Puffy and Sticking Out" Collaborative Image Classification with Kids

Yubing Tian*, Yilin Zeng* and Stefania Druga*

* Joint first authors from University of Washington, Seattle, USA

Abstract

In this pilot study, we are exploring how best to design explainable image classification algorithms for children. We ran two co-design sessions with KidsTeam UW, which aim to understand the way children (aged 6 – 11 years old) approached visual classification games, collaborated with their peers, and compared themselves to machines doing similar tasks. We present our initial findings around opportunities to engage children in playful ways with a critical reflection of image classification algorithms. We also propose a series of design recommendations for future curriculum on explainable image classification. In the upcoming months, we plan to digitize the study design and extend the collaborative paper-based activities to a blended learning environment. We hope this workshop would be an excellent avenue for further refining our next steps in this study by exploring digital tools and platforms that can create a shared space for children to share their reflections on image classification algorithms.

Author Keywords

Interaction Design and Children; Explainable Image Classification; Games for Learning; Collaborative sense-making.

12.1 Motivation and background

Children in the modern digital information era are rapidly engaging with technologies that are powered by "artificial intelligence" (AI) (Statista 2019). AI systems show great promise in helping children and families through improved online search quality, increased accessibility via advances in digital voice assistants, and AI-supported learning. However, AI systems can also amplify bias, sexism, racism (Ferguson 2012, Gebru 2019, O'neil 2016), and other forms of discrimination, particularly for those in marginalized communities (Buolamwini 2018, Angwin 2016). Promoting a critical understanding of AI for children is of critical importance in this context. Prior studies have addressed the topic of explainable AI (XAI) for adults (Gunning 2017, Holzinger 2018), but little research has been done on explainable AI for children. In order to address this gap, we build on prior studies analyzing co-design games for learning with children and collaborative sense-making in order to propose tools and curriculum for explainable image classification for children.

With STEAM education going mainstream, coding has become a digital literacy for children. It is being integrated into school curriculums and teacher training programs around the world. Initiatives like "Hour of Code" and "Scratch Days" are currently reaching tens of millions of students in 180+ countries (Code.org, 2020). This raises the opportunity to not only teach children how to code but also how to understand how computers use machine learning to classify information such as object recognition or image prediction.

HCI studies analyzing adults' mental models of AI technologies found that even a short tutorial with an experimenter (i.e., 15 min) can significantly increase the soundness of participants' mental models. This phenomenon was consistent in Kulesza et al.'s study on intelligent music recommender systems and Bansal et al.'s study on the effect of different kinds of AI errors (Kulesza 2012, Bansal 2019). More so than users' explicit mental models, research on AI systems in HCI has focused on explainability and trust. Rutjes et al. argue for capturing a user's mental model and using it while generating explanations (Rutjes 2019). In our study design we tried to leverage this by encouraging children to engage in collaborative sense-making with their peers when participating in a series of image classification challenges.

Within this frame, we define sense-making as a process by which people come across situations or contexts that are unfamiliar but need to process and understand to move forward (Klein 2006). When thinking about children's sense-making practices regarding AI, we need to consider not only how they use smart technologies and devices (Ingold 2001), but also recognize that youth interaction with AI applications is grounded in an attentive and perceptual involvement with these devices and their peers or families. A more recent study on children's home interaction with voice assistants also argues for the importance of considering the ecology in which these devices are used by kids (Sciuto 2018).

12.2 Research Design

Our research questions were as follows:

1. What strategies do children employ in visual classification games?
2. How do children collaborate with their peers and explain their choices?
3. Finally, how can collaborative activities foster critical understanding of image classification algorithms?

To address our research questions, we employ qualitative and design methods. We ran two sessions with University of Washington’s KidsTeam, a design team composed of 11 children aged 6 to 11 years old. Overall, 10 children participated in our sessions. The two sessions focused on distinct elements of image classification, the first sessions examined label creation and the second session on anchor selection.

In the first session, children formed groups, and each group was given a set of approximately 20 images depicting either healthy or unhealthy corals. The children were told that there should be two piles in each set of images, and they had to first, guess what the two piles (and their corresponding labels) were, and second, to place individual images into the correct piles (see Figure 1). The session concluded with discussions around whether children thought machines could perform the image classification task better.

In the second session, children, in groups, were given the same set of images and an A4 cardboard for the anchor selection activity. The cardboard had a 5x5 cm cutout in the center to replicate anchors used by machine learning image classifiers (see Figure 2). In this activity, children were told to select the area of the image that best represents the entire image. The anchor activity also had an additional competitive component: once the groups had finished selecting anchors for all images, each group selected the top two anchors that they thought would beat a computer-generated anchor of the same image. Then, children blindly voted in rounds on which anchor they thought was a better representation of the entire image. Children could not vote for anchors generated by their own group.

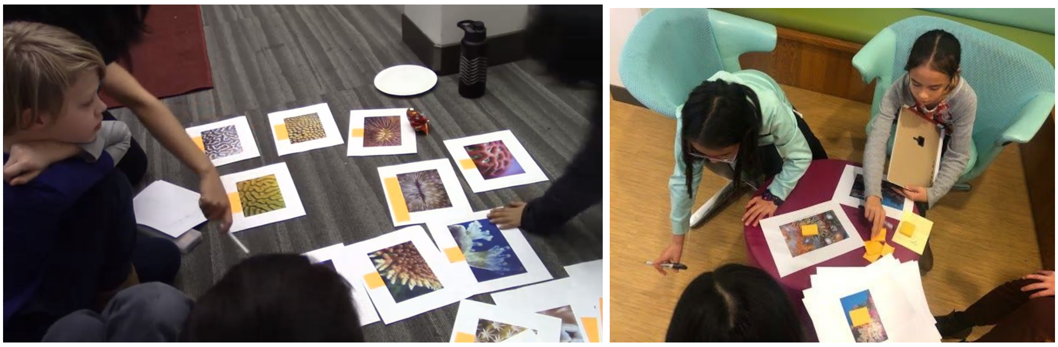

Figure 1. Study participants grouping images and discussing classification labels

12.3 Pilot Results

Initial qualitative observations suggest that children use a variety of collaborative sense-making strategies in approaching the visual classification games, and continued exposure to such activities could help foster a greater understanding of and foster critical reflections on image classification algorithms. During the two sessions we observed how children's collaborative sense-making strategies, self-comparison between their group's results and other groups' results, and a shift in perception of machine's image classification competence.

12.3.1 "*Puffy and Sticking Out"* - Kids classification practices

We observed children employing different strategies for image classification. The most common strategy was color-based image classification. Children agreed that 'colorful' and 'non-colorful' corals go to different piles. Another strategy was based on the patterns of corals in the images. For example, children identified that some corals were 'pokey' and 'smooth'. Other pattern-based labels included: 'mazes', 'fireworks', and 'puffy and sticking out'. During the discussion at the end of the session, children unanimously agreed that machines could perform the classification task better than they could. D., a 9-year-old girl, explained that it was because "they [referring to AI systems] know because they were programmed to know." K., a 10-year-old boy, shared his reasoning of why machines might be more knowledgeable: "it is called the search engine."

12.3.2 "*Where is the fish?*" - Finding image anchors

In the anchor-selection session, we observed children in all three groups employing a similar strategy. Children selected anchors that were deemed representative of the image at hand. Children selected anchors that included corals, water, and fish when possible. When asked why they selected a particular anchor, R., an 11-year-old boy, replied: "there are a lot of corals and a lot of fish". Ak, a 6.5-year-old boy, said: "because there's water, corals and pink fish."

12.3.3 "*We beat it!*" - Kids vs Machines Game

At the end of the game, we observed a shift in children's perception of machine's image classification competence. During the discussions around children’s outputs versus machines’ outputs, children voted for children-generated anchors over machine-generated anchors in seven of the eight rounds. Children also no longer thought that machines could perform the classification task better than humans. During the discussions, we observed that children demonstrated an understanding that the performance of the algorithm is dependent on the images in the training set. Ad., an

11-year-old boy, shared that "on one [image] there was bright white coral and we didn't know it was alive or not, so it [referring to AI systems] could think that this one is alive if the other one [meaning image] was alive or not". Children pointed out that even when machines classify images correctly if the conditions (of the ocean) changed then the classifications could be wrong. Ad. said, "maybe they beat people, and then maybe the area changes and that it [the classification] could not be true anymore."

12.4 Coraland: Gamified Image Classification App

Currently, we are working on digitizing the study design. We introduce a beta version of Coraland, a gamified machine learning and training application with image classification activities. The primary objective of the digital app is to create an inclusive digital environment that fosters children's AI literacy through gamified activities where children can gain an understanding of how computer vision algorithms are designed in order to foster critical thinking through collaborative activities.

Figure 2. Overview Coraland Digital Platform:Classification Game

The activities are designed to elicit children's critical reflections through game mechanics such as a dual voice and text interface prompting children's reasoning and reflections. Further, the digital application can promote sustainable and scalable design through diversifying players' composition and game mechanics. It also has the potential to bridge the gaps between traditional learning to e-learning and help to develop children's communication and collaboration skills in an online environment.

Before starting the app development, we surveyed education platforms, AI learning platforms, and machine learning games and identified features that promote children's critical thinking, while soliciting their analysis, reflection, and feedback gathered from others on their work (see Table 1).

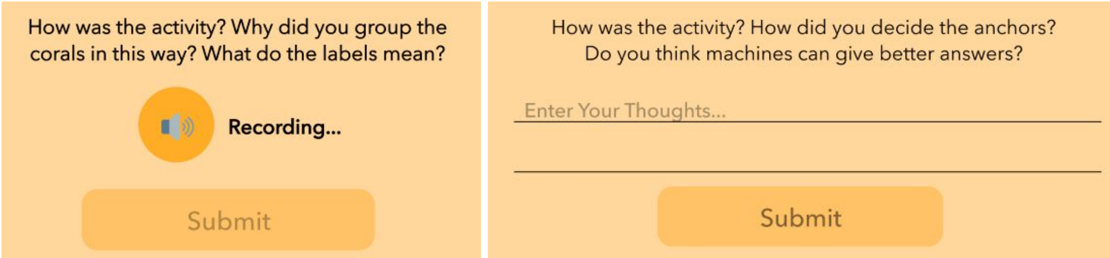

Figure 3. Coraland features:eliciting children's reflections on the game

For our Coraland design, we decided to implement specific features based on observations from the initial co-design sessions with kids. Assigning labels to different groups of images during the classification game contributed to children's critical reflection. For one image set, L., a 10-year-old boy, suggested:

"I would take the normal corals together and others go to the other pile."

In another round, A., an 8-year-old boy reached out to grab an image and said:

"That goes in the pokey [image group]".

Ak., a 6.5-year-old boy and L. agreed and handed the picture over to A. who put the picture in the 'pokey' pile. We introduced the re-label feature, where users can change their mind and rename the labels they use for an image set as they reflect on the grouping throughout the game.

During the classification game, children changed their perspectives over time and often revised their classification decisions. L. moved closer to the images and said "No wait, no wait, I see something" and took one of the images that Ak. was holding and started sorting and organizing some of the images.

Based on this observation, we decided to allow users to make changes to their grouping decisions with the same flexibility in the app interface. Users are also able to preview all images in a given group as thumbnails. During the anchor game, children often rotated the cardboard tools when selecting the anchor that they thought best represented the entire image. For example, when Ad., an 11-year-old boy, placed the anchor in the middle of the image, K., a 10-year-old boy, rotated the anchor's angle slightly so that the anchor would include the ocean. Based on this observation, we allow users to customize the anchor points with the options to move, resize, and rotate them in the app.

12.5 Conclusion

In this pilot study, we present how children classify images, engage in image anchor detection, and compete with computer vision algorithms. Addressing the question of how we can ensure that children are given the opportunities to critically reflect on the design of increasingly pervasive technologies, such as image classification algorithms, through collaborative activities. In our next steps, we aim to address how digital tools and platforms can create a shared space for children's reflections.

12.6 Acknowledgement

Thank you to all the children in University of Washington's KidsTeam, all the undergraduate and graduate students that helped with transcripts, and Jacobs Foundation for the support.

Name	Type	Features For Critical Reflections	Notes
Curiosity Machines [2]	A hands-on AI learning competition. Children create a final project that solves a problem in their community using AI through 10 lessons.	1. Test your knowledge. 2. Share your feedback. 3. Further Learning.	1. After completing a lesson, children can take a mini quiz to see how much they have learned. 2. Examples of prompts: "What did you think about this lesson?" "Do you have any ideas on how to make it better?" 3. Children can share their thoughts with their families through a table of exploratory questions.
Flipgrid [4]	A video-based social learning platform for educators, learners, and families. Users create grids to spark discussions around different topics.	1. Children share their ideas, stories, and work. 2. Families are invited to the discussions.	1. Children record short, spontaneous videos to share their thoughts and build upon diverse opinions of other people. 2. Children can share videos with families, create video collections with families, or invite families to join video conferences.
The Gopher Game [10]	A mobile social game prompting players to tag location data in the real world through gameplay.	Peer-reviewed content assessment.	Players evaluate, vote, and reflect on others' data submissions through the "rate the mission" interface. Players vote on (i) has the task been completed? (ii) how difficult was it? and (iii) which individuals helped the most?
HerdIt [8]	A competitive, multi-player, web-based game collecting tags for music and social data.	1. In-game chat. 2. Group feedback.	1. Players can chat with others during the games. 2. At the end of each game, players can compare their answer with the group's answers.
ClassDojo [1]	A platform connecting teachers with students and parents to build online classroom communities.	1. Give students a voice. 2. Share moments with parents.	1. Students showcase and share photos and videos of them learning through digital portfolios. 2. Students can share their classroom moments with their parents.
Common Voice [5]	An initiative to help teach machines how real people speak.	Validate human voice clips.	Users help create quality open-source voice data by listening to voice clips of other people and by evaluating the accuracy of speech with a "yes" or "no" button.

Table 1. AI Literacy, Machine Learning Algorithms Training, and Educational Platforms Analysis

References

[1] 2011. ClassDojo. https://www.classdojo.com.

[2] 2012. AI Education for Families https://www.curiositymachine.org.

[3] 2013. Code.org.https://studio.code.org/s/oceans.

[4] 2016. Flipgrid | Empower Every Voice. https://info.flipgrid.com/

[5] 2017. Common Voice. https://voice.mozilla.org/en.

[6] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine bias. ProPublica, May 23 (2016), 2016.

[7] Gagan Bansal, Besmira Nushi, Ece Kamar, Daniel S Weld, Walter S Lasecki, and Eric Horvitz. 2019. Updates in human-ai teams: Understanding and addressing the performance/compatibility tradeoff. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33. 2429–2437.

[8] Luke Barrington, Damien O'Malley, Douglas Turnbull, and Gert Lanckriet. 2009. User-centered design of a social game to tag music. In Proceedings of the acm sigkdd workshop on human computation. 7–10.

[9] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency. 77–91.

[10] Sean Casey, Ben Kirman, and Duncan Rowland. 2007. The gopher game: a social, mobile, locative game with user generated content and peer review. In Proceedings of the international conference on Advances in computer entertainment technology.9–16.

[11] Andrew Guthrie Ferguson. 2012. Predictive policing and reasonable suspicion.

Emory LJ 62 (2012), 259.

[12] Timnit Gebru. 2019. Oxford Handbook on AI Ethics Book Chapter on Race and Gender. arXiv preprint arXiv:1908.06165 (2019).

[13] David Gunning. 2017. Explainable artificial intelligence(xai). Defense Advanced Research Projects Agency (DARPA), 2nd Web 2 (2017).

[14] Andreas Holzinger. 2018. From machine learning to explainable AI. In 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA). IEEE,55–66.

[15] Tim Ingold. 2001. Beyond art and technology: the anthropology of skill. Anthropological perspectives on technology (2001), 17–31.

[16] Gary Klein, Brian Moon, and Robert R Hoffman. 2006. Making sense of sensemaking 2: A macrocognitive model. IEEE Intelligent systems 21, 5 (2006), 88–92.

[17] Todd Kulesza, Simone Stumpf, Margaret Burnett, and Irwin Kwan. 2012. Tell me more? The effects of mental model soundness on personalizing an intelligent agent. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1–10.

[18] Cathy O'neil. 2016. Weapons of math destruction: How big data increases inequality and threatens democracy . Broadway Books.

[19] Heleen Rutjes, Martijn Willemsen, and Wijnand IJsselsteijn. 2019. Considerations on explainable AI and users' mental models. (2019).

[20] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I Hong. 2018. Hey Alexa, What's Up?: A mixed-methods studies of in-home conversational agent usage. In Proceedings of the 2018 Designing Interactive Systems Conference. ACM, 857–868.

[21] Statista. 2019. Digital assistant and voice assistant adoption 2019.