Using Mathematically-Grounded Metaphors to Teach AI-Related Cybersecurity

Bart P. Knijnenburg, Nicole Bannister and **Kelly Caine**
Clemson University
{bartk, nbannis, caine}@clemson.edu

Abstract

This position paper describes our research project to improve middle school students' use of security "best-practices" in their day-to-day online activities, while enhancing their fundamental understanding of the underlying security principles and math concepts that drive AI and cybersecurity technologies. The project involves the design and implementation of a time- and teacher-friendly learning module that can be readily integrated into existing middle school math curricula. We plan to deploy this module at a high-needs, rural-identifying middle school in South Carolina that serves underrepresented students.

1 Introduction

The cybersecurity implications stemming from the increasingly pervasive use of Artificial Intelligence (AI) directly impact some of our nation's most vulnerable people. With the protections of the Children's Online Privacy Protection Act (COPPA) ending at age 13, it is crucial for adolescents to develop AI-related cybersecurity literacies so that they may effectively and responsibly take ownership of their digital identities. We propose to improve middle school students' use of security "best-practices" in their day-to-day online activities, while enhancing their fundamental understanding of the underlying security principles and math concepts that drive AI and cybersecurity technologies.

Our project makes a trailblazing effort to link AI and cybersecurity principles to their mathematical underpinnings in a way that middle school students will understand. To this end, we plan to develop a time- and teacher-friendly learning module that can be readily integrated into existing middle school math curricula. Under the moniker explainable AI (xAI, cf. [1]), the field has produced a recent but substantial body of work attempting to explain its operations to the end-user. Most of this work, however, is focused on explaining the provenance of AI-based inferences with the aim of supporting judgments about efficacy and/or fairness. In contrast, hardly any work exists that explains AI from a cybersecurity perspective. We conjecture that this is a challenging task, as it requires a more fundamental understanding of the mathematical principles behind AI.

Our proposed module fills this gap by relying on the educational principle of "metaphors as reification" [11,12] to teach AI-related cybersecurity. Metaphorical reasoning has had only limited success in cybersecurity training [3]–a problem we aim to solve by grounding the metaphors in mathematical principles. If successful, the mathematically-grounded metaphors approach contributes a key advance in the state-of-the-art in cybersecurity training.

2 Related Work

Our research involves developing cybersecurity materials for children, using an analysis of their "folk models" to find common misconceptions and mathematically-grounded metaphorical mental models to repair these misconceptions. We describe the existing research on each of these topics below.

2.1 Cybersecurity Materials for Children

Most existing cybersecurity education programs are geared toward training employees to detect and avoid cybersecurity vulnerabilities in corporate settings. This pattern is reflected in the cybersecurity education literature, as most studies of cybersecurity education were conducted in tertiary education settings in the US [13]. A notable exception is the CSP project (teachingsecuirty.org), which produced lesson materials on threat modeling and authentication. These lessons integrate with the AP Computer Science curriculum, with focus on preparing future software developers and engineers. Other commercially available materials from initiatives that focus on personal implications of cybersecurity include Data Detox x Youth (datadetox.myshadow.org/detox), and Garfield Cyber Safety Adventures (cybersafetykits.org), and Balbix Cybersecurity Activity Book for Kids (balbix.com/resources/kids-cybersecurity-activity-book).

Besides a general scarcity of cybersecurity training initiatives for middle school-aged children—only Data Detox x Youth specifically targets early adolescents—the general area of cybersecurity training also lacks a foundational, theory-based pedagogical approach that promotes an in-depth understanding of cybersecurity principles [5]. Our goal is to introduce such a theory-based pedagogical approach.

AI-Cybersec Concepts	Mathematical Concepts	Metaphorical Mental Model	Example
AI inferences & identifiability	Entropy, probability & information gain	Guessing games	AI can infer things you didn't tell it. Teach students to avoid answering questions that have a high information gain.
Risk of data sharing & recombination	Exponential growth	Spread of infections	Teach students about information brokers. Every time your data is shared and re-shared to *n* others, the risk grows *n*-exponentially.
Detecting deep fakes	Generative Adversarial Networks (GANs)	Faking handwriting	Show how difficult it is for human to create "fakes", but how easy it is for a GAN.
AI inferences & collective data privacy	Graph theory & collaborative filtering	Herd immunity	AI can learn things about you by studying people who are like you. Disclosure can negatively affect others even if it does not impact you.
Forecasting & racial profiling	Difference between estimation & exploration	Fortunetelling, racial profiling (by a person)	AI is good at estimation but not at extrapolation. Teach students the danger of spurious inferences (e.g. profiling).
AI acting as humans	Markov chains	Impersonation	Show students how a rudimentary "Twitter bot" and/or conversational agent works.
Photo obfuscation	Encode/decode geometric features	Wearing a disguise	Have student use face paint to trick facial recognition software vs. classmates.

Table 1: Mathematically grounded metaphors for AI-related cybersecurity concepts.

2.2 Folk Models of AI-Related Cybersecurity

Researchers from the National Institute of Standards and Technology (NIST) outlined a vision that provides a starting point for our work. Their key recommendation is to tailor cybersecurity education efforts to user perceptions [5] since end-users typically have "folk models" of cybersecurity that are incorrect and/or incomplete. Folk models often result in poor decision making and ineffective privacy and security protective behaviors [2,14].

An important first step in challenging and changing students' existing cybersecurity practices thus involves recognizing the predictable preconceptions that are inherent to the folk models students have about AI-related cybersecurity [9,10]. For example, in their study of online behavioral advertising, Yao et al. [15] identified four folk models held by participants, and each one was connected with different user behaviors and preconceptions about tracking. Whereas folk models of "home computing" security have been studied extensively, no such body of research exists for AI-related cybersecurity beyond Yao et al.'s [15] study. Likewise, there is a dearth of learner-centered educational approaches for cybersecurity, with most work focusing on behavioral adjustments through training rather than empowering users with cybersecurity fundamentals [8,9]. In response to these gaps, we propose to investigate a learner-centered educational approach that is responsive to middle school students' preconceptions and folk models of AI-related cybersecurity.

2.3 Teaching with Metaphorical Mental Models

A potential educational mechanism that has convenient parallels with the "folk models" approach is the use of metaphorical mental models from areas the user is more familiar with (e.g., disease risk, physical security risk, criminal behavior risk) to demonstrate their resemblance to AI-related cybersecurity risks. While metaphors have been hailed as an effective tool for education [7], Brase et al. [3] show that such metaphors fail to impact users' cybersecurity behaviors. A potential reason for this is that metaphors as a proxy for relational understanding (reification) is difficult to achieve in abstract disciplines, especially when students have a shaky understanding of the foundational concepts that drive the metaphor [7,8].

In AI-related cybersecurity, most prevailing metaphors have a mathematical basis (e.g. exponential growth, graph theory, entropy). Hence, purposefully integrating metaphor-based cybersecurity education into a math curriculum would result in synergies, where the mathematical concepts provide a basis to improve students' understanding of the cybersecurity-AI metaphors, and the cybersecurity-AI metaphors in turn provide relevant and relatable examples that can improve students' understanding of the underlying mathematical concepts (see Table 1). An added benefit of the integration of cybersecurity-AI into a math curriculum is that an understanding of the mathematical principles will allow students to generatively reason (i.e., reasoning about cases that are beyond the scope of the original metaphor). This is particularly important in AI-related security, where risks evolve at an accelerating pace.

3 Research Plan

We propose a mixed methods, exploratory research study [4] to investigate middle school students' AI-related cybersecurity competencies relative to their mathematics knowledge and behaviors, which we will use to develop a time- and teacher-friendly learning module that can be readily integrated into existing middle school math curricula. To optimize our impact, we will implement the proposed module at a high-needs, rural-identifying middle school in South Carolina that serves underrepresented students. While this is a challenging task—the students involved in this project likely perform under the national average, especially in the aftermath of the COVID-19 pandemic—our project is positioned to yield high rewards: transformative school-based experiences that will improve underrepresented students' learning, middle school math curricula, identifiable cybersecurity competencies and practices, and a more ethical AI. The project will be evaluated using pre/post-tests of students' cybersecurity knowledge, behavioral intention surveys, math affinity, observed decision-making in a "cybersecurity drill," and module-specific test scores. The following subsections describe the steps in our research plan.

3.1 Exploring Students' Folk Models (completed)

We have conducted a qualitative interview study with 33 middle school students to investigate their folk models of AI-related cybersecurity issues and underlying mathematics. We recruited participants for this study by administering a survey asking 118 students at the collaborating middle school 21 questions about their attitudes towards mathematics. We conducted a Confirmatory Factor Analysis (CFA) on the data and found three distinct factors measuring whether math was a) fun, b) useful, or c) a waste of time and effort. We recruited participants who varied maximally on these three scales, and added students to increase the diversity of our sample.

The goal of this study was to gather a list of topics to be covered in the educational module. As such, we organized our interviews around the following research questions:

- *How do middle school-aged children think about AI and cybersecurity?* What are their opinions, expectations, and fears about online interactions? What privacy-enhancing behaviors do they engage in? What is their strategy? How do they rationalize it?
- *What are the (cyber-)interests of middle school-aged children?* How do they assess the risks of the online activities they engage in (cf. [6])?
- *How do middle school-aged children engage with mathematics, both inside and outside the classroom?* How do they characterize what it means to be "smart" in mathematics and who can be good at it? Which aspects of mathematics do they find particularly interesting, useful, boring and/or challenging? How do they cope with challenging concepts? Do they feel that they "belong" in mathematics? What connections do they see between math and using technology?

The results of our interviews suggested that students perceived AI as robots or non-playable game characters, but they did not consider prediction algorithms to be AI. When asked about how a streaming service would be able to make music/movie recommendations or how online advertisements are personalized, several students imagined that a real person would be "using Google behind the scenes."

Unsurprisingly, then, while students were very familiar with cybersecurity threats that could be perpetrated by their social environment (e.g. revenge porn, cyberbullying), and somewhat familiar with threats from unknown individuals (e.g. social engineering, hacking, and ransomware), most students were *not* familiar with the cybersecurity threats that emanate from online algorithms (e.g. data collection and inference, filter bubbles, fake news propagation). We hope that our metaphorical mental models can help students understand how AI drives online algorithms and what the cybersecurity implications of such algorithms are. A full paper with detailed results of the interview study is forthcoming.

3.2 Developing the Module (in progress)

Using our findings from the interview study and feedback from teachers, we will develop the middle-school level education module. The outcome of this step will be a deployable, evidence-based, theory-driven module that includes activities for each grounded metaphor (Table 1) and has been tailored to each middle school grade level. This step consists of:

1. revising and expanding our grounded metaphors based on the interview study outcomes and formative evaluation feedback;
2. devising grade level-appropriate educational tasks around these grounded metaphors, using input from middle school math teachers and our external evaluator;
3. developing themed lesson materials for each task, leveraging the (cyber-)interests of middle school-aged children.

3.3 Deploying the Module (projected Fall 2021)

The deployment of the module will take place at a middle school in the area. Approximately 300 fifth through eighth grade students attend the school. The deployment will involve a parallel collaboration between the five math teachers and the technology teacher, and spans the following phases:

1. preparation period in which we introduce the teachers to the module;
2. pre-test to evaluate students' pre-existing math and cybersecurity knowledge;
3. student engagement in the module, which includes frequent check-ins with teachers;
4. post-test to evaluate the effects of the module;
5. qualitative exit interview with 20 students;
6. debriefing interviews with teachers.

The outcome of this step will be empirical evidence of factors and conditions associated with students' folk models of cybersecurity relative to their understanding of related mathematical ideas.

3.4 Evaluating the Module (end of Fall 2021)

Following accepted standards in HCI and Education research, we will evaluate the education module as follows:

- We will evaluate the effect of the education module on students' understanding of AI-related cybersecurity principles using pre/post comparisons of their folk models (comparing the results of the initial qualitative interview with the exit interview), a quantitative pre/post-test of their AI-related cybersecurity knowledge, and a pre/post-test of their cybersecurity-related statistical reasoning performance (as per [3]).
- We will evaluate the potential effect of the education module on students' cybersecurity-related behaviors using a pre/post behavioral intention survey, and an unannounced "cybersecurity drill" where students will be exposed to a fake AI-related cybersecurity threat, and we will observe their reactions. Note that we will work with our institution's IRB experts to design the drill, with ethical standards for research with children guiding each methodology decision.
- We will evaluate the effect of the module on students' math learning outcomes using a pre/post-test of their math knowledge (tailored to their grade level).

The outcomes of this step will be a theoretical explanation for the factors and conditions associated with students' folk models of cybersecurity relative to their understanding of related mathematical ideas, as well as a well-specified conceptual framework that supports this theorization.

3.5 Revising the Module (projected Spring 2022)

Based on our exit interviews with teachers and summative feedback from our external evaluator, we will revise the module with the aim of increasing its ease of use and delightfulness (optimizing both student and teacher enjoyment) and its effectiveness (optimizing positive learning outcomes), with the goal of producing an updated education module that can be readily deployed at any middle school without our direct involvement. The projected outcome of this step is a determination of the type of future study that comes next (e.g., design and development, efficacy study, or foundational/early-stage exploratory) based on the empirical evidence and conceptual framework.

4 Conclusion

This paper outlines how we plan to develop a middle school math module to teach AI-related cybersecurity using mathematically-grounded metaphorical mental models. Beyond our implementation at a regional middle school, we plan to release our materials for public use. Furthermore, we hope that our work can serve as an inspiration for other teams.

Acknowledgments

This work is supported by NSF award no. 2039616. We thank Khushbu Singh, Sushmita Khan, Mehtab (Shahan) Iqbal, Channing Smith, Philip Nelson and Kyra Derrick for their support of this work.

References

1. Amina Adadi and Mohammed Berrada. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6: 52138–52160.
2. Farzaneh Asgharpour, Debin Liu, and L. Jean Camp. 2007. Mental Models of Security Risks. *Financial Cryptography and Data Security*, Springer, 367–377.
3. Gary L. Brase, Eugene Y. Vasserman, and William Hsu. 2017. Do Different Mental Models Influence Cybersecurity Behavior? Evaluations via Statistical Reasoning Performance. *Frontiers in Psychology* 8.
4. John W. Creswell and Vicki L. Plano Clark. 2011. *Designing and Conducting Mixed Methods Research*. SAGE Publications.
5. Susanne Furman, Mary Frances Theofanos, Yee-Yin Choong, and Brian Stanton. 2012. Basing Cybersecurity Training on User Perceptions. *IEEE Security Privacy* 10, 2: 40–49.
6. Alice E Marwick and danah boyd. 2014. Networked privacy: How teenagers negotiate context in social media. *New Media & Society* 16, 7: 1051–1067.
7. Robert Moses and Charles E. Cobb. 2002. *Radical Equations: Civil Rights from Mississippi to the Algebra Project*. Beacon Press.
8. National Academies of Sciences, Engineering, and Medicine. 2018. *How People Learn II: Learners, Contexts, and Cultures*.
9. National Research Council. 2005. *How Students Learn: History, Mathematics, and Science in the Classroom.* .
10. Martina Angela Sasse, Debi Ashenden, D. Lawrence, L. Coles-Kemp, I. Fléchais, and P. Kearney. 2007. Human vulnerabilities in security systems.
11. Anna Sfard. 1991. On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics* 22, 1: 1–36.
12. Anna Sfard. 1994. Reification as the Birth of Metaphor. *For the Learning of Mathematics* 14, 1: 44–55.
13. Valdemar Švábenský, Jan Vykopal, and Pavel Čeleda. 2020. What Are Cybersecurity Education Papers About? A Systematic Literature Review of SIGCSE and ITiCSE Conferences. *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, Association for Computing Machinery, 2–8.
14. Rick Wash. 2010. Folk models of home computer security. *Proceedings of the Sixth Symposium on Usable Privacy and Security*, Association for Computing Machinery, 1–16.
15. Yaxing Yao, Davide Lo Re, and Yang Wang. 2017. Folk Models of Online Behavioral Advertising. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, Association for Computing Machinery, 1957–1969.