UNIVERSITY OF MARYLAND 18 56

HUMAN-COMPUTER INTERACTION LAB

HCIL

43rd Annual Symposium

May 29, 2026

Presentation Summaries and Other Information

Human-Computer Interaction Lab (HCIL)
University of Maryland, College Park

The Human-Computer Interaction Lab (HCIL), launched in 1983 at the University of Maryland, has a rich history of transforming the experience people have with new technologies. From understanding user needs to developing and evaluating these technologies, the lab's faculty, staff, and students have been leading HCI research for more than 40 years.

The HCIL was established as an interdisciplinary effort within the University of Maryland's Institute for Advanced Computer Studies (UMIACS). Today, HCIL participants include faculty, staff, and students from the following units on campus: INFO College, Computer Science, Psychology, Education, English, Engineering, Journalism, and American Studies. In addition, HCIL faculty are members in a number of centers and institutes on campus, including UMIACS, the Maryland Institute for Technology in the Humanities (MITH), the Applied Research Laboratory for Intelligence and Security (ARLIS), the Maryland Cybersecurity Center (MC2), the Social Data Science Center (SoDa), the Artificial Intelligence Interdisciplinary Institute at Maryland (AIM), the Institute for Trustworthy AI in Law & Society (TRAILS), and the Dingman Center for Entrepreneurship.

This booklet contains Symposium presentation summaries and selected highlights of our faculty and students' news and accomplishments over the past year.

Please visit our website (https://hcil.umd.edu) for more information about the research happening in the HCIL.

If you would like more information, please contact:

Human-Computer Interaction Lab
2105 Hornbake Building, South Wing
4130 Campus Drive
College Park, Maryland 20742
hcil-info@umd.edu

HCIL Director's Letter

HCIL's 43rd Annual Symposium | May 2026

We (Joel and Jessica) have been working together for a long time, making this symposium even more special, as it is Jessica's last day – and Joel's first day – as HCIL director. So before we get into our reflections on the work being done in the lab over the last year, please allow Jessica to indulge in a little meta reflection.

****

I joined the INFO College at UMD in 2012, having earned a PhD in Communication but knowing it wasn't the best fit for my research. I'd recently had my first paper accepted to the ACM CHI conference but didn't really know how deep UMD was connected to the origins of HCI research. Lucky for me, my office was right next to the lab and I joined this amazing group of people immediately.

I don't know if I've ever considered myself a "true" HCI scholar – I love to share that I have never taken an HCI class in all my schooling. That said, the HCIL has always framed itself as a big tent, and with HCI being inherently interdisciplinary, I never felt unwelcome in the lab. Rather, I have been regularly inspired by my colleagues who work in the lab, and I have learned so much about research, collaboration, and design thinking.

It has been a privilege to be part of the leadership of the HCIL for the last decade. I love this lab and all the amazing folks who have come through it. And I am even more excited for the future of the lab because I think Joel is going to be a **much** better director than I ever was. He has the vision and the passion (and energy!) to lead the HCIL into our next chapter as we approach our 50th anniversary.

I may be leaving the HCIL and UMD officially, but this space will always hold a special place in my research-loving heart.

****

And now onto the celebration part of our letter!

HCIL's faculty and students are doing important work across a wide range of topics, and today we celebrate work done in the last year. Many HCIL members have recently returned from sharing their work at the ACM Conference on Human Factors in Computing Systems (CHI), held

this year in Barcelona, Spain. UMD had yet another impressive showing this year; by our count, more than two dozen students and faculty presented 19 papers, as well as workshops and posters. Notably, HCIL authors are included on two honorable mentions (top 5% of papers). An impressive showing indeed!

One of the core things that makes the HCIL special is our wonderful diversity. Today you'll hear from many of our students and faculty on a wide range of research topics. In total, there will be three plenary talks, one panel, and 27 lightning talks spanning design methods, visualization, usable privacy and security, sensemaking and collaboration, well-being and personal informatics, and much more. We hope you leave today with fresh inspiration and ideas, and maybe a new friend or collaborator for the coming year and beyond!

There is much to celebrate this year, and we encourage you to check out the back of the program, which includes a list of some of the awards earned by HCIL faculty and students. We'd also like to call out a few extra-noteworthy items. First, this year we welcomed our newest HCIL faculty member, Keke Wu, to campus. Second, a big congrats to our HCIL faculty who were recently promoted! Ge Gao, Caro-Williams-Pierce, and Huaishu Peng were promoted to Associate Professor and awarded tenure, and Sheena Erete and Eun Kyoung Choe were promoted to Full Professor. Promotion in academia is a very long and involved process, and promotions are a testament to the research impact they have in HCI. We also wish a hearty congratulations and best wishes to the graduating PhD and master's students – some of them are listed in the back of the program.

It takes a village to run this symposium, so we want to offer our gratitude to those who have supported us, including through financial support (VEX Robotics, funding agencies, campus units), as well as the facilities team at Iribe, and the events and communications teams in INFO. Special thanks goes to Paris Lane, the HCIL Coordinator!

Finally, we want to thank the students and faculty who make up the HCIL. You inspire us every day with your curiosity, ingenuity, and perseverance. We're excited to share with you some of the amazing work they've been up to over the last year.

Welcome to the 43^{d} annual HCIL symposium!

Jessica Vitak, Outgoing Director

Joel Chan, Incoming Director

Human-Computer Interaction Lab (HCIL)
University of Maryland, College Park

43rd Annual Human-Computer Interaction Lab (HCIL) Symposium

May 29, 2026

Brendan Iribe Center, University of Maryland

Detailed Program

Time (ET)	Details
9:00am	**Registration Opens (atrium)**
	Check in and grab your name tag, some swag, and a coffee.
9:30am	**Welcome Address**
	Jessica Vitak and Joel Chan – Outgoing and Incoming Directors of the HCIL
9:50am	**Plenary Talks**
	• *David Weintrop & Rotem Israel-Fishelson: API Can Code: An Interest-driven Data Science Curriculum* • *Fengfeng Ke: From Practice Partner to Co-Learning: Designing Role-Fluid AI Partnership for Skills Development* • *Caro Williams-Pierce and Khondaker Mushfiqur Rahman: HCI and Sludge*
10:50am	**Lightning Talks: Session 1**
	12 short talks on usable privacy & security, data vis, and design methods
12:30pm	**Lunch**
	Grab a drink and a boxed lunch and make some new friends.
1:30pm	**Panel**
	DetectAIves: Intergenerational Learning with Community Members about AI • ***Participants:*** *Caro Williams-Pierce, Galina Reitz, Tamara Clegg, Kibron Tesfatsion, Samarth Ashwathanarayana, Sarah Bamba*
2:15pm	**Lighting Talks: Session 2**
	9 short talks on sensemaking and collaboration
3:30pm	**Coffee Break**
	Coffee, tea, and light snacks
4:00pm	**Lightning Talks: Session 3**
	8 short talks on personal informatics and well-being
5:45pm	**Reception (atrium)**
	Stay for cake and conversation, plus our demo and poster reception.

Sponsorships

The research that happens in the lab—and the lab itself—cannot happen without financial support, and we are grateful to our corporate, campus, and research sponsors, especially those who have contributed to HCIL and its members in the last year.

Industry Sponsor:

VEX Robotics

Campus Sponsors:

College of Information
College of Education
Computer Science Department
Artificial Intelligence Interdisciplinary Institute at Maryland (AIM)

Research Sponsors:

Army Research Office
Chan-Zuckerberg Initiative
Google Research
Meta
National Aeronautics and Space Administration (NASA)
National Institutes of Health (NIH)
National Science Foundation (NSF)
Navigation Fund
Sloan Foundation
Teaching and Learning Transformation Center (TLTC)
University of Maryland
US Defense Advanced Research Projects Agency (DARPA)
US Department of Defense (DoD)
US Food & Drug Administration (FDA / M-CERSI)

PLENARY SPEAKERS
&
PANEL DESCRIPTION

Human-Computer Interaction Lab
43rd Annual Symposium

API Can Code: An Interest-driven Data Science Curriculum

David Weintrop[1,2], Rotem Israel-Fishelson[2]
[1] College of Information; [2] College of Education
weintrop@umd.edu, rotemisf@umd.edu

OVERVIEW

In our data-driven world, students must develop the ability to interpret, analyze, and reason with data to actively engage in daily life and be an informed citizen [1]. Data science education has therefore emerged as a critical area within K-12 education, aiming to equip students with the skills needed to transform raw data into actionable insights [2]. This work presents API Can Code, an interest-driven curriculum designed to lay the computational foundations of data science in high school classrooms while situating it within students' lived experiences [3].

The API Can Code curriculum is informed by interest-driven learning theory [4] and grounded in the data science cycle, encouraging students to ask questions that matter to them, identify and evaluate relevant data sources, programmatically access, manipulate, and analyze the data, and communicate findings using visualization. By positioning students' interests as the entry point into learning, the curriculum fosters deeper engagement and personal relevance. Students explore data on topics they care about, ranging from social issues to entertainment, sports, and environmental concerns, allowing them to see data science not as a set of technical skills but as a way to understand the world around them [5].

The curriculum is organized into three sequential units (Figure 1) that scaffold students' understanding of data science concepts and practices.

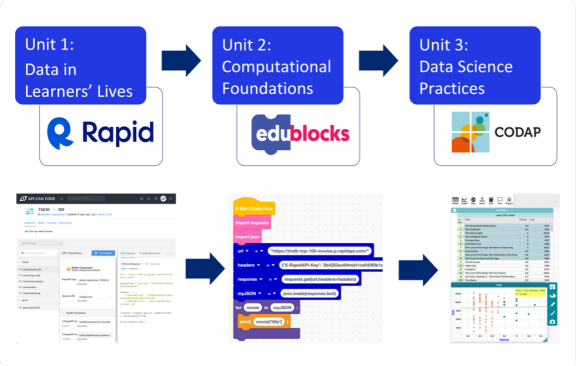


Figure 1 – The API Can Code instructional units.

The first unit introduces students to the role of data in their everyday lives, exploring how data is collected, interpreted, and used, while also engaging with issues such as data ethics and bias. Students also learn to distinguish between primary and secondary sources, collect and sample data, and critically evaluate datasets using the 5Vs for K-12 framework [6] (Figure 2).

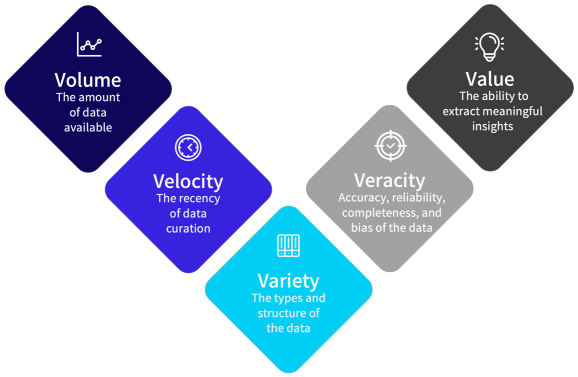


Figure 2 – The 5Vs for K-12 Framework [6]

The second unit focuses on the computational foundations of data science. Students first learn about data science and its importance, then practice retrieving and manipulating data programmatically using publicly available data from various publicly available sources via application programming interfaces (APIs). This is done using EduBlocks, a block-based programming tool designed to introduce text-based programming languages, such as Python, in a user-friendly and engaging manner. Working with various APIs is a key component of the curriculum, as they serve as accessible gateways to authentic, real-time, and personally meaningful data. Using a Use-Modify-Create approach [7], students learn to run predefined programs, modify and query data, and create new programs to request and access APIs that interest them.

The third unit emphasizes data analysis and visualization, guiding students in using CODAP, a free online data analysis platform, to further explore the datasets, create and interpret a variety of summary plots, perform statistical tests, and draw conclusions. This progression supports a coherent learning trajectory from conceptual understanding to

computational practice and, ultimately, to independent data-driven inquiry.

The curriculum culminates in a final project (Figure 3) that invites students to perform their own investigations and synthesize the concepts and skills introduced in the three units. Students identify APIs that align with their interests, formulate research questions, build end-to-end computational programs to clean and manipulate their data, create visualizations to answer their questions, and, lastly, communicate their results to their peers using oral presentations.

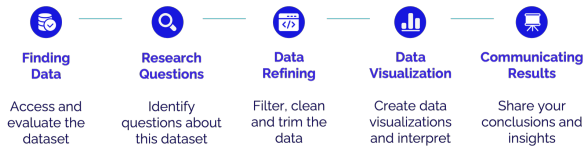


Figure 3 – The structure of the final project

PROJECT WEBSITE

To read more about the project, visit the API Can Code Website: https://apicancode.umd.edu/. The full API Can Code curriculum is available at: https://apicancode.umd.edu/curriculum.html

ACKNOWLEDGEMENTS

This research is supported by the National Science Foundation (Award # 2141655).

PAPERS

[1] T. LaMar and J. Boaler, "The importance and emergence of K-12 data science," *Phi Delta Kappan*, vol. 103, no. 1, pp. 49–53, Sep. 2021, doi: 10.1177/00317217211043627.

[2] M. H. Wilkerson, "Mapping the Conceptual Foundation(s) of 'Data Science Education,'" *Harvard Data Science Review*, vol. 7, no. 3, Jul. 2025, doi: 10.1162/99608f92.9ac68105.

[3] R. Israel-Fishelson and D. Weintrop, "API Can Code: Laying the computational foundations of data science in high school classrooms," in *Proceedings of the 57th ACM Technical Symposium on Computer Science Education V.1*, in SIGCSE TS 2026, vol. 1. New York, NY, USA: Association for Computing Machinery, Feb. 2026, pp. 505–511. doi: 10.1145/3770762.3772533.

[4] K. A. Renninger and S. Hidi, *The power of interest for motivation and engagement*. Routledge, 2015.

[5] D. Weintrop and R. Israel-Fishelson, "Bringing students' lives into data science classrooms," *Harvard Data Science Review*, vol. 6, no. 3, 2024, doi: https://doi.org/10.1162/99608f92.6d2aec03.

[6] R. Israel-Fishelson and D. Weintrop, "5Vs for K-12: A framework for helping students evaluate data," Under Review.

[7] D. Franklin *et al.*, "An analysis of Use-Modify-Create pedagogical approach's success in balancing structure and student agency," in *Proceedings of the 2020 ACM Conference on International Computing Education Research*, Virtual Event New Zealand: ACM, Aug. 2020, pp. 14–24. doi: 10.1145/3372782.3406256.

Human-Computer Interaction Lab
43rd Annual Symposium

From Practice Partner to Co-Learning: Designing Role-Fluid AI Partnership for Skills Development

Fengfeng Ke, Rosalyn Shin
College of Education/Department of Teaching and Learning, Policy and Leadership
fke@umd.edu, rhjshin@umd.edu

OVERVIEW

Artificial intelligence (AI), defined as computer systems capable of tasks typically requiring human intelligence, such as perception, reasoning, learning, and decision-making (Russell & Norvig, 1995), is increasingly embedded across domains to augment human performance in data processing, analytical reasoning, and creative content generation. Accordingly, discourse on human–AI relations has shifted from a "man versus machine" paradigm to a "man with machine" paradigm, emphasizing *augmentation* or *collaboration* over competition (Dellermann et al., 2019; Shneiderman, 2020). Within this view, AI is conceptualized as an *assistant* or a *collaborator* that extends human cognitive and creative capacities (Seeber et al., 2020; Lai, 2023).

Although research on human-AI interaction has advanced, particularly in supporting decision-making and analytical reasoning, there is limited understanding of how to design or scaffold *novice–AI collaboration* that promotes skill development and sustained learning rather than cognitive offloading or short-term task completion (Fragiadakis et al. 2024; Song et al. 2024). Moreover, *continuous co-learning*, in which humans and AI systems mutually adapt and improve through iterative interaction, remains underexamined, especially in dynamic, context-dependent tasks requiring social, emotional, or interpretive nuance (Lu et al. 2025, Schoonderwoerd et al. 2022).

Building on the Dynamic Relational Learning-Partner (DRLP) framework for human-AI collaboration (Mossbridge, 2024) and the perspective of appropriate AI reliance (Cao et al., 2024; Chen et al., 2023; Vasconcelos et al., 2023), we propose a conceptual design model of a *role-fluid, AI-driven practice-learning partner* to support novices' skill development in dynamic, context-dependent tasks. The model integrates recent advances in large language model (*LLM) generated personas* with *agent-based modeling*—an established approach for simulating interactions among autonomous agents in social and ecological systems. Using a multiple-case, mixed-methods design, we examined its implementation and feasibility across three domains: (1) virtual students as practice partners for pre-service teachers in simulation-based training for responsive teaching; (2) simulated patients as practice partners for nursing students in transitional care training, and (3) AI co-learners collaborating with novice researchers in multimodal qualitative data coding.

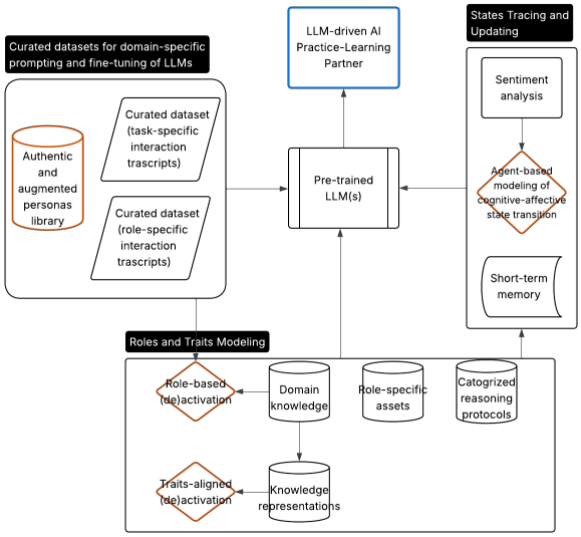


Figure 1 – Proposed Architecture for Operationalizing ALPHA

ALPHA—CONCEPTUAL MODEL OF AI-DRIVEN PRACTICE-LEARNING PARTNER

We propose a conceptual design model called *ALPHA*—an Adaptable (AI) Learning Partner for Human Advancement—that can dynamically shift its epistemic stance, role, and communicative behavior depending on the learner's needs. This is richer than a performance assistant or support tool: it's an interactive co-agent for learning and skill formation, grounded in enactive, role-sensitive, and perspective-shifting dynamics. ALPHA-supported enactive learning emphasizes that skills formation is fostered through dynamic engagement, not merely transmitted or retrieved. Learning is enacted through the evolving interaction, sense-making, reflection-in-action, and adaptation between the learner and their environment (or the practice situation and

partner). Rather than functioning as a static tutor or content provider, the AI system operates as an *interactive co-participant* capable of dynamically shifting roles in response to the learner's intent and situational demands. It may act as a *practice partner*, supporting repeated skill application and real-time feedback through shared, hands-on engagement, or as a *learning partner*, facilitating skill formation and conceptual understanding by collaboratively navigating challenges. This *role fluidity* enables the AI to adapt its level of guidance, interaction style, and focus of support—affording both practice-based reinforcement and co-constructive learning experiences. In this way, the learner and the AI partner mutually shape the learning process through sense-making, adaptation, and action-in-context.

A proposed architecture for operationalizing ALPHA to develop an AI-driven practice-learning partner is illustrated in Figure 1. The architecture comprises three core modules: (a) domain-specific adaptation, which fine-tunes pretrained LLMs (e.g., GPT-4, Google Gemini 2.5) using curated localization datasets that combine authentic, task- and role-specific interaction transcripts with exemplars from both real-world and LLM-generated persona libraries; (b) roles and traits modeling, which simulates situation-specific roles and trait-driven behaviors through adaptive knowledge activation, representation of role-relevant skill assets (e.g., instructional noticing in a simulated peer teacher), and structured reasoning types; and (c) dynamic state tracking and modeling, which continuously traces and updates role-related states across iterative human-AI interactions to enable context-sensitive adaptation and co-learning. Together, these modules enable the AI system to fluidly transition between *practice* and *learning* partner roles while dynamically adapting to domain-specific contexts and role expectations—adjusting its interaction style, guidance level, and reasoning focus in response to users' evolving goals and situational demands, thereby embodying the principles of role fluidity, domain sensitivity, and continuous co-learning central to the ALPHA framework.

A Multiple-Case Mixed-Methods Study of ALPHA

We conducted a multiple-case study to examine ALPHA's implementation and feasibility across three skill-development domains: responsive teaching, transitional care, and multimodal qualitative data coding. These domains were chosen because they involve iterative interactions with complex, dynamic environments such as students, patients, or datasets, requiring analytical reasoning, context-sensitive judgment, and attention to social, emotional, or interpretive nuances. In the first two domains, LLM-driven AI agents acted as practice partners—simulated students and patients—for pre-service teachers and nursing students. In the third domain, the AI agent functioned as a co-learning partner, collaborating with novice researchers on multimodal data coding and analysis.

Summary of Current Findings.

Current study findings indicate that ALPHA supports role-fluid adaptation and domain-specific customization across diverse learning contexts, and can facilitate continuous skills development through iterative, context-sensitive interactions. In Case 1, results from paired-sample *t*-tests indicated that interacting with the ALPHA-driven virtual students in the teaching simulation led to statistically significant improvements in participants' teaching knowledge and skills, $t(35) = 2.40$, $p = .02$, Cohen's $d = 0.40$. Similarly, participants demonstrated significant gains in teaching self-efficacy, $t(35) = 4.48$, $p < .001$, Cohen's $d = 0.74$. Qualitative analyses triangulated across conversation transcripts, behavioral observations, and interview responses revealed that the AI-powered simulation supported flexible, divergent, and dynamic enactments of target responsive teaching practices. In Case 2, both internal and user testing indicated positive perceptions of the virtual patients' potential to enhance experiential learning in adaptive patient engagement and conversational skills. Participants also highlighted the need for clearer patient-analytics presentation and a more intuitive interface to better support reflection in action with virtual patients. In Case 3, AI coders showed limited flexibility in dynamically determining units of analysis, tending to standardize coding segments rather than adaptively ranging from phrases to extended passages, and struggled to select an appropriate interpretive perspective, such as shifting between the focal participant and their interlocutor, when analyzing conversational data. However, they demonstrated a more systematic and rigorous approach in inductive and deductive coding, incorporating broader contextual knowledge into the analytic process. This suggests that AI partners may serve as a stable, knowledge-rich reference point in collaborative coding, complementing the adaptability and context-sensitive judgment of human coders. Together, these findings highlight key design needs, including greater interactional realism, adaptive reasoning flexibility, and interpretable feedback to balance human adaptability with AI's data-driven systematicity in designing and deploying novice-AI collaboration.

PAPERS

Fengfeng Ke, Nuodi Zhang, Chaewon Kim, Alex Barrett, Rosalyn Shin, Tusher Mondol, Xin Yuan, James Whyte. From practice partner to co-learning: Designing role-fluid AI partnership for skills development. *In Proceedings of the 20th International Conference of the Learning Sciences-ICLS 2026*. In Press. (2026).

Human-Computer Interaction Lab
43rd Annual Symposium

HCI and Sludge: Death by a Thousand Clicks

Caro Williams-Pierce
College of Information
carowp@umd.edu

OVERVIEW

I take Sunstein's definition of sludge as "frictions that prevent you from doing what you want to do or from going where you want to go" [1, p. 1], and advance it into our HCI-infused modern lives. Sludge is sometimes necessary – for example, multi-factor authentication is a form of sludge that disrupts your flow, but it is *also* a necessary cybersecurity measure. Sometimes, though, sludge is intentionally designed into systems to make it harder for you to accomplish what you want. For example, if you call an insurance company, hoping for a sludge-free helpful human picking up the phone, you're actually going to spend 20 minutes navigating options only to sign up for a call-back in 2 days. (And when you miss that call-back, back to Step 1.)

BACKGROUND

Sunstein originally developed his definition of *sludge* shortly after working as the administrator of the White House office of Information and Regulatory Affairs for President Obama. His focus was primarily on paperwork, broadly writ: how people manage the red tape, bureaucracy, and requirements of navigating a modern world. Then COVID-19 hit, and everyone began shoving all their services, documents, and previously human interactions online – 'paperwork' suddenly didn't involve paper, and even the word 'work' seems too hefty for something that can sometimes just be a click.

But while a single click is small, it can have a hugely disproportionate impact on your life. A single click can: determine whether or not you receive the food stamps you are eligible for and desperately need; indicate your agreement as an end user to never ever sue a multinational company or any of its holdings, no matter what, and maybe give them your firstborn child as well; determine whether you can access your own very important personal information in varied online portals; submit your job application on time; and *the list goes on and on*. Added to that, we do many of these single clicks in a single day, and some are simple little clicks, and some have hugely disproportionate impacts, and often they're all mixed together so you can't even tell which is which [5].

METHODS: FINDING SLUDGE IN THE WILD

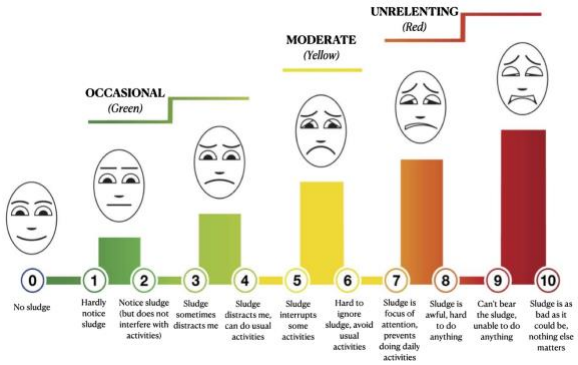


Figure 1 – Sludge Pain Scale.

Since sludge can be complex, multilayered, and multimodal, the Sludge Team conducted a reflective auto-aggro-ethnography [2] to capture different types and experiences of sludge. Reflective auto-aggro-ethnography is a methodology that embraces and validates moments of anger and frustration with technology, and recognizes the tremendous impact even tiny digital moments can have on our physical, mental, and emotional lives.

We also sought to perceive forms of sludge in our lives that *weren't* accompanied by rage, such as ubiquitous sludge (like a familiar Norman door [3], sludge can become normalized [1]), and potential sludge. Identifying ubiquitous sludge is easier when encountering counter-examples – for example, previously normalized frustration at trying to find the Chat button in a crowded Zoom bar becomes more emotionally visible when you're suddenly on a beautifully simple Cal.com call. Identifying potential sludge, on the other hand, entails following a much quieter emotion in often harried moments: *huh, I wonder why that happened – was it supposed to happen?* We adapted a modern pain scale [4] to illustrate the wide variety of sludge moments we experienced, ranging from the increasingly rare "No sludge" to the unfortunately frequent "Sludge is as bad as it could be, nothing else matters" – and I'll give examples from each of these categories in the plenary.

ABSTRACTING: FOUR TYPES OF SLUDGE

The Sludge Team captured hundreds of different sludge experiences, and then abstracted four distinct patterns of sludge. *Experiential* sludge refers to moments where a

person *experiences* sludge, such as through a conflict of use case goals, expectations, and results. *Designed for Good* sludge – MFA and cookies, for example, are meant to give us more control over our data and accounts. *Designed for Manipulation* sludge is carefully designed to push us in the direction the product or company wants us to go. See Figure 2 for an illustrated graphic of a sludge experience, with sludge designed for good (that turns out to be sludge designed for manipulation).

Figure 2 – **Experiential Sludge for Good – oh wait, nope, for Manipulation!**

Layered sludge refers to the sludge that is regularly created by multiple digital products interacting badly with each other – no one tool is responsible for the sludge, but you're stuck with it anyway. For example, if you're using Figma on Chrome, and Figma says, "Hey, use these hotkeys to design faster!", there's an even chance that those hotkeys make Chrome do something *completely* different. Just like Figma, Chrome added those hotkeys to reduce friction and increase your speed, but put together, you are definitely experiencing layered sludge!

CONCLUSION

Sludge is everywhere, and identifying its modern iteration is the first step towards 1) avoiding it in your design and development work, and 2) telling government and companies to STOP IT. Sludge is profitable – users keep paying for Amazon Prime, and people eligible for food stamps are unable to get them. Sludge can *seem* like a small problem, a hiccup in our goal-seeking – but there's a fine line between sludge that slows you down, and sludge that flat out *prevents you* from accomplishing your goal. If it takes 10 clicks to unsubscribe or 10 pages of paperwork to get your food stamps, is that bad design or intentional sludge? If it takes 100 clicks and 100 pages of paperwork, that's probably just intentional manipulation sludge, right? Well, what about 25 clicks and 25 pages? If I pay $10 for a videogame, and 100 clicks later, I'm still jumping through hoops to get the game – do I actually *have* what I paid for?

In order to answer these questions, we are continuing to identify sludge types and nuance our categories of sludge, with the goal of developing a modern day sludge audit. This sludge audit will use our sludge types and established design principles to document the worst and most sludgy offenders. Our long term goal is to provide a tool that can be used at local, state, and federal policy levels to determine when a promised service or level of access is being denied by intentional sludge.

ACKNOWLEDGEMENTS

Many thanks to Nony Mushfiqur Rahman for his role in and co-authorship of pre-plenary goofiness! Thanks also to Phillip Thompkins for script-writing assistance and never-ending breadsticks. This conceptualization of sludge was developed with Amanda Lazar, Beth Barnett, Ruipu Hu, and Sumedh Tirodkar, with additional insights through conversations with Samarth Ashwathanarayana and Maithreyi Pejathaya.

PAPERS

1. Sunstein, C. (2022). *Sludge*. MIT Press. https://doi.org/10.7551/mitpress/13859.001.0001
2. Toups Dugas, P. O., Seetharaman, B., Fleming, R. J., Liang, S., Elvitigala, D. S., Rode, J. A., & LaLone, N. (2025). Mobile Maps Continue to Fail Pedestrians: Synthesised Reflective Auto-Aggro-Ethnographies of Walking. *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1–12.
3. Norman, D. (2013). *The design of everyday things*. NY: Basic Books.
4. Polomano, R. C., Galloway, K. T., Kent, M. L., Brandon-Edwards, H., Kwon, K. N., Morales, C., & Buckenmaier, C., 3rd (2016). Psychometric Testing of the Defense and Veterans Pain Rating Scale (DVPRS): A New Pain Scale for Military Population. *Pain medicine (Malden, Mass.)*, *17*(8), 1505–1519.
5. Williams-Pierce, C. (2025). Musings on the brutality of email. In *Proceedings of the Extended Abstracts of the 2025 CHI Conference on Human Factors in Computing Systems*, (Paper 639, p. 1-9). Yokohama, Japan: ACM.

Human-Computer Interaction Lab
43rd Annual Symposium

DetectAIves: Intergenerational Learning with Community Members about AI

Panelists: Samarth Ashwathanarayana[1], Kibron Tesfatsion[1], Alice Mitchell, Sarah Bamba, Ron Nicholson, Galina Reitz[1]
Additional Collaborators: Unlam "Lexie" Leong[2], Navya Gunna[3], Courtney Ray[1], Caro Williams-Pierce[1], Tamara Clegg[1], Galina Reitz[1], Amanda Lazar[1]
[1] College of Information, [2] Department of Computer Science, [3] School of Business, [4] BCAUSEICAN Inc.
sashwath@umd.edu

OVERVIEW

This panel will bring together organizers and participants from the Intergenerational DetectAIves workshop. The panel will provide a space for participants to share their experiences in the workshop, what they learned, and answer questions of symposium attendees.

The workshops brought together seven adults age 50 and over from Prince George's County with the help of our off-campus partner BCAUSEICAN Inc., and five undergraduate interns that were hired through the Computing Catalyst's Sprinternship program during the Winter 2026 Session.

The Sprinterns worked for 40 hours for three weeks, taking on their individual technical projects, helped put together presentations and activities for the workshops and worked with the older adult participants to get feedback on their projects.

The workshop took place over six 3-hour workshop sessions over the 3 weeks. Our curriculum, adapted from the AI4K12 framework [1], combined structured lectures, hands-on activities, and interest-driven projects to understand AI concepts. The curriculum focused on:

1. **Week 1:** Introductions, interests and Perceptions - How computers perceive the world around us.
2. **Week 2:** Representation and reasoning, and how computers learn from data.
3. **Week 3:** How agents are designed for natural interactions with humans, and the broader societal impact of AI.

In conjunction, students developed projects through interacting regularly with adult participants. Projects focused on technical exploration of something that is impactful or necessary to know for community members, creating a tool, demonstration, visualization, or activity that helps make AI concepts accessible. At the end of the workshops, the Sprinterns presented their projects to all the participants, showcasing an introductory video set to help people understand what AI is, a data analysis project on AI generated content in the news, a resume builder using AI, an "AI eye" to help identify AI generated content, and an educational video on how AI is used to scam people.

PANELIST INFORMATION

Panelist 1 - Alice Mitchell, PhD was a community member participant in the workshop. She is an affiliate faculty member for the Student Affairs graduate program at the University of Maryland after retiring from full time administrative work as the Director of the Testing Office at the University. She earned her PhD from the University of Maryland in 1997. Her career has included positions at Arizona State University, Bowling Green State University (OH), Ohio Northern University, Heidelberg College (OH), Allegheny College (PA), and Longwood University (VA). She is the editor of two books and has held multiple elected and appointed leadership positions with the American College Personnel Association, an international association in her field

Panelist 2 - Mr Ron Nicholson is a Co-Founder and currently serves as Executive Director for BCAUSEICAN INC(our collaborators in designing the workshop). BCAUSEICAN is a 501c3 nonprofit located in Greenbelt, MD that provides STEM education to students and adults in underserved communities including adult seniors. BCAUSEICAN conducts computer coding classes for grades K-12 and Artificial Intelligence classes to grades 6-12. BCAUSEICAN also conducts Adult Digital Literacy classes including Artificial Intelligence Literacy in English and Spanish. BCAUSEICAN is a member of the Prince George's County Artificial Intelligence Task Force. Mr.

Nicholson spent most of his 30-year computer engineering career working for Georgetown University in the Main Campus Information Technology department. His Operational Team was responsible for bringing virtualization technology to Georgetown University in 2008. Mr. Nicholson graduated with a B.S. in Computer Science from N.C. State University.

Panelist 3 - Dr. Galina Reitz was the faculty facilitator of the workshop. Dr. Reitz is a strategic leader in higher education with over two decades of experience driving innovation, advancing digital accessibility, and fostering inclusive student engagement. She currently serves as Director of the Master's in Human-Computer Interaction (HCIM) program at the University of Maryland and as Associate Director for Higher Education with the Maryland Initiative for Digital Accessibility (MIDA). Previously, she led the Information Science program at the Universities at Shady Grove, where she achieved record enrollment growth and launched new interdisciplinary minors. Dr. Reitz holds a PhD in Human-Centered Computing, with a research agenda focused on accessible interface design for individuals with disabilities, particularly in voice-based technologies.

Panelist 4 - Sarah Bamba was a Sprintern participant in the workshop, producing a project and actively engaging in learning and discussing artificial intelligence products, their applications and impact alongside the older adult community members.

Panelist 5 - Samarth Ashwathanarayana was a lead graduate student facilitator of the workshop, Sprinterns mentor, and research team member.

Panelist 6 - Kibron Tesfatsion was a lead graduate student facilitator of the workshop, Sprinterns mentor, and research team member.

PROJECT WEBSITE

To learn more about our program and view student projects, please visit https://go.umd.edu/IntergenAI

ACKNOWLEDGEMENTS

Partial funding for this research was provided by UMD's Artificial Intelligence Interdisciplinary Institute at Maryland (AIM) via a Track A Community-Based Seed Proposal (Lazar, PI).

PAPERS

AI4K12. 2020. The Artificial Intelligence (AI) for K-12 Initiative. AAAI and CSTA. Retrieved from https://ai4k12.org/. *HCIL Tech Report HCIL-2007-9*. Under submission. (2007).

LIGHTNING TALKS: ABSTRACTS
[talks listed in order of presentation]

Clusters:

- Design Methods
- Usable Privacy and Security
- Data Visualization
- Sensemaking & Collaboration
- Well-Being and Personal Informatics

Human-Computer Interaction Lab
43rd Annual Symposium

As Content and Layout Co-Evolve: TangibleSite for Scaffolding Blind People's Webpage Design through Multimodal Interaction

Jiasheng Li[1]*, Zining Zhang[1]*, Zeyu Yan[1], Matthew Wong[1], Arnav Mittal[1], Ge Gao[2]†, Huaishu Peng[1]†
[1]Department of Computer Science, University of Maryland, College Park
[2]College of Information, University of Maryland, College Park
jsli@umd.edu
*Co-first authors †Co-last authors

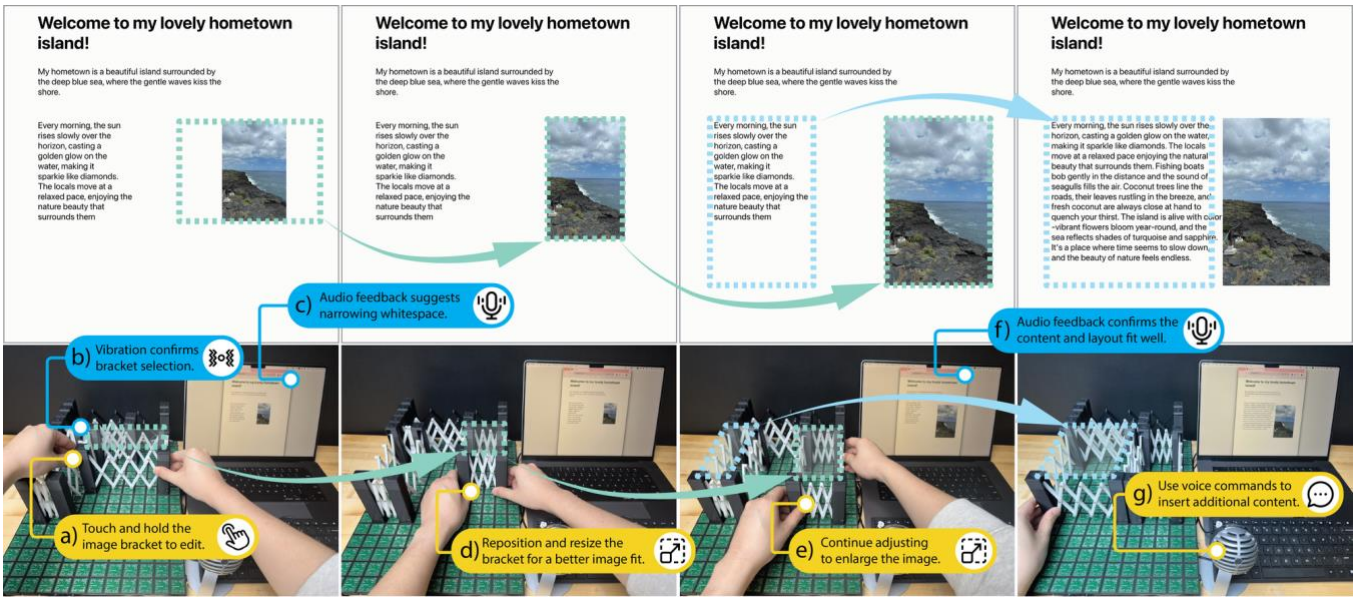


Figure 1 – Iterative webpage design by blind people with TangibleSite via multimodal feedback.

OVERVIEW

Creating webpages requires generating content and arranging layout while iteratively refining both to achieve a coherent design, a process that can be challenging for blind individuals. To understand how blind designers navigate this process, we conducted two rounds of co-design sessions with blind participants, using design probes to elicit their strategies and support needs. Our findings reveal a preference for content and layout to co-evolve, but this process requires external support through cues that situate local elements within the broader page structure as well as multimodal interactions. Building on these insights, we developed TangibleSite, an accessible web design tool that provides real-time multimodal feedback through tangible, auditory, and speech-based interactions. TangibleSite enables blind individuals to create, edit, and reposition webpage elements while integrating content and layout decisions. A formative evaluation with six blind participants demonstrated that TangibleSite enabled independent webpage creation, supported refinement across content and layout, and reduced barriers to achieving visually consistent designs.

BACKGROUND AND CO-DESIGN STUDY

Web page design is a process of exploration and refinement. For sighted designers, the iterative cycle is greatly supported by rapid prototyping tools that provide instant visual feedback (e.g., Figma) or coding environments with built-in previews (e.g., Visual Studio Code). For blind individuals, however, perceiving a webpage, identifying differences, and applying changes is undeniably challenging. Screen readers convey limited spatial information about web elements and provide limited access to visual formatting. Existing

assistive technologies provide limited support for an end-to-end workflow in which a webpage can be created from scratch and iteratively refined through coherent editing of both content and layout.

To ground our work in blind designers' strategies and support needs, we conducted two rounds of co-design sessions with three blind participants using a tangible design probe inspired by TangibleGrid. Three findings shaped our system design:

F1: Participants used flexible workflows in simple designs, but this flexibility diminished as complexity increased. When designs were simple, participants alternated naturally between editing content and rearranging layout elements. As layouts became more complex, this flexibility diminished and participants expected more external cues to stay oriented. This suggests a need for systems that preserve structure across edits so blind designers can verify changes without reconstructing the page from memory.

F2: Hands serve as the primary channel for monitoring, confirming, and planning. Touch played a central role in inspecting elements, confirming system state, and planning next steps. Because similar touch actions often served different functions, touch interactions should clearly distinguish intents such as inspecting, selecting, editing, and moving through touch-coupled feedback.

F3: Audio feedback should be actionable on content-layout fit. Participants expected explicit, actionable audio feedback on content fit and element relationships within the layout. Systems need to report how content fits within available space, highlight spatial relationships, and offer concrete options for resolving mismatches so that blind designers can make informed decisions about their layouts.

THE TANGIBLESITE SYSTEM

Building on these findings, we designed TangibleSite, an accessible web authoring environment that supports webpage creation as content and layout co-evolve. TangibleSite comprises a tangible baseboard and a set of shape-adjustable brackets that represent the webpage canvas and elements, along with a companion browser interface that supports content entry (via voice or keyboard) and provides audio feedback. The browser interface communicates with the tangible baseboard to track user actions, detect design issues, save content, render a live page preview, and maintain an up-to-date representation of the evolving page.

Tangible interactions for spatial grounding. TangibleSite retains a tactile grid and snap-in brackets, but extends their function by linking every physical bracket to a persistent digital record of its type, size, and associated content. When a bracket is moved, resized, or temporarily lifted from the board, the system preserves its content and layout metadata, giving blind designers a reliable spatial anchor they can return to at any time without reconstructing the page from memory.

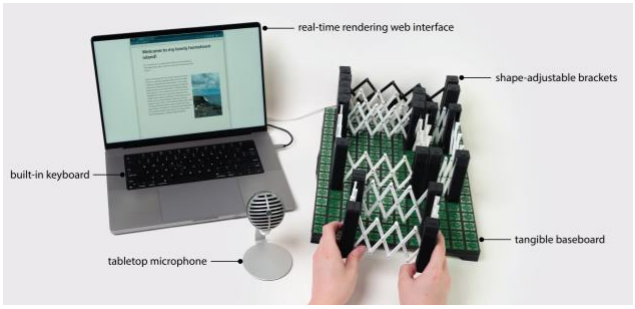


Figure 2 - System overview. TangibleSite supports accessible webpage authoring by coupling tangible layout manipulation with real-time multimodal feedback. Shape-adjustable brackets on a baseboard represent webpage elements and preserve structure across edits. Touch and audio feedback operate at both element and page lebels, supporting inspection, verification, and overall layout understanding.

Touch and haptic cues for clarifying system state. A brief vibration confirms that a bracket has been selected and that subsequent speech commands will be routed to it. A long touch triggers a state readback that announces the bracket's type, size, and any stored content or media. These immediate, touch-aligned cues reduce uncertainty about whether the system registered an action and help users maintain orientation as the layout changes.

Audio guidance for content–layout relationships. TangibleSite provides audio feedback that explains how content and layout interact, paired with concrete adjustment options (e.g., "Image inserted; consider narrowing the bracket to reduce whitespace."). A "Check" command retrieves a structured summary of the entire page—bracket counts and the percentage of whitespace not occupied by brackets—turning audio from passive narration into a decision-making resource.

PROJECT WEBSITE

For more details, see the project page: https://doi.org/10.1145/3772318.3791895

PAPERS

1. Jiasheng Li, Zining Zhang, Zeyu Yan, Matthew Wong, Arnav Mittal, Ge Gao, Huaishu Peng. As Content and Layout Co-Evolve: TangibleSite for Scaffolding Blind People's Webpage Design through Multimodal Interaction. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*, April 13–17, 2026, Barcelona, Spain. ACM (2026).

Human-Computer Interaction Lab
43rd Annual Symposium

"From the Neighborhood to the NCAA: Infrastructuring Sustained Participatory Design in Communities"

Tamara Clegg
College of Information, University of Maryland
tclegg@umd.edu

OVERVIEW

Participatory design or co-design is a common methodology in technology design particularly for democratizing design for diverse users, especially those who are often historically marginalized or excluded from the design process of the technologies that govern their worlds. Often, in studying participatory/co-design within such contexts, researchers focus on the design insights that come out of them and the methods and techniques used in the sessions (what Bodker, Dindler & Iversen (2017) call the front stage). But this leaves out the critical back stages of these experiences - i.e., the ways they are developed within broader infrastructures and the ways they re-shape and/or integrate into existing systems. Limited studies and understanding of back stage aspects of participatory design obscures impactful structures, practices, routines, and cultures that shape/influence the design process and resulting designs.

In this talk, I will take a "back stage" approach to looking at some of the participatory design (co-design) infrastructures I have been apart of or led over the years. We will look at several vignettes/cases of the processes, partnerships, and infrastructures leading up to and within these participatory design infrastructures. I will highlight key back stage characteristics and aspects of different models of participatory design and some of the impacts they can have on design and the design process. We will close by looking at one of my current co-design projects with NCAA athletes at two Division I Power 5 institutions. I will explore how the all-consuming context of big time collegiate athletics influenced the process of setting up the infrastructure, the resulting infrastructure of participatory design within this context and ways in which participatory design thrives and is constrained within two very different institutions (academia and athletics) within a university.

Figure 1: KidsTeam participatory design group.

Figure 2: Co-design with collegiate athletes in the DataGOAT project.

Human-Computer Interaction Lab
43rd Annual Symposium

Designing a Narrative-Based Game for Suicide Prevention Education

Phillip Thompkins
College of Information
pthomp@umd.edu

INTRODUCTION

In recent years, mental health-related topics have become more mainstream through inclusion in news media, books, movies, television, and video games. This more open discussion of mental health has led to success in humanizing specific diagnoses and experiences, such as depression. While many digital mental health products take the form of therapeutic interventions or support treatment itself, there is an untapped potential to leverage narrative and interactivity as tools for psychoeducation and to encourage at-risk individuals to consider or seek treatment. There is a demonstrable gap in narratives depicting suicidal ideation and suicidality as pervasive, systemic phenomena that individuals struggle with chronically, as opposed to just moments of severe crisis when someone is at a breaking point. This disparity is especially clear when comparing what knowledge is available to members of the lay public and mental healthcare providers. Both populations want to help those who are struggling with suicidality, but members of the public are often unsure of how to proceed or worry they may cause harm, and mental healthcare providers are not given consistent training and guidance when it comes to suicide prevention and crisis response.

METHODS

In an effort to bridge gaps between best clinical and supportive practices, public discomfort around discussing suicide, and diverse potential audiences with similar information needs, I chose to develop an interactive, choose-your-own-adventure narrative game in the Twine engine. The game's themes, structure, and tone were informed by interviews with 7 diverse experts with backgrounds encompassing academia, healthcare, advocacy, writing, game design, and members of the general public with lived experience (Mean age = 40 years; 42% female; 42% trans/non-binary; 71% white; 28% non-American origin).

RESULTS

Expert interviews were distilled into 7 priority areas for educational design and 7 considerations for narrative and experience design. "Don't Quit" is the resulting interactive narrative with over 20,000 words of content across 13 scenes, each with their own educational purpose relative to suicide prevention, anti-stigma, and mental health awareness. 5 characters were developed, including the player avatar, to discuss different aspects of suicidality and mental health such as socioeconomic stability, compassion fatigue, substance use, media coverage of suicides, and means safety.

CONCLUSIONS

Interview participants across demographics and backgrounds were enthusiastically willing to contribute to the development of fictional narrative experiences that can be used to educate and support others. Co-designed storytelling based on diverse lived experiences and expertise could be a valuable way to build public-private partnerships and elevate the voices of lived experience perspectives as complementary to best practices in mental healthcare. With further research and refinement, this could lead to the creation of accessible experiences to bridge gaps between providers and individuals seeking treatment.

LIMITATIONS

This abstract is limited to discussing the interviews and game/narrative development, with separate testing of the educational value to occur as future research. Additionally, due to the variety of interview participants, insights skewed towards layperson and lived experiences perspectives over reinforcing current best clinical practices.

PROJECT WEBSITE

To view prior symposium programs, visit the HCIL Events Archive: https://hcil.umd.edu/event-archive/

ACKNOWLEDGEMENTS

Funding for this research was provided by the College of Information's Research Improvement Grant (RIG).

Human-Computer Interaction Lab
43rd Annual Symposium

Evaluating and Co-Designing Speculative XR Use Cases to Surface the Ethical Priorities of Powerful Male Youth Targeted for Racial Oppression

Elana Blinder, Beth Bonsignore, Tammy Clegg
University of Maryland, College of Information
eblinder@umd.edu

INTRODUCTION

As costs have fallen and technology has improved, extended reality (XR) technologies—which encompass virtual reality (VR), augmented reality (AR), and mixed reality (MR)—have become increasingly prevalent in our daily lives. While this recent proliferation of XR tools and experiences has arguably reduced friction across many everyday tasks and offered a variety of novel learning opportunities, concerns regarding the safety, feasibility, accessibility, and impact of these and other emerging technologies have been and continue to be raised. This is especially true in the case of children, whose unique developmental needs and status as a vulnerable population necessitate careful attention to the ethical implications of how these technologies are designed and implemented in research and authentic applied settings.

Recent work has explored the concerns of parents, educators, pediatricians, child psychologists, and—though to a lesser extent—youth themselves regarding children's current and prospective use of XR [1, 2]. However, much of this research in the US has focused on primarily rural White populations, with less known about the specific ethical priorities of Black and other youth targeted for racial oppression. We build on these efforts by directly engaging with and seeking a better understanding of the perspectives of such communities, whose voices have been historically marginalized in the conceptualization and design of new technology and who are among the most likely to experience its adverse effects [3]. In this study, we therefore asked: *What ethical considerations do powerful male youth targeted for racial oppression apply when speculating about the pros and cons of future applications of XR technologies?*

METHODS

This study engaged 17 local, primarily Black, middle school boys, who participated in a 2-week summer co-design program called Tech on the Block, in two activities to surface their ethical perspectives about XR. In the first activity, *line-judging*, participants individually evaluated a series of hypothetical XR scenarios, as depicted in short 1-minute videos, by physically positioning themselves along a giant Likert scale taped to the floor where they then discussed and debated their responses. In the second *speculative design* activity, they worked in small groups with near peer mentors and members of our research team to design their own utopian and dystopian visions of what the future of XR might look like across various domains, such as family life, education, and health.

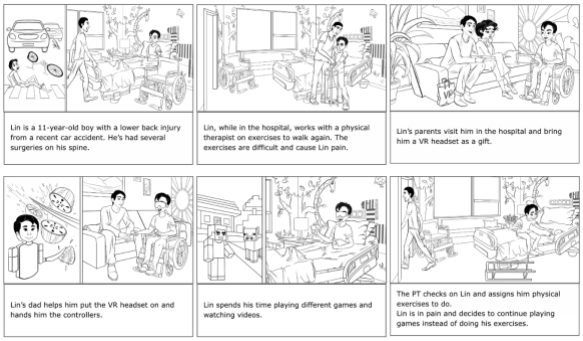


Figure 1 – Stills from a line-judging scenario video [2].

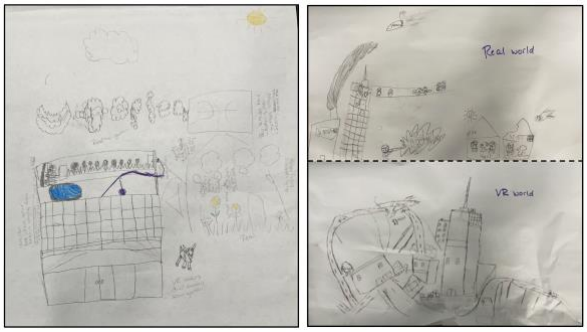


Figure 2 – Examples of boys' utopia and dystopia designs.

FINDINGS

The first line-judging activity surfaced participants' commitment to certain **personal values and ethical priorities**, such as *exercising self-discipline and self-care*, and *maintaining skepticism about the potential misleading nature of immersive virtual content*. During our second speculative design activity, boys' futuristic visions and

accompanying presentations highlighted their **societal values and broader belief systems** which emphasized themes of *overcoming entrapment and exercising community agency* and *promoting equity and access*. Based on these findings, we highlight how our participants' discussions and designs subtly push back against stereotypes commonly ascribed to Black male youth, while celebrating community action and resilience. We complement these insights with a set of design and methodological recommendations to better integrate the perspectives of male youth targeted for racial oppression into XR innovation to ensure these tools are ethical, inclusive, and responsive to the communities they serve.

REFERENCES

1. Juan Pablo Hourcade, Summer Schmuecker, Delaney Norris, and Flannery Hope Currin. 2024. Understanding Adult Stakeholder Perspectives on the Ethics of Extended Reality Technologies with a Focus on Young Children and Children in Rural Areas. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference* (IDC '24), 455–468. https://doi.org/10.1145/3628516.3655811
2. Qiao Jin, Saba Kawas, Stuti Arora, Ye Yuan, and Svetlana Yarosh. 2024. Is Your Family Ready for VR? Ethical Concerns and Considerations in Children's VR Usage. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference (IDC '24)*, 436–454. https://doi.org/10.1145/3628516.3655804
3. Linnet Taylor. 2017. What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society 4*, 2. https://doi.org/10.1177/2053951717736335

ACKNOWLEDGEMENTS

This work is part of a consortium working to develop ethical guidelines for XR research with children. It was funded by an unrestricted gift from Reality Labs Research.

PUBLICATION

Elana B Blinder, Elizabeth Bonsignore, and Tamara Clegg. 2026. "Even though you feel like it's good, it isn't": Surfacing the Ethical Priorities of Powerful Male Youth Targeted for Racial Oppression through Evaluating and Co-Designing XR Utopian and Dystopian Use Cases. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain*. ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3772318.3791604

PROJECT WEBSITE

XR for Youth Ethics Research Consortium: https://homepage.cs.uiowa.edu/~hourcade/projects/xrforyouthethics/

Human-Computer Interaction Lab
43rd Annual Symposium

Trust & Playful Learning

Hayley Park, Nihal Katirci*, Uhjin Sim, & Caro Williams-Pierce
College of Information, *MoNE, Türkiye
hp00@umd.edu, nhl.nihal@gmail.com, uhjim@umd.edu, carowp@umd.edu

OVERVIEW

Trust is a fundamental yet underexplored component of learning. This paper introduces a conceptual framework that positions trust as an anchoring factor in play-based learning within higher education contexts. Drawing on perspectives from education, philosophy, and psychology, and grounded in radical constructivism, our framework delineates cognitive, social, and dispositional dimensions of trust between students and instructors. This conceptual work offers early insights into how trust manifests in classrooms and how it may foster students' willingness to take risks and engage in deeper learning. By positioning trust as both an analytical construct and a pedagogical principle, we contribute to the dialogue on how learning environments can be intentionally designed to cultivate trust and, in turn, support transformative learning experiences.

BACKGROUND

At the 2025 HCIL Symposium, we presented the Formal Education Trust Model (Figure 1) for the first time, along with our initial results from Year 1 [1]. In Year 1, we had data covering each part of the Model except the *Motivational and Behavioral Implications* component, which required longer-term data collection. Now, with Year 2 under our belt, we are excited to share the preliminary longitudinal results!

THEORETICAL FRAMING

Trust plays a key role in human interactions and society as "the fabric of our social universe" [2, p. 95], and is studied differently across disciplines. For example, education research may refer to elements of trust in terms of students or teachers feeling safe in a variety of ways; philosophers discuss trust as a foundational but deeply complex part of the human condition; and psychologists tend to focus on the role of trust in experimental lab-based studies. Our goal is to bring together various components of trust from these different fields and incorporate learning through the theory of radical constructivism [3] in order to ask: *How can we conceptualize trust as a contributing factor to learning?* We specifically focus on trust in play-based higher education classes, both as a form of action research [4] and because we posit that trust and playful learning have a special relationship. Namely, we posit that *increased trust can lead to increased willingness to experiment and take intellectual risks, which is a key part of productive playful learning* [5, 6].

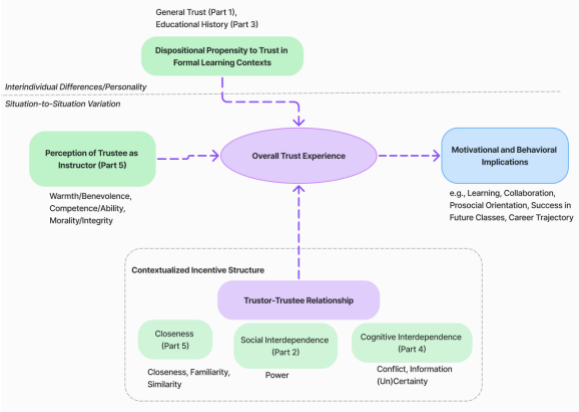


Figure 1 – Formal Education Trust Model

METHODS & ANALYSIS

During Year 1, survey data was collected in two sections of the same core Human-Computer Interaction masters course, both taught by the last author during the Fall semester. She adapted the course to use her Read, Watch, Play pedagogy [6]: every week, students investigated a core question in HCI through readings, presentation videos from leaders in the field, and a 'play' assignment. Each week, different parts of the survey were administered (Figure 1 illustrates which part aligns with which component in the Model), with some parts being administered multiple times.

Year 2 data collection involved recruiting from the same students during the Spring semester, nearly two years later. All parts of the same survey were administered, with additional items designed to capture information about the longer-term motivational and behavioral outcomes. We are still collecting the longitudinal data as of the writing of this abstract, and we will have preliminary results to present at the Symposium. We will also put up our preliminary findings and slide deck here after our presentation! https://go.umd.edu/trustplaylearn

ACKNOWLEDGEMENTS

Endless thanks to the students who filled out surveys week after week, year after year – we could not have done this without you <3 And many thanks to Nony Mushfiqur Rahman and Azure DeVico for their feedback on the Year 2 survey!

PAPERS

1. Williams-Pierce, C., Sim, U., & Park, H. (2025, May). The Shape of Trust in Formal Education. Lightning Talk presented at the Human-Computer Interaction
2. Weiss, A., Michels, C., Burgmer, P., Mussweiler, T., Ockenfels, A., & Hofmann, W. (2021). Trust in everyday life. Journal of Personality and Social Psychology, 121(1), 95–114. https://doi.org/10.1037/
3. von Glasersfeld, E. (1995). *Radical constructivism: A way of knowing and learning.* London: The Falmer Press.
4. Herr, K. (1995). Action Research as Empowering Practice. *Journal of Progressive Human Services*, *6*(2), 45–58. https://doi.org/10.1300/J059v06n02_04
5. Williams-Pierce, C. (2019). Designing for mathematical play: Failure and feedback. Information and Learning Sciences, 120(9–10), 589–610. https://doi.org/10.1108/ILS-03-2019-0027
6. Williams-Pierce, C. (October 20, 2021). *Loud Gleeful Learning: Supporting and Provoking Adult Play in Formal Contexts*. Published in Sparks of Innovation: Stories from the HCIL. https://medium.com/hcil-at-umd/loud-gleeful-learning-86f603f27497

Human-Computer Interaction Lab
43nd Annual Symposium

The Illusion of Email Verification

Ayushi Mishra[1], Miuyin Yong Wong[1], Joseph A. Calandrino[2], Michelle L. Mazurek[1]
[1] Department of Computer Science, University of Maryland, College Park
[2] School of Computer Science, Carnegie Mellon University
Contact: amishr13@umd.edu

OVERVIEW

Today, most Internet users manage multiple online accounts across various services, making email addresses a central part of digital identity. Yet, a subtle and familiar problem persists: many users have received emails intended for a stranger, or have themselves mistyped an address during registration. These incidents are easy to dismiss as minor inconveniences, but they point to a deeper gap. When an email address is misused or misassigned during account creation, the consequences can affect both the account creator and the unintended recipient in ways that are not immediately apparent.

This work builds on that everyday experience to examine how websites verify that an email address truly belongs to the person creating an account. We explore how minor breakdowns in this process can lead to real-world risks. By analyzing current practices and their consequences, we aim to surface the broader implications of email verification failures and explore how verification practices can be improved to reduce these risks while preserving usability.

METHODS

Our methodology consists of three stages: domain selection, verification workflow classification, and privacy risk evaluation. We will sample websites from the Tranco List [1] to ensure both representativeness and diversity, combining highly ranked domains with randomly selected lower-ranked domains. We will retain only domains that are accessible, offer English-language interfaces, include an account registration workflow, avoid shared login systems, and do not require sensitive personal or financial information during sign-up. For each selected domain, we will classify email verification workflows into three categories: **verification required**, where accounts remain unusable until the email is verified; **no verification required**, where accounts are fully accessible immediately after registration; and **optional verification**, where accounts can be used prior to verification but with some restricted functionality.

We will then evaluate privacy and security risks under a two-role model consisting of an account creator (AC) and an email owner (EO). For domains requiring verification, we will test whether verification can be bypassed through direct login attempts, navigation manipulation, or URL modification, and if unsuccessful, whether the EO can later take over the account through delayed verification. For optional or non-required verification workflows, we will analyze what actions are possible using unverified accounts, including credential changes, enabling security settings, adding payment information, purchases, or user interactions.

We will also assess whether unverified accounts reserve email addresses and block legitimate registration attempts over time. Finally, we will examine verification and welcome emails for information leakage, such as usernames, IP addresses, device or location details, transaction data, and link expiration policies. We map the observed behaviors and vulnerabilities to usercentric privacy harms using established frameworks such as Daniel J. Solove's taxonomy of privacy harms [2].

We categorize harms across dimensions such as economic, reputational, psychological, and autonomy-related impacts, and examine how system design choices in verification workflows contribute to these risks. This analysis enables us to systematically connect technical weaknesses to real-world consequences experienced by users.

Given the potential impact of identified vulnerabilities, we will follow a responsible disclosure process for affected platforms. For systems where verification weaknesses lead to significant security or privacy risks, we will notify the corresponding service providers with detailed reports describing the issue and reproduction steps.

This work aims to uncover systemic weaknesses in widely deployed email verification mechanisms and demonstrate how these flaws translate into real user harms.

REFERENCES

1. "Tranco list," 2026. [Online]. Available: https://tranco-list.eu/list/ 2NZW9/1000000
2. D. J. Solove, "A taxonomy of privacy," U. Pa. l. Rev., vol. 154, p. 477, 2005.

Human-Computer Interaction Lab
43rd Annual Symposium

The Impact of Emerging AI Practices on the Cybersecurity Workforce

Miuyin M. Yong Wong, Alan F. Luo, Yunze Zhao, Shubham Bhatnagar, Michelle L. Mazurek
Department of Computer Science, University of Maryland
miuyin@umd.edu

OVERVIEW

Large language models (LLMs) are increasingly being applied in many disciplines, and cybersecurity is no exception. Extensive research has evaluated LLM effectiveness on cybersecurity tasks, such as reverse engineering [1], malware detection [2], and capture-the-flag exercises [3]. Though fully automated solutions are still limited, LLMs may still offer benefits to defenders' productivity and efficacy, which could prove especially important as cyberattacks continue to proliferate [4] while cybersecurity professionals remain in short supply [5].

However, there is limited understanding of how cybersecurity practitioners are adopting LLMs in real world settings and how this adoption may be shaping professional practice. Early work examines query data from one Security Operations Center (SOC) to understand how LLMs are used in daily work, finding that SOC analysts primarily used LLMs for sensemaking, context building, and low-level task support, while retaining human decision authority [6]. Other research examines practitioner perspectives on LLM use within software engineering workflows [7, 8]. Klemmer et al. [9] report mistrust in AI outputs and a lack of standardized verification practices, whereas Giray et al. [7] highlight perceived benefits such as reduced cycle time and improved quality, despite limited objective productivity measures. Although these studies offer valuable insights, it remains unclear whether their findings extend to the broader cybersecurity landscape, such as incidence response, threat intelligence, compliance, and digital forensics. Examining LLM adoption across a wider set of roles is essential to identifying potential systemic shifts in professional norms, collaboration patterns, and knowledge transfer that may not be visible from a single vantage point.

Moreover, many groups of cybersecurity practitioners rely on norms of openness, collaboration, and mentorship [10], as demonstrated through competitions such as capture-the-flag tournaments [11], bug report publishing [12], and participation in online communities [13]. This makes it critical to understand how LLM use affects social dynamics in cybersecurity, such as how practitioners communicate about LLM use with their peers, how they perceive others' use of such systems, and the informal guidance and advice emerging around this technology.

To fill this gap, we conducted semi-structured interviews with 28 cybersecurity practitioners across 26 distinct organizations, with differing amounts of cybersecurity experience and different job roles. We aim to answer the following research questions:

RQ1. What policies and guidance for LLM use are organizations providing to their cybersecurity practitioners?
RQ2. What challenges and risks do cybersecurity practitioners encounter when using LLMs, and what strategies do they employ to address or mitigate them?
RQ3. How do peers' shared perceptions, advice, and informal guidance regarding LLMs shape cybersecurity practitioners' adoption and use of LLMs?

Our findings indicate that many organizations lack formal policies on LLM use, and informal guidance by peers is often inconsistent. Yet, participants still reported that their organizations encouraged the use of these models due to expected increases in productivity. Additionally, our participants reveal cybersecurity-specific obstacles associated with LLMs, such as guardrails restricting "offensive" tasks like penetration testing and red-teaming or models being unable to keep up with the quickly-evolving cybersecurity landscape. To our surprise, few participants expressed concern regarding potential adversarial threats, such as data poisoning or prompt injection. This suggests security practitioners trained to anticipate system failures may be just as susceptible to emerging AI vulnerabilities as other users.

Our most significant finding is a psychological barrier that cybersecurity practitioners are experiencing: Participants voiced a fear of being judged for using AI tools, with potential harm to their professional reputations. While this anxiety rarely stops practitioners from using LLMs, we begin to observe a shift in social and interpersonal dynamics within the community. These tensions highlight a pressing need for the security community to engage in a broader

dialogue about the role of generative AI, to ensure that we preserve the collaborative practices that have long supported the field.

REFERENCES

1. Artur Boronat and Jawad Mustafa. MDRE-LLM: A Tool for Analyzing and Applying LLMs in Software Reverse Engineering. In 2025 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER), pages 850–854, Montreal, QC, Canada, March 2025. IEEE

2. Ruirui Feng, Hui Chen, Shuo Wang, Md Monjurul Karim, and Qingshan Jiang. LLM-MalDetect: A Large Language Model-Based Method for Android Malware Detection. IEEE Access, 13:81347–81364, 2025.

3. Yuwen Zou, Yang Hong, Jingyi Xu, Lekun Liu, and Wenjun Fan. Leveraging Large Language Models for Challenge Solving in Capture-the-Flag. In 2024 IEEE 23rd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pages 1541–1550, December 2024. ISSN: 2324-9013.

4. Olivér Gulyás and Gábor Kiss. Impact of cyber-attacks on the financial institutions. Procedia Computer Science, 219:84–90, January 2023.

5. William Crumpler and James A. Lewis. The Cybersecurity Workforce Gap. Technical report, Center for Strategic and International Studies (CSIS), 2019.

6. Ronal Singh, Shahroz Tariq, Fatemeh Jalalvand, Mohan Baruwal Chhetri, Surya Nepal, Cecile Paris, and Martin Lochner. LLMs in the SOC: An Empirical Study of Human-AI Collaboration in Security Operations Centres. In 2026 IEEE Symposium on Security and Privacy (SP), pages 2083–2102, May 2026.

7. Gorkem Giray, Onur Demir.rs, Marcos Kalinowski, and Daniel Mendez. An empirical study of generative ai adoption in software engineering. arXiv preprint arXiv:2512.23327, 2025.

8. Jan H Klemmer, Stefan Albert Horstmann, Nikhil Patnaik, Cordelia Ludden, Cordell Burton Jr, Carson Powers, Fabio Massacci, Akond Rahman, Daniel Votipka, Heather Richter Lipford, et al. Using ai assistants in software development: A qualitative study on security practices and concerns. In Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security, pages 2726–2740, 2024.

9. Jan H Klemmer, Stefan Albert Horstmann, Nikhil Patnaik, Cordelia Ludden, Cordell Burton Jr, Carson Powers, Fabio Massacci, Akond Rahman, Daniel Votipka, Heather Richter Lipford, et al. Using ai assistants in software development: A qualitative study on security practices and concerns. In Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security, pages 2726–2740, 2024.

10. Nathan Fisk, Nicholas M. Kelly, and Lori Liebrock. Cybersecurity communities of practice: Strategies for creating gateways to participation. Computers & Security, 132:103188, September 2023.

11. Tyler Balon and Ibrahim (Abe) Baggili. Cybercompetitions: A survey of competitions, tools, and systems to support cybersecurity education. Education and Information Technologies, 28(9):11759–11791, September 2023.

12. Daniel Votipka, Rock Stevens, Elissa Redmiles, Jeremy Hu, and Michelle Mazurek. Hackers vs. Testers: A Comparison of Software Vulnerability Discovery Processes. In 2018 IEEE Symposium on Security and Privacy (SP), pages 374–391, May 2018. ISSN: 2375-1207.

13. Daniel Votipka, Mary Nicole Punzalan, Seth M. Rabin, Yla Tausczik, and Michelle L. Mazurek. An Investigation of Online Reverse Engineering Community Discussions in the Context of Ghidra. In 2021 IEEE European Symposium on Security and Privacy (EuroS&P), pages 1–20, September 2021.

Human-Computer Interaction Lab
43rd Annual Symposium

Sports Clip Retrieval via Text, Audio, and Touch

Archit Ramanasai Kambhametttu, Zhicheng Liu
University of Maryland, College Park; Department of Computer Science
architk@umd.edu

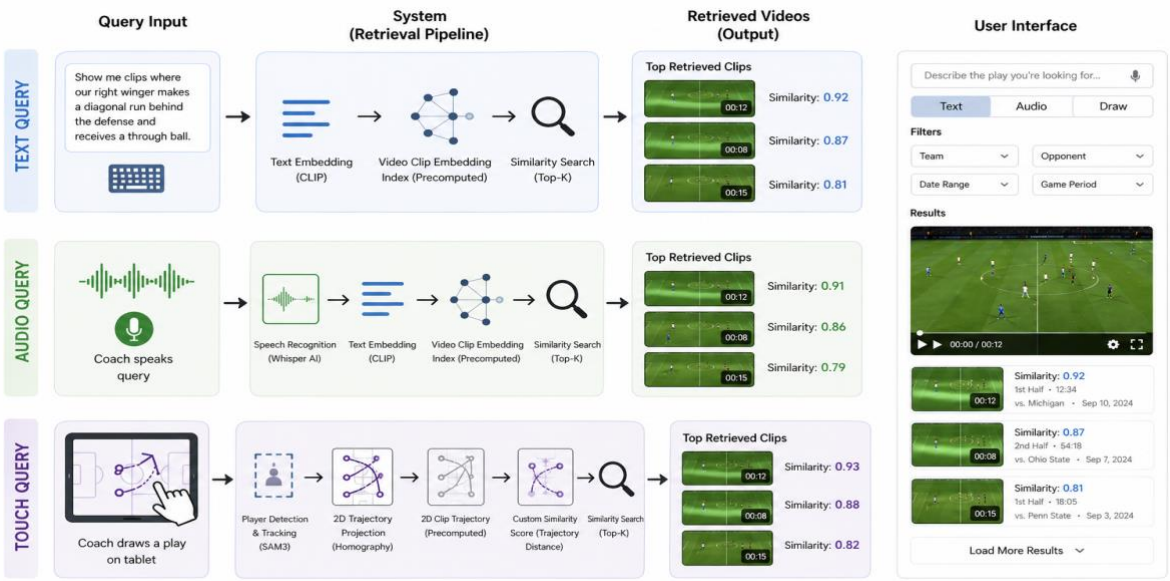


Fig 1: Proposed pipeline, detailing the text, audio, and touch modalities for retrieving soccer video clips.

OVERVIEW

Film analysis is a central component of sports performance evaluation; however, the increasing volume of recorded video footage creates a significant operational bottleneck for coaches. Existing video retrieval workflows rely heavily on manual tagging or predefined action labels (e.g., passes, shots, and fouls), which are often too broad to capture nuanced tactical patterns. As a result, retrieving relevant clips, particularly under time constraints such as during halftime, remains inefficient and cognitively demanding. In preliminary discussions with the University of Maryland men's soccer team, we observed that coaches struggle to locate relevant footage in real time and often spend more effort searching for clips than analyzing them. This limitation reduces the effectiveness of video-assisted decision making during time-sensitive phases of the game.

RELATED WORK

Research in soccer analytics has advanced considerably over the past few years. Since 2021, the SoccerNet challenge, organized by the CVPR Workshop on Computer Vision in Sports (CVSports), has established foundational benchmark challenges for video, field, player, and game understanding. These tasks include action spotting, video captioning, camera calibration, player re-identification, and game state reconstruction, among others.

To facilitate model development, the organizers have released extensive, publicly available soccer video datasets with corresponding labels. Vision-Language Models (VLMs) are uniquely suited for our proposed clip retrieval task because they facilitate the direct alignment of high-dimensional video features with semantic natural language embeddings. This cross-modal capability enables the retrieval of complex soccer scene descriptions that traditional action recognition models cannot identify due to their rigid and finite taxonomies. Consequently, VLMs provide the flexibility necessary to interpret and execute a coach's highly specific queries within a single unified framework. VLMs, however, are trained to be generalist models rather than domain experts, and thus off-the-shelf models are unable to properly solve the retrieval task. Domain adaptation has been an increasingly active research

area, and prior work has shown that adaptation techniques applied to VLMs can improve soccer reasoning by a considerable margin. While model performance has improved substantially, existing research objectives are often rigid, prioritizing standardized metrics over practical, real-world use cases. Our research aims to shift the focus from purely model-centric performance to human-centered design, creating an intuitive retrieval system that translates complex coaching intent into high-value video evidence.

METHODS

We will conduct interviews and observation studies with the University of Maryland soccer coaches to understand how they describe retrieval criteria and leverage various modalities such as video footage, natural language, audio, and touch during their halftime film analysis. This formative study will guide our framework for the clip retrieval process.

Querying via Text

We aim to develop a two-step Vision-Language Model (VLM) based semantic retrieval framework. This framework will first project game clips to a high dimensional CLIP (Contrastive Language-Image Pre-training) embedding space, then retrieve game clips most similar to the natural language prompt entered by the user. This VLM-style architecture facilitates the direct alignment of visual features and text queries within a shared latent space, enabling the system to calculate a similarity score (e.g., cosine similarity) to identify the most relevant clips. We plan on evaluating the performance of both off-the-shelf VLMs and domain-adapted VLMs as the backbone for the task. To achieve this domain adaptation effectively, we will implement Low-Rank Adaptation (LoRA) on the VLM backbone to enable the model to learn specialized soccer language and its relation to video clips.

Querying via Audio

We aim to expand on our Vision-Language semantic retrieval framework to enable coaches to query videos via audio as well. To do this, we will record the coaches describing a certain clip they want to retrieve, and leverage Whisper to translate it to natural language. Then we will use our previously established framework to calculate similarity scores between the clip and the spoken query, and retrieve the most similar clips.

Querying via Touch

While natural language text and audio are very expressive, coaches sometimes have difficulty in describing a specific type of motion or play in explicit language. When developing these plays and teaching them to players, coaches often draw out the play on a whiteboard or tablet for the players to visualize. These drawings depict the motion trajectories of each of the players on the field, indicating the locations each player should move to throughout the progression of the play. Building on this idea, we aim to develop a method for coaches to draw out plays on a tablet and retrieve clips with similar player motion trajectories. To do this, we aim to employ the SAM3 model to segment all the soccer players on the field, then track them throughout the clip. Then, we will perform a field homography to convert the clip to a 2D motion trajectory image. We will save this 2D representation for all the clips. Then, at retrieval time, the coaches will draw a specific play motion on an image of a soccer field with their tablet. Using a custom similarity score, we will find the most similar image to the query.

EVALUATION

We plan to have a multi-stage evaluation framework. First, we will use preexisting benchmarks to rigorously evaluate whether our models accurately understand soccer concepts. We will also construct a held-out test set of clips with detailed descriptions, and evaluate whether the models can both output the description of the video and retrieve the video given the description. Furthermore, we will ensure our methods work in practice by integrating them into the University of Maryland Soccer Team halftime analysis. We will evaluate whether our methods are accurate, reduce cognitive load, and are easy to use so they are practical and effective for the team.

ACKNOWLEDGEMENTS

We would like to thank the University of Maryland Athletics Department for collaborating with us on this project.

Human-Computer Interaction Lab
43rd Annual Symposium

Bringing Data to Life: Designing Data Characters for the Evolving Self

Diego Abarcar Calugay[1], Isabella Amador[2,3], Keke Wu[2]
[1]Department of Computer Science, [2]College of Information, [3]College of Behavioral and Social Sciences
diegoac@terpmail.umd.edu, iamador1@terpmail.umd.edu

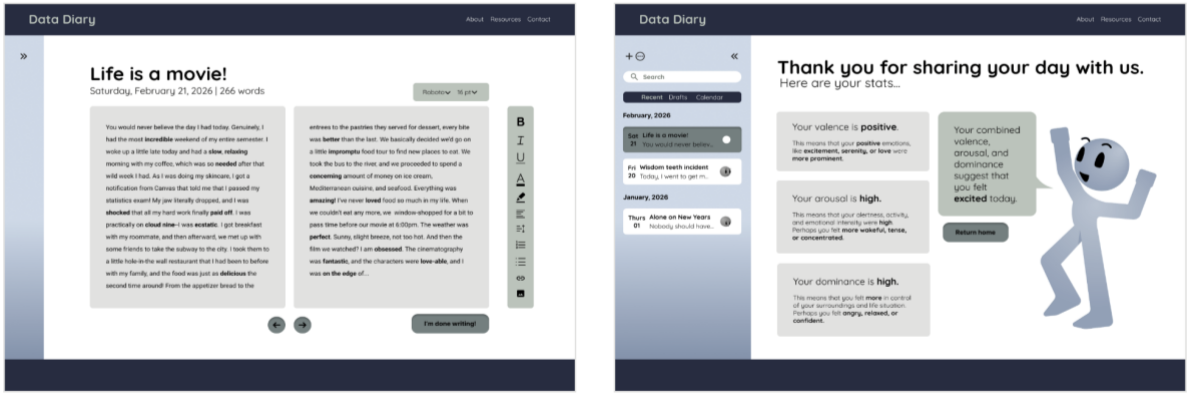


Figure 1 – We built a proof-of-concept system that embeds a data character into a digital journaling interface to provide at-a-glance understanding of one's daily affective state. A data character is an anthropomorphic visualization that encodes statistical patterns, such as emotion, into expressive visual features to support understanding and reflection of data [1]. This system uses natural language processing to extract affective cues from journaling passages that map to the data character's physical attributes, including facial expression and pose.

OVERVIEW

Journaling helps people process feelings, supports goal setting, and promotes overall well-being. It also plays an important role in mental health: qualitative self-tracking can document rich, evolving pictures of health, helping individuals recall experiences and enabling more engaged patient-provider collaboration in psychological settings [3]. However, journaling generates large amounts of unstructured, text-heavy data that are difficult to revisit, synthesize, and communicate. Identifying patterns or retrieving relevant experiences can be cognitively demanding and time-consuming, limiting its value for both personal and clinical use. Data visualization offers a promising approach by transforming journal entries into structured, interpretable representations, making long-term patterns easier to see, reflect on, and share.

Conventional approaches to visualizing textual data, however, face several challenges. Many operate at the word or sentence level (e.g., word clouds or phrase nets) and rely on frequency, overlooking context and emotional nuance [2]. As a result, they can misrepresent the richness of free-form responses and lived experience. More structured methods, like thematic coding, organize text into discrete categories (e.g., "happy" or "sad"); while they introduce clarity, they can oversimplify nuanced experiences, reducing the depth and subtlety of human emotion [2].

In response, we propose *data characters*—anthropomorphic visual entities that encode and convey themes and messages in journaling data—as a novel approach to textual visualization [1]. By grounding abstract patterns in familiar, human-like forms, data characters make text more interpretable and intuitive. They support understanding of personal patterns and changes over time by evolving to represent continuity and transformation, integrating complex insights into a cohesive whole, and fostering sustained engagement through expressive, relatable interfaces that invite reflection [1].

PROOF-OF-CONCEPT SYSTEM

To operationalize this idea, we developed a proof-of-concept system that encodes emotional nuance into data characters. We focus on emotion as an initial step, as it is both central to journaling and difficult to represent with conventional

techniques. This concept allows us to explore how data characters can capture affective complexity through dynamic, embodied change. Future work will extend this approach to other aspects of journaling data (e.g., memory recall, emotional regulation, and social context).

Our system maps users' **pleasure** (positive-negative affect), **arousal** (level of alertness), and **dominance** (sense of control) to the facial features and body language of data characters, based on the Pleasure-Arousal-Dominance (PAD) model [4]. The mapping of PAD dimensions to character attributes follows a heuristic design approach based on established cues in non-verbal communication. Pleasure is encoded with mouth curvature to leverage high-signal markers of valence. Arousal is reflected through mouth opening and head height, simulating states of high and low activation. Dominance is mapped to eyebrow positioning and eye shape, serving as a secondary modifier to distinguish between emotional states, such as the transition from fear to anger. These mappings are normalized on a scale of -50 to 50, with neutral states at 0, and intermediate values calculated via linear interpolation to ensure a continuous representation along the three dimensions. Using Gemini Flash 2.5 as our natural language processing model, we extract scale values for pleasure, dominance, and arousal from the journal entry. A data character is then generated alongside a textual summary to support interpretation. Users can also revisit entries through a calendar view, where each day is represented by a character snapshot. This longitudinal view enables quick recognition of patterns and changes over time, supporting reflection and retrieval in text-based journaling data.

FUTURE DIRECTIONS

We plan to evaluate data characters through empirical studies. One direction is a controlled experiment comparing them with conventional qualitative visualizations (e.g., word clouds) to assess differences in preference, comprehension, and emotional response; through this design, we may investigate both cognitive and affective impact. We also envision a longitudinal diary study, in which participants use a journaling interface with data characters over time, providing more ecologically valid insights into everyday use, journaling practices, and evolving perceptions. More broadly, future work may explore how data characters can extend beyond representing emotional patterns to capture other aspects of journaling data, such as memory recall, events, and social context, examining multiple dimensions of self-growth and well-being. Ultimately, data characters offer a promising approach for making journaling data more accessible, enabling researchers and practitioners to identify meaningful patterns in textual data.

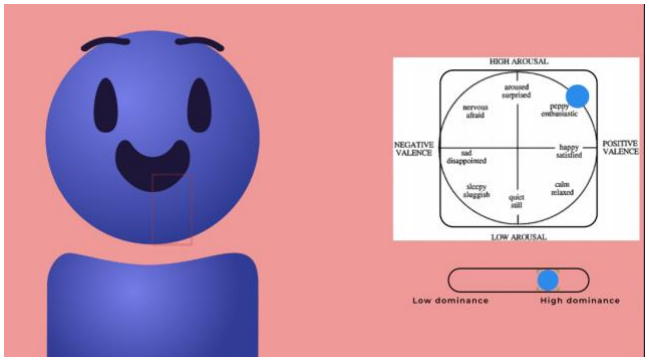


Figure 2 – The system uses the PAD model's dimensions of pleasure, arousal, and dominance to represent emotions through the character's dynamic facial features [4].

REFERENCES

1. Keshav Dasu, Yun-Hsin Kuo, and Kwan-Liu Ma. 2024. Character-Oriented Design for Visual Data Storytelling. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (January 2024), 98-108. https://doi.org/10.1109/TVCG.2023.3326578
2. Stuart Henderson and Eden H. Segal. 2013. Visualizing Qualitative Data in Evaluation Research. *New Directions for Evaluation* 2013, 139 (September 2013), 53-71. https://doi.org/10.1002/ev.20067
3. Christina Kelley, Bongshin Lee, and Lauren Wilcox. 2017. Self-tracking for Mental Wellness: Understanding Expert Perspectives and Student Experiences. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '17),* May 6 - 11, 2017, Denver, Colorado. SIGCHI. New York, NY, 629-641. https://doi.org/10.1145/3025453.3025750
4. Albert Mehrabian. 1996. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament. *Current Psychology* 14, (December 1996), 261 - 292. https://doi.org/10.1007/BF02686918

Human-Computer Interaction Lab
43rd Annual Symposium

ViveChart: Authoring Interactive Visualizations through Visual Programming and Natural Language Descriptions

Yutian Zhang, Aadesh Kabra, Zhicheng Liu
Department of Computer Science, University of Maryland
yzhan127@umd.edu

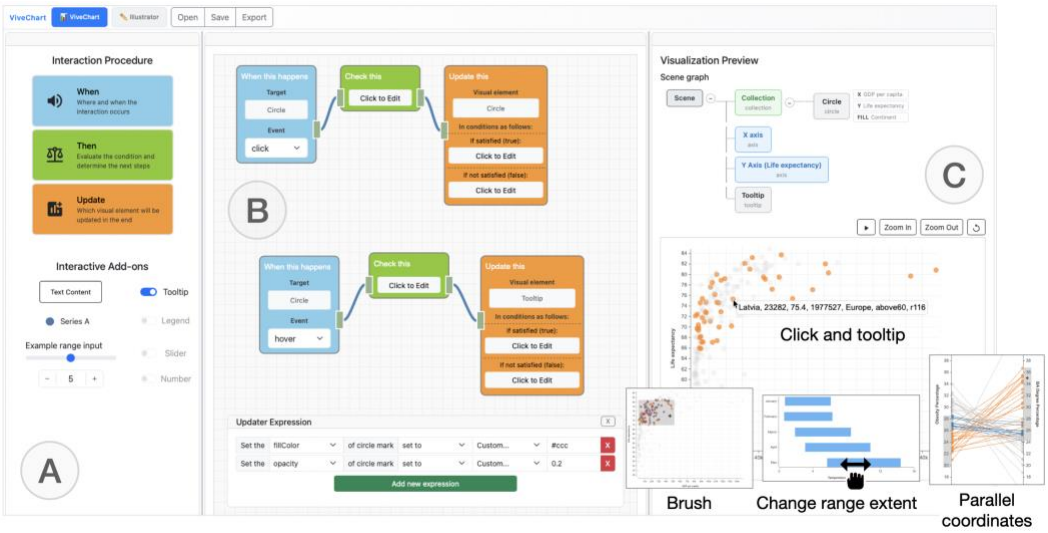


Figure 1 – The interface of ViveChart: (A) *Control Panel* provides essential interaction components. (B) *Canvas* allows users to author interactive data visualizations via visual programming. (C) *Visualization Preview* renders the scene, enabling users to review interaction effects in real time.

OVERVIEW

Interactive data visualization is pivotal for effective data analysis and knowledge extraction. However, creating interactive charts remains challenging due to technical barriers and time investment. To fill this gap, we present ViveChart, a novel interactive interface that enables users to author interactive visualizations without coding expertise. By leveraging scene-centric visualization library, ViveChart allows users to define interactive components and specify underlying logic through a combination of visual programming and natural language expressions. Our approach provides an effective and user-friendly workflow for non-programmers to create and refine diverse interactive data visualizations.

METHODS

Scene-centric Visualization Library: Built upon a scene-centric design principle, Mascot.js [1] provides a semantic representation of visualization components, enabling structured reasoning about marks, encodings, and data bindings. By leveraging this mechanism, interactions are decomposed into four fundamental components: *Trigger*, *Responder*, *Evaluator* and, *Updater*, which provides a robust foundation for the authoring of interactive charts.

Natural Language Expressions: We curated a corpus of natural language descriptions for interactive visualizations. Following a "When-If-Update" structure, we formulate component-specific expressions that enables users to translate their intents into actionable interaction logic.

Visual Programming Interface: We design a glyph-based visual programming interface to lower the technical barrier for non-programmers. The interface represents underlying interaction logic and constraints as composable, editable blocks, allowing users to construct interactions through intuitive drag-and-drop manipulations.

PAPERS

1. Zhicheng Liu, Chen Chen, John Hooker. Manipulable Semantic Components: a Computational Representation of Data Visualization Scenes. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 31(1), 732-742. (2024).

Human-Computer Interaction Lab
43rd Annual Symposium

Analyzing the Functions of Visual Components in Information Graphics

Xiaoyu Liu, Vishnu Sreekanth, Tanya Singh, Riya Thapar, Zoey Tsai, Fumeng Yang, Zhicheng Liu
Department of Computer Science, University of Maryland
Contact: dasqxyl@umd.edu

OVERVIEW

Data and information are two fundamental concepts for visualization, and extensive research has investigated the relationship between them [1]. Traditionally, studies categorize data types in visualization based on their dimensionality, unit and hierarchical structure [2, 3], which draws upon the implicit assumption that all input data are inherently structured. However, the proliferation of information graphics (infographics) has seen visualization of real-world data that are more heterogeneous and granularized. Many existing taxonomies for infographics focus on their design aspects [4] or specific narrative strategies [5], but neglect the more nuanced data or information types that they can illustrate in general. Our study aims to bridge this gap in infographic understanding by constructing a visual function framework that answers the following questions:

(a) What types of information are shown by infographics, and how are they mapped to visual elements?
(b) What narrative roles do visual elements in infographics play in relation to their information types?

METHOD

Corpus & Sampling Strategy. To gather a comprehensive corpus of infographics for subsequent analysis, we identified three categories of image sources: paper, book and online. Paper source includes 357 images extracted from 111 visualization / HCI papers published between 1971 and 2025. Book source includes 557 images extracted from 5 infographic-themed books. Online source incorporates major infographic datasets (e.g., InfographicVQA [6] and ChartGalaxy [7]) into a total of ~60K images after availability and de-duplication checks. These images were predominantly collected via search engines, websites and online media outlets. Due to the significantly unbalanced distribution of images across the three sources, we adopted a stratified batch sampling approach to ensure representation of all sources and internal subsets.

Visual Function Taxonomy. We developed an initial taxonomy of visual functions based on iterative discussion and manual annotation of visual elements in 33 sampled images. Inspired by the verb-noun formulation used by Stokes *et al*. in their text function analysis [8], we propose a 3-tuple syntax to express the function of visual elements used in infographics: [*subject*: visual element] [*verb*: narrative role] [*object*: information type]. The subject specifies the visual element's type of graphical primitive (text, mark, or group). The verb identifies the relationship between the visual element and its conveyed information (e.g., encodes, explains, compares). The object describes the type of the conveyed information, classified as either a primitive (e.g., number, date/time, concept) or a composite (e.g., multivariate item, causality, process). Altogether, the function of a visual element may be captured as "*shape encodes number*" or "*text explains concept*". A tentative listing of visual elements, narrative roles and information types is shown in Figure 1.

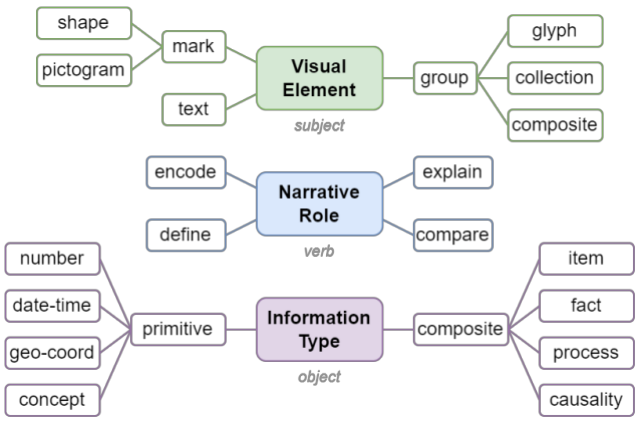


Figure 1 – Visual elements, narrative roles and information types currently adopted by the visual function taxonomy.

Visual Element Segmentation & Annotation. We will employ vision models specializing in visual element detection tasks [9] to perform preliminary image segmentation, followed by human verification and correction. To facilitate the annotation process, we

developed a semi-automated labeling tool that enables intuitive inspection of visual element context and group hierarchy. We will annotate all segmented elements on the basis of our visual function taxonomy and other relevant semantic fields, e.g., number of instances, semiotic type and level of visual detail. We will refine our visual function taxonomy through the annotation process and identify patterns in the correlation between visual elements, narrative roles, information types and related semantic fields.

APPLICATION

To evaluate our visual function taxonomy and to demonstrate its practical utility, we discuss three application scenarios.

Infographic taxonomy building. Our taxonomy may be applied to content-based clustering of infographics by identifying the most prevalent functions used among visual elements. The results may be further augmented by design-oriented classification approaches to produce more fine-grained infographic taxonomies.

Prompt assistance for infographic generation. Our taxonomy may be used as a design reference when prompting large language models to generate new infographic designs. Using source data and desired narrative as input, authors can further actualize their design intents by expressing them using commonly observed visual functions. Compared with direct prompting of data and narratives, we expect the added guidance of visual functions to yield more controllable and verifiable designs.

Information extraction from infographics. Conversely, our taxonomy may also facilitate infographic reverse-engineering by reducing the task to a visual element level. We believe that the effective reconstruction of infographic's underlying data and narratives will lead to a better understanding of their composition and more contextualized reuse of existing designs.

REFERENCES

1. M. Chen, *et al*. Data, Information, and Knowledge in Visualization. *IEEE Computer Graphics and Applications*, 29(1), 12-19. (2009).
2. D. A. Keim. Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1), 1-8. (2002).
3. B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. *The Craft of Information Visualization*, 364-371. (2003).
4. H. C. Purchase. A classification of infographics. *International Conference on Theory and Application of Diagrams*, 210-218. (2018).
5. Y. Guo, *et al*. Unpacking Visual Metaphors in Infographics: A Design Space. *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, 447, 1-24. (2026).
6. M. Mathew, *et al*. InfographicVQA. *arXiv*. (2021). DOI: 10.48550/arXiv.2104.12756
7. Z. Li, *et al*. ChartGalaxy: A Dataset for Infographic Chart Understanding and Generation. *arXiv*. (2025). DOI: 10.48550/arXiv.2505.18668
8. C. Stokes, *et al*. An Analysis of Text Functions in Information Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 32(1), 769-779. (2026).
9. P. Yan, *et al*. Context-Aware Chart Element Detection. *arXiv*. (2023). DOI: 10.48550/arXiv.2305.04151

Human-Computer Interaction Lab
43rd Annual Symposium

Evaluating Accuracy and Safety of Election Chatbot

Megan Lu, Vashawn Robinson, Ryan Yu, Lina Romero-Fabien, Johnny Zhu, Do Won Kim, Giovanni Luca Ciampaglia, Cody Buntain
College of Information, University of Maryland
mlu12349@terpmail.umd.edu

OVERVIEW

Trust in democratic institutions and confidence in elections are declining. Although reliable election information is available on state and local government websites, it is often difficult for voters to locate and interpret. In contrast, misleading or false claims frequently spread more rapidly and are easier to access than official sources.

Can an AI chatbot grounded in official government resources improve how citizens access, understand, and trust election information? We developed Voter Information from Official Local Election Trusted Sources (VIOLETS), a RAG chatbot that grounds responses in a curated knowledge base of official election resources from the Maryland State Board of Elections and the Montgomery County Board of Elections. For more information on the chatbot architecture, please visit here.

Here, we report results from a pre-deployment test of VIOLETS. Specifically, we ask: **(RQ1)** Does VIOLETS improve factual accuracy and reduce hallucinations compared to a baseline model? **(RQ2)** And does VIOLETS improve the safe handling of adversarial, misleading, and sensitive queries compared to a baseline model?

We compare responses from VIOLETS against those from the baseline model (GPT-4o-mini). First, a participant LLM (RQ1) generates queries that a voter would likely ask about election procedures, while an attacker LLM (RQ2) generates adversarial queries. Then, both VIOLETS and the baseline model generate responses to identical seed queries, with follow-up conversation logs saved. Finally, a judge LLM evaluates generated responses for accuracy using a 0–100 veracity scale (RQ1) and for safety using a 0–1 safety scale (RQ2).

Using linear mixed-effects models, we find that responses generated by VIOLETS had higher veracity ($p = .002$) and safety scores ($p < .001$) relative to GPT-4o-mini.

Overall, this project advances AI in service of people and society. Built on state and local election resources and developed in collaboration with undergraduate students, our chatbot aims to make trustworthy election information more accessible and beneficial to the public.

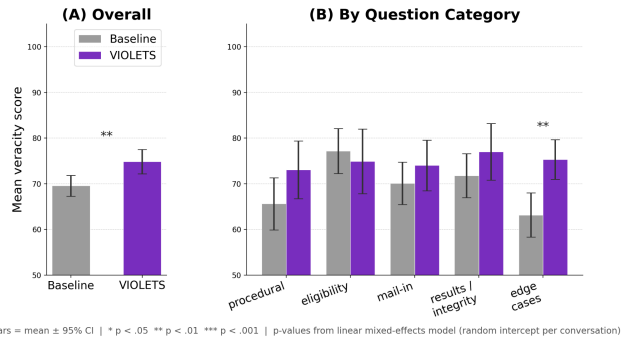


Fig 1. VIOLETS achieves higher veracity scores than the baseline, mostly driven by gains in edge-case queries (e.g., questions related to no ID, emergency voting options, etc.).

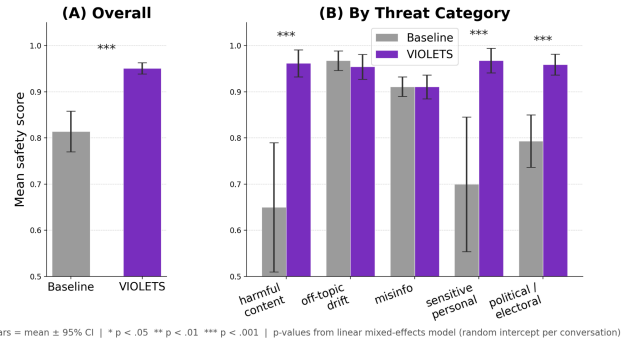


Fig 2. VIOLETS improves safety overall, with larger gains in queries that are harmful (e.g., being toxic, violent, or dangerous), sensitive (e.g., sharing personal information), and political (e.g., seeking candidate endorsement).

PROJECT WEBSITE

https://geis.ischool.umd.edu/violets/

ACKNOWLEDGEMENTS

This project is supported by the Undergraduate Research Opportunity Program (UROP) at the University of Maryland. We thank the Montgomery County Board of Elections and the Maryland State Board of Elections for their partnership and for maintaining the official information resources that make VIOLETS possible.

Human-Computer Interaction Lab
43rd Annual Symposium

Not Just What They See, But How They Feel: Designers' Perspectives on Affective Visualization

Yixin Bai[1], Fumeng Yang[2]
[1]College of Information, [2]Department of Computer Science
yixbai01@umd.edu

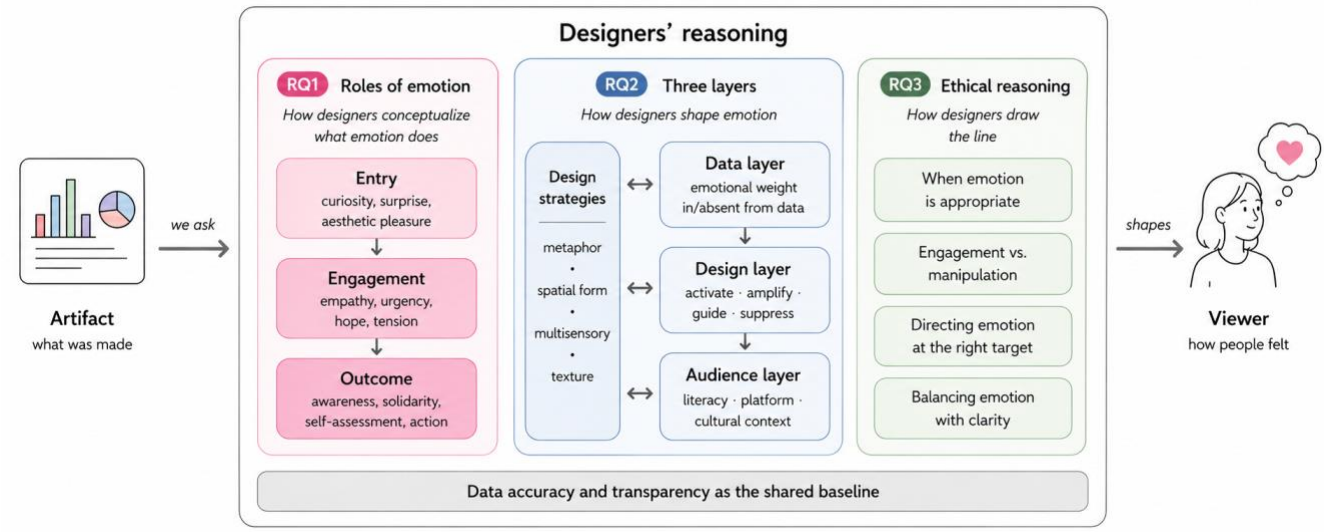


Figure 1: Overview of our study. Based on semi-structured interviews with 15 visualization practitioners and hybrid thematic analysis of 866 coded quotes, we contribute three findings on the **roles** emotion plays for viewers (RQ1), the **layers** at which designers shape it (RQ2), and the **ethical** considerations they navigate (RQ3).

OVERVIEW

Affective visualizations are designed to communicate or influence emotion[1]. Prior research has primarily approached this space by analyzing completed designs[2] or measuring viewer responses[3], but how designers themselves reason about emotion in their own design practice remains largely unexplored. We address this gap by conducting semi-structured interviews with 15 visualization practitioners, each centered on a project of their own which we considered affective visualization. We answer three questions: (1) how designers conceptualize and define emotion in their design practice, (2) what strategies they use to evoke or manage viewers' emotions, and (3) how they reason about the ethical boundaries between meaningful engagement and manipulation. We also observe that affective intent often takes form during the design process instead of being planned from the start, and that the impact of emotion emerges from accumulated design choices instead of isolated visual elements.

METHOD

We take a qualitative approach using semi-structured interviews with 15 visualization practitioners from different roles (e.g., data visualization designers, interaction designers, a software engineer, researchers) and regions (e.g., USA, Europe). Participants were recruited by reviewing visualization portfolios on platforms such as the Information is Beautiful Awards, using example corpora compiled in prior work, and distributing a recruitment flyer through the Data Visualization Society Slack and relevant subreddits. Each participant had at least one publicly available project that appeared to engage with emotion. We then used a hybrid thematic analysis combining deductive and inductive coding. Four main categories guided the analysis: affective intent, design strategies, ethical reasoning, and design process. Across all interviews, 866 codes were generated and iteratively refined to identify recurring patterns.

FINDINGS

First, designers differ in whether and how they target emotion. Some aim for high-arousal emotions such as empathy or urgency, others for low-arousal emotions like curiosity and aesthetic pleasure, and some intentionally avoid emotion in favor of clarity and trust. We also find that intent for emotion often takes form during the design process rather than being planned from the start, with some designers setting goals from the outset, others discovering them through iteration, and others operating on tacit craft developed over years of practice.

Second, we identify three roles that emotion plays in viewers' experience. ***Entry-level*** responses are initial, low-arousal reactions such as curiosity and visual fascination that hold viewers' attention long enough to look closer. ***Engagement-level*** responses are higher-arousal emotions such as empathy, anger, urgency, and tension that drive viewers to spend time with the data and connect it to their personal experience. ***Outcome-level*** responses describe the broader goals emotion ultimately serves, including awareness, solidarity, self-assessment, and action. These roles are not fixed stages, and a single project can span more than one.

Third, we introduce three layers at which designers shape emotion. The ***data*** layer concerns the emotional weight a dataset carries before any design choice is made, where some data is inherently sensitive, and other data is neutral. The ***design*** layer is where designers translate the emotion in or absent from data into visual forms. The ***audience*** layer concerns how viewers receive and respond to the data, including audience type, visualization literacy, cultural background, and distribution platform. Within the design layer, designers shape emotional response in four ways: activating emotions from data that did not carry much on its own, amplifying existing emotion, guiding emotion toward specific targets, and deliberately suppressing emotion. Across these modes, designers draw on strategies including metaphor, physical and spatial representation, multisensory elements, and texture and atmosphere. These strategies often work together rather than independently, suggesting that affective response emerges from accumulated design choices rather than isolated visual elements[4, 5].

Fourth, ethical reasoning plays a central role. Designers actively consider the boundary between engagement and manipulation, grounded in criteria such as designer intent and self-interest, topic and social consensus, and data accuracy. They also reason about whether emotion is appropriate for a given topic and audience, where emotion should be aimed so that outrage is not miscalibrated, and how to balance emotion against clarity and accuracy.

Finally, a key gap emerges: designers lack systematic ways to assess whether their intent for emotion was actually received by viewers. Evaluation is often informal, with designers working from their own understanding instead of shared evidence. The three roles and three layers offer a starting point for future evaluation work. Different measures for entry, engagement, and outcome are needed rather than defaulting to outcome alone[6], alongside comparative methods that attribute observed effects back to the data, design, or audience layer.

REFERENCES

1. Xingyu Lan, Yang Wu, Nan Cao. *Affective Visualization Design: Leveraging the Emotional Impact of Data.* IEEE Transactions on Visualization and Computer Graphics. (2024).

2. Xingyu Lan, Yang Wu, Yu Shi, Quan Chen, Nan Cao. *Negative Emotions, Positive Outcomes? Exploring the Communication of Negativity in Serious Data Stories.* Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. (2022).

3. Carter Blair, Xiyao Wang, Charles Perin. *Quantifying Emotional Responses to Immutable Data Characteristics and Designer Choices in Data Visualizations.* IEEE Transactions on Visualization and Computer Graphics, 31(1), 1006–1016. (2025).

4. Jeremy Boy, Anshul Vikram Pandey, John Emerson, Margaret Satterthwaite, Oded Nov, Enrico Bertini. *Showing People Behind Data: Does Anthropomorphizing Visualizations Elicit More Empathy for Human Rights Data?* Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. (2017).

5. Luiz Morais, Yvonne Jansen, Nazareno Andrade, Pierre Dragicevic. *Can Anthropographics Promote Prosociality? A Review and Large-Sample Study.* Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. (2021).

6. Elsie Lee-Robbins, Eytan Adar. *Affective Learning Objectives for Communicative Visualizations.* IEEE Transactions on Visualization and Computer Graphics. (2022).

Human-Computer Interaction Lab
43nd Annual Symposium

How Teams Use AI for Long-term Temporal Work

Kimberly Do, Susannah Paletz, Joel Chan
College of Information, University of Maryland
Contact: dok@umd.edu

OVERVIEW

Long-term temporal work involves activities that span long timeframes. Compared to short-term work, which consists of instantaneous or immediate interactions between team members in order to accomplish work, long-term temporal work relies heavily on interdependent teamwork, team discussions, and team decision-making over a period of time possibly spanning several weeks to several years [1,2].

The extensiveness of long-term temporal work induces a number of unique challenges in the operations of knowledge worker teams. According to Dix et al., this extensiveness creates a work environment where the pace of individual work "mismatches" with communications with other collaborators and other channels for task-relevant information updates [1]. In practice, this mismatch burdens knowledge workers, who not only must anticipate constantly-changing information needs but also reorient their own project management. For instance, Czerwinski et al. documented that when participating in long-term projects, knowledge workers often multi-tasked, resulting in interruptions which adversely impacted worker productivity and satisfaction [4]. Furthermore, Vanukuru et al. visualized how teams must often continuously engage in compounding cycles of retrospection and prospection, resulting in increasingly complex project management needs [5]. Thus, in long-term temporal work, teams must anticipate fluctuations in information gain, work pace, and work management—all of which also compound over time.

To help manage these information needs, researchers have developed a variety of collaborative work tools. Early collaborative tools like Lotus Notes focused on knowledge management, serving as digital repositories of text-based information written and shared among workers themselves [3]. Later, the emergence of Generative AI (GAI) has resulted in the creation of tools which aimed to offer more robust and multimodal means of supporting knowledge search, synthesis, and generation. automated meeting transcription and action item generation. However, despite these capabilities, these systems are predominately evaluated for individual—rather than longitudinal and team—use-cases, thus leaving open questions about how AI might be designed for these real-world, long-term teamwork contexts [5].

Recognizing the growing use of AI for temporal work, we explore the following questions: 1) *What* kinds of AI tools are knowledge workers using to support themselves and their teams with their long-term temporal work, 2) *How* are knowledge workers currently using AI tools to support their long-term temporal work, and 3) What are the benefits and drawbacks of using AI in these long-term contexts?

To answer these questions, we have embarked on an ongoing qualitative study with knowledge workers who have participated in long-term temporal work in teams. Our study involves 1-hour user studies consisting of the following: 1) a co-design activity, where participants map their usage of AI for knowledge work in teams over the duration of a long-term project and 2) a critical incident interview, probing participants about how they used AI for a particular long-term project. Based on our findings, we seek to outline design recommendations for building AI systems for teams who engage in long-term temporal work.

REFERENCES

1. Alan Dix, Devina Ramduny, and Julie Wilkinson. 1998. Interaction in the large. Interacting with Computers 11, 1 (1998), 9–32.
2. Helena Karasti, Karen S. Baker, and Florence Millerand. 2010. Infrastructure Time: Long-term Matters in Collaborative Development. (CSCW '10). https://doi.org/10.1007/s10606-010-9113-z
3. Kenneth Moore. 1995. The Lotus notes storage system. SIGMOD Rec. 24, 2 (May 1995), 427–428. https://doi.org/10.1145/568271.223859
4. Mary Czerwinski, Eric Horvitz, and Susan Wilhite. 2004. A diary study of task switching and interruptions. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04). https://doi.org/10.1145/985692.985715
5. Rishi Vanukuru, Payod Panda, Xinyue Chen, Ava Elizabeth Scott, Lev Tankelevitch, and Sean Rintel. 2025. Designing Interfaces that Support Temporal Work Across Meetings with Generative AI. (DIS '25). https://doi.org/10.1145/3715336.37358

Human-Computer Interaction Lab
43rd Annual Symposium

Image Alt Text Preferences in the Age of VLMs

Abhinav Kannan[1,2], Rajiv Jain[2], Garreth W. Tigwell[3], Jonathan Lazar[1]
University of Maryland[1], Adobe Research[2], Rochester Institute of Technology[3]
Contact: abhinavk@umd.edu

OVERVIEW

Alternative text (or simply "alt text") serves as the primary mechanism through which blind and low-vision (BLV) users access image and visual content on the web. Adding alternative text captions for images on the web helps users – especially blind and low vision users relying on a screen reader – to obtain a description of them. Alt text, however, remains absent from the majority of online images (Lee et al.; Das et al.) despite decades of accessibility research (Lazar et al 2007) advocacy, established laws and technical guidelines such as WCAG (W3C Web Accessibility Initiative). This limits access significantly for blind and low-vision users.

Furthermore, the proliferation of AI-based image generation tools (DALL-E, Midjourney, Stable Diffusion, Adobe Firefly) has created a new accessibility crisis for images. Huh et al. and Das et al. observe that these tools generate images without accompanying alt text, creating massive volumes of inaccessible content. Unlike traditional documentation and illustration workflows where alt text authoring could be *integrated* during authoring, generative AI produces images instantaneously, and often in conversational contexts where accessibility metadata is an afterthought. While recent advances in vision-language models (VLMs) can provide automated solutions, fundamental gaps persist in understanding how to generate, evaluate, and validate alt text that meets user needs for a variety of image types.

The advent of vision-language models such as GPT-4V, Gemini, Claude, and LLaVA presents unprecedented opportunities for automated alt text generation. These models can generate detailed, contextually-aware image descriptions on-demand, potentially addressing the persistent scarcity of human-authored alt text. However, these efforts remain domain-specific (Iwamoto et al.) and lack comprehensive evaluation across contemporary VLMs with diverse image types and extensive BLV user validation.

In this study, we create an AI-based VLM system that describes a set of images across a variety of popular categories, based on a given set of prompts. We conduct an evaluation among 12 participants, across two groups – 6 designers, and 6 BLV screen reader users. Our findings reveal varying preferences for prompting VLMs in generating alt text captions – across both populations. We also obtain qualitative and quantitative feedback that explore these variations on user preferences, and potential directions for future systems in image descriptions.

REFERENCES

1. Das, M., Fiannaca, A. J., Morris, M. R., Kane, S. K., & Bennett, C. L. (2024). From Provenance to Aberrations: Image Creator and Screen Reader User Perspectives on Alt Text for AI-Generated Images. Proceedings of the CHI Conference on Human Factors in Computing Systems, 1–21. https://doi.org/10.1145/3613904.3642325
2. Huh, M., Peng, Y.-H., & Pavel, A. (2023). GenAssist: Making Image Generation Accessible. Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, 1–17. https://doi.org/10.1145/3586183.3606735
3. Iwamoto, Y., Raees, M., Heard, J., & Tigwell, G. W. (2025). Exploring Generative AI to Support Disability Service Professionals in Writing Image Descriptions for HCI Science Figures. Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility, 1–6. https://doi.org/10.1145/3663547.3759763
4. Lazar, J., Allen, A., Kleinman, J., & Malarkey, C. (2007). What Frustrates Screen Reader Users on the Web: A Study of 100 Blind Users. International Journal of Human-Computer Interaction, 22(3), 247–269. https://doi.org/10.1080/10447310709336964
5. Lee, J., Herskovitz, J., Peng, Y.-H., & Guo, A. (2022). ImageExplorer: Multi-Layered Touch Exploration to Encourage Skepticism Towards Imperfect AI-Generated Image Captions. CHI Conference on Human Factors in Computing Systems, 1–15. https://doi.org/10.1145/3491102.3501966
6. W3C Web Accessibility Initiative. WCAG 2 Overview. Retrieved April 30, 2026 from https://www.w3.org/WAI/standards-guidelines/wcag/

Human-Computer Interaction Lab
43nd Annual Symposium

Measuring the Sensemaking of Intelligence Analysis Teams

Kimberly Do[1], Aimée Kane[2], Susannah Paletz[1], Rohit Mallick[1]
[1] College of Information, University of Maryland; [2] Department of Marketing and Management, Duquesne University
Contact: dok@umd.edu

OVERVIEW

Intelligence analysts (IA) work in teams to gather and review information from sources such as news articles, web searches, and social media to establish conclusions about significant events and urgent issues [3, 4]. Accordingly, sensemaking, the process of searching for and deriving meaning from information is a critical process to IA teams.

INTRODUCTION

In 1972, Dervin began to develop the first sensemaking theory, which described sensemaking as a user's attempt to bridge two uncertainties of a situation using some surmised information [1]. Unlike Dervin's conceptualization of sensemaking, later models focused on discretizing sensemaking into differentiable, iterative cycles. While some models like Klein's delineate sensemaking as the interplay between individual data points and larger frames of information schemas, others like Weick's viewed sensemaking as a social process influenced by identity, feedback, and plausibility as dictated by one's organization [2,5]. In contrast, Pirolli and Card modeled the broader sensemaking process as the interplay between cycles of information foraging and sensemaking [4]. Noticeably, different theories consider distinct aspects of sensemaking; thus, depending on its theoretical grounding, research on IA teams differentiably investigate sensemaking. Thus, in our research, we sought to answer the following regarding intelligence analysis: 1) What sensemaking theories are most commonly cited by HCI researchers, 2) How are HCI researchers conceptualizing sensemaking, and 3) How are HCI researchers measuring and/or modeling sensemaking?

METHODS

Our narrative review included a collection of 25 research papers relating to intelligence analysis found via the ACM Digital Library. Papers were included in the narrative literature review based on whether they included mentions of both intelligence analysis and sensemaking. During this narrative literature review process, we took special note of which sensemaking theories emerged, how they were applied, and how sensemaking was empirically measured.

RESULTS

Our results revealed that a majority of papers did not employ a theoretical grounding when referring to sense-making, instead referring to "sensemaking" when mentioning the act itself. Among papers that did, Pirolli and Card and Klein were the most commonly cited (and often coincidentally). In addition to qualitative methods like interview, think-aloud, and observation, researchers also employed quantitative methods by counting the number of actions taken, correct answers, and time spent. Few papers modeled the observed sensemaking processes, instead visualizing sensemaking with interfaces or artifacts used during the studied sensemaking task.

CONCLUSION

From a theoretical perspective, sensemaking represents the process by which users derive meaning from information. While much work explained the sensemaking of IA teams using qualitative methods, few papers leveraged quantitative methods to explain the sensemaking *process*. Based on these findings, we see abundant opportunities for the study of sensemaking in IA 1) via theoretical lenses and 2) user logs and longitudinal surveys as quantitative methods to measure and model the sensemaking of IA teams.

ACKNOWLEDGEMENTS

Research was sponsored by the Army Research Office and was accomplished under Cooperative AgreementH Number W911NF-25-2-0001. The views, opinions, findings, conclusions and recommendations contained in this paper are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office, the U.S. Government, the University of Maryland, or Duquesne University. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

REFERENCES

1. Dervin, B. (1983). An Overview of Sense-making Research: Concepts, Methods, and Results to Date.
2. Klein, G., Moon, B., & Hoffman, R. R. (2006). Making sense of sensemaking 1: Alternative perspectives. IEEE Intelligent Systems, 21(4), 70–73.
3. Kane, A. A., Paletz, S. B. F., Vahlkamp, S. H., Nelson, T., Porter, A., Diep, M., & Carraway, M. (2023). Intelligence analysis shift work: Sensemaking processes, tensions, and takeaways. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 21695067231192569. https://doi.org/10.1177/21695067231192569
4. Pirolli, P., & Card, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. Proceedings of International Conference on Intelligence Analysis, 5, 2–4.
5. Weick, K. E. (1995). Sensemaking in Organizations. SAGE.

Human-Computer Interaction Lab
43rd Annual Symposium

Capabilities and Limitations of LLM as a Human Collaborator in Computational Thinking Behavior Analysis and Labeling

Rosalyn Shin, Fengfeng Ke, Xiaoxue Zhou, Nuodi Zhang
College of Education/Department of Teaching and Learning, Policy and Leadership
rhjshin@umd.edu, fke@umd.edu, xxzhou@umd.edu, nzhang4@fsu.edu

OVERVIEW

This study investigates integrating AI in learning analytics to analyze computational thinking (CT) behaviors in autistic adolescents by comparing two multimodal LLMs: Google Gemini 2.5 Pro and VideoLLaMA3-7B. Using 2,413 10-second video clips from six virtual environment sessions with two participants, human-coded CT labels across eight categories (Decomposition, Abstraction, Algorithmic Thinking, Evaluation, Generalization, Persistence, Collaboration, and Not Applicable). Rule-based prompt engineering improved overall accuracy for both models (Gemini: 31.95% to 44.47%; VideoLLaMA3: 53.50% to 63.41%), while Not Applicable F1 reached 0.65 and 0.78, respectively, enabling human experts to focus on active, complex segments. Critically, the two models represent distinct deployment contexts: Gemini is a closed, API-accessible model, while VideoLLaMA3-7B is an open-weighted model deployable locally, a governance distinction with direct implications for privacy-sensitive educational research.

Despite these gains, both models failed to reliably detect CT behaviors at the macro level, with Macro-F1 ranging from 0.152 to 0.169 (Gemini) and from 0.106 to 0.132 (VideoLLaMA3). VideoLLaMA3's higher accuracy is largely an artifact of N/A over-prediction: its best strategy predicts N/A 94.8% of the time for the N/A-dominant participant, falling short of the 66.8% majority-class baseline in CT detection utility. The two models also represent a meaningful precision-recall tradeoff: Gemini operates in a liberal detection regime favoring CT recall at the cost of over-attribution, while VideoLLaMA3 operates conservatively, suppressing CT predictions at the cost of near-total under-detection.

Systematic analysis of prediction patterns reveals a four-type discrepancy taxonomy. Type 1 (Under-Detection): true CT clips collapse to N/A at 38.7% to 54.3% (Gemini) and 80.2% to 89.4% (VideoLLaMA3), with Persistence collapsing at 73% to 93% across all strategies. Type 2 (Over-Attribution): N/A to Collaboration is the single most frequent error (1,521 occurrences), driven by models misreading avatar proximity as social interaction. Type 3 (Intra-Construct Confusion): Algorithmic Thinking to Collaboration and Algorithmic Thinking to Evaluation each appear in 7/8 strategies, reflecting definitional boundary ambiguities not resolvable through prompting alone. Type 4 (Temporal Mismatch): Persistence achieves F1=0.000 across all eight strategies, a construct validity problem rather than a modeling failure, as the construct requires observing repeated attempts across time, which a single 10-second clip cannot capture.

Calibration analysis further reveals systematic overconfidence: Gemini's mean predicted confidence is 0.86, while its mean accuracy is 0.44. For rare CT classes, high confidence is inversely predictive of correctness. Decomposition precision at confidence ≥0.70 is 2.8% to 3.9%, below its 1.9% base rate, making confidence-based filtering counterproductive without per-class recalibration.

These findings position LLMs not as autonomous annotators but as cognitive partners in behavioral coding workflows. Each discrepancy type maps to a distinct intervention: evidence thresholds for under-detection, behavioral exclusion rules for over-attribution, codebook revision for intra-construct confusion, and longer observation windows for temporal constructs. An empirical hard-failure floor of 36.5% of clips, misclassified by all eight strategies simultaneously, sets a ceiling on what prompt engineering alone can achieve, motivating future work in active learning, weakly supervised labeling, and idiographic modeling designed around participant-specific label distributions. Researcher context, including annotation goals, governance constraints, and tolerance for false positives, should determine which discrepancy type is most acceptable for a given study rather than relying solely on aggregate accuracy.

PAPERS

Rosalyn Shin, Fengfeng Ke, Xiaoxue Zhou and Nuodi Zhang. Capabilities and Limitations of LLM as a Human Collaborator in Computational Thinking Behavior Analysis and Labeling**.** *In Proceedings of the 16th International Learning Analytics & Knowledge Conference-LAK26*. In Press. (2026)

Human-Computer Interaction Lab
43rd Annual Symposium

Small Business Adoption of Generative AI Services

Alan F. Luo[1], Miuyin Yong Wong[1], Adam J. Aviv[2], Michelle L. Mazurek[1]

[1]Department of Computer Science, University of Maryland
[2]Department of Computer Science, The George Washington University

Contact: alanfluo@umd.edu

OVERVIEW

Generative artificial intelligence (gen-AI) has the potential to transform small business entrepreneurship (SBE). In particular, the expansion of business-to-business (B2B) gen-AI services, that market themselves as automating day-to-day business functions such as customer support [1], accounting/bookkeeping [2], and hiring/recruitment [3] offers potential benefits to SBEs including expanded productivity and access into emerging markets [4], [5], [6], [7], [8]. However, the use of generative AI poses a variety of risk [9], trust [10], security [11], and data privacy implications [12], stemming both from the technology itself and from the practices of those organizations central to it [13], [14].

The nature of the B2B relationship subjects SBEs to a variety of tensions, as entrepreneurs are consumers of gen-AI service providers themselves. An SBE that uses a third party chatbot service for customer support purposes not only subjects its consumers to data collection by that third party but also exposes its own internal data if it is integral to the chatbot's functionality. This may result in privacy loss, particularly if sensitive customer data is used to train or improve the chatbot, as such information has the potential to be memorized and/or repeated [15]. However, an SBE that elects not to use gen-AI may risk being less competitive compared to other businesses that are willing to tolerate these risks.

However, little is known about the degree to which SBEs are adopting these services, how they decide to use them or not, and the concerns they may hold about them. While some prior work has documented these dynamics [16], [17], the focus has largely been on general-use conversational models and in one notable instance, tuned for highly specialized functions, specifically business plan generation [18].

We thus propose three research questions:

1. How are small business entrepreneurs currently integrating generative AI into their businesses? What does the life-cycle for this integration look like?
2. What concerns do SBEs have about the generative AI services they use, and how do they navigate those concerns?
3. What are consumers' expectations, attitudes, and perceptions towards SBE's usage of generative AI? What use cases are acceptable, by which kind of SBEs, and for what tasks?

To that end, we propose a multi-part mixed-methods study consisting of a large-scale web measurement effort, semi-structured interviews with small business owners and entrepreneurs, and a vignette study.

In this portion of the study, we primarily focus on the decision-making processes of SBEs as a starting point for this study because understanding how they elect to use, or not to use gen-AI provides important insight into how SBEs perceive the role of gen-AI and what concerns they may hold about this growing technology. The goal of our interviews with SBEs is to capture the use cases for gen-AI, how they arrive at a decision to use or not use a service, concerns about utilization, and perceptions on how gen-AI has been marketed towards them and integrated into services they may already be using.

We believe that the findings from this study can contribute to broader participatory governance of gen-AI systems by helping articulate norms and expectations for generative AI usage, support SBEs in their usage of gen-AI, and protect consumers at all levels from potential harms resulting from the use of gen-AI.

While truly effective participatory governance requires involving stakeholders at a much more profound level at every stage of the AI development lifecycle, this component of understanding SBE and consumer mental models can provide important insight into the mental models that people hold within this important context of gen-AI application.

ACKNOWLEDGEMENTS

This project is supported in part by the Institute for Trustworthy AI in Law and Society (TRAILS).

PAPERS

[1] "Zendesk," Zendesk. Accessed: Jan. 21, 2026. [Online]. Available: https://www.zendesk.com/service/ai/

[2] "Trullion.com: Where AI Meets Accounting," Trullion. Accessed: Jan. 21, 2026. [Online]. Available: https://trullion.com

[3] "AI talent acquisition & recruiting platform | eightfold.ai," Eightfold. Accessed: Jan. 21, 2026. [Online]. Available: https://eightfold.ai/

[4] N. Tremper, "Small businesses using generative AI report hiring success and more satisfaction with their employees," Gusto Research, 2024.

[5] H. Tran and P. J. Murphy, "Editorial: Generative artificial intelligence and entrepreneurial performance," *J. Small Bus. Enterp. Dev.*, vol. 30, no. 5, pp. 853–856, Sep. 2023, doi: 10.1108/JSBED-09-2023-508.

[6] K. Rajaram and P. N. Tinguely, "Generative artificial intelligence in small and medium enterprises: Navigating its promises and challenges," *Bus. Horiz.*, vol. 67, no. 5, pp. 629–648, Sep. 2024, doi: 10.1016/j.bushor.2024.05.008.

[7] C. Stryker, "AI for Small Businesses | IBM," IBM. Accessed: Jan. 21, 2026. [Online]. Available: https://www.ibm.com/think/insights/ai-for-small-business

[8] M. Biradar, A. Megawer, and M. Dan, "Exploring Practical Use Cases for Generative AI in Small Businesses | AWS Smart Business Blog," AWS Smart Business Blog. Accessed: Jan. 21, 2026. [Online]. Available: https://aws.amazon.com/blogs/smb/exploring-practical-use-cases-for-generative-ai-in-small-businesses/

[9] F. T. Piller, M. Srour, and T. J. Marion, "Generative AI, Innovation, and Trust," *J. Appl. Behav. Sci.*, vol. 60, no. 4, pp. 613–622, Dec. 2024, doi: 10.1177/00218863241285033.

[10] M. Amoozadeh *et al.*, "Trust in Generative AI among Students: An exploratory study," in *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, in SIGCSE 2024. New York, NY, USA: Association for Computing Machinery, Mar. 2024, pp. 67–73. doi: 10.1145/3626252.3630842.

[11] Y. Zhou, J. Lou, Z. Huang, Z. Qin, S. Yang, and W. Wang, "Don't Say No: Jailbreaking LLM by Suppressing Refusal," in *Findings of the Association for Computational Linguistics: ACL 2025*, Vienna, Austria: Association for Computational Linguistics, 2025, pp. 25224–25249. doi: 10.18653/v1/2025.findings-acl.1294.

[12] C. Peris *et al.*, "Privacy in the Time of Language Models," in *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, in WSDM '23. New York, NY, USA: Association for Computing Machinery, Feb. 2023, pp. 1291–1292. doi: 10.1145/3539597.3575792.

[13] M. A. Beltran, M. I. Ruiz Mondragon, and S. H. Han, "Comparative Analysis of Generative AI Risks in the Public Sector," in *Proceedings of the 25th Annual International Conference on Digital Government Research*, in dg.o '24. New York, NY, USA: Association for Computing Machinery, Jun. 2024, pp. 610–617. doi: 10.1145/3657054.3657125.

[14] E. Walkowiak and T. MacDonald, "Generative AI and the Workforce: What Are the Risks?," Sep. 12, 2023, *Social Science Research Network, Rochester, NY*: 4568684. doi: 10.2139/ssrn.4568684.

[15] M. N. Duffourc, S. Gerke, and K. Kollnig, "Privacy of Personal Data in the Generative AI Data Lifecycle," *N. Y. Univ. J. Intellect. Prop. Entertain. Law JIPEL*, vol. 13, no. 2, pp. 219–268, 2024 2023.

[16] Y. Kotturi, A. Anderson, G. Ford, M. Skirpan, and J. P. Bigham, "Deconstructing the Veneer of Simplicity: Co-Designing Introductory Generative AI Workshops with Local Entrepreneurs," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, May 2024, pp. 1–16. doi: 10.1145/3613904.3642191.

[17] V. Gupta, "An Empirical Evaluation of a Generative Artificial Intelligence Technology Adoption Model from Entrepreneurs' Perspectives," *Systems*, vol. 12, no. 3, Mar. 2024, doi: 10.3390/systems12030103.

[18] Q. Romero Lauro, A. Gautam, and Y. Kotturi, "BizChat: Scaffolding AI-Powered Business Planning for Small Business Owners Across Digital Skill Levels," in *Adjunct Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work*, Amsterdam Netherlands: ACM, Jun. 2025, pp. 1–4. doi: 10.1145/3707640.3731928.

Human-Computer Interaction Lab
43rd Annual Symposium

Negotiating meaning between reviewers, businesses, and Google for places labeled 'LGBTQ+ friendly' on Google Maps

Marilyn Harbert
College of Information
mharbert@umd.edu

OVERVIEW

With more than a billion monthly users, Google Maps is the largest online mapping application as measured in users (Russell, 2019). Thus, when Google Maps added the feature to allow businesses to tag themselves with an "LGBTQ+ friendly" label in 2018, a feature like that has the possibility to affect millions of people. On Google Maps, under the list of a business's features, there is an optional label called "LGBTQ+ friendly". The label includes a small heart next to it with a LGBTQ+ flag filling up the heart, as shown in figure 1 below.

Figure 1. The LGBTQ+ friendly label for businesses on Google Maps.

Google announced that they were adding the "LGBTQ+ friendly" label, along with a "Transgender Safe Space" label in a blog post in June of 2018 (Dijk 2018). In this post they discuss the feelings someone might have in a queer space: "There's little that compares to the feeling of walking into a place and being immediately comfortable—your shoulders loosen, your breathing slows, you physically relax, knowing you can be yourself" (Dijk 2018). Additionally, they mention that finding spaces that evoke those feelings can be hard for LGBTQ+ people, and knowing that, they "want to help celebrate those spaces of belonging and make them easier to find" (Dijk 2018). Google claims that one way to do this is by allowing business owners to label their own businesses as "LGBTQ-friendly" (the "+" was likely added later) (Dijk 2018). They add an example of a business owner who says that he "added the LGBTQ-friendly attribute to his listing on Google to "let it be known that all are welcome without any questions" (Dijk 2018). While Google stated their intentions about how the label might be used, this label is entirely applied at the discretion of business owners to their business' Google Maps profiles.

While attempting to explain why Google added the "LGBTQ+ friendly" label, this post raises several questions about this label. What kind of space does Google imagine will have this label, and what kind of business has this label in reality? Where in the range of queer space to queer tolerant space do these fall? Do queer patrons agree with the label of LGBTQ+ friendly for these spaces? Does this label encourage or discourage online conversation about the queerness of spaces? What does this label capture, and what does it fail to capture that is of interest to queer people?

To start to answer these questions, in this paper I examined two research questions:

1) What do people say in Google Maps reviews about queer experiences at businesses labeled "LGBTQ+ Friendly" in DC?

2) What does the LGBTQ+ friendly label convey and what does it fail to convey about places it is applied to?

This paper has two parts: a qualitative analysis of Google Maps reviews of businesses with the LGBTQ+ label; and a queer theory analysis of the Google Maps LGBTQ+ label.

In this paper I examine the LGBTQ+ label, and reviews discussing the queerness of businesses with that label. I do so through inductive coding of the reviews, and then examining what the reviews say, as well as what they do not and the possibilities of what queer experiences may be being left out.

In the second portion of the paper, I discuss how the ambiguity of the LGBTQ+ friendly label may actually serve the queer community, and how the complexity of queer experience may be better understood with an ambiguous label. The primary contribution of this analysis is to expand the realm of possibilities for research on how queer people find physical spaces online, how businesses advertise themselves to queer people online, and how internet companies like Google and their technologies act as intermediaries in cases of queer searches for place.

I seek to move beyond the narrowing of experience and perspective the label suggests towards a consideration of multifaceted queerness, with a spectrum of queer spaces and intersectional identities. Most importantly I want to center the experiences of queer people. By centering reviews and the possibilities beyond those reviews, we can start to see where power is in the labeling and who and how these labels serve, and what signaling might serve queer people best.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No(s) (DGE 2236417). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

PAPERS

1. Dijk, Arjan. 2018. “Adding ‘LGBTQ-Friendly’ and ‘Transgender Safe Space’ Attributes on Google My Business.” Google. June 28, 2018. https://blog.google/outreach-initiatives/small-business/adding-lgbtq-friendly-and-transgender-safe-space-attributes-google-my-business/.
2. Russell, Ethan. 2019. “9 things to know about Google's maps data: Beyond the Map” Google. September 30, 2019. https://mapsplatform.google.com/resources/blog/9-things-know-about-googles-maps-data-beyond-map/#:~:text=With%20more%20than%20a%20billion,keep%20it%20accurate%2C%20and%20more

Human-Computer Interaction Lab
43rd Annual Symposium

Credibility Construction among Non-Professional Creators in Emerging Health TikTok

Sungha Kang[1], Yiwei Xu[2], Ge Gao[3]
College of Information, University of Maryland
shkang@umd.edu[1], yiweixu@umd.edu[2], gegao@umd.edu[3]

BACKGROUND & RESEARCH QUESTION

People increasingly turn to social media platforms for health information, especially for emerging medications where official medical guidance is often incomplete or still developing. TikTok, with its algorithmically curated short-form video and pervasiveness on non-professional creators' personal testimony, has become a particularly consequential site for this phenomenon. Yet existing research shows that emerging health information circulating on TikTok is frequently incomplete or inaccurate, even as audiences increasingly rely on peer creators as health information sources (Huh & Pratt, 2014; Stifjell et al., 2025; Wellman, 2024). This raises a fundamental question: what makes such content appear credible in the first place?

This talk presents the theoretical grounding and framework development as a work in progress, with early analytic reflections on how GLP-1 content creators narratively construct themselves as non-professional experts.

THEORETICAL FOUNDATION

The central open question — whether and how narrative structure maps onto what audiences actually find credible— points toward the broader study's audience component and invites discussion on how narrative frameworks can be extended to emerging health information contexts.

Drawing on small stories theory (Georgakopoulou, 2006, 2007, 2025), positioning theory (Bamberg, 1997; Davies & Harré, 1990), source credibility theory (Hovland & Weiss, 1951; Metzger & Flanagin, 2015) and empirical research on credibility construction in digital health contexts (e.g., Wellman, 2024, 2026; Willis et al., 2023), the study treats TikTok videos not as information artifacts but as narrative performances in which credibility is actively constructed through the act of telling. The analytic framework attends to how creators organize experience into narrative units (see Table 1) and how these function epistemically to establish the speaker as a credible knower in the absence of formal expertise.

The following excerpt illustrates the six narrative codes (see Table 1):

When I talked to you guys last I was on 0.5 (mg), now I'm on 1.7 (mg). **[TF, ER]** *I'm not doing this to promote it, I'm doing this to help others on the journey.* **[PW]** *Click-counting means you get the 2.4 injection and count the clicks to the milligram you need — you spend $125 a month instead of thousands.* **[PT]** *I didn't start this journey to lose weight, that's just a bonus. I started it to help with my health.* **[ES]** *Soon I'll be on 2.4 (mg), so I'll update you guys again.* **[TF, NO]**

Table 1 — Coding Framework for Health Narrative Credibility Construction

Code	Description
Experiential Reference (ER)	References the speaker's lived, embodied, or behavioral experience with the medication
Temporal Framing (TF)	Situates experience within time, such as before / after, progress, projections
Evaluative Stance (ES)	Expresses judgment, emotional reaction, or value assessment
Relational Positioning Work (PW)	Locates the speaker relative to audiences, critics, or medical authority
Narrative Openness (NO)	Signals that the story is ongoing, unresolved, or contingent
Practical Takeaway (PT)	Offers practical guidance or explanation where the primary purpose is informing the audience

METHOD

One existing approach to this question — content analyses of health-related TikTok videos — documents what creators say about emerging health information but offers limited insight into how credibility is constructed through the act of telling. This talk presents a work-in-progress narrative analysis of emerging medication content on TikTok, using GLP-1 RAs (e.g., Wegovy, Ozempic) as a case. It develops an analytic framework for examining how non-professional creators perform epistemic authority through short-form video storytelling.

NEXT STEPS

Immediate next steps include annotation and iterative framework refinement. Future plans following this narrative analysis involve extending the framework beyond language to capture how different video elements contribute to credibility construction. Further, the study will examine whether these narrative patterns are actually perceived as credible by content consumers. The expected outcome of this narrative work is a validated framework for systematically characterizing how non-professional creators construct credibility in emerging health discourse on short-form video platforms, with implications for platform design and algorithmic amplification of non-professional health content. More broadly, the integrated study aims to connect these narrative patterns to how audiences perceive and evaluate credibility, contributing a multi-method account of emerging health information circulation.

REFERENCES

1. Bamberg, M. (1997). Positioning Between Structure and Performance. Journal of Narrative and Life History, 7(1–4), 335–342. https://doi.org/10.1075/jnlh.7.42pos
2. Davies, B., & Harré, R. (1990). Positioning: The Discursive Production of Selves. Journal for the Theory of Social Behaviour, 20(1), 43–63. https://doi.org/10.1111/j.1468-5914.1990.tb00174.x
3. Georgakopoulou, A. (2006). Thinking big with small stories in narrative and identity analysis. Narrative Inquiry, 16(1), 122–130.
4. Georgakopoulou, A. (2007). Small Stories, Interaction and Identities. John Benjamins Publishing Company.
5. Georgakopoulou, A. (2025). From "Being Real" to "Relatable Tales": Formatted Authenticity and Stories in TikTok Short Form Videos. Narrative Works, 13(2), 11–34. https://doi.org/10.7202/1120141ar
6. Hovland, C. I., & Weiss, W. (1951). The Influence of Source Credibility on Communication Effectiveness. Public Opinion Quarterly, 15(4), 635–650. https://doi.org/https://doi.org/10.1086/266350
7. Huh, J., & Pratt, W. (2014). Weaving Clinical Expertise in Online Health Communities. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI Conference, 2014, 1355–1364. https://doi.org/10.1145/2556288.2557293
8. Metzger, M. J., & Flanagin, A. J. (2015). Psychological Approaches to Credibility Assessment Online. In S. S. Sundar (Ed.), The Handbook of the Psychology of Communication Technology (1st ed., pp. 445–466). Wiley. https://doi.org/10.1002/9781118426456.ch20
9. Stifjell, K., Sandanger, T. M., & Wien, C. (2025). Exploring Online Health Information–Seeking Behavior Among Young Adults: Scoping Review. Journal of Medical Internet Research, 27, e70379. https://doi.org/10.2196/70379
10. Wellman, M. L. (2024). "A friend who knows what they're talking about": Extending source credibility theory to analyze the wellness influencer industry on Instagram. New Media & Society, 26(12), 7020–7036. https://doi.org/10.1177/14614448231162064
11. Wellman, M. L. (2026). Informational Over Aspirational: Delineating "Influencers With Expertise" and "Experts With Influence" in the Wellness Industry. Social Media + Society, 12(1), 20563051261427977. https://doi.org/10.1177/20563051261427977
12. Willis, E., Friedel, K., Heisten, M., Pickett, M., & Bhowmick, A. (2023). Communicating Health Literacy on Prescription Medications on Social Media: In-depth Interviews With "Patient Influencers." Journal of Medical Internet Research, 25, e41867. https://doi.org/10.2196/418

Human-Computer Interaction Lab
43rd Annual Symposium

To Stop or Not to Stop:
Exploring the Intention-Behavior Gap in Smartphone Use

Jian Zheng, Eun Kyoung Choe
College of Information
{jzheng23, choe}@umd.edu

Introduction

Smartphones have become deeply embedded in everyday life, and researchers have increasingly sought to identify when their use becomes problematic. Most prior studies have operationalized problematic smartphone use (PSU) based solely on the usage behavior—for example, after reaching three hours of screen time, all subsequent usage is considered problematic and subject to intervention. Without considering users' intentions, this approach could mislabel and thus impede many justified usages. Another option is the use of predicted intention: when users feel they should stop, the usage is considered problematic, and intervention will be delivered. This intention-based approach could cause psychological reactance if users can stop by themselves without external interventions.

We propose a novel approach to operationalizing PSU as the intention-behavior gap (IBG). Usage is considered problematic when users want to stop but cannot do so independently. Interventions are genuinely needed and more likely to be welcomed in such cases. In our exploration of conceptualizing PSU as IBG, we aimed to describe, explain, and predict IBG in smartphone usage.

Method

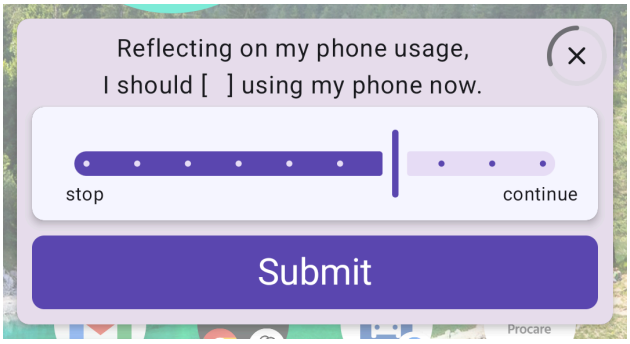


Fig. 1. The ESM question to measure intention.

We collected intention and behavior data of phone usage from 37 participants over two weeks. The intention data were collected using the experience sampling method (ESM). Participants answered whether they should stop or continue using their phones (see Fig. 1) after a random delay ranging from 5 seconds to 30 minutes following screen unlock events, for up to 16 times per day. The primary behavioral measure was the remaining session length after answering the ESM question. We invited 19 participants to participate in data engagement interviews, where we showed them visualizations reconstructing their phone usage and asked them to provide details, context, and explanations for observed intention-behavior gaps.

Results

We propose to operationalize PSU through IBG on a continuous scale (see Fig. 2, right) in contrast to a binary classification (see Fig. 2, left).

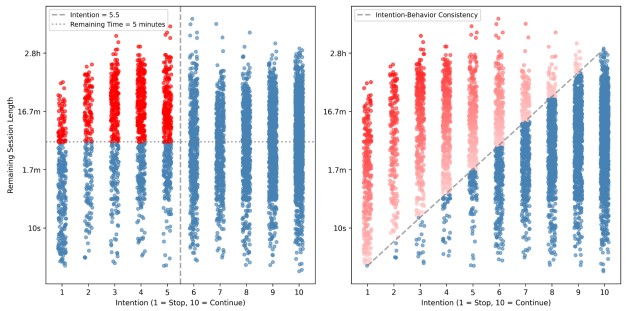


Fig. 2. Strip plots showing remaining session length across usage intention.

We normalized the highly skewed behavior data using non-linear transformations, standardized intention and behavior data to calculate the IBG, and then standardized the IBG (see Fig. 3 for the full pipeline).

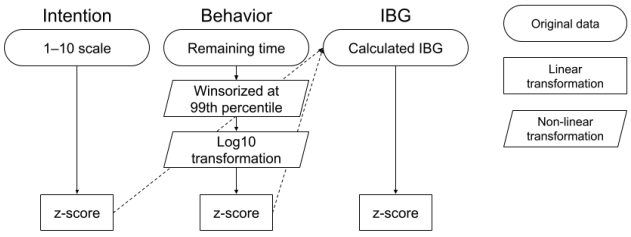


Fig. 3. The data transformation and calculation pipeline

To explore which factors are associated with IBG, we fitted linear mixed-effects models to the intention, behavior, and IBG data. We found that gender, time, app category, and input interactions can explain the variance in IBG. The results are generally consistent with previous research on PSU, providing support for the validity of conceptualizing PSU as IBG. For example, previous studies have found that PSU can cause sleep loss. We also observed a larger gap during sleep hours. Besides, from the interviews, we learned that loss of sleep may lead to PSU. Many participants use their phones late at night because they cannot fall asleep and hope that scrolling on their phones can make them feel tired and sleepy.

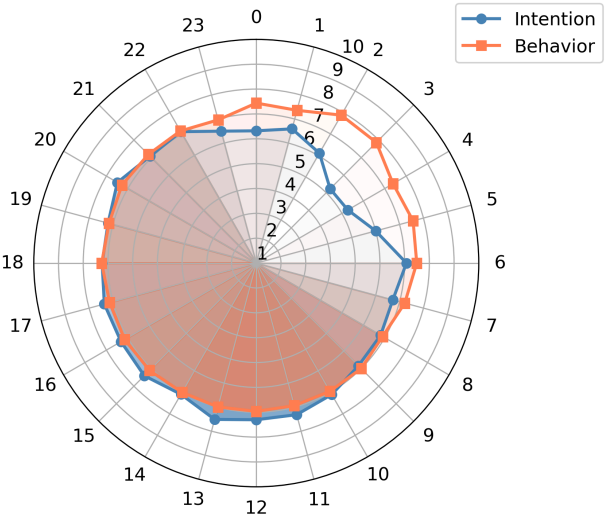


Fig. 4. Intention-behavior gap through 24 hours

To inform the design of just-in-time interventions, we need to predict IBG in real-time during phone usage. We trained random forest regression models to predict intention, behavior, and IBG. Intention was predicted most accurately with only personal data, whereas behavior and IBG were predicted most accurately with the combination of personal and global data.

Table 1. The best-performing models for intention, behavior, and IBG

Target	Model	MAE (Z-score)	MAE (original scale)
Intention	Personal	0.528	1.514 (1–10 scale)
Behavior	Combined	0.660	10.3 minutes
IBG	Combined	0.599	—

Discussion

Our findings can inform the design of future intervention tools with better timing and adaptive intensity. We propose two components as improvements over current interventions: triggering by IBG and continuous intensity. Although we expect the combination of those two components to work best, it is possible to decouple them and apply just one component.

The IBG-based trigger can be integrated with most existing interventions by replacing the current trigger, if there is one, with the predicted IBG, without altering the intensity schema. For example, if an intervention sends a blocking page whenever the phone is unlocked, we can change the trigger from unlocking the screen to the predicted IBG reaching a threshold.

Most current interventions take an all-or-nothing approach. We believe the intervention intensity should be calibrated to the severity of PSU, which could be measured as session length, daily screen time, or predicted IBG, among other options. The intervention intensity could take many formats, for example, the number of digits to be typed correctly or the delay before an app can launch, the monetary reward if users abstain from using the phone for some time, the strength of vibration if users keep using their phones, the frequency of reminders, the reduced saturation of color (full reduction to grayscale), and so on.

Our next step is to empirically test the effect of intervention triggers (i.e., intention, behavior, or IBG) and intensity schema (i.e., binary vs. continuous) in assisting self-control of smartphone usage.

References

Tran, J. A., Yang, K. S., Davis, K., & Hiniker, A. (2019). Modeling the engagement-disengagement cycle of compulsive phone use. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.

Sheeran, P. (2002). Intention-behavior relations: A conceptual and empirical review. European Review of Social Psychology.

Brehm, S. S., & Brehm, J. W. (2013). Psychological Reactance. Elsevier Science.

Zheng, J., Seo, S., & Choe, E. K. (2025). From awareness to action: Ambient display and customizable attention signals for self-regulated smartphone usage. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 9(3), 1–28.

Human-Computer Interaction Lab
43rd Annual Symposium

Adaptive Ambient Display based on Intention-Behavior Gap for Self-Regulated Smartphone Usage

Jian Zheng, Eun Kyoung Choe
College of Information
{jzheng23, choe}@umd.edu

Introduction

Smartphones have become deeply embedded in everyday life, and many people use them more than they intend to. Researchers have developed various interventions to help people control their usage, but two limitations remain:

1. Interventions are usually triggered by either behavior or intention alone, and each approach has costs. Behavior-based triggers could impede justified usage, while intention-based triggers may cause psychological reactance when users are already capable of self-regulating.
2. Interventions typically operate in an all-or-nothing manner. We argue that the problematic severity varies along a continuum; therefore, intervention strength should also vary accordingly.

In this study, we aim to develop and evaluate a novel intervention tool that uses the intention-behavior gap (IBG) as the trigger and varies its intensity continuously according to the IBG. We will test two hypotheses:

1. Interventions triggered by IBG are more effective than those triggered by behavior or intention.
2. Interventions with continuous intensity are more effective than those with binary intensity.

Method

We will integrate our previous work on ambient display and IBG to develop an adaptive ambient display system. The ambient display of the current session length could change its visibility along a continuum according to the predicted IBG (see Fig. 1).

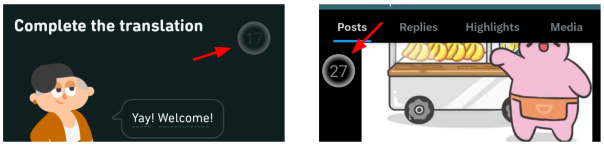


Fig. 1. Adaptive ambient display. The display on the right is more salient than that on the left.

We will compare our proposed continuos-IBG with three other intervention conditions, varying in trigger and intensity schemas.

- Continuous-behavior. The longer the ongoing use, the stronger the intervention.
- Continuous-intention. The stronger the predicted intention to stop, the stronger the intervention
- Binary-IBG. The intervention is triggered when IBG is above a threshold.

We aim to collect data from 40 participants over six weeks.

- Week 1. Additional data collection. We will use the additional data to personalize the models trained in our previous study.
- Week 2. Baseline without intervention.
- Weeks 3–6. Intervention of four conditions, counterbalanced.

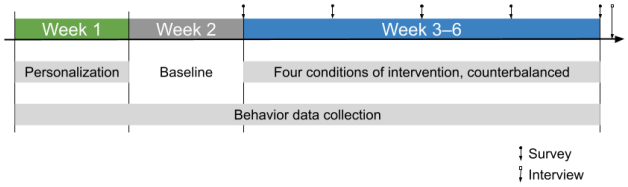


Fig. 2. Study procedure

Data collection includes experience sampling method responses (Week 1 only), phone usage behavior, weekly surveys measuring participants' autonomy and capacity in controlling smartphone usage during the intervention period, and an exit interview. This is an ongoing project, and results are not yet available.

References

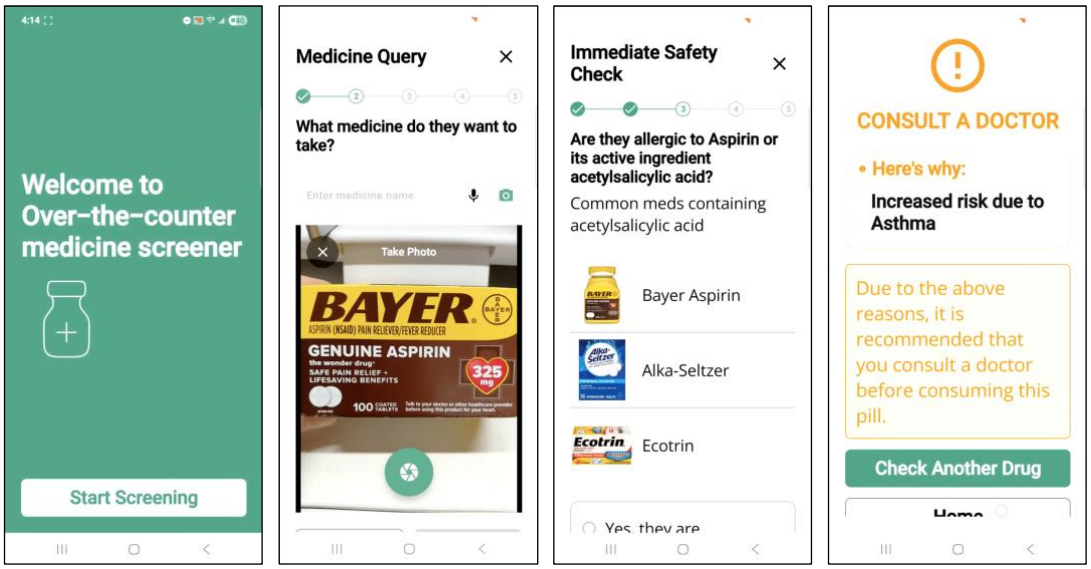
Zheng, J., Seo, S., & Choe, E.K. (2025). From awareness to action: Ambient display and customizable attention signals for self-regulated smartphone usage. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 9(3), 1–28.

Zheng, J., Choe, E.K. (under review). To Stop or Not to Stop: Exploring the Intention-Behavior Gap in Smartphone Usage.

Human-Computer Interaction Lab
43rd Annual Symposium

Aidara: Supporting Over-the-Counter Medication Decision-Making for Older Adults and Caregivers

Mengying Li[1], Harshitha Kommaraju[1], Seongjae Bae[2], Jarrett G.W. Lee[1], Siqiao Austin Ao[1], Bongshin Lee[2], Eun Kyoung Choe[1]
[1]University of Maryland, [2]Yonsei University
Contact: myl99629@umd.edu



(a) (b) (c) (d)

Figure 1. Prototypes of Aidara's mobile app. After initiating the screening process (a), users can begin a medication query using text, voice, or camera input (b). Based on the Drug Facts label, the app generates required screening questions, such as allergy checks (c), and provides a recommendation informed by users' responses: Consult a Doctor (d), Consult a Doctor/Pharmacist, Safe For Use, or Do Not Use.

INTRODUCTION

Over-the-counter (OTC) medications are widely used for everyday health needs. Unlike prescriptions, OTCs often require users to evaluate safety-relevant information by themselves. These decisions can be especially challenging for older adults as well as informal caregivers of individuals with cognitive impairments, as they may need to interpret complex medication labels while accounting for chronic conditions, polypharmacy, and another person's health context. The FDA Drug Facts label provides standardized safety information for OTC medications, but its information is often dense and difficult to translate into concrete decisions. Rather than expecting users to read and interpret the full label, mobile decision-support systems may help by transforming label content into structured, step-by-step screening questions. In this project, we examined how AI-supported decision guidance for OTC medications might be delivered through a mobile-based system. We built an interactive prototype that translates Drug Facts label content into structured decision logic, and used it to explore how older adults and caregivers perceive and respond to this form of OTC decision support.

METHOD

We developed Aidara, an Android-based mobile app prototype designed to support OTC medication decision-making workflows (Figure 1). Users could initiate a medication query through three input modalities: *text, voice,* and *camera*. Voice and camera inputs used LLM-based prompting to extract medication names, while downstream safety guidance was based on structured FDA Drug Facts label content. We manually translated Drug Facts labels into

rule-based decision trees for four common OTC medications: Bayer® (aspirin), Tylenol® (acetaminophen), Rolaids® (calcium carbonate and magnesium hydroxide), and Benadryl® (diphenhydramine). These decision trees encoded key eligibility questions related to intended use, ingredient allergies, health conditions, and concurrent medications.

We conducted an in-lab/in-home user study with 19 participants aged 50-80, including 7 caregivers. Participants first received a guided tutorial on the app's input modalities and question-and-answer workflow. They then completed four scenario-based OTC decision-making tasks using Aidara. For each task, participants used the app to determine whether the medication was safe to use, completing the task when they reached the app's final recommendation. During the task session, they were encouraged to think aloud while the researcher observed their interactions. After the tasks, participants completed a post-task survey on usability and decision confidence, followed by a semi-structured debriefing interview.

FINDINGS

Participants generally viewed Aidara as a useful tool for OTC medication decision-making. Many described the app as easy to follow and less difficult than expected. The step-by-step question-and-answer workflow helped participants focus on decision-relevant factors.

Error Analysis

Task interactions revealed several categories of errors (Table 1). Medication name alignment errors occurred when participants were unsure what forms of input the system could recognize, leading them to enter brand names, descriptive phrases, or misspelled medication names. Medication attribute selection errors occurred when participants selected an incorrect formulation, variant, or strength. We also observed interaction-level breakdowns, such as accidental taps on the microphone or camera icons, and scenario understanding errors, where participants misinterpreted app questions or made unsupported assumptions about the persona's medical context.

Flexible Input Needs

Typing was familiar and often served as the default because it felt controlled and routine, but participants recognized that typing could introduce errors, especially for medication names that were difficult to spell. Voice input was seen as helpful for reducing manual effort, particularly for users who may have difficulty typing, but participants raised concerns about background noise, pronunciation, accents, and speaking aloud in public settings. Camera input was often viewed as accurate and convenient because it reduced the need to spell or pronounce medication names. However, participants expected camera input to do more than identify the medication name, such as recognizing the exact product, formulation, or barcode.

Table 1. Error categories identified during the task session.

Category	Subcategory	Example
Medication name alignment errors	Non-recognizable input	"Bayer," "genuine aspirin"
	Typo-induced failure	"Asprin," "Tylelnol"
	Misaligned input strategy	Barcode scanning
Medication attribute selection errors	Formulation mismatch	Capsule instead of tablet
	Variant-level mismatch	"Extra Strength" vs. "Regular Strength"
Interaction-level breakdowns	Accidental UI actions	Clicked delete or microphone icon by mistake
Scenario understanding errors	Task instruction misinterpretation	Misunderstood "health condition" selection
	Uncertainty in medical context	Unsure of ingredients in their medications
	Incorrect medical inference	Mistook Claritin for a prescription

Design Implications

Our findings suggest two directions for future OTC decision-support systems. First, more sophisticated multimodal input could help capture more than medication names. Specifying product variants, formulations, and strengths was a frequent source of error, and tools such as barcode scanning, package recognition, or label-image extraction could help map users directly to the exact product and reduce manual selection errors. Second, future systems should consider how output information is presented. Participants sometimes overlooked reasons for recommendations or potential risks, suggesting a need for clearer visual hierarchy, concise risk summaries, icons, audio readouts, or interactive "why" explanations.

ACKNOWLEDGEMENTS

This work is supported by the Maryland Center of Excellence in Regulatory Science and Innovation (M-CERSI) and the National Institute on Aging grant P30AG073107-05.

Human-Computer Interaction Lab
43rd Annual Symposium

Older Adults Personalizing Wearable Foundation Models

Hossein Khayami[1], Sungjin Hwang[2], Xiyang Wu[1], Eshed Ohn-Bar[3], David E. Conroy[4], Amanda Lazar[2], Eun Kyoung Choe[2], Hernisa Kacorri[2]
[1]Department of Electrical and Computer Engineering, UMD,
[2]College of Information, UMD, [3]Boston University, [4]University of Michigan
Contact: khayami@umd.edu

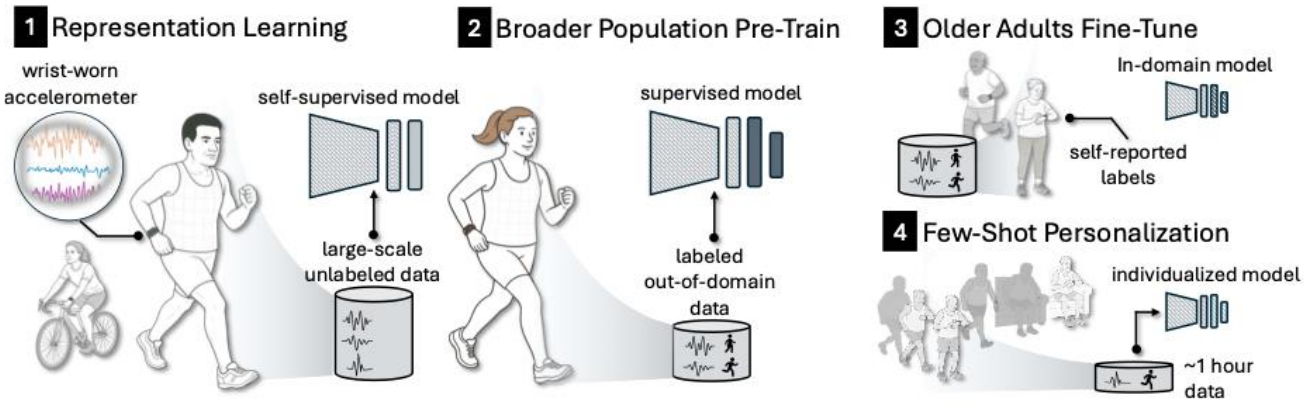


Figure 1- How transferable are wearable foundation models to older adults, and to what extent can they be customized given limited user data? We follow these four stages of personalization to answer this question.

OVERVIEW

Wearable activity trackers can benefit everyone, including older adults, by enabling self-monitoring and supporting overall well-being. However, most trackers are not trained for older users. Personalization could bridge this gap, yet training a model from scratch would place an impractical burden on older adults, who would need to collect and label a large amount of data.

METHODS

We investigate the benefits of model pre-training and personalization with limited individual data for this population, which has not been studied previously. We present a multi-stage personalization framework that leverages self-supervised learning on large-scale unlabeled sensor streams, medium-sized labeled activity data from younger populations, and our small size in-the-wild older adults activity data. We used a ResNet architecture for our representation layer and a Multi-Layer Perceptron (MLP) and a Transformer as our classification heads on top of the ResNet representation. The framework enables ablation studies, allowing us to evaluate the contribution of each training stage from self-supervised pretraining to few-shot personalization.

RESULTS

Personalization

Our results show that using this system, older adults can personalize their activity tracker with as few as a single example per activity, achieving an F1 score of 0.7. They can improve the performance to 0.75 with 50 minutes of data per activity. We see that the ResNet+Transofrmer have better off-the-shelf baseline performance, but the ResNet+MLP achieves a better personalized results with as few as single shot in our full pipeline.

Figure 2- Few-shot performance for our two architectures.

Error Analysis

Error analysis reveals that the model performs best on well-segmented, isolated activities (e.g., ``walking" or ``standing") but struggles with composite activities that involve ambiguous postures such as ``gardening" or ``stretching," where multiple sub-activities or posture transitions obscure clear labeling.

PAPERS

1. Hossein Khayami, Sungjin Hwang, Xiyang Wu, Eshed Ohn-Bar, David E. Conroy, Amanda Lazar, Eun Kyoung Choe, Hernisa Kacorri. Older Adults Personalizing Wearable Foundation Models: Feasibility and Challenges *The 32nd Annual International Conference on Mobile Computing and Networking*. Under submission. (2026).

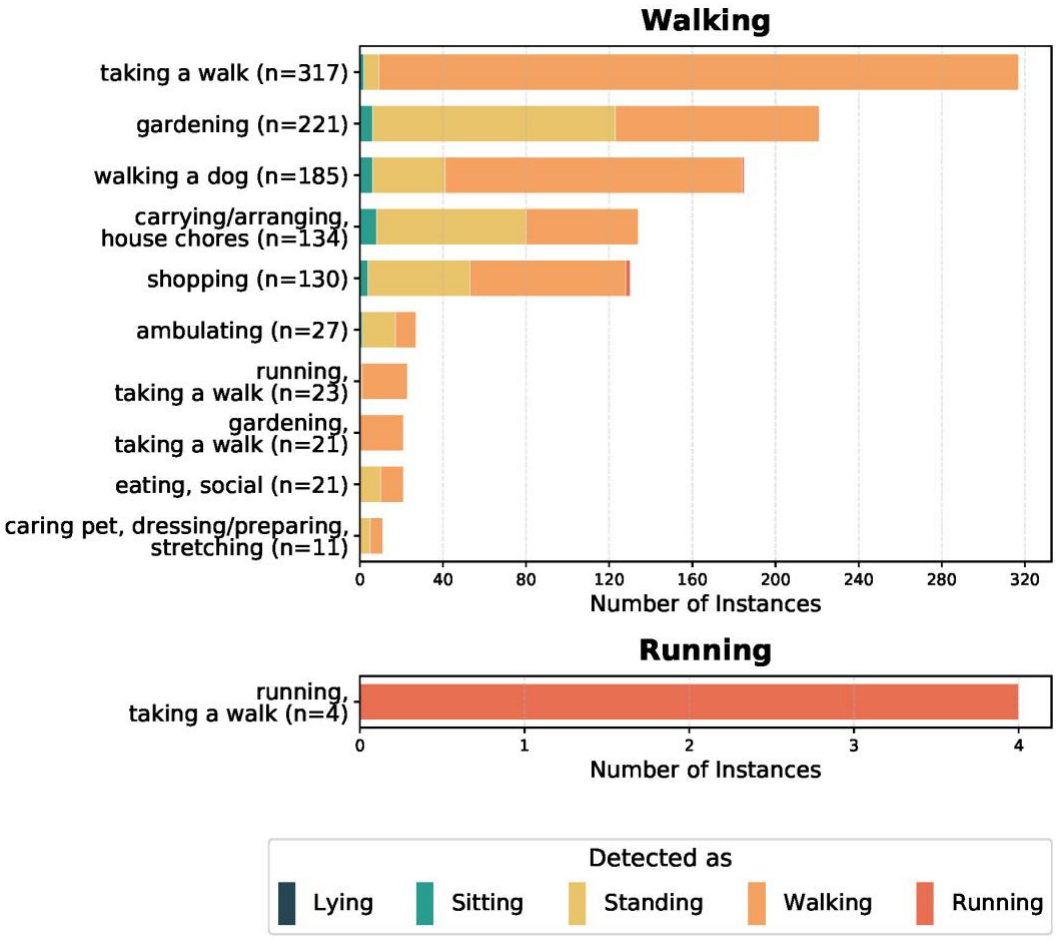


Figure 3- Detected postures using full pipeline personalized models across top 10 reported semantic activities in *Walking* and *Running* instances. The colors represent the detected activity.

Human-Computer Interaction Lab
43rd Annual Symposium

Co-Designing Teachable Activity Trackers with Older Adults with Cognitive Impairment

Yiwen Wang, Kibron Tesfatsion, Hernisa Kacorri, Amanda Lazar, Eun Kyoung Choe
College of Information, University of Maryland, College Park
Contact: yw7615@umd.edu
Project Website: https://mymove-collective.github.io/

INTRODUCTION

Tracking technologies often underserve older adults with cognitive impairment, as they are rarely included in the design and data collection process—leaving their accessibility needs and activity patterns unaccounted for. Machine teaching [1]—where non-expert users create training examples and labels—offers promise for personalization, yet remains largely inaccessible to people with cognitive impairment. The goal of this study is to inform the design of a teachable activity tracker that enables older people with cognitive impairment to train and fine-tune their personalized models. As part of a broader project aimed at supporting older adults' diverse and individualized active lifestyles, this work examines *how to enable meaningful and accessible engagement with teachable systems for activity labeling among older adults with cognitive impairment*.

METHODS

We conducted 9 co-design workshops with older adults diagnosed with dementia or mild cognitive impairment, and their care partners. We provided design scaffolds to help participants envision using teachable interfaces to label and track their seated activities. Our prior work shows that seated activities occupy a nuanced position for older adults: they can encompass a wide range of meaningful experiences (e.g., doing puzzles or computer work [2]) while also being viewed as behaviors to reduce due to prolonged sedentary risks. This tension makes seated activities a particularly rich context for personalized tracking.

FINDINGS

Through codebook thematic analysis [3], we found that participants saw value in approaching machine teaching as an evolving process, given fluctuations in their condition (P2,5). Five (P3-7) valued capturing contextual nuances of activities, especially seated ones, whereas three (P1,8,9) did not see machine teaching as personally relevant. Our findings revealed a range of preferred labeling mechanisms, diverse accessibility needs related to cognition, vision, hearing, speech, and emotional state. For example, participants found in-the-moment labeling—collecting and verifying labels while seated or immediately after standing up—more feasible for managing short-term lapses. They also envisioned machines compensating for their long-term memory by adapting to changes in their colloquial language use over time (P2) or providing repetitive exercises that help establish unconscious associations between activities and labeling (P3,7). Additionally, care partners can support labeling when participants welcome their involvement—e.g., by marking end times (P6), assisting with word-finding, reviewing labeled data, and nudging their counterparts to label (P7). However, some felt uncomfortable with care partners or caregivers labeling their seated activities (P2,3), while others need support outside of their current care networks (P5, C8).

CONCLUSION

Our findings show how cognitive impairment shapes interactions with teachable machines, the strategies participants proposed to make labeling manageable, and the role care partners and others can play in sustaining engagement. We translate these insights into design considerations for teachable activity trackers that flexibly accommodate short- and long-term memory loss, fluctuating sensory needs, and care dynamics, while accounting for the contextual variations of seated activities.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation awards #1955568.

REFERENCES

1. Xiaojin Zhu, et al. An overview of machine teaching. *arXiv preprint arXiv:1801.05927*. (2018).
2. Yiwen Wang, et al. Redefining Activity Tracking Through Older Adults' Reflections on Meaningful Activities. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems.* (2024).
3. Virginia Braun and Victoria Clarke. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative research in psychology 18.3.* (2021).

Human-Computer Interaction Lab
43rd Annual Symposium

Genetic Ancestry Testing, Technoracial Authenticity, & Codesigned Assemblages

Chioma Agoh, Reza Amanahu, Silvana Martinez, Jayla Parrott, Daniel Radvany, and S. Nisa Asgarali-Hoffman
College of Information
Corresponding Contact: sasgara1@umd.edu

OVERVIEW

As algorithmic technologies are increasingly embedded in our daily lives, it is vital to ensure that all people are empowered to understand how algorithmic models analyze and use data, and how this can impact our identity formation. Project GATTACA addresses this need. The primary goal of this project is to bring together people that have taken genetic ancestry tests and INFO students to collaborate on designing a new set of interfaces for displaying genetic ancestry testing data that challenges the algorithmic (i.e. quantitative, reductionist) logics involved in technologically-determined racial authentication. The guiding research questions for this project are:

1. How do users of commercial genetic ancestry tests understand/conceptualize the relationship between algorithms, genetic data, identity, and privacy?
2. How can INFO design students engage with the public in co-design workshops to critique/contest/reimagine algorithmically-driven genetic data practices and design alternative, ethical, and creative engagements with genetic ancestry data?

METHODS

By bringing together commercial genetic ancestry test (GAT) users and INFO students in a series of co-design workshops, we work to discover how users understand/conceptualize identity through the lens of their genetic data reports. INFO students use the insights gathered from codesign workshops to engage in an iterative design process with GAT users that eschews industry-standard genetic data representations for explorative, creative, and storytelling avenues to identity sense-making. In these designs we deprioritize a clean, modern design aesthetic to instead emphasize friction and narrative in exploring the relationship between genetic data and identity. We utilize game design, GIS, and static web design as alternative methods for communicating GAT results and broader contextual information. This project is ongoing.

Figure 1 – Thumbnail sketch of interface.

NEXT STEPS

The project will culminate with the collection of our alternative designs, research findings, participant quotes, and provocations into a publicly accessible assemblage, an online module designed to not just educate the public but encourage their active participation in technology design while promoting algorithmic literacy. These imagined interfaces seek to challenge technoracial authenticity, asking GAT users to critically engage with the ambiguities, limitations, and discrepancies inherent in current test result interfaces.

UNDERGRADUATE RESEARCH OPPORTUNITY PROGRAM (UROP)

This study is conducted and funded as part of the Undergraduate Research Opportunity Program (UROP), a new initiative directed by Dr. Eun Kyun Choe that allows students in the College of Information to collaborate with doctoral students to gain hands-on research experience. In this study, UROP students have equal footing in all research design decisions, ensuring they experience every level of the design and implementation process. UROP is a unique program that has allowed students to contend with epistemologies and methods outside of their core curriculum while engaging in a research topic pertinent to their interests.

Human-Computer Interaction Lab
43rd Annual Symposium

Exploring Integration of Technology Into Play Therapy Through Participatory Design

Nitya Jani, M.S.[1], Urvi Ashturkar, M.S.[1], Anushka Verma, M.S.[1], Rhea Stokes, M.S.[1], Anushka Chimote, M.A.[1], Evelyn Lee[1], Elizabeth Bonsignore, Ph.D.[1], Rachel Altvater, Psy.D. RPT-S[2]

[1]College of Information
njani1@umd.edu, urviash@umd.edu, ebonsign@umd.edu

[2]Creative Psychological Health Services
dr.rachelaltvater@gmail.com

OVERVIEW

Play therapy is a form of expressive psychotherapy that uses play as a language for children [4] to practice adaptive social emotional skills and behaviors and regulate challenging emotions [6]. Play is an integral part of human well-being [1, 2], serving as the basis for trained mental health practitioners, known as play therapists, to support children's psychological and emotional health [5]. Familiar, everyday toys afford children the opportunity to explicitly explore, verbalize, and role play specific themes and life events [3]. Play therapy is gaining momentum, especially now that the number of children with mental health disorders is at an all time high [7]. Also, more children are using technology as a medium for play. However, there is little research about if, and how, technology could be applied to expressive therapies. This leads to play therapists avoiding technology out of fear of its dangers and limitations. Generation Alpha is growing up with digital devices [1] and many children will have expressive needs related to technology, but play therapists are not always able to meet the children where they are at.

Our research aimed to answer the following research question:

- How might technology help play therapists and their clients work towards meaningful therapeutic goals?

We hoped to gain insight on the views and ideas of therapists and children, and how (or if) technology can benefit children therapeutically. Our research revolved around the Therapeutic Powers of Play [9], a framework that explains how play can be therapeutic and emotionally beneficial to children. We aimed to discover if the Therapeutic Powers of Play emerged during technology-based play sessions between children and therapists.

METHODS

Figure 1: Children and an adult partner design with Bags of Stuff.

We started by exploring the needs and wants of both therapists and children through co-creation of digital playrooms using art materials in the Participatory Design (co-design) framework. In particular, we had participants use Bags-of-Stuff, a low-fidelity prototyping technique that includes arts and crafts materials.

In the first session, we conducted co-design with children (n=9, ages 5-12). In the second series of sessions, we conducted co-design with various play therapists (n=6). Participants designed ideal playspaces and shared their desires, perspectives, and ideas. The researchers then created virtual playspaces on a digital whiteboard (FigJam) based on the children's and therapists' models. Next, the play

therapists completed a survey about their clinical context and attitudes and behaviors related to technology. Finally, we conducted an observational session with therapists and children together (children: n=8, ages 5 -11, therapists: n=4). The therapists and children played together using the virtual playspaces, followed by reflection in group discussion.

PRELIMINARY FINDINGS

We found that some therapeutic methods and themes appeared during virtual play therapy sessions. Therapists discovered ways of implementing their skills and approaches into a session that was formatted differently than they are used to.

Shifting Control

Therapists experienced reduced control due to a lack of familiarity and competency with the virtual environment. The children guided the therapists with confidence and empathy, which the therapists appreciated.

Challenges with Visibility/Awareness

Due to lack of visual cues, therapists struggled to track children's actions, leading to more questioning and concern about disrupting play.

Maintained Personal Connection

Children co-attented to digital content, even through separate devices. Children showed therapists their screens and physically co-interacted with their device.

Figure 2: Child and therapist play in a digital playspace.

We also observed design implications for technology-based play, including designing for digital-native interactions and supporting certain favored interactions. The adults had designed spatial play environments that mimicked typical in-person playrooms. However, the children treated the whole environment as an interactive object; It was not necessary for the designed objects and space to replicate an actual playspace. The device itself became the playspace.

ACKNOWLEDGEMENTS

Thank you to our co-design partners - Rhea Stokes, Evelyn Lee, Anushka Verma & Anushka Chimote. Photo credits: Dr. Rachel Altvater. This research was financially supported by Reality Labs Research & by Arts for All, University of Maryland.

REFERENCES

1. Ginsburg, K. and the Committee on Communications, and the Committee on Psychosocial Aspects of Child and Family Health. (2007). The Importance of Play in Promoting Healthy Child Development and Maintaining Strong Parent-Child Bonds. Pediatrics 119, 1 (Jan. 2007), 182–191. doi:10.1542/ peds.2006-2697
2. Huizinga, J. (2016). Homoludens: a study of the play-element in culture. Angelico Press, Kettering, OH.
3. Landreth, G. (2012). The Playroom and Materials. In Play Therapy: The Art of the Relationship (3 ed.). Routledge, New York, 149–173. doi:10.4324/ 9780203835159
4. Landreth, G. (1993) Child-centered play therapy. Elementary School Guidance & Counseling 28, 1(1993), 17–29. Place: US Publisher: American School Counselor Association.
5. Lilly, J.P et. al, [n.d.]. Play Therapy Makes a Difference. https://www.a4pt.org/page/ PTMakesADifference.
6. Pedro-Carroll, J. and Jones, S. (2005). A Preventive Play Intervention to Foster Children's Resilience in the Aftermath of Divorce. Empirically based play interventions for children. American Psychological Association, Washington, DC, US, 51–75. doi:10.1037/11086-004
7. Rise of Mental Health Issues in Children. (2022). Childrens.com; Children's Health. https://www.childrens.com/health-wellness/rise-of-mental-health-issues-in-children

HCIL Honors and Awards

On the following pages, we celebrate some of the awards and honors that HCIL faculty and students have received in the last year.

Promotion & Tenure	
Ge Gao	Received tenure in 2026 (INFO)
Caro Williams-Pierce	Received tenure in 2026 (INFO)
Huaishu Peng	Received tenure in 2026 (CS)
Sheena Erete	Promoted to full professor in 2026 (INFO)
Eun Kyoung Choe	Promoted to full professor in 2026 (INFO)

Awards for research papers	
Elana Blinder, Tammy Clegg, Jessica Vitak	Best Paper Honorable Mention, "Creating and evaluating privacy and security micro-lessons for elementary school children," ACM CSCW
Connie Siebold, Susannah Paletz, and other	Paper awards for DEI and Impact. "Disability meets modality: A sociotechnical approach to team meetings," ACM CSCWX
Huaishu Peng, Zeyu Yan, SuHwan Hong and others	Best Paper Award and Special Recognition for Sustainability in Demo, “DissolvPCB: Fully Recyclable 3D-Printed Electronics with Liquid Metal Conductors and PVA Substrates,” ACM UIST
Stephanie Valencia and others	Best Poster, “CustomSight: Enhancing LLM-Powered Visual Assistance for Blind Individuals using Goal-Directed Dynamic Filters,” ACM UIST
Stephanie Valencia and others	Best Demo Jury Choice Award, “Why So Serious? Exploring Timely Humorous Comments in AAC Through AI-Powered Interfaces”, ACM CHI
Zeyu Yan	Inaugural Ben Shneiderman Outstanding Thesis Award in Human-Computer Interaction (HCI)

Other Awards & Achievements	
Jessica Vitak	Elected Steering Committee Chair, ACM CSCW
Susannah Paletz and Connie Siebold	UMD Do Good Award, "Team Meeting Accessibility on Campus"

Jessica Vitak	ACM Distinguished Member Inductee
Eun Kyoung Choe	ACM Distinguished Member Inductee
Michelle Mazurek	Technical Program Chair, IEEE S&P 2027
Michelle Mazurek	Track Chair, Usability and Measurement Track, ACM CCS 2026
Huaishu Peng	Subcommittee Chair, CHI 2026
Eun Kyoung Choe	Technical Program Chair, UbiComp 2026
Cody Buntain	Award, MURI on Offline-to-Online influence
Cody Buntain	Co-PC Chair for ICWSM 2026

HCIL Student Graduation

Finally, we celebrate the many HCIL students who have graduated over the last year. Students are an integral part of the lab's success, and these students have worked with faculty on a range of important research projects in recent years. Below is a partial list. We offer congratulations and good luck to all students graduating who have been a part of the HCIL during their studies!

Student Name	Advisor	Grad Date	Degree
Utkarsh Dwivedi	Hernisa Kacorri	May 2026	PhD, INFO
Sneha Gathani	Leo Zhicheng Liu	Dec 2025	PhD, CS
Sunyup Park	Jessica Vitak	May 2026	PhD, INFO
Md Main Uddin Rony	Naeemul Hassan	May 2026	PhD, INFO
Nathan Reitinger	Michelle Mazurek	June 2025	PhD, CS
Rachel Wood	Jonathan Lazar	May 2026	PhD, INFO
Zeyu Yan	Huaishu Peng	May 2026	PhD, CS
Yongle Zhang	Ge Gao	May 2026	PhD, INFO
Kazi Tasnim Zinat	Leo Zhicheng Liu	May 2026	PhD, CS
Xinyi Hu	mols sauter	May 2025	HCIM
Lei Mao	Stephanie Valencia	May 2025	HCIM
Jia-Yun (Joan) Li	Stephanie Valencia	May 2026	HCIM
Phillip Thompkins	Caro Williams-Pierce	December 2025	HCIM