International Society of the Learning Sciences

The Human-Interpreter Problem in Youth Encounters with AI

Eric Greenwald, University of California, Berkeley's Lawrence Hall of Science, eric.greenwald@berkeley.edu
Maxyn Leitner, University of Southern California Institute for Creative Technologies, leitner@ict.usc.edu
Ning Wang, University of Southern California Institute for Creative Technologies, nwang@ict.usc.edu

Abstract: Artificial Intelligence's impact on society is increasingly pervasive. While innovative educational programs are being developed, there is yet little understanding of how pre-college aged youth engage with and construct understanding of core AI concepts and strategies. In this paper, we discuss emerging findings from a cognitive interview study with middle school and high school students to better understand how students learn AI concepts. This research was supported by the National Science Foundation under Grant 1842385.

Introduction: The human interpreter

When a novice programmer encounters a debugging problem in computer science, the novice sometimes imbues the program with a human's ability to interpret ambiguous instructions (Spohrer & Soloway, 1986). This may be especially common where the correct interpretation is obvious, or incorrect interpretations are exceedingly unlikely (Van Someren, 1990). The implicit expectation is that there is a human interpreter between the computer and the code mediating implementation of the code. Across a set of extended cognitive interviews with middle and high school aged youth, we observed evidence of an AI variation on the human interpreter problem: when a novice works to figure out how an AI system might solve a real-world problem, they sometimes imbue the system not only with computational capacity, but also with capacity to infer and reason about human motivations. We hypothesize that youth may begin with a working theory of AI that assumes general intelligence for the system, including the capacity to recognize and reason from human motivations.

Methods

We conducted cognitive interviews with a convenience sample of eight students (12 to 17 years old) from a private school located in the western United States as they worked through five AI problems. As youth worked through the problem set, the cognitive interview employed a semi-structured protocol to elicit student thinking as each student initially encountered each problem, attempted to solve the problem, and in reflection after they had settled on a solution. In-the-moment scaffolding was provided throughout each interview to enable students to reveal thinking across each step of the solution, surface and test emerging ideas about why a student might be stuck, and to disambiguate between superficial challenges and conceptual difficulties (see, Greenwald, Leitner & Wang, 2021, for a full description of methods). The sample was racially and ethnically diverse and also higher-resourced than that of the average public school. Given persistent inequities in computer science academic and career pathways, as well as broader questions of access and privilege across society, the sample presents limitations for generalizability that will be important to address in future studies.

To analyze the video-recorded interview data, we created excerpts of each student's work on each problem, and applied codes directly to the video excerpts according to the problem type using Dedoose (2018), a mixed methods data analysis tool. We then completed two passes through the data. With the first pass we viewed each interview in sequence, generating and applying a set of broadly applicable codes (e.g., problem type, challenges, scaffolds); then, we compiled the excerpts by problem type and viewed the variety of student responses on each problem together. In this second pass through the data, we drew from the principles of grounded theory (Glaser, 1992), to iteratively introduce new codes as themes emerged, which we then added to and revised through successive passes through the data, continually comparing the emergent codes against the data. These initial codes were revised for consistency, then applied systematically across the data.

Initial findings: Insight into students' working theories of AI systems

As previously reported, students in our sample were unfamiliar with parsing the world in terms an AI system can operate on (Greenwald, Leitner, & Wang, 2021). A challenge for all students interviewed, including those with advanced mathematical skills, was recognizing how a problem in the world could be made amenable to the computational power of an AI system. That is to say, students needed support in conceiving a problem space in a way that would enable an AI system to solve it. Students' initial uncertainty, however, did shed light on possible working theories students brought into the AI problem solving tasks.

For the decision tree problem, students were presented with a data table about past behavior and tasked with helping a friend decide whether or not they would wait to get a seat at a restaurant. With this problem,

multiple students approached the task of finding patterns in the presented data set by first speculating on the underlying motivations that may have led to the data itself. For example, when presented with the data table and the overarching problem of imagining how an AI system might leverage the data to make decisions, students typically began by inferring from their own experiences and intuitions about eating out to propose reasons that a family might wait or not wait at a table. The dialogue excerpted below occurs after the variables and values in the data table were explained by the interviewer:

> Interviewer: If the goal is to predict whether or not the family is going to wait, how could this data table be used to make that prediction?
> Student: All the ones that waited, said that they were hungry; they had a reservation, and it was a pretty expensive restaurant, so they waited And it was a short wait, so they probably...and they were hungry, so they wouldn't want to leave their reservation; same with this one, they also had a reservation and they were hungry, and there was no alternative, so they probably would...they waited as well.

Similar student responses across the interviews suggest that when considering how AI systems use data to make decisions, students begin by drawing on prior experience to suggest underlying motivations within the decision space, rather than attending to features of the data themselves. A possible interpretation is that the student was attempting to understand the family's motivated reasoning for waiting or not waiting as a path to devising an AI solution that could leverage that reasoning. While building understanding by drawing on prior experience is often a productive move, we note that each of the student's conjectures in the excerpt above (hunger level, reservation, price) conflicts with the data as presented in the table. For multiple students, we also observed that this reasoning about diners' motivations as a path for AI system decision-making often persisted after explicit call-backs to the observable features of the data table.

Discussion: A call to explore working theories of AI among youth

Findings suggest that students may be including notions of intent and motivation in their working theories of AI systems: that because *human* intelligence includes the capacity to infer and make use of one's reasoning behind a decision, perhaps then artificial intelligence shares this capacity. The version of AI common in science fiction and popular culture is often that of an intelligent system more or less similar to (or advanced beyond) the intelligence of a human: a yet-to-be realized instantiation of artificial general intelligence, or artificial super intelligence. Evidence across several interviews suggests that students may be drawing on these models of AI when encountering AI problems. It is also likely that students' initial focus on the rationale behind the data dovetails with long-documented difficulties students have analyzing and interpreting data more broadly (Curcio, 1987).

Further research is needed to better understand students' working theories about AI systems, and how those theories may mediate students' developing understanding within the learning experience. Interviews suggest a related need to support students toward a more generalized understanding of how AI systems can be applied to problems, and how problems can be reimagined to be solvable by AI systems. This finding adds weight to efforts aimed at promoting "explainable AI" (Gunning & Aha, 2019) that makes the decision-making of AI algorithms transparent to users. Through its transparency, explainable AI can create opportunities to make AI concepts accessible, in part by supporting youth in developing working theories about how such systems function.

References

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. Journal for research in mathematics education, 382-393.

Dedoose (2018) Version 8.0.35, web application for managing, analyzing, and presenting qualitative and mixed method research data. Los Angeles, CA: SocioCultural Research Consultants, LLC www.dedoose.com.

Glaser, B. G., & Strauss, A. L. (1967). Discovery of grounded theory: Strategies for qualitative research. Routledge.

Greenwald, E., Leitner, M., and Wang, N. (2021). Learning Artificial Intelligence: Insights into How Youth Encounter and Build Understanding of AI Concepts. Proceedings of AAAI-21, February 2-9. AAAI Press, Palo Alto, California USA

Gunning, D. & Aha, D. W. (2019). DARPA's explainable artificial intelligence program. AI Magazine, 40(2), 44-58.

Spohrer, J. C., & Soloway, E. (1986). Novice mistakes: Are the folk wisdoms correct?. Communications of the ACM, 29(7), 624-632.