz methodology. This novel prototyping approach replaces the knowledge researcher/experimenter on the back-end (the ”Wizard”) with a naïve participant. We then randomly assign each participant in the pair as either a teacher or a learner. Further, both the teacher and learner are provided detailed instructions on how to interact with both their environment and one another. We believe this approach should enable researches to evaluate interactions of hypothetical systems in a cost effective manner. Our methodology assumes naïve human participants serve as a suitable proxy for an arbitrary tutor system. Though this assumption poses limitations to the applicability of AI tasks, we believe this is a reason-

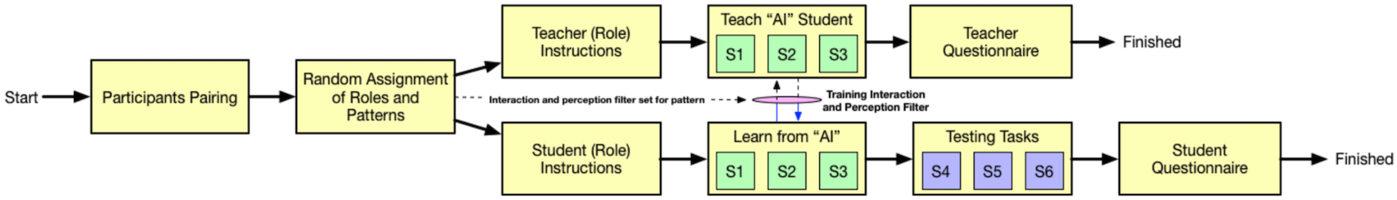
Start
Participants Pairing
Random Assignment of Roles and Patterns
Teacher (Role) Instructions
Student (Role) Instructions
Interaction and perception filter set for pattern
Training Interaction and Perception Filter
Teach "AI" Student
S1
S2
S3
Learn from "AI"
S1
S2
S3
Teacher Questionnaire
Finished
Testing Tasks
S4
S5
S6
Student Questionnaire
Finished

Figure 2: The Dual-Sided, Restricted Perception Wizard-of-Oz Experimental Design.

able assumption when dealing with knowledge components to solve trivial tasks. This WoZ approach will enable us to cost effectively prototype different interaction patterns for teachable agents to support tutor authoring. We aim to identify the best interaction patters before fully implementing a teachable AI system for tutor authoring.

Natural Training Interactions Testbed

This novel variant of WoZ is, in conception, similar to prior work on using WoZ experiments to study social interactions (Sequeira et al. 2016). However, our interest lies in applying this approach to further investigate the design of interactive machine learning systems, such as teachable agents to support tutor authoring. While WoZ studies generally promote rapid iterations during the experimentation process, setting up these WoZ experiments can be cumbersome. Replicating a WoZ experiment online may be fast and cheap, but still may take extensive setup and development time. Our approach is to decouple the actual experiment from the participant interaction layer. Doing so enables us to swap experiments to be able to run more studies efficiently.

Further, to test this assumption, we engineered the Natural Training Interactions Testbed for conducting dual-sided restricted-perception WoZ experiments. We created a training user interface (UI) which enables us to provide tasks to participants. Once paired participants adhere to the onboarding instructions, one participant is randomly assigned as the teacher, and the other participant is assigned as the learner. The training UI supports different modalities including clickable buttons, text based chat message box, and clickable gestures. The NTI Testbed is capable of loading a Unity/WebGL, and is agnostic to the state of the unity object. Each pairing of participants will assigned a random interaction pattern which will dictate how knowledge is transferred between the teacher and the learner. These interaction patterns include: Operant Conditioning, Direct Instruction, and Apprentice Learning (Table 1).

We engineered the Natural Training Interactions (NTI) testbed to allow novel tasks to be studied in a WoZ-experiment without having to start from scratch, given an implementation of the testbed interface. Our goal with the NTI Testbed is to enable dual-sided, restricted-perception WoZ experiments with a variety of possible experiments to better understand the interaction patterns that are most natural for human-to-human knowledge transfer.

To test the feasibility of this approach, we plan to conduct a pilot study on Amazon's Mechanical Turk Platform (Mturk). Mturk is an online marketplace for human intelligence tasks. Accessible via programmatic API, workers can be requisitioned at any time of the day, and from a wide variety of sub-populations according to experimental need. Human participants on will be randomly paired using Mturk and one participant will be assigned as the teacher role, and the other as the learner role.

See Figure 2 for an overview of our experimental approach. After pairing, each participant must read specific instructions as to what their role may entail. Based on the interaction pattern utilized in the experiment, the teacher may approve student's tasks, demonstrate how to complete tasks, or simply mark these tasks as completed. During the experiments, participants will need to transfer knowledge to each other necessary for completing simple tasks, games, or puzzles, while we manipulate which interactions and patterns are available for them to use. These experiments will help us identify and test interaction patterns used in human-to-human knowledge transfer.

Proposed Research Questions

Our experiments will produce novel data regarding the patterns and modalities that make tutor authoring natural and efficient. The output of this research will introduce a framework for interactive teachable AI technologies that can support the creation of AI-driven tutoring systems. To achieve this, we will tackle three questions.

RQ1. What interaction patterns do people naturally employ when teaching an AI system? Before building a complex AI system, it is crucial to first understand what makes human-to-human knowledge transfer effective. The Natural Training Interaction (NTI) framework (MacLellan et al. 2018) decomposes cognitive system training instructions into patterns, types, and modalities - which support the transfer of different types of knowledge. In this study, we will utilize Amazon Mechanical Turk (MTurk), to conduct a variant of WoZ experiments. Human participants will be randomly paired using Mturk and one participant will be assigned as the teacher, and the other as the learner. During the experiments, participants will need to transfer knowledge to each other necessary for completing simple tasks, games, or puzzles, while we manipulate which interactions and patterns are available for them to use. These experiments will help us identify and test interaction patterns used in human-to-human knowledge transfer.

RQ2. Can we build a Teachable AI (TAI) agent that supports patterns naturally used by humans? Once we

Table 2: Interaction Patterns to Technique Mapping

Patterns	Technique
Teacher provides feedback	Reinforcement Learning
Teacher provides worked examples	Learning by Demonstration
Teacher provides language based guidance	Learning from Language
Patterns discovered in RQ1	Novel machine learning techniques discovered in RQ2

identify the teaching patterns that are natural for humans, we will extend the Apprentice Learning system (MacLellan and Koedinger 2020) to support these interactions. Previous work on TAI agents outline a wide range of technical approaches for supporting different kinds of interactions, such as learning from mixed-modality demonstrations, feedback, and language. To address RQ2, we will leverage the Apprentice Learning architecture to map these existing techniques to patterns we identify in RQ1. Table 2 outlines a hypothetical mapping between learning approaches and the patterns they might support.

RQ3. How does human-to-human knowledge transfer compare to human-to-AI knowledge transfer? Our human-centered experiments will pave the way for building TAI agents that humans can naturally teach and interact with. To evaluate our TAI system (developed in RQ2), we will run experiments where humans teach the system directly and measure its correctness, efficiency, and naturalness. These experiments will focus on the same tasks as RQ1, which will let us compare human learner versus TAI model performance.

Conclusion and Future Work

Intelligent tutoring systems (ITS) consistently improve students' educational outcomes when used alone or in combination with traditional instruction. However, authoring AI tutoring systems is non-trivial, requiring both extensive time and expertise. To enable the wider use of AI tutoring systems, we propose making AI Tutors which can be taught rather than programmed. In order to understand how to teach AI systems through interactions that are natural to humans, we must first understand which patterns and modalities are effective in human-to-human knowledge transfer.

These teachable AI (TAI) technologies can be used more broadly in the education technology domain to enable non-programmers such as K-12 teachers to naturally teach AI systems. Beyond tutor authoring, teachable AI technology can be used to build personal cognitive systems such as AI assistants to facilitate with learning in the K-12 environment. Further, to test the abilities of tutoring systems, teachable AI technology can be utilized to test existing AI tutors. Doing so would help us identify areas of improvement in existing pedagogical tutoring systems.

References

Aleven, V.; McLaren, B. M.; Sewall, J.; and Koedinger, K. R. 2009. A new paradigm for intelligent tutoring systems: Example-tracing tutors. *International Journal of Artificial Intelligence in Education* 19(2): 105–154. ISSN 15604292.

Dahlbäck, N.; Jönsson, A.; and Ahrenberg, L. 1993. Wizard of oz studies-why and how. *International Conference on Intelligent User Interfaces, Proceedings IUI* Part F1275.

Li, N.; Schreiber, A. J.; Cohen, W. W.; and Koedinger, K. R. 2012. Efficient Complex Skill Acquisition Through Representation Learning. *First Annual Conference on Advances in Cognitive Systems* 2: 149–166. ISSN 2324-8416. URL http://www.cogsys.org/paper/paper-3-2-104.

Maclellan, C. J.; Harpstead, E.; Wiese, E. S.; Zou, M.; Matsuda, N.; Aleven, V.; and Koedinger, K. R. 2015. Authoring tutors with complex solutions: A comparative analysis of Example Tracing and SimStudent. *CEUR Workshop Proceedings* 1432: 35–44. ISSN 16130073.

MacLellan, C. J.; and Koedinger, K. R. 2020. *Domain-General Tutor Authoring with Apprentice Learner Models*. International Journal of Artificial Intelligence in Education. ISBN 4059302000214. doi:10.1007/s40593-020-00214-2.

MacLellan, C. J.; Koedinger, K. R.; and Matsuda, N. 2014. Authoring Tutors with SimStudent: An Evaluation of Efficiency and Model Quality. In Trausan-Matu, S.; Boyer, K. E.; Crosby, M.; and Panourgia, K., eds., *Intelligent Tutoring Systems*, 551–560. Cham: Springer International Publishing. ISBN 978-3-319-07221-0.

MacLellan, E.; Harpstead, C. J.; Marinier Iii, R. P.; and Koedinger, K. R. 2018. Towards Natural Cognitive System Training Interactions: A Preliminary Framework .

Matsuda, N. 2010. SimStudent : Building an Intelligent Tutoring System by Tutoring a Synthetic Student. *International Journal of Artificial Intelligence in Education.* (2).

Murray, T. 1999. Authoring Intelligent Tutoring Systems: An analysis of the state of the art. *International Journal of Artificial Intelligence in Education (IJAIED)* 10: 98–129.

Pane, J. F.; Griffin, B. A.; McCaffrey, D. F.; and Karam, R. 2014. Effectiveness of Cognitive Tutor Algebra I at Scale. *Educational Evaluation and Policy Analysis* 36(2): 127–144. URL https://doi.org/10.3102/0162373713507480.

Sequeira, P.; Alves-Oliveira, P.; Ribeiro, T.; Di Tullio, E.; Petisca, S.; Melo, F. S.; Castellano, G.; and Paiva, A. 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*, HRI '16, 197–204. IEEE Press.

Weitekamp, D.; Harpstead, E.; and Koedinger, K. R. 2020. An Interaction Design for Machine Teaching to Develop AI Tutors. *Conference on Human Factors in Computing Systems - Proceedings* 1–11. doi:10.1145/3313831.3376226.

Yang, Q.; Steinfeld, A.; Rosé, C.; and Zimmerman, J. 2020. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. *Conference on Human Factors in Computing Systems - Proceedings* 1–13.