How to Make Automata Examples and Definitions More Natural: From Explainable AI to Explainable Automata Theory

Ian Bautista Ambriz[1], Monet Nevarez Sanchez[1], Christian Servin[2],
Olga Kosheleva[3], Vladik Kreinovich[1], and Nguyen Hoang Phuong[4]
[1,3]Departments of [1]Computer Science and [3]Teacher Education
University of Texas at El Paso
500 W. University, El Paso, Texas 79968, USA
ibautistaambr@miners.utep.edu, mnevarezsanch@miners.utep.edu, vladik@utep.edu
[2]Computer Science and Information Technology Systems Department,
El Paso Community College, 919 Hunter, El Paso, TX 79915, USA, cservin1@epcc.edu
[4]Thang Long University, Hanoi, Vietnam, nhphuong2008@gmail.com

In computer science – as well as in many other disciplines – researchers and practitioners have come up with many creative ideas and techniques. Computing examples are:

- how to sort numbers faster that we would usually do, how to multiply numbers faster than what we learn in school, how to multiply matrices faster that we learn in most matrix algebra classes,
- how to create a safe encoding that is difficult to break,
- how to prove that it is not possible to predict when a program will halt, etc.

In each such case, coming up with the corresponding method was a stroke of genius.

We use these methods, we proudly teach them to students – because they need to use them. But these methods are not easy for students to learn – because they do not naturally follow from the problem, each of them looks like an alien trick.

To students, this sometimes start feeling like Harry Potter's Hogwarts school, where students have to memorize incantations like Abracadabra without any understanding of how to explain their work. Similarly, students have to memorize many such tricks.

In many cases, it later turns out that it is possible to explain such a method – after the fact – in a more natural way.

The need to make the material more understandable is similar to one of the main challenges of modern AI: how to make AI results more understandable. In both cases – when trying to understand AI results and when trying to learn Automata material – we have complex difficult-to-understand-why ideas generated by a "black box": AI's "brain" or a human brain. In both cases, we want to better understand where these idea came from, and how to make their appearance more natural.

In this talk, on the example of several basic Automata concepts and proofs, we show that the use of invariances can help to make many such ideas more natural.