International Society of the Learning Sciences

Doors, Decisions, and Discovery: An Interactive Smart Door Lock System to Promote Children's Understanding of AI Classification

Aayushi Dangol, University of Washington, adango@uw.edu
Robert Wolfe, University of Washington, rwolfe3@uw.edu
Rotem Landesman, University of Washington, roteml@uw.edu
Jason C. Yip, University of Washington, jcyip@uw.edu
Julie A. Kientz, University of Washington, jkientz@uw.edu

Abstract: Understanding AI's decision-making process is essential for children as they navigate a world increasingly influenced by AI-driven technologies. Despite frequently interacting with AI-powered systems, children often have limited knowledge about how these systems make classification decisions. We present *AI-Smartlock*, an educational platform designed to teach young learners about the various parameters that influence AI classification, including classification rules, confidence scores, and decision thresholds. Through image classification tasks that simulate a smart door lock, children can interactively explore these parameters. This paper describes how we designed the system and our results from a participatory design session with a group of seven children (ages 7-12).

Introduction

From early childhood, humans rely on classification to make sense of the world, a process central to what classification theory describes as essential for survival and adaptation (Smith & Medin, 1981). By categorizing people, objects, and experiences, we build mental frameworks, or schemas, that guide our understanding throughout life. Originating from Bartlett's (1932) work, schemas are not mere memory storage but reconstructive frameworks that adapt based on context, experience, and existing knowledge. In this view, people's schemas are inherently dynamic; they evolve as individuals encounter new information that either assimilates into existing structures or requires accommodation through modification (Derry, 1996). Yet, as Bowker and Star (1999) emphasize, many classification systems function as an "invisible infrastructure"—structures that shape thought and decision-making but often remain unnoticed and unexamined.

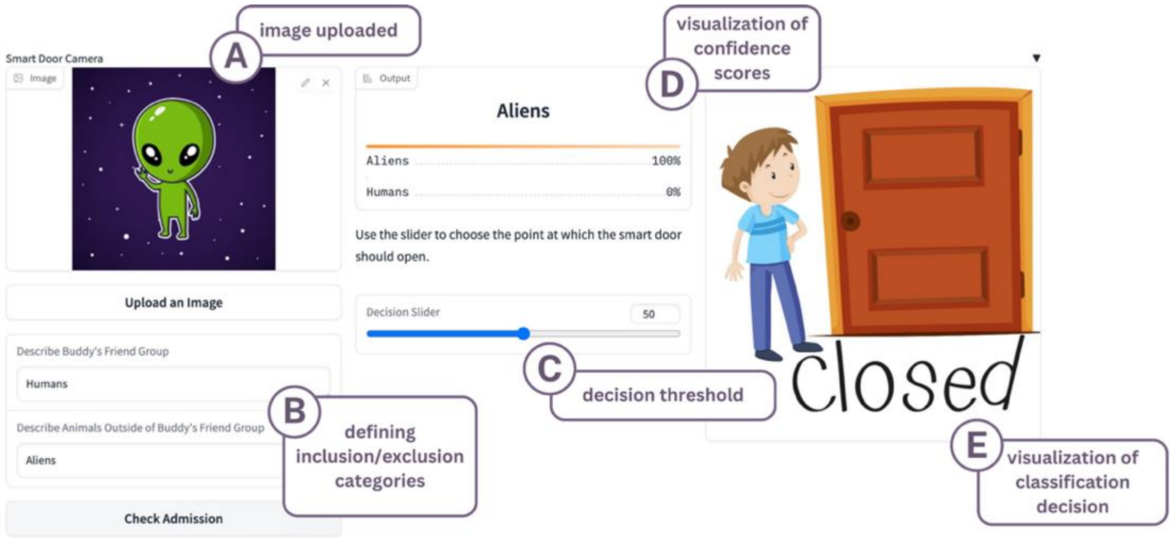
In the world of AI and technology, we argue that there is importance in helping children make sense of how machines classify data and information for several reasons. First, most AI systems operate using predefined classification parameters that evaluate data against set rules and features (Krizhevsky et al., 2012). The classification schemas that humans use to understand the world do not always align with the schemas used by AI systems, and prior research has shown that it is often unclear to children how these technical systems categorize information (Druga & Ko, 2021). As a result, without a clearer understanding of these technical schemas, children may form inaccurate mental models of AI systems, often overestimating its capabilities (Dietz et al., 2023). Second, human schemas and classifications are far from neutral; they are shaped by cultural, social, and individual experiences, which can lead to biased classification (Bowker & Star, 1999). Similarly, AI systems, trained on these imperfect human classifications, can inherit and even amplify these biases (Wolfe et al., 2024). Understanding these parallels is crucial because AI systems, much like human systems, are only as fair as the data and underlying decisions they are built on (Buolamwini & Gebru, 2018).

Building on this theoretical foundation, our study introduces *AI-Smartlock*, an interactive tool designed to support children's understanding of AI classification. Findings from our pilot study with seven children (ages 7–12) show that as they interact with the system to classify data in various settings, they engage in building and refining their understanding of the conceptual, technical, and societal dimensions of AI classification. Social interaction and collaboration further enrich schema-building, as children engage with peers and encounter diverse perspectives, prompting them to expand or reframe their mental models to better understand AI classification systems (Vgyotsky, 1978). Additionally, by comparing classification decisions made by AI to their own and those of their peers, children engage in metacognitive reflection (Flavell, 1979) and gain insights about their own internal schemas, making visible the typically hidden role these systems play in everyday life (Bowker & Star, 1999).

AI Smartlock: System Design

AI-Smartlock is designed to simulate a smart door lock, where access decisions—whether to lock or unlock the door—are determined by classifications made by the AI system. The metaphor of a locking door resonates with children's everyday experiences of being included or excluded from group activities or navigating physical spaces where access is granted or restricted. This provides a relatable way to introduce different parameters influencing AI classification and help children reflect on the real-world implications of algorithmic decision-making. Figure 1 illustrates the system components in action, with the resultant classification output.

Figure 1
Screenshot of AI-Smartlock with added annotations highlighting system components



<u>Interactive Learning with Classification Rules and Confidence Scores</u>
AI-Smartlock enables learners to establish the classification rules for the AI to follow, framing their understanding of how AI systems make decisions. Children first define the inclusion and exclusion categories for the *AI-Smartlock*, deciding which types of images will lock or unlock the door. For example, they may choose "Humans" as an inclusion category (unlock the door) and "Aliens" as an exclusion category (keep the door locked). They then upload an image to the system and observe how the AI classifies it based on the predefined inclusion and exclusion categories. Prior research highlights the importance of quick data inspection, noting that children engage more meaningfully with AI systems when they can inspect data easily (Dwivedi et al., 2021). In *AI-Smartlock*, children use images for classification as it provides a visually intuitive way to observe how the AI applies classification rules, helping them understand how various features within the image influence outcomes. To deepen their understanding, the system also presents confidence scores for each category, expressed as percentages, reflecting how confident the AI is in its classification (e.g., 100% for "Aliens" and 0% for "Humans" in Figure 1). This aligns with prior AI literacy research, highlighting the importance of displaying confidence scores to help learners interpret the AI's level of certainty in its predictions (Dwivedi et al., 2021).

<u>Adjustable Decision Thresholds for Understanding AI Uncertainty</u>
Learners can also experiment with the system's decision threshold, an adjustable parameter that controls how confident *AI-Smartlock* must be before it opens the door. For example, if the threshold is set at 70%, the system will only unlock the door if it is at least 70% confident the image belongs to the inclusion category. This feature allows children to experiment with how the AI handles uncertainty by adjusting thresholds and observing the impact on its decision-making. Through this hands-on interaction, children can understand the concept of decision boundaries and the trade-offs between sensitivity (detecting true positives) and specificity (avoiding false positives), which are key considerations in real-world AI systems. Moreover, when interacting with *AI-Smartlock*, learners bring prior expectations about the classification outcomes—whether the door should open or stay locked—based on the categories they have defined. Immediate feedback from the system allows them to quickly recognize when the AI's classification does not match their expectations. Such instances of AI error prompt children to engage in deeper cognitive processing as they compare their own schemas with the AI's knowledge representation, reflected through confidence scores. By exploring these discrepancies, children can evaluate why the AI made an incorrect classification, examine the confidence scores, and experiment with adjusting the decision threshold to refine the AI's behavior.

Preliminary Findings

We conducted a participatory design session with KidsTeam UW, an intergenerational co-design group (Druin, 1999) based at our university. Seven child participants (ages 7–12) joined five adult facilitators (researchers and research assistants) for a 1.5-hour session. To build rapport and encourage open dialogue, we began with a 15-minute informal discussion of commercially available smart door lock systems, prompting participants to consider, “How does AI know when to unlock a door for someone?” Next, during a 60-minute hands-on activity, we introduced *AI-Smartlock* through a narrative featuring a character named Max and his dog, Buddy. In the story, *AI-Smartlock* granted Buddy access but blocked some of his friends. Children worked in groups to help *AI-Smartlock* grant access to Buddy’s friends while keeping out unapproved visitors. We closed the session with a 15-minute group discussion to reflect on children's experiences with *AI-Smartlock*. In the findings that follow, we present representative vignettes that illustrate children’s learning processes, embedding our analysis within each vignette.

Classification Rules, Confidence Scores & Decision Thresholds

As children interacted with *AI-Smartlock*, they discovered that the door’s behavior, whether it opened or stayed closed, depended on the system’s confidence scores in relation to the categories they had chosen. For example, at the beginning of the session, Hana (girl, age 8) and Niko (boy, age 8) set “hamsters” as their inclusion category and “dogs” as their exclusion category. When they uploaded a hamster image, *AI-Smartlock* displayed its confidence scores as “100 percent hamster” and “0 percent dog.” When asked about what was happening, Niko pointed to the hamster’s confidence score, saying, “this will let it in,” and then to the dog’s confidence score, adding, “this definitely won’t let it in.” Niko’s statements reflect his understanding of the relationship between confidence scores and the AI’s decision-making process. By saying “this will let it in,” Niko shows he recognizes that a high confidence score in the inclusion category prompts the door to open. Conversely, when he says, “this definitely won’t let it in,” he demonstrates his understanding that a high score in the exclusion category would keep the door closed. Together, these statements reveal Niko’s grasp of confidence scores as a mechanism that either permits or restricts access based on alignment with chosen categories.

Children also applied the concept of decision thresholds, which involves setting a specific point at which an outcome is classified as one way or another. By experimenting with different cutoff points, they observed how changing the threshold affected whether the door opened or stayed closed. Dylan (boy, age 9) and Riley (girl, age 12) exemplified this process. After selecting “cats” as their inclusion category and “shrimps” as their exclusion category, they uploaded an image of a “cat shrimp”—a coiled cat that resembled a shrimp (see Figure 2). To explore how the decision threshold worked, they initially set the parameter to 0 percent. *AI-Smartlock* classified the image as 56% shrimp and 44% cat but still opened the door. Reflecting on this, Riley remarked, “It thinks that because of the slider,” showing her awareness of the threshold’s impact on the AI’s behavior. Curious, she then asked, “If it’s at 50, how would it change?” Dylan adjusted the slider to 50%, and this time, the door stayed closed. Observing this, she noted, “50 does not open it.” They continued testing various thresholds between 50% and 100%, but the door remained closed each time. When Riley lowered the threshold to 30%, the door opened, prompting her excited exclamation, “AAH!” Trying again at 40%, the door still remained opened. She then gradually increased the threshold until the door closed. Riley concluded, “Everything from here” (pointing to the AI’s confidence score for cat) “causes the door to close,” revealing her growing understanding that while confidence scores indicate how much the AI “believes” an image fits a category, the decision threshold ultimately determines if this confidence is sufficient to trigger an action, such as opening the door.

Figure 2
Screenshots of decision threshold exploration in AI-Smartlock.

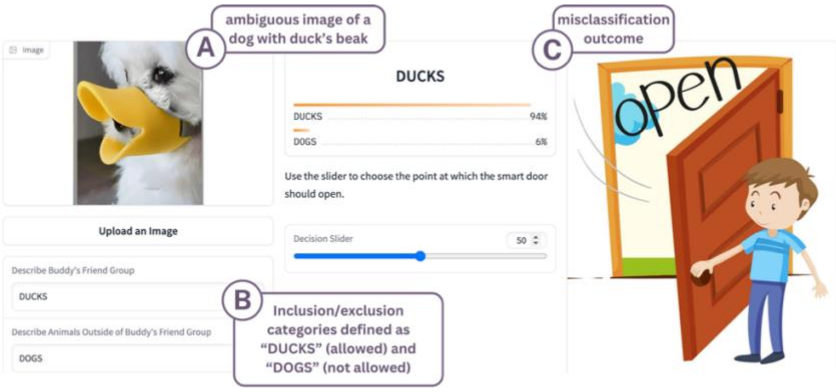
(2a) A "cat shrimp" image is uploaded. (2b) Setting a 30% threshold opens the door. (2c) Raising the threshold to 80% closes the door, demonstrating the impact of thresholds on AI decisions.

Ambiguous inputs can trick AI

Children explored the concept of ambiguity in AI classification, examining how AI systems respond to inputs that do not fit neatly within predefined categories. By intentionally selecting images that blurred the lines between categories, children gained insight into the challenges AI faces with deceptive inputs. For example, Hana (girl, age 8) and Niko (boy, age 8) expanded on their exploration of inclusion and exclusion categories by devising an experiment to trick the system. Since dogs were set as their exclusion category, in their previous explorations, *AI-Smartlock* had blocked all the images of dogs from entering. Curious to see if they could trick the AI, they uploaded an image of a dog wearing a duck beak (see Figure 3) to see if the disguise might bypass the system. Before testing the AI, they discussed why the AI might misclassify the image as a duck. Hana observed that while the image had the "fur of a dog and the nose of a dog… the fake duck mask [could] confuse the AI." When the AI indeed classified the image as a duck, Niko reflected on the difference between human and AI perception. He remarked, "it's not the same thing as human…AI thought the image was a duck because somebody put something like a beak."

Figure 3

Screenshot demonstrating how ambiguous inputs can lead to misclassification with added annotations highlighting children's input and the AI's classification output.



Discussion

Humans naturally rely on their own classification systems to organize and interpret information, a process deeply rooted in classification theory and schema development (Smith & Medin, 1981). These systems are constructed through experiences and cultural contexts, allowing individuals to make decisions and judgments about the world around them. In parallel, AI technologies use their own embedded classification systems and shape countless societal decisions such as determining eligibility for loans, ranking job applications and optimizing healthcare diagnostics. Yet their mechanisms remain largely inaccessible to users, including children. This raises a critical question: *how should children navigate situations where their own classification schemas interact, or conflict, with the opaque classification logic of AI systems?* Addressing this question is vital for equipping children with the skills needed to critically engage with AI technologies in their daily lives (Long & Magerko, 2020). Our findings demonstrate that open-ended exploration with *AI-Smartlock* enabled children to compare and contrast their own decision-making processes with those of the AI. Through iterative experimentation with various classification parameters, children actively hypothesized about AI's decision-making processes, tested their predictions, and refined their understanding about AI classification. Moreover, children's experiments to trick the AI, sparked meaningful dialogue around how "AI thinks" and helped them recognize that AI, while powerful, has limitations and requires thoughtful consideration of the assumptions it carries. Future studies could scaffold this progression by first focusing on personal, relatable tasks (e.g., smart locks), then gradually introducing real-world examples of AI's impact (e.g., algorithmic bias in hiring or facial recognition). This progression could help learners contextualize AI's role in society (Dangol & Newman et al., 2024) while retaining the hands-on, inquiry-based engagement that *AI-Smartlock* offered. Additionally, in future iterations of the system design, we plan to incorporate visual markers, such as heatmaps, to make AI's decision-making processes more transparent. By visually highlighting the areas the AI focuses on when classifying inputs, children may gain deeper insights into the underlying assumptions and potential biases of AI systems.

References

Bartlett, F. C. (1995). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

Bowker, G., & Star, S. L. (1999). *Sorting things out. Classification and its consequences.* MIT Press.

Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency* (pp. 77-91). PMLR.

Dangol, A., Newman, M., Wolfe, R., Lee, J. H., Kientz, J. A., Yip, J., & Pitt, C. (2024, July). Mediating Culture: Cultivating Socio-cultural Understanding of AI in Children through Participatory Design. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (pp. 1805-1822).

Derry, S. J. (1996). Cognitive schema theory in the constructivist debate. *Educational psychologist*, 31(3-4), 163-174.

Dietz, G., Outa, J., Lowe, L., Landay, J. A., & Gweon, H. (2023). Theory of AI Mind: How adults and children reason about the "mental states" of conversational AI. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45, No. 45).

Druga, S., & Ko, A. J. (2021). How do children's perceptions of machine intelligence change when training and coding smart programs? In *Proceedings of the 20th Annual ACM Interaction Design and Children Conference* (pp. 49-61).

Druin, A. (1999). Cooperative inquiry: developing new technologies for children with children. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 592-599).

Dwivedi, U., Gandhi, J., Parikh, R., Coenraad, M., Bonsignore, E., & Kacorri, H. (2021, October). Exploring machine teaching with children. In *2021 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* (pp. 1-11). IEEE.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34(10), 906.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems, 25.*

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-16).

Smith, E. and Medin, D. (1981). *Categories and concepts*. Harvard University Press, Cambridge, MA.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (Vol. 86). Harvard university press.

Wolfe, R., Dangol, A., Howe, B., & Hiniker, A. (2024, October). Representation Bias of Adolescents in AI: A Bilingual, Bicultural Study. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (Vol. 7, pp. 1621-1634).