CS-CAN | INFO-CAN

2026 ABSTRACTS

CAN-CWIC 2026

THE CANADIAN CELEBRATION OF WOMEN IN COMPUTING

THE COMPLETE COLLECTION

Submissions from CAN-CWIC

WEST EAST CENTRAL

INSIDE THIS COLLECTION

Posters

Visual presentations of research projects and case studies.

Presentations & Lightning Talks

Sharing research, ideas, and discoveries with the community

CANADIAN CELEBRATION OF WOMEN IN COMPUTING | November 20-21, 2026

CAN-CWIC 2026 - University of Waterloo, Waterloo, ON

CAN-CWIC East 2026 - Halifax, NS

CAN-CWIC West 2026 - Simon Fraser University & University of British Columbia, Vancouver, BC

CAN-CWIC 2026

University of Waterloo

Abstracts for Presentations and Lightening Talks

Session: Artificial Intelligence (AI)

Investigating How Different Fractal Properties Affect Exploration and Exploitation Mechanics in Search Algorithms

Jasleen Baher, Athabasca University, jbaher1@learn.athabascau.ca
Dr. Jason Bernard, Athabasca University, cbernard@athabascau.ca

Format: Lightning Talk

ABSTRACT

Fractal decomposition works by dividing a search space iteratively into regions using a fractal shape. By partitioning and sampling the regions, search algorithms can be induced to exploit promising regions or search in other unexplored regions. Fractal decomposition often uses simple shapes, like a hypersphere. The specific fractal structure determines which regions are emphasized via its geometry; hence, a simple structure potentially provides poor separation by creating naïve regions. Other fractal shapes have more complex properties; however, the effect of these properties on the search behaviour is not yet well understood.

Most challenging optimization problems also involve high-dimensional deceptive search spaces, where local optima can trap the search algorithm [1]; hence, understanding the relationship between fractal structure and search mechanisms as dimensionality increases provides an opportunity for improving search performance. This research investigates how different properties of fractal shapes influence their use as exploration and exploitation mechanisms, especially when extended to higher-dimensional search spaces. The long-term goal is to use fitness-landscape features to adaptively optimize fractal selection.

This research selects different fractal structures based on complexity, space-filling behaviour, locality, coverage, and their ability to generalize to higher dimensions. The selected fractals will be evaluated on fitness-landscape features for their effect on search performance. The experiments will examine whether particular fractal properties are better suited to specific fitness-landscape features and whether these relationships consistently influence search performance across different optimization problems. This is done to observe how effectively they create separable regions for exploration and exploitation, thereby allowing an algorithm to focus adaptively on either behavior. Adaptivity allows a search algorithm to escape detected local optima via exploration while otherwise concentrating effort toward promising areas [2]. A better understanding of these relationships works toward optimal fractal selection, thereby improving the effectiveness of fractal-based search methods across a range of different optimization problems.

A selection of fractals covering different properties has been integrated into the Zellij optimization framework to generate, evaluate, and compare search points across successive fractal iterations under comparable search conditions. This initial implementation provides a basis for comparing additional fractal structures and examining how their properties affect performance across different search conditions. These comparisons help identify relationships between fractal characteristics and fitness-landscape features, ultimately moving fractal selection from a fixed parameter toward a landscape-informed search strategy.

Keywords: Fractal-Based Decomposition, Exploration and Exploitation, Fitness Landscape, Optimization Algorithms, Fractal Search

References

[1] Das, R., & Whitley, L. D. 1991. The Only Challenging Problems Are Deceptive: Global Search by Order-1 Hyperplane Sampling. In R. K. Belew & L. B. Booker (Eds.), Proceedings of the Fourth International Conference on Genetic Algorithms, pp. 166–173. Morgan Kaufmann.

[2] Goldberg, D. E., & Richardson, J. 1987. Genetic Algorithms with Sharing for Multimodal Function Optimization. In J. J. Grefenstette (Ed.), Proceedings of the Second International Conference on Genetic Algorithms and Their Applications, pp. 41–49. Lawrence Erlbaum Associates.

CoMix: Type-Aware Oversampling for Tabular Imbalanced Regression

Jiaxin Yang; Western University; jyan682@uwo.ca
Mike Domaratzki; Western University; mdomarat@uwo.ca

Format: Lightning Talk

ABSTRACT

Data imbalance has been widely studied in classification, where underrepresented cases are typically identified by minority-class labels. In regression, however, target values are continuous, making rare regions less straightforward to identify and address. When these regions contain too few examples, machine learning models may struggle to predict their outcomes accurately. [1] Such imbalances arise in applications including housing, insurance and environmental forecasting. Addressing this imbalance is particularly challenging in tabular data, where generating additional examples requires accounting for relationships among numerical features, categorical features and target values. [2] One way to address this challenge is oversampling, which increases the representation of underrepresented target regions, we introduce CoMix, a type-aware oversampling framework that emphasizes underrepresented target values while generating realistic combinations of numerical and categorical features. Evaluated on 15 public regression datasets against 10 baseline methods, CoMix outperforms state-of-the-art oversampling methods on both dense and rare target values, demonstrating its effectiveness for addressing the challenge of imbalanced regression.

In this talk, we will consider an example of house-price prediction in a city where most houses sell for $200,000 to $600,000, but only a few exceed $1.5 million. A model trained on this data has limited examples of expensive properties to learn from. Simply duplicating these houses adds little diversity for the machine learning algorithm to learn from, while randomly combining their characteristics may create implausible examples. For instance, generating a $1.8 million property involves more than assigning a high price: its numerical features, such as floor area and number of bedrooms, should be consistent with its categorical features, such as property type. The goal is therefore to generate diverse examples of rare properties while maintaining realistic relationships among their characteristics.

CoMix approaches this problem by first generating target values through relevance-weighted sampling, favoring rare observations and introducing small variation to create new continuous target values. [3] To construct complete training examples, CoMix then generates different combinations of numerical features corresponding to each target value. For instance, given a generated house price near $1.8 million, the model can produce different combinations of floor area, bedroom count and building age. CoMix then assigns categorical features, such as property type, by drawing on patterns observed in similar real houses. Together, these components provide a unified oversampling approach to generating reliable synthetic data for imbalanced regression while accounting for the distinct characteristics of numerical and categorical features.

Keywords: Imbalanced Regression; Tabular Data Generation; Conditional Diffusion Models; Copula Modeling; Data Augmentation

References

[1] Ribeiro, R. P., & Moniz, N. 2020. Imbalanced Regression and Extreme Value Prediction. *Machine Learning*, 109(9–10), 1803–1835.

[2] Kotelnikov, A., Baranchuk, D., Rubachev, I., & Babenko, A. 2023. TabDDPM: Modelling Tabular Data with Diffusion Models. *Proceedings of the 40th International Conference on Machine Learning*, 17564–17579.

[3] Branco, P., Torgo, L., & Ribeiro, R. P. 2017. SMOGN: A Pre-Processing Approach for Imbalanced Regression. *Proceedings of the First International Workshop on Learning with Imbalanced Domains: Theory and Applications*, 36–50.

Positive-Unlabeled Learning for Weakly Supervised Placental Whole-Slide Image Analysis

Tünde Lapohos, Carleton University, tundelapohos@cmail.carleton.ca
Adrian D. C. Chan, Carleton University, adrianchan@cunet.carleton.ca

Format: Lightning Talk

ABSTRACT

Cardiovascular disease (CVD) is a leading cause of mortality among women, yet women have historically been underrepresented in cardiovascular research, contributing to gaps in CVD diagnosis and treatment [1]. Pregnancy provides an opportunity for early cardiovascular risk identification, as associated placental abnormalities can reveal underlying maternal vascular dysfunction. In particular, placental abnormalities associated with maternal vascular malperfusion (MVM) have been linked to postpartum cardiovascular risk [2]. Placental histopathology could therefore help identify women who may benefit from cardiovascular risk screening. However, automated histopathological analysis presents several challenges. Whole-slide images (WSIs) are extremely large and contain heterogeneous tissue in which diagnostically relevant lesions may be sparse and unevenly distributed. Furthermore, identification of placental lesions can be subjective, with intra- and inter-observer variability among pathologists contributing to uncertainty in diagnostic labels [3]. Deep learning models typically divide WSIs into smaller image tiles for computational analysis [4], while labels are generally available only at the slide level. However, assigning these labels to all tiles can introduce substantial label noise: an MVM-positive WSI can contain tiles without MVM-associated lesions, while localized MVM-associated regions may occur within an MVM-negative WSI.

This work investigates positive-unlabeled (PU) learning as a strategy for addressing this weak-label problem in placental WSI analysis [5]. Exhaustive expert annotation of every tile would be extremely time-consuming; however, identifying a smaller number of confidently MVM-positive regions is more feasible. These regions can provide a set of positive training examples, while the remaining tiles are treated as unlabeled rather than negative. This is important because the unlabeled set is expected to contain a mixture of true positive and negative tiles. PU learning provides a framework for learning without assuming that all unannotated tissue is MVM-negative [5].

The proposed work will investigate incorporating PU learning into deep learning models for MVM classification, emphasizing robust learning despite limited positive annotations and uncertain labels. We will also explore confident learning [6] and curriculum learning as complementary strategies to identify potential label errors and progressively incorporate less certain examples during training. We will also explore attention-based architectures, including Vision Transformers, to capture heterogeneous and spatially dispersed pathological patterns [7].

Ultimately, this work aims to determine whether PU learning can reduce reliance on exhaustive expert annotation while improving automated analysis of heterogeneous placental WSIs. More broadly, it investigates a common challenge in medical imaging: developing effective deep learning models when reliable labels are scarce, imperfect, and expensive to obtain.

Keywords: positive-unlabeled learning, weakly supervised learning, whole-slide imaging, digital pathology, deep learning, placental histopathology

References

[1] World Health Organization, "Cardiovascular diseases (CVDs)," WHO Fact Sheet, Jun. 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] S. J. Benton, E. E. Mery, D. Grynspan, L. M. Gaudet, G. N. Smith, and S. A. Bainbridge, "Placental pathology as a tool to identify women for postpartum cardiovascular risk screening following preeclampsia: A preliminary investigation," *J. Clin. Med.*, vol. 11, no. 6, p. 1576, Mar. 2022, doi: 10.3390/jcm11061576.

[3] T. Y. Khong, E. E. Mooney, I. Ariel, N. C. M. Balmus, T. K. Boyd, M.-A. Brundler, *et al.*, "Sampling and definitions of placental lesions: Amsterdam Placental Workshop Group consensus statement," *Arch. Pathol. Lab. Med.*, vol. 140, no. 7, pp. 698–713, Jul. 2016, doi: 10.5858/arpa.2015-0225-CC.

[4] P. Patnaik, A. Khodaee, G. Vasam, A. Mukherjee, S. Salsabili, E. Ukwatta, *et al.*, "Automated detection of microscopic placental features indicative of maternal vascular malperfusion using machine learning," *Placenta*, vol. 145, pp. 19–26, Jan. 2024, doi: 10.1016/j.placenta.2023.11.005.

[5] J. Bekker and J. Davis, "Learning from positive and unlabeled data: a survey," *Machine Learning*, vol. 109, pp. 719–760, 2020, doi: 10.1007/s10994-020-05877-5.

[6] C. Northcutt, L. Jiang, and I. Chuang, "Confident learning: Estimating uncertainty in dataset labels," *Journal of Artificial Intelligence Research*, vol. 70, pp. 1373–1411, May 2021, doi: 10.1613/jair.1.12125.

[7] H. Xu, Q. Xu, F. Cong, J. Kang, C. Han, Z. Liu, *et al.*, "Vision transformers for computational histopathology," *IEEE Reviews in Biomedical Engineering*, vol. 17, pp. 63–79, 2024, doi: 10.1109/RBME.2023.3297604.

FOD, FOND, and LTL Automated Planning for Narrative Generation

Nisha Simon, Queen's University, nisha.simon@queensu.ca

Format: Lightning Talk

ABSTRACT

Today we rely on intelligent generative Artificial Intelligence (AI) systems such as Large Language Models (LLMs) that automate both routine and high-level tasks for us to meet our desired goals. However, we still know relatively little about the manner in which those systems actually make their decisions and plan sequences of actions that fulfill those goals. Nevertheless, when we use these systems to generate narratives, their ability to tell a coherent, logical, relevant story can be considered a valid measure of the 'intelligence' and reliability of these language generation systems.

Story *(Narrative) Generation* is the process of creating a sequence of logical steps that are taken by an agent (the entity that takes the action) or a character in order to achieve an end goal. *Automated Planning* (or simply Planning) is a subfield of AI, and it is the logical process of thinking before acting in order for an *'agent'* to move from a given initial state of the world to a goal state, within the constraints of a specified environment or domain.

We hypothesize that the use of Automated Planning will allow LLMs to maintain coherence and believability over longer narratives. The intersection of Natural Language Generation (NLG), specifically Narrative (Story) Generation, and Automated Planning is therefore a subject with great research potential. As stated by Young et al., '*plan-based methods for narrative generation hold a key to the capabilities for rich narrative content creation*. [8]'

In numerous applications it becomes necessary to convert plans to Natural Language instructions and vice versa [3, 5, 7, 1, 2]. Planning can be applied to text generation for the purpose of creating narratives (stories), and it also has applications in various educational and entertainment fields such as '*choose-your-own-adventure*' games [6]. Natural language instructions can also be created from given plans in order to make the plan more understandable to a human user. Automated Planning can therefore be used for storytelling, and text generation can be viewed as a planning task [4].

The purpose of this research is to demonstrate how Automated Planning including FOD, FOND, and LTL, can be applied to Natural Language text generation in order to create believable and engaging narratives (stories). Planners create rigid, stilted, and mechanical but coherent output. Generative AI systems such as Large Language Models (LLMs) create creative, exciting, but incoherent language. By taking the best of both worlds, and simultaneously reducing each system's flaws, we create *TattleTale*, a novel and improved Storytelling system.

Keywords: Automated Narrative Generation, Automated Planning, Large Language Models

References

[1] Wenfeng Feng, Hankz Hankui Zhuo, and Subbarao Kambhampati. "Extracting action sequences from texts based on deep reinforcement learning". In: *arXiv preprint arXiv:1803.02632* (2018).

[2] Patrik Haslum et al. *An Introduction to the Planning Domain Definition Language*. Morgan & Claypool, 2019. ISBN: 9781627058759. URL: http://www.morganclaypoolpublishers.com/catalog_Orig/product_info.php?products_id=1384.

[3] Thomas Hayton et al. "Narrative Planning Model Acquisition from Text Summaries and Descriptions". In: *The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)*. 2020.

[4] Alexander Koller and Ronald PA Petrick. "Experiences with planning for natural language generation". In: *Computational Intelligence* 27.1 (2011), pp. 23–40.

[5] Shivam Miglani and Neil Yorke-Smith. "NLtoPDDL: One-Shot Learning of PDDL Models from Natural Language Process Manuals". In: *ICAPS'20 Workshop on Knowledge Engineering for Planning and Scheduling (KEPS'20). ICAPS*. 2020.

[6] Mark Riedl. "An Introduction to AI Story Generation". In: *The Gradient* (2021).

[7] Mauricio Steinert and Felipe Meneguzzi. "Planning Domain Generation from Natural Language Step-by-Step Instructions". In: *Proceedings of the Twentieth International Conference on Automated Planning and Scheduling (ICAPS 2020)*. 2020.

[8] R Michael Young et al. "Plans and planning in narrative generation: a review of plan-based approaches to the generation of story, discourse and interactivity in narratives". In: *Sprache und Datenverarbeitung, Special Issue on Formal and Computational Models of Narrative* 37.1-2 (2013), pp. 41–64.

CAN-CWIC 2026

University of Waterloo

Abstracts for Presentations and Lightening Talks

Session: AI Ethics

THE SYSTEM IS MANIPULATED. ARE YOU? ASSESSING THE DOWNSTREAM IMPACT OF MANIPULATED LLM DECISION SUPPORT

Yee Man Choi, University of Waterloo & Vector Institute, ymchoi@uwaterloo.ca
Abdul Basit Ali, University of Waterloo, b3ali@uwaterloo.ca
Krzysztof Czarnecki, University of Waterloo, krzysztof.czarnecki@uwaterloo.ca
Freda Shi, University of Waterloo & Vector Institute, fhs@uwaterloo.ca

Format: Lightning Talk

ABSTRACT

AI is becoming a common source of information, with 60% of U.S. adults reporting that they read AI-generated search summaries [1]. As people increasingly rely on large language models (LLMs) to evaluate evidence, compare alternatives, and make decisions, the reliability of these systems becomes critical. However, LLMs can also be manipulated to present information selectively, omit relevant evidence, or favor particular conclusions. Because AI-generated responses are often perceived as credible and persuasive [2, 3], such manipulation may influence decisions without producing obviously false information. At scale, these risks extend to advertising, propaganda, information control, and censorship [4].

Existing security research examines attacks such as data poisoning (corrupting information used by AI systems) and prompt injection (inserting malicious instructions to influence model behavior), but often focuses on whether an attack changes the AI system's output rather than whether it changes the user's final decision [5]. This distinction is important for autonomous AI agents that retrieve information and make decisions with limited human oversight [6].

We investigate these risks in retrieval-augmented generation (RAG) systems, which improve LLM responses by retrieving information from external sources before generating an answer [7]. Specifically, we examine a coordinated attack that combines manipulated retrieved information with system-level instructions designed to favor a predetermined option. We develop a controlled experimental framework to measure how these attacks influence decisions made by both humans and AI agents, while separately evaluating the contribution of each attack method.

Our key finding is that **coordinated manipulation substantially reduces initial decision accuracy**, by 56.3 percentage points among 64 human participants and by 44.4 percentage points in the main AI-agent setting. A second key finding is that **humans and AI agents respond differently** when given opportunities to reconsider or identify manipulation. Among initially incorrect decisions, agents correct their answers more frequently than humans after receiving additional evidence (85.2% versus 28.6%). Although agents also identify manipulation more frequently after disclosure (72.8% versus 46.9%), they are more likely to incorrectly flag legitimate information as manipulated (59.1% versus 21.9%). These findings demonstrate that coordinated attacks on RAG systems can meaningfully distort human and AI-agent decisions, while revealing important differences in their ability to recover from and recognize manipulation.

Keywords: Retrieval-Augmented Generation (RAG), human-AI decision-making, LLM auditing, AI Governance

References

[1] J. Gottfried, W. Bishop, M. Anderson, M. Faverio, E. Park, and C. McClain, "Americans' views on ai chatbots, smart devices and ai's impact," Jun 2026.

[2] E. Dönmez and A. Falenska, ""I understand your perspective": LLM persuasion through the lens of communicative action theory," in Findings of the Association for Computational Linguistics: ACL 2025 (W. Che, J. Nabende, E. Shutova, and M. T. Pilehvar, eds.), (Vienna, Austria), pp. 15312–15327, Association for Computational Linguistics, July 2025.

[3] Y. Zhang and R. Gosline, "Human favoritism, not ai aversion: People's perceptions (and bias) toward generative ai, human experts, and human–gai collaboration in persuasive content generation," Judgment and Decision Making, vol. 18, p. e41, 2023.

[4] S. Ball and P. Hackemann, "Position: The alignment community is unintentionally building a censor's toolkit," in Forty-third International Conference on Machine Learning Position Paper Track, 2026.

[5] W. Yang, X. Bi, Y. Lin, S. Chen, J. Zhou, and X. Sun, "Watch out for your agents! investigating backdoor threats to llm-based agents," in Proceedings of the 38th International Conference on Neural Information Processing Systems, NIPS '24, (Red Hook, NY, USA), Curran Associates Inc., 2024.

[6] OpenClaw, "Openclaw - personal ai assistant." https://openclaw.ai/, 2026. Personal AI assistant framework. Accessed: 2026-04-06.

[7] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive nlp tasks," in Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS '20, (Red Hook, NY, USA), Curran Associates Inc., 2020.

Same Question, Different Answer? A Cross-Cultural Evaluation of LLM Privacy and Security Advice

Shanza Shanza, University of Waterloo, sshanza@uwaterloo.ca
Juliann Zhu, University of Waterloo, j478zhu@uwaterloo.ca
Dr. Leah Zhang-Kennedy, University of Waterloo, lzhangkennedy@uwaterloo.ca

Format: Lightning Talk

ABSTRACT

People increasingly turn to large language models for privacy and security advice [1], and how much they trust that advice shapes how they use it [2]. Most usable-security research, though, still assumes a user who is resourced, literate, and living in a single stable jurisdiction. We tested whether that assumption holds for users it wasn't built around: immigrants, gig workers, people in coercive or economically controlling relationships, and households where accounts and devices are shared across generations.

We built a dataset of 100 question-answer pairs across fifteen categories, moving from general topics like password management and phishing through prompts grounded in cross-cultural folk models of security, and ending in scenarios involving gendered vulnerability and marginalized populations. Each response from Claude and ChatGPT was scored against an expert-authored golden answer for comprehensiveness, actionability, and technical accuracy.

Across 99 scored questions, Claude's mean score was 9.26 out of 10 for each and ChatGPT's was 7.29, with a wider spread. The two models scored similar on well-established topics such as password hygiene, phishing, and device security. The gap widened on subjective and jurisdictionally ambiguous questions, particularly in the categories built around gendered vulnerability and marginalized populations. In those two categories, 83% of ChatGPT's scored rows named a Canadian institution or law even when the question described a different country or gave no location, and five cases carried an unprompted assumption that the user was a minor with no age indicator in the question. On prompts about covertly monitoring another person, Claude declined to treat the request as a straightforward how-to question, consistently reframing the user's underlying concern toward transparency and consent-based alternatives. Conversely, ChatGPT rarely pushed back on the ethical dimension of the request, providing more neutral responses and walking through the general considerations of the situation without clearly naming the legal exposure, safety risk, or consent violation embedded in the question.

Evaluation research on cultural alignment in LLMs has shown that small changes in how you test a model can swing the results as much as real differences between populations do [3]. Our results point to a related but distinct pattern: reliability depends less on the model itself and more on whether a question maps onto an existing policy or professional framework, and the individuals most dependent on these inquiries are often the least equipped to evaluate the accuracy or relevance of the responses. Our work outlines the key considerations for developing a culturally aware evaluation standard for AI safety advice.

Keywords: usable privacy and security, large language models, cross-cultural evaluation, demographic equity, human-computer interaction

References

[1] Ali, M., Arunasalam, A., & Farrukh, H. (2025). Understanding Users' Security and Privacy Concerns and Attitudes Towards Conversational AI Platforms. In 2025 IEEE Symposium on Security and Privacy (S&P).

[2] Leschanowsky, A., Rech, S., Popp, B., & Bäckström, T. (2024). Evaluating privacy, security, and trust perceptions in conversational AI: A systematic review. Computers in Human Behavior, 159, 108344.

[3] Khan, A., Casper, S., & Hadfield-Menell, D. (2025). Randomness, Not Representation: The Unreliability of Evaluating Cultural Alignment in LLMs. In Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25).

Abstract

Alejandra Zambrano

September 20, 2026

Generative AI chatbots are increasingly used for mental health support, despite having been developed as general-purpose systems rather than regulated clinical tools. This creates a significant governance challenge, particularly for vulnerable users who may place excessive trust in chatbot responses or fail to recognize biased, incomplete, or unsafe advice. This calls for investigating both the individual factors associated with AI over-reliance and a practical transparency intervention designed to promote more critical engagement with chatbot outputs.

Approximately 200 adults will participate in a two-phase experimental study. Participants will first complete measures of AI use and attitudes, attachment style, epistemic trust, social support, loneliness, and mental health symptoms. They will then interact with chatbots exhibiting different conversational biases, including over-validation or sycophancy, positive reframing, normalization, and action-focused responses. Outcomes will assess perceived helpfulness, trust, acceptance, reliance on chatbot advice, selective reliance on accurate versus flawed responses, and users' ability to detect bias.

In the second phase, the study will evaluate whether increasing transparency can reduce over-reliance. Before each chatbot response, users will be shown a reasoning checklist indicating which aspects of the situation the response addressed and which it omitted. This intervention is intended to make the chatbot's response process more transparent, reduce anthropomorphic perceptions, and encourage users to critically evaluate potentially biased or incomplete suggestions.

The project will identify psychological and contextual factors associated with vulnerability to AI over-trust while experimentally evaluating a scalable safeguard for responsible AI use. The findings will inform evidence-based guidelines for developers, clinicians, and policymakers and contribute to safer and more transparent deployment of generative AI in mental health and other high-stakes domains.

Understanding Parents' Privacy Perceptions of AI

Prachi Sadhwani, York University, prachi19@yorku.ca
Sana Maqsood, York University, smaqsood@yorku.ca
Yan Shvartzshnaider, York University, yansh@yorku.ca

Format: Lightning Talk

ABSTRACT

AI technologies are becoming commonplace in households, with families using them for various purposes, such as learning and entertainment [1]. For example, children are using AI to help them with schoolwork and for advice on interpersonal and medical issues, which are responsibilities associated with trusted, influential roles in a child's life such as their parents, teachers, and friends [1-2]. These people use AI to assist them in their responsibilities; for example, some parents seek advice about their children's health from generative AI chatbots [1].

As people think about what they should share with AI, they refer to common rules and practices (i.e. social norms) in different contexts, such as education and the household, to guide them. Parents may project norms they associate with influential roles in a child's life to AI when it performs the same roles, shaping what information they consider appropriate for the technology to collect and use [3]. However, AI may not always follow these norms which leaves children vulnerable to privacy harms. While previous research has shown that parents are increasingly worried about their children's data privacy when using AI, less is known about what parents consider to be appropriate information sharing practices across the different roles that AI could play in their child's life [3-5]. We therefore examine how parents' privacy expectations change when AI assists an influential role or takes over their responsibilities entirely. We explore privacy expectations through the theory of contextual integrity, which posits that privacy expectations are tied to established contextual social norms in which information is shared [6].

In this lightning talk, we will present preliminary findings from a survey with 200 Canadian parents about the norms they associate with the influential roles they assign to AI technologies and their privacy expectations of AI performing each role. We hope our findings can help understand the emerging contextual privacy norms and perceptions of AI-use practices.

Keywords: human-computer interaction, privacy, artificial intelligence, parents, children

References

[1] M. Madden, A. Calvin, A. Hasse, and A. Lenhart, "The Dawn of the AI Era: Teens, Parents, and the Adoption of Generative AI at Home and School" Common Sense Media, San Francisco, CA, USA, 2024. https://www.commonsensemedia.org/research/the-dawn-of-the-ai-era-teens-parents-and-the-adoption-of-generative-ai-at-home-and-school

[2] Y. Yu, Y. Liu, J. Zhang, Y. Huang, and Y. Wang, "Youth-Centered GAI Risks (YAIR): A Taxonomy of Generative AI Risks from Empirical Data," in Proceedings of the Twenty-First Symposium on Usable Privacy and Security, Seattle, WA, USA: USENIX, 2025, pp. 146--168. https://www.usenix.org/conference/soups2025/presentation/yu

[3] M. Eira, A. Rasouli, and V. Charisi, "Parents' Perceptions About the Use of Generative AI Systems by Adolescents," in Proceedings of the 24th Interaction Design and Children, in IDC '25.

New York, NY, USA: Association for Computing Machinery, Jun. 2025, pp. 927–931. doi: 10.1145/3713043.3731508.
[4] O. Collins Oguine, A. Alvarado Garcia, M. Muller, and K. Badillo-Urquiola, “Who Gets to Define Safety? A Systematic Review of How Generative AI Research Addresses Youth Online Safety,” in Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems, in CHI ’26. New York, NY, USA: Association for Computing Machinery, Apr. 2026, pp. 1–19. doi: 10.1145/3772318.3791346.
[5] Y. Yu, F. Mohi, A. Debroy, X. Cao, K. Rudolph, and Y. Wang, “Principles of Safe AI Companions for Youth: Parent and Expert Perspectives,” in Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems, in CHI ’26. New York, NY, USA: Association for Computing Machinery, Apr. 2026, pp. 1–21. doi: 10.1145/3772318.3793265.
[6] H. Nissenbaum, Privacy in Context: Technology, Policy, and the Integrity of Social Life. Stanford Law Books, 2009

On Experts and Algorithms: Who Decides When AI Decides?

Matya Stavnitzky, University of Waterloo, mstavnit@uwaterloo.ca

Format: Quick Talk

ABSTRACT

With the growing capabilities of Artificial Intelligence (AI) models, we seem more inclined to treat these models as if they were experts. Pulling on literature in the philosophy of science I argue that despite increased capacity in a variety of tasks, AI cannot replace scientific experts in decision- making. I will explore the application of LLMs to produce scientific assessments, specifically focusing on the Intergovernmental Panel on Climate Change (IPCC). Establishing that AI models diverge from scientific methodology because they focus on knowledge output rather than the process of its production. I demonstrate that assumptions underlying AI reliability depend on human values, but unlike scientific practices, the production of AI models does not address these values in clear, transparent, or communal ways.

Keywords: Artificial-Intelligence, expertise, decision-making, philosophy of science, epistemology

CAN-CWIC 2026

University of Waterloo

Abstracts for Presentations and Lightening Talks

Session: Bioinformatics & Software Engineering

Can AI Embeddings Reveal Your DNA?

Reem Al-Saidi, Postdoctoral Researcher, School of Computer Science, University of Windsor, alsaidir@uwindsor.ca

Format: Lightning Talk

ABSTRACT

Large language models are increasingly being adapted to an unexpected language: DNA. Genomic language models learn patterns from nucleotide sequences and encode them as numerical vectors called embeddings. Because embeddings support downstream analyses without exposing raw sequences, they are often treated as a safer artifact to exchange in genomic data sharing, federated learning, and machine-learning-as-a-service settings. But does converting DNA into numbers make it private?

This work shows why the answer can be no, using a single DNA-to-embedding-to-reconstruction example drawn from my doctoral research. A DNA sequence is tokenized and encoded by a genomic language model; the attacker never sees the sequence, only the embedding. A reconstruction attack then predicts the nucleotide at each position, and the recovered sequence is compared with the original, base by base. The example makes one point visible: an embedding is not a one-way transformation, because it retains information about individual nucleotides that an attacker can learn to recover. Such leakage matters because genomic data is permanent, uniquely identifying, shared with biological relatives, and revealing of ancestry and disease susceptibility. Unlike a password, a leaked genome cannot be reset.

Repeating this example across seven general-purpose and DNA-specific transformer models shows that leakage is not uniform: it depends on model architecture, tokenization, and position within the sequence. Counterintuitively, fine-tuning, typically performed only to improve task performance, can make embeddings more resistant to reconstruction in some architectures.

These patterns point to a practical defense. Vulnerability-Enhanced Selective Privacy Adaptation (VESPA), our position-aware approach, concentrates protection where embeddings are most exposed rather than treating every position identically. In our evaluation, it reduced reconstruction attack success by 57–64% while retaining more than 96% downstream-task accuracy.

The takeaway is simple: an embedding is not automatically a privacy shield. As foundation models increasingly handle genomic and health data, understanding what their representations retain, and what an attacker can recover from them, is essential to building responsible AI. For anyone who shares or publishes genomic embeddings, that means assessing reconstruction risk before release and giving embeddings the same privacy scrutiny as the data they are derived from.

Keywords: genomic privacy, genomic language models, embeddings, reconstruction attacks, privacy-preserving machine learning, responsible AI

Incorporating Sequence Diagrams into Communicating StateCharts

Shriya Borad, McMaster University, borads2@mcmaster.ca
Cindy Zhu, University, of Toronto, cindyy.zhu@mail.utoronto.ca
Sheida Emdadi, STaBL Foundation, sheida.emdadi@stablfoundation.org
Christopher Anand, McMaster University, anandc@mcmaster.ca
Anya Tafliovich, University of Toronto, anya@cs.utoronto.ca

Format: Lightning Talk

ABSTRACT

Concurrent applications are hard for beginners to understand [1, 2, 3] because events can happen in any order, and cannot be captured by a single flowchart. Communicating StateCharts (CSC) were introduced to capture the distributed state of a multi-user application. Statecharts model isolated actors only communicating through established message channels. While this helps beginners by enforcing separation of concerns and supporting design at a high level of abstraction, CSC doesn't provide a visualization of communication protocols.

"Work on multiple external representations (MERs) shows that well-aligned visualizations can reduce cognitive load and support deeper understanding." [4, *vide* 5] This suggests that adding Sequence Diagrams (SDs) could deepen students' understanding of concurrency. In fact, Otero et al [6] found that "a higher semantic comprehension" in the case of UML state/sequence diagram pairs, but not for real-time reactive systems where the statecharts alone were superior. Note that Harel and Kugler [7] found statecharts and live sequence charts are actually equivalent external representations.

We present a learning framework that pairs CSC and SDs based on the MER principles, using a simulation of a simplified elevator with multiple users to illustrate the different views of system states.

Our immediate future work is to conduct think-aloud studies, in which we will present the simulation tool. We will use different versions (CSC-only, CSC+SD, SD-only) of this tool to understand how the different visual components aid novices in understanding a simple concurrent system, and shed light on the negative result of [6]. The results of this study will guide the integration of SDs into the CSC Model-Driven Development tool.

Keywords: sequence diagrams, statecharts, concurrency, event driven programming, multiple external representations, visualization

References

[1] M. Ben-Ari, "Teaching concurrency and nondeterminism with Spin," in *Proc. 12th Ann. SIGCSE Conf. Innov. Technol. Comput. Sci. Educ. (ITiCSE '07)*, New York, NY, USA: ACM, 2007, pp. 363–364, doi: 10.1145/1269900.1268936.

[2] J. Lönnberg and A. Berglund, "Students' understandings of concurrent programming," in *Proc. 7th Baltic Sea Conf. Comput. Educ. Res. (Koli Calling '07)*, vol. 88, Darlinghurst, NSW, Australia: Australian Computer Society, 2007, pp.77–86.

[3] F. Strömbäck, L. Mannila, M. Asplund, and M. Kamkar, "A student's view of concurrency – A study of common mistakes in introductory courses on concurrency," in *Proc. 2019 ACM Conf. Int. Comput. Educ. Res. (ICER '19)*, New York, NY, USA: ACM, 2019, pp. 229–237, doi: 10.1145/3291279.3339415.

[4] N. Sibia et al., "From state to structure: Towards abstraction support in CS2," in *Proc. 25th Koli Calling Int. Conf. Comput. Educ. Res. (Koli Calling '25)*, New York, NY, USA: ACM, 2025, Art. no. 17, pp. 1–12, doi:10.1145/3769994.3769998.

[5] S. Ainsworth, "DeFT: A conceptual framework for considering learning with multiple representations," *Learn. Instr.*,vol. 16, no. 3, pp. 183–198, 2006, doi: 10.1016/j.learninstruc.2006.03.001.

[6] M. C. Otero and J. J. Dolado, "Evaluation of the comprehension of the dynamic modeling in UML," *Inf. Softw. Technol.*, vol. 46, no. 1, pp. 35–53, 2004, doi: 10.1016/S0950-5849(03)00108-3.

[7] D. Harel and H. Kugler, "Synthesizing state-based object systems from LSC specifications," *Int. J. Found. Comput. Sci.*, vol. 13, no. 1, pp. 5–51, 2002, doi: 10.1142/S0129054102000935.

Scalable Detection and Validation of Copy Number Variations from Whole-Genome Sequencing Data

Mame Seynabou Diop, Université de Montréal and CHU Sainte-Justine Research Center, mame.seynabou.diop@umontreal.ca
Sylvie Hamel, Université de Montréal, hamelsyl@iro.umontreal.ca
Sébastien Jacquemont, CHU Sainte-Justine Research Center, sebastien.jacquemont@gmail.com

Format: Lightning Talk

ABSTRACT

Copy Number Variations (CNVs) are deletions or duplications of genomic regions. They can affect a single gene or many genes at once. CNVs are associated with disorders such as autism and schizophrenia. Detecting them accurately is important for both genomic research and clinical diagnosis.

For many years, CNVs were detected using genotyping arrays. These technologies are fast and well established, but their limited resolution makes smaller CNVs difficult to detect. Whole-genome sequencing (WGS) offers much higher resolution and enables the detection of both small and large CNVs. The challenge is scale: many WGS-based methods rely on sequence files that are enormous, more than 80 GB per person, and take hours of computation each. A recent study needed 72 days and 7.3 million CPU hours to analyze 470,727 genomes [1]. Most laboratories cannot afford this.

My PhD project addresses two complementary challenges: how to detect CNVs efficiently across very large WGS datasets, and how to determine whether the detected CNVs are reliable.

First, I developed gVCF2CNV [2], a scalable Nextflow pipeline for CNV detection. Instead of reprocessing these large sequence files, gVCF2CNV uses a compact format already generated during standard WGS analysis, which still contains the information needed for CNV detection. The pipeline transforms this information into signals that can be processed by fast, well-established algorithms originally developed for genotyping arrays. We first evaluated it in 12,509 individuals from the SPARK cohort, then applied it to 414,824 genomes from the All of Us cohort, where the analysis was completed in four days.

Scalability, however, is only part of the problem. Large-scale CNV detection can generate false-positive calls, and available benchmark datasets are limited in size and completeness. To address this second challenge, I developed MCNV2 [3], a reproducible validation framework based on inheritance in parent-offspring families. Most true CNVs are inherited, so we measure how many CNVs found in a child are also present in at least one parent. MCNV2 formalizes this principle through Mendelian Precision, providing a systematic way to assess CNV call quality and to identify filtering strategies that improve reliability.

In this talk, I will present how we addressed the challenges of scalability and validation in large-scale CNV analysis. Together, gVCF2CNV and MCNV2 make CNV analysis more reliable and more accessible to laboratories without specialized computing infrastructure.

Keywords: Scalability, Copy Number Variations, Whole-genome Sequencing, Detection, Validation, Bioinformatics

References

[1] Zou, X. Z., Hu, F., Lou, H., Burren, O. S., Li, X., Megy, K., et al. (2026). Phenome-wide analysis of copy number variants in 470,727 UK Biobank genomes. Nature, 652(8110), 675-685. https://doi.org/10.1038/s41586-025-10087-x

[2] Diop, M. S., Bénitière, F., Kumar, K., Clark, B., Jean-Louis, M., Saci, Z., Huguet, G., Hamel, S., & Jacquemont, S. (2026). gVCF2CNV: a scalable pipeline for CNV detection from whole-genome sequencing data. bioRxiv, 2026.09.02.748565. https://doi.org/10.64898/2026.09.02.748565

[3] Diop, M. S., Lemacon, A., Kumar, K., Clark, B., Huguet, G., Bénitière, F., Jean-Louis, M., Hamel, S., & Jacquemont, S. (2026). MCNV2 (Mendelian CNV Validation): Mendelian Precision for CNV quality assessment. bioRxiv, 2026.04.29.721462. https://doi.org/10.64898/2026.04.29.721462

A Model-Driven Formalization of Prompt Patterns

Vennila Sooben, DIRO – Université de Montréal, vennila.sooben@umontreal.ca
Eugene Syriani, DIRO – Université de Montréal, syriani@iro.umontreal.ca

Format: Lightning Talk

ABSTRACT

A prompt is an instruction or input given to a large language model (LLM), such as ChatGPT, to guide its response. Prompt patterns provide reusable structures for designing prompts, but existing catalogs typically describe them through informal prose, examples, or loosely structured templates. This limits their systematic reuse, detection, comparison, and analysis. In this talk, we propose a model-driven formalization of prompt patterns based on a taxonomy. We define a textual domain-specific modeling language (DSL) that specifies prompt patterns using a well-defined notation. The taxonomy and pattern definitions are transformed into a graph representation stored in a graph database. This representation makes prompt patterns machine-processable and supports their instantiation in prompts, their detection in concrete prompt instances, and the analysis of relationships among patterns, prompt components, and prompting strategies. We validate the operational taxonomy on a repository of industrial prompts.

Keywords: Model-driven engineering, prompt engineering, domain-specific language, prompt patterns, graph models, model transformations

References *(if any)*

[1] Ian Arawjo, Chelse Swoopes, Priyan Vaithilingam, Martin Wattenberg, and Elena L. Glassman. 2024. ChainForge: A Visual Toolkit for Prompt Engineering and LLM Hypothesis Testing. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 304, 18 pages. doi:10.1145/3613904.3642016

[2] Stephen H. Bach, Victor Sanh, Zheng-Xin Yong, Albert Webson, Colin Raffel, Nihal V. Nayak, Abheesht Sharma, Taewoon Kim, M. Saiful Bari, Thibault Fevry, Zaid Alyafeai, Manan Dey, Andrea Santilli, Zhiqing Sun, Srulik Ben-David, Canwen Xu, Gunjan Chhablani, Han Wang, Jason Alan Fries, Maged S. Al-shaibani, Shanya Sharma, Urmish Thakker, Khalid Almubarak, Xiangru Tang, Dragomir Radev, Mike Tian-Jian Jiang, and Alexander M. Rush. 2022. PromptSource: An Integrated Development Environment and Repository for Natural Language Prompts. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. Association for Computational Linguistics, 93–104. doi:10.18653/v1/2022.acl-demo.9

[3] Antonio Bucchiarone, Marco Panciera, Antonio Cicchetti, Nadia Mana, Carlotta Castelluccio, and Lee Stott. 2024. PromptDeck: A No-Code Platform for Modular Prompt Engineering. In Proceedings of the ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems (Linz, Austria) (MODELS Companion '24). Association for Computing Machinery, New York, NY, USA, 895–904. doi:10.1145/3652620.3688336

[4] Jeffery Cao, Lampros Flokas, Yujian Xu, Eugene Wu, Xu Chu, and Cong Yu. 2025.

Prompt Editor: A Taxonomy-driven System for Guided LLM Prompt Development in Enterprise Settings. In Companion of the 2025 International Conference on Management of Data (SIGMOD/PODS '25). ACM, 59–62. doi:10.1145/3722212.3725124
[5] Boqi Chen, Fandi Yi, and Dániel Varró. 2023. Prompting or Fine-tuning? A Comparative Study of Large Language Models for Taxonomy Construction. In 2023 ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C). 588–596. doi:10.1109/MODELS-C59198.2023.00097
[6] Oluwole Fagbohun, Rachel M. Harrison, and Anton Dereventsov. 2024. An Empirical Categorization of Prompting Techniques for Large Language Models: A Practitioner's Guide. Technical Report 2402.14837. arXiv. doi:10.48550/arXiv.2402.14837
[7] Angela Fan, Beliz Gokkaya, Mark Harman, Mitya Lyubarskiy, Shubho Sengupta, Shin Yoo, and {Jie M.} Zhang. 2023. Large Language Models for Software Engineering: Survey and Open Problems. In Proceedings - 2023 IEEE/ACM International Conference on Software Engineering (Proceedings - 2023 IEEE/ACM International Conference on Software Engineering: Future of Software Engineering, ICSE-FoSE 2023). Institute of Electrical and Electronics Engineers Inc., 31–53. doi:10.1109/ICSE-FoSE59343.2023.00008
[8] My M. Mosthaf and Andrzej Wasowski. 2024. From a Natural to a Formal Language with DSL Assistant. In Proceedings of the ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems (Linz, Austria) (MODELS Companion '24). Association for Computing Machinery, New York, NY, USA, 541–549. doi:10.1145/3652620.3687811
[9] Yuetian Mao, Junjie He, and Chunyang Chen. 2025. From Prompts to Templates: A Systematic Prompt Template Analysis for Real-world LLMapps. In Proceedings of the 2025 ACM International Conference on the Foundations of Software Engineering (FSE Companion '25). ACM, 75–86. doi:10.1145/3696630.3728533
[10] Sergio Morales, Robert Clarisó, and Jordi Cabot. 2025. Impromptu: A Framework for Model-Driven Prompt Engineering. Software and Systems Modeling 24, 6 (2025), 1627–1645. doi:10.1007/s10270-024-01235-4
[11] M. Z. Naser. 2025. A Review of Prompt Engineering Techniques for Large Language Models. International Journal of Human–Computer Interaction 0, 0 (2025), 1–35. arXiv:https://doi.org/10.1080/10447318.2025.2607553 doi:10.1080/10447318.2025.2607553
[12] Pranab Sahoo, Ayush Kumar Singh, Sriparna Saha, Vinija Jain, Samrat Mondal, and Aman Chadha. 2024. A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications. Technical Report 2402.07927. arXiv. doi:10.48550/arXiv.2402.07927
[13] Kurt Sandkuhl, Balbir S. Barn, and Souvik Barat. 2026. LLM Support for Enterprise Modelling Based on Integrating Prompt and Enterprise Modelling Meta-Models. Discover Artificial Intelligence 6, 303 (2026). doi:10.1007/s44163-026-01056-y
[14] Sander Schulhoff, Michael Ilie, Nishant Balepur, Konstantine Kahadze, Amanda Liu, Chenglei Si, Yinheng Li, Aayush Gupta, HyoJung Han, Sevien Schulhoff, Pranav Sandeep Dulepet, Saurav Vidyadhara, Dayeon Ki, Sweta Agrawal,

Chau Pham, Gerson Kroiz, Feileen Li, Hudson Tao, Ashay Srivastava, Hevander Da Costa, Saloni Gupta, Megan L. Rogers, Inna Goncearenco, Giuseppe Sarli, Igor Galynker, Denis Peskoff, Marine Carpuat, Jules White, Shyamal Anadkat, Alexander Hoyle, and Philip Resnik. 2025. The Prompt Report: A Systematic Survey of Prompting Techniques. Technical Report 2406.06608. arXiv. doi:10.48550/arXiv.2406.06608

[15] Vennila Sooben and Eugene Syriani. 2026. A Taxonomy of Single-Turn Textual Prompt Patterns. Technical Report 2607.00043. arXiv. doi:10.48550/arXiv.2607.00043

[16] Eugene Syriani, Lechanceux Luhunu, and Houari Sahraoui. 2018. Systematic Mapping Study of Template-based Code Generation. Computer Languages, Systems Structures 52, 1 (2018), 43–62. doi:10.1016/j.cl.2017.11.003

[17] Shubham Vatsal and Harsh Dubey. 2024. A Survey of Prompt Engineering Methods in Large Language Models for Different NLP Tasks. Technical Report 2407.12994. arXiv. doi:10.48550/arXiv.2407.12994

[18] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C. Schmidt. 2023. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. Technical Report 2302.11382. arXiv. doi:10.48550/arXiv.2302.11382

[19] Yujing Yang, Boqi Chen, Kua Chen, Gunter Mussbacher, and Dániel Varró. 2024. Multi-step Iterative Automated Domain Modeling with Large Language Models. In Proceedings of the ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems (Linz, Austria) (MODELS Companion '24). Association for Computing Machinery, New York, NY, USA, 587–595. doi:10.1145/3652620.3687807

CAN-CWIC 2026

University of Waterloo

Abstracts for Presentations and Lightening Talks

Session: Computing Science Education

HelperCell: Leveraging AI to Provide Timely Beginner Focused Qualitative Feedback For Programing Tasks

Heather Meatherall, Ontario Tech University, heathermeat7@gmail.com

Format: Lightning Talk

ABSTRACT

This Lighting Talk presents my Undergraduate Honours Thesis, which investigates how to use AI to provide timely beginner focused qualitative feedback on programming tasks and address limitations faced by traditional methods. In this talk I present HelperCell, a Jupyter Notebook extension that leverages AI to provide beginner programmers with timely qualitative feedback. I share how I evaluated HelperCell in a user study to gain insights as to how AI feedback affects correctness, how students use the system, and how they perceive the feedback. The results from this study revealed that AI-based feedback can support beginner programmers by correcting errors and scaffolding their problem-solving process.

Keywords: Artificial Intelligence, Education, Human-Computer Interaction

Out of Scope? Approaching Responsible Computing through Epistemic Diversity

Rutwa Engineer, University of Toronto Mississauga, rutwa.engineer@utoronto.ca

Format: Lightning Talk

ABSTRACT

Historically, the computing curriculum has been organized around three core traditions, which are a mathematical tradition, a scientific tradition, and an engineering tradition [1]. These traditions give computing its rigour, but they can also narrow what is considered legitimate knowledge in the discipline. Through a set of practices that Breslin [2] describes as technical rendering, computing students learn to filter and translate complex scenarios into computable representations such as code, data, and models. As a result, the social, cultural, political, and historical dimensions of a problem are often treated as out of scope. These are the questions that responsible computing depends on, but the way we often teach computer science can make them feel a bit out of place.

This lightning talk is an invitation to re/envision what could happen when we broaden the technical frame by treating more forms of knowledge as legitimate in computing. I use the term epistemic diversity to describe an approach where multiple perspectives shape how we define problems, choose data, design systems, and evaluate solutions. It brings in lived experience, community knowledge, and work from other fields, and it recognizes communities and affected users as producers of computing knowledge. I posit that responsible computing, where we design, build, and deploy technologies while considering who is affected, how harms and benefits are distributed, and who is accountable, depends on epistemic diversity. In this approach, the technical work stays central, and rigour can also include whether the assumptions in the code match the reality being modelled.

To illustrate, I share two small changes that we made to core computer science courses at the University of Toronto Mississauga. In a second-year software design course, students build an object-oriented paint program and then propose and implement one accessibility feature, which they must justify using Ability-Based Design. In a second-year computer organization course, students connect the narrative of processor upgradability to the social and environmental effects of e-waste by answering questions about responsibility, fairness, and repair from the perspective of a specific stakeholder.

I hope attendees will leave with a working definition of epistemic diversity in computing, an understanding of how it can support responsible computing, and one call-to-action and strategy that they can try in their context.

Keywords: responsible computing, epistemic diversity, computing education, curriculum

References

[1] Erkki Sutinen, Matti Tedre, and Matti Tedre, “Three traditions of computing: what educators should know,” *Computer Science Education*, Sept. 15, 2008. https://www-tandfonline-com.myaccess.library.utoronto.ca/doi/full/10.1080/08993400802332332

[2] S. Breslin, “The making of computer scientists: rendering technical knowledge, gender, and entrepreneurialism in Singapore,” *Scholaris.ca*, Aug. 2018. https://memorial.scholaris.ca/items/5e12b228-ab0c-4728-a951-114b9494b67d

Students' Attitudes Toward CS Ethics: The Role of Embedded Ethics Modules, Experience, and Demographics

Emma Tomiuk, McGill University, emma.tomiuk@mail.mcgill.ca
Hugo Cossette-Lefevbre, McGill University, hugo.cossette-lefebvre@mail.mcgill.ca
Armin Yazdani, McGill University, armin.yazdani@mcgill.ca
Golnoosh Farnadi, McGill University & Mila Quebec AI Institute, farnadig@mila.quebec

Format: Lightning Talk

ABSTRACT

Harms caused by AI and technology, including discriminatory AI recruiting tools [1], data centers' growing environmental footprint [6], and privacy concerns with home devices [3], are now prominent in public discourse. Computer scientists involved in technology development must be aware of such ethical issues and need to have the skills to anticipate, mitigate, and prevent technological disasters. However, they rarely receive formal ethics training.

To address this need, we piloted an Embedded Ethics (EE) program in the computer science (CS) department at McGill in the Fall 2025 and Winter 2026 terms. Inspired by similar models at Harvard [2], Stanford [7], and the University of Toronto [4], our interdisciplinary group of CS, philosophy, law, and education professors and students designed and taught ethics modules for seven undergraduate CS courses, reaching approximately 2500 students. In a 1.5-hour session, students learn about a course-related ethical issue (e.g. fairness issues in facial recognition systems) through technical and philosophical lenses. As students attend modules throughout their studies, they are expected to develop ethical reasoning abilities to be better equipped to tackle various ethical issues encountered later in the workforce [2, 5].

This lightning talk will present our EE initiative and key findings from our program evaluation project. We address whether (1) students' attitudes towards ethics (interest, value, self-efficacy, responsibility) improve after EE modules and (2) how these attitudes differ between different demographics.

Survey results comparing students who attended EE modules (n=139) with those who did not (n=102) revealed that module attendees had 13% higher self-efficacy scores than non-attendees. Additional ethics experiences beyond our modules further increased self-efficacy. Interest and value in ethics were higher for students who attended a module, but only when they were previously exposed to ethics. Interest, value, and self-efficacy also increased as students advanced in their studies. Together, these results suggest that attitudes towards ethics in CS, particularly ethics confidence, can develop with time and experience with ethics, with our EE modules being one possible source of this experience. We therefore recommend that efforts should be made to provide ethics training opportunities for CS students.

Finally, women displayed 10% higher interest in ethics than men regardless of whether they attended a module. This finding makes us reflect on the role that EE modules and ethics education more generally could play in attracting and keeping more women in CS and diversifying the technological workforce that is needed to develop ethical systems [8].

References

[1] Jeffrey Dastin. 2018. Amazon ditches AI recruiting tool that didn't like women. *Global News*. Retrieved September 21, 2026 from https://globalnews.ca/news/4532172/amazon-jobs-ai-bias/

[2] Barbara J. Grosz, David Gray Grant, Kate Vredenburgh, Jeff Behrends, Lily Hu, Alison Simmons, and Jim Waldo. 2019. Embedded EthiCS: integrating ethics across CS education. *Commun. ACM* 62, 8 (July 2019), 54–61. https://doi.org/10.1145/3330794

[3] Scharon Harding. 2025. Everything you say to your Echo will be sent to Amazon starting on March 28. *Ars Technica*. Retrieved October 6, 2025 from https://arstechnica.com/gadgets/2025/03/everything-you-say-to-your-echo-will-be-sent-to-amazon-starting-on-march-28/

[4] Diane Horton, Sheila A. McIlraith, Nina Wang, Maryam Majedi, Emma McClure, and Benjamin Wald. 2022. Embedding Ethics in Computer Science Courses: Does it Work? In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education*, February 22, 2022. ACM, Providence RI USA, 481–487. https://doi.org/10.1145/3478431.3499407

[5] Louie Ortiz, Jenny Yang, and Anne Newman. 2026. Evaluating the Role of Ethics Education in Shaping College Students' Ethical Engagement in Computing Technology. In *Proceedings of the 57th ACM Technical Symposium on Computer Science Education V.2* (*SIGCSE TS 2026*), February 17, 2026. ACM, New York, NY, USA, 1459–1460. https://doi.org/10.1145/3770761.3777295

[6] Simmone Shah. 2026. Both Democrats and Republicans Are Increasingly Worried About AI's Environmental Impact. *Time*. Retrieved September 18, 2026 from https://time.com/article/2026/09/16/majority-americans-worried-ai-environment-impact/

[7] Stanford Embedded Ethics. Embedding Ethics in Computer Science. Retrieved October 17, 2025 from https://embeddedethics.stanford.edu/

[8] Kevin A. Wortman, Aakash Gautam, Sarah Hug, Paul Salvador Inventado, Ayaan M. Kazerouni, Jane Lehr, Kanika Sood, and Zoë Wood. 2025. Reflecting on Practices to Integrate Socially Responsible Computing in Introductory Computer Science Courses. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1* (*SIGCSETS 2025*), February 18, 2025. ACM, New York, NY, USA, 1253–1259. https://doi.org/10.1145/3641554.3701962

Keywords: Embedded Ethics, Computer Science Education, Ethics Education, Ethical Issues in CS, Demographics

CAN-CWIC 2026

University of Waterloo

Abstracts for Presentations and Lightening Talks

Session: Data

EviPlan: Metadata-Guided Experimental Planning

Zhengyuan Dong, University of Waterloo, zhengyuan.dong@uwaterloo.ca

Format: Lightning Talk

ABSTRACT

Experimental design and execution are critical stages of AutoResearch, but execution can be costly: after proposing an idea, an agent must select baselines, benchmarks, metrics, and protocols, then run the experiments needed to support or challenge its central claim. An open question is whether selecting these resources in light of published results can reduce the new experiments needed to test the claim. Answering it requires retrieving experimental results obtained under the same evaluation settings from a lake of prior papers. We present EviPlan, a table-centered, evidence-aware experimental-design module that can operate independently or within an AutoResearch system. Given a research idea and a prepared paper lake, EviPlan retrieves and reranks experimental tables to determine which previously reported results can be reused and how much of the planned evaluation they cover. It recommends resource settings with greater reuse potential, together with supporting papers, results, and available code for reuse, reproduction, or adaptation, and designs the remaining main and ablation experiments. Our evaluation on a temporally filtered paper corpus measures the extent of reusable evidence and examines whether recommended resource settings require fewer new experiments than author-reported settings under comparable evaluation requirements.

Keywords: automated research, table discovery, scientific data lake, experimental design

Data Profiling for Multi-Attribute Change Rules

Nishttha Sharma, McMaster University, sharmn99@mcmaster.ca
Fei Chiang, McMaster University, fchiang@mcmaster.ca

Format: Lightning Talk

ABSTRACT

Data rarely stays still. Patients' lab results change, weather conditions fluctuate, energy consumption rises and falls, and organizations continually update their records. Yet when we look at a dataset, we often focus on the values themselves rather than the *changes between them*. This makes it difficult to answer questions such as: *When one thing changes, what else tends to change? Are these changes part of a recurring pattern? And when is a change unusual?* Existing database systems have limited support for change management, relying on statistics, triggers, and constraints. Traditional data dependencies, such as Order Dependencies (ODs) [1], Sequential Dependencies (SDs) [2], and Differential Dependencies (DDs) [3], can describe relationships between values but often do not capture how changes evolve across consecutive observations or how a change in one attribute relates to a change in another.

To address this, our previous work introduced **Change Rules (CRs)**, a new type of data quality rule for describing relationships between changes in data. CRs capture bounded changes in an antecedent and a consequent attribute over an ordered sequence of records.

In this talk, we focus on efficiently discovering multi-attribute Change Rules by identifying combinations of antecedent attributes and change intervals that collectively explain changes in a consequent attribute. Many real-world changes depend on multiple factors rather than a single attribute. For instance, a change in energy consumption may be associated with simultaneous changes in temperature and humidity. However, discovering such rules introduces challenges due to the potentially large number of attribute combinations and candidate change intervals.

We will illustrate these challenges through an example that motivates the main contribution of this work, followed by an overview of our discovery framework. The framework uses joint change information to identify promising candidates, incrementally expands change intervals, and prunes redundant candidates while preserving meaningful rules. Rather than exhaustively exploring all possible attribute combinations and intervals, the approach aims to navigate the search space efficiently. We evaluate the framework on real-world datasets, examining its ability to discover meaningful multi-attribute change patterns and its scalability as dataset size and the number of attributes increase.

By making relationships between changes explicit and discoverable, Change Rules provide an interpretable foundation for understanding how data evolves. Efficient discovery of multi-attribute CRs further enables the identification of complex relationships that may remain hidden when changes are analyzed independently.

Keywords: change exploration, data profiling, change rules

References

[1] Szlichta, Jaroslaw, et al. "Effective and complete discovery of order dependencies via set-based axiomatization." arXiv preprint arXiv:1608.06169 (2016).

[2] Golab, Lukasz, et al. "Sequential dependencies." Proceedings of the VLDB Endowment 2.1 (2009): 574-585.

[3] Song, Shaoxu, and Lei Chen. "Differential dependencies: Reasoning and discovery." ACM Transactions on Database Systems (TODS) 36.3 (2011): 1-41.

Data-Driven Planning of EV Charging Infrastructure for Multi-Unit Residential Buildings in Canada

Tashneet Kaur, Wilfrid Laurier University, Waterloo, Canada, kaur6910@mylaurier.ca
Sukhjit Singh Sehra, Wilfrid Laurier University, Waterloo, Canada, ssehra@wlu.ca

Presentation Format: Lightning Talk

ABSTRACT

The adoption of electric vehicles (EVs) continues to accelerate worldwide, underscoring the need for reliable and equitable access to charging infrastructure. Unlike residents of single-family homes, residents of multi-unit residential buildings (MURBs) often face barriers to home charging due to shared parking facilities, limited electrical capacity, retrofit costs, and complex governance arrangements. As research on residential EV charging has expanded, studies have examined charging demand, infrastructure planning, smart charging, user behaviour, policy, and energy integration from diverse disciplinary perspectives. However, this evidence remains fragmented, limiting a holistic understanding of how technical, behavioral, regulatory, and planning factors collectively influence the large-scale deployment of residential charging infrastructure. [1, 2, 3].

This lighting talk introduces **EIPRE**, a modular decision-support tool for EV infrastructure planning and return on investment assessment in Canadian MURBs. The framework integrates long-term EV-adoption forecasting in a building, stochastic charging-demand estimation, charging-infrastructure sizing and phased deployment, building electrical-capacity assessment, and lifecycle techno-economic analysis.

EIPRE incorporates EV registration data and demographics, building information, electrical infrastructure data, electricity costs, policy information, and infrastructure cost information. These inputs are translated into practical planning outputs, including projected charging demand, recommended charger deployment over time, electrical-capacity requirements, capital and operating expenditures, total cost of ownership, payback period, net present value, and return on investment under alternative EV-adoption scenarios. With Ontario serving as the starting point for development and validation, the framework is aimed at showing how a data-driven approach can forecast, model, and optimize EV infrastructure planning from isolated technical analyses toward an integrated decision-support process.

The talk will highlight how EIPRE connects future EV adoption with building-level infrastructure and investment decisions, also providing an approach to generalize the framework. The analytics are intended to support property managers, developers, utilities, charging service providers, and policymakers in determining when and how much charging infrastructure should be deployed, whether existing electrical capacity is sufficient, and whether proposed investments remain financially viable under alternative future adoption scenarios.

Keywords: electric vehicle charging, multi-unit residential buildings, infrastructure planning, data-driven modelling, techno-economic analysis

References

[1] Abdulaziz Almutairi. Optimizing ev charging infrastructure in multi-unit residential buildings for sustainable energy management. *Sustainability*, 18(6):3051, 2026.

[2] M Kuby, A Cordova-Cruzatty, NC Parker, and DA King. Ev charging for multifamily housing: Review of evidence, methods, barriers, and opportunities. *Renewable and Sustainable Energy Reviews*, 210:115253, 2025.

[3] Noah Horesh, Yan Zhou, and Jason Quinn. Home charging for all: Techno-economic and life cycle assessment of multi-unit dwelling electric vehicle charging hubs. *Journal of Cleaner Production*, 383:135551, 2023.

Exploring Challenges in Multi-Table Retrieval for Open-Domain Question Answering

Siyuan Zhao, University of Waterloo, s56zhao@uwaterloo.ca

Format: Lightning Talk

ABSTRACT

Large collections of structured tables contain valuable information for answering complex questions, but identifying the right tables remains challenging. Existing open-domain table question answering systems commonly retrieve tables based on the relevance between a natural-language question and representations of individual tables [1]. This becomes particularly challenging in multi-table settings, where answering a question may require retrieving complementary information distributed across multiple tables and capturing relationships that are not evident from query–table similarity alone [2].

In this ongoing work, we investigate failure modes of table retrieval across existing open-domain table question answering benchmarks. Our preliminary experiments and error analysis reveal recurring failures caused by semantic mismatches between natural-language queries and table representations. Relevant concepts may be expressed using different terminology in the query and the underlying tables, while noisy, redundant, or weakly normalized schemas can further obscure relevant information. These observations suggest that direct query–table similarity may not fully capture the signals needed for robust retrieval, particularly in multi-table settings.

Motivated by these findings, we are exploring retrieval and reranking approaches that can better capture relationships between natural-language queries and heterogeneous tables. In this lightning talk, I will present representative retrieval failure modes, preliminary empirical findings, and ongoing directions toward more robust multi-table retrieval for open-domain question answering.

Keywords: table retrieval, multi-table retrieval, table question answering, information retrieval

References

[1] Herzig, J., Müller, T., Krichene, S., & Eisenschlos, J. 2021. Open Domain Question Answering over Tables via Dense Retrieval. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 512–519.

[2] Chen, P. B., Zhang, Y., & Roth, D. 2024. Is Table Retrieval a Solved Problem? Exploring Join-Aware Multi-Table Retrieval. Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2687–2699.

CAN-CWIC 2026

University of Waterloo

Abstracts for Presentations and Lightening Talks

Session: Gender Dynamics

Leading Without Code? Accommodation, Maternalism, and Hollowed Authority in Software Production

Alejandra Moreno-López, Universidad Nacional De Córdoba, Argentina
[alejandra.moreno.lopez@mi.unc.edu.ar]

Kelly Thomson, York University, Canada
[thomsonk@yorku.ca]

Lucía Jaimez, Universidad Nacional De Córdoba, Argentina
[luciajaimez@unc.edu.ar]

Format: Presentation

ABSTRACT

In the tech industry, corporate narratives often portray management and technical roles as gender-neutral. However, our digital ethnography tracking daily interactions across two high-tech software teams in Argentina reveals a different reality. Examining one domestic outsourcing team and one global offshoring team, this study demonstrates how operating without code mastery forces female leaders into a precarious negotiation of power. In the local team, authority is hollowed through dependence: the leader enacts a protective maternalism to shield juniors from client pressure, but constantly relies on male developers for "technical rescues". Conversely, in the offshore group, authority is hollowed through exclusion: the leader strategically accommodates by avoiding client emergencies, allowing male coders to bypass her formal position.

Ultimately, these tactics reveal that operating from the sidelines leaves entrenched male-dominated cultures intact. Through field observation, we uncover how functional segregation persists across both environments: while hard technical coding remains masculinized within male-dominated decision-making "bunkers", women are systematically channeled toward operational management and "glue work," such as office logistics or taking meeting minutes. This distancing from core technical tasks directly hollows their formal leadership. Rather than being passive subjects, these leaders actively negotiate these structural boundaries, deploying survival strategies that entail complex political trade-offs.

Appointing women to management does not automatically disrupt workplace inequality. This presentation offers industry professionals practical insights into recognizing invisible barriers, equipping them to move beyond performative diversity toward structural inclusion in their own teams.

Keywords: Gender dynamics, Software production, Glue work, Technical authority, Micro-inclusion

Who Gets to Lead in Tech? Claiming Space from Zero to One

YinNing Zhu, Dalhousie University, chloez@dal.ca

Format: Lightning Talk

ABSTRACT

Who gets to lead in tech, and who decides? After representing Dalhousie University at CAN-CWiC 2025, I returned to Halifax with no student cloud communities to bridge our campus with the wider industry. Seeing that gap, I founded the first AWS Student Builder Group in Atlantic Canada. Despite Halifax being far from a traditional tech hub, the group rapidly grew to over 200 members and became the third largest of its kind across Canada within three months.

Starting from zero, I mapped community needs, engaged local stakeholders, and launched our event. Despite a harsh Halifax snowstorm, dozens of passionate students and professors braved the weather to pack the room. Seeing that genuine hunger for technical community firsthand deeply moved me, transforming an initial initiative into an absolute conviction that our students deserved the same opportunities as peers in major tech hubs. Driven to bridge our geographic isolation, I forged relationships with external leaders and independently established a regional presence at the AWS Summit in Toronto to build external pipelines. Soon, these efforts brought industry professionals directly to our students for hands-on technical workshops and career mentorship.

Yet beneath that visible momentum lay real cultural resistance. As a queer woman leading a technical community, I navigated systemic skepticism about my technical authority, harassment within male-dominated peer spaces, and resistance to diverse student leadership. Growing an organization while navigating environments that signal you do not belong is an exhausting gender dynamic in computing.

Instead of retreating, I countered skepticism through technical execution, redefining what authority looks like in community building. Rather than adopting rigid, hierarchical leadership styles, our team implemented modern, distributed workflows to automate operations, analyze engagement, and decentralize knowledge. Upskilling peers alongside these shared systems transformed the group into an autonomous, self-sustaining community. This demonstrated that inclusive, decentralized governance can effectively dismantle traditional gatekeeping.

Attendees will gain actionable strategies for navigating gender and identity dynamics in early-career leadership, alongside scalable models for community-driven technical initiatives in peripheral regions. Ultimately, this session stands as a testament to resilience in the face of adversity. It offers practical frameworks for establishing boundaries, building authentic allyship, and unapologetically claiming space in rooms not initially built for us.

Keywords: Gender Dynamics in Computing, Inclusive Governance, Women in Tech Leadership, Community Building

The Making of Tech Leaders: Gender, Career Advancement, and the Ideal Tech Worker

Helena Wright, University of Toronto, helena.wright@mail.utoronto.ca

Format: Lightning Talk

ABSTRACT

The tech industry is characterized by entrepreneurialism, creativity, risk-taking, and hustle. Emerging from Silicon Valley's countercultural movements, the "California Ideology" brought together technological development, individual freedom, entrepreneurialism, and market-oriented approaches to work (Barbrook & Cameron, 1996; Neff, 2012; Irani, 2019). These values have contributed to the development of an idealized tech worker who is entrepreneurial, continuously developing their skills, and deeply committed to their work (Neff, 2012; Peters, 2000). This presentation examines how these expectations inform tech work culture and shape career advancement and leadership within the technology industry. Tech workers embody an entrepreneurial spirit, becoming their own "venture capitalists" by investing in ongoing skill development, networking, and professional mobility (Neff, 2012). Within this culture, long hours and continuous self-optimization can become markers of professional value. Leadership practices such as Amazon's "bias towards action" further reinforce expectations that workers should proactively identify opportunities, take action, and move quickly in response to changing markets and technologies (Irani, 2019). However, the expectations associated with the ideal tech worker are not experienced equally. Research on Silicon Valley describes a work ethic characterized by "15-hour days and 7-day weeks," while identifying this culture as particularly discriminatory toward women who are tasked with child-rearing and domestic labour (Duff, 2016). This reflects a broader problem with treating the worker as separate from the social and domestic conditions in which work takes place. As Acker (1990) argues, organizational structures are themselves gendered, producing assumptions about what an ideal worker looks like and what kinds of bodies, lives, and responsibilities can be accommodated within the organization. Rather than treating leadership as an individual problem of skills, confidence, or professional development, this presentation examines how organizational culture, workplace structures, and historically constructed ideas about tech work shape the conditions under which women can advance into executive-level tech leadership. By examining the entrepreneurial tech worker as a cultural and organizational ideal, this presentation argues that increasing women's representation in technology leadership must be accompanied by organizational reform. Creating more equitable pathways into leadership requires rethinking the workplace norms, structures, and expectations that determine whose labour is valued and what forms of leadership are recognized.

Keywords: Tech Work, Women in Leadership, Gendered Organizations, Silicon Valley Culture

References

Acker, J. (1990). Hierarchies, jobs, bodies: A theory of gendered organizations. *Gender & Society, 4*(2), 139–158.

Barbrook, R., & Cameron, A. (1996). The Californian ideology. *Science, Technology, & Culture, 6*(1), 44–72. https://doi.org/10.1080/09505439609526455

Duff, A. S. (2016). Rating the revolution: Silicon Valley in normative perspective. *Information, Communication & Society, 19*(11), 1605–1621. https://doi.org/10.1080/1369118X.2016.1142594

Irani, L. (2019). *Chasing innovation: Making entrepreneurial citizens in modern India*. Princeton University Press.

Neff, G. (2012). *Venture labor: Work and the burden of risk in innovative industries*. MIT Press. https://doi.org/10.7551/mitpress/9780262017480.001.0001

Peters, T. (2000, May 22). What will we do for work. *TIME*. https://time.com/archive/6954840/what-will-we-do-for-work-3/

CAN-CWIC 2026

University of Waterloo

Abstracts for Presentations and Lightening Talks

Session: Human Computer Interaction (HCI) and Accessibility

Towards a Cross-Cultural Framework for Dark Patterns: Why the Same Interface Affects Users Differently

Parinda Rahman, University of British Columbia, parinda1@student.ubc.ca
Ifeoma Adaji, University of British Columbia, ifeoma.adaji@ubc.ca

Format: Lightning Talk

ABSTRACT

Dark patterns are manipulative interface designs that exploit cognitive biases and information asymmetries to steer users toward outcomes benefiting providers at users' expense, and appear throughout e-commerce, social media, and account management [1]. HCI research has spent a decade naming and classifying them [2, 3], but these taxonomies remain descriptive and assume similar user responses. They tell us what dark patterns are, but not why they work, on whom, or under what conditions. Because much of this work comes from Western, English-language settings [4], we cannot explain why one mechanism reads as coercive to one user and routine to another.

This lightning talk presents a six-phase research plan for building a Cross-Cultural Framework for Dark Patterns, treating manipulation as something produced jointly by an interface mechanism and the cultural orientation of the person interpreting it [5]. Two reviews ground a four-layer schema linking pattern type, cultural susceptibility mapped onto Hofstede's dimensions [6], linguistic and visual-semiotic mechanisms, and harm severity.

The first review synthesises 61 studies, shifting from taxonomy to mechanism. Definitions remain fragmented around intent and harm, while sixteen psychological mechanisms map onto five pattern clusters, led by scarcity bias, cognitive overload, default effects, loss aversion, framing, and fear of missing out. The second synthesises 128 studies from 2015 to 2025. Pattern types retain their structure across regions, but their force depends on interpretation. In high power-distance contexts, for example, forced action and obstructive flows may be interpreted as routine procedure rather than coercion. Vulnerability is situated, surfacing where comprehension burden, time pressure, or refusal friction meets authority framing or incomplete localization.

Then came an exciting finding. Our eleven-rater expert content validation challenged a key schema assumption: experts could agree that a cultural cue was present while disagreeing about whether it belonged to the "high" or "low" end of a cultural dimension. That disagreement was revealing. Instead of forcing cultural signals into two opposing categories, we are now restructuring the cultural layer to represent them more flexibly. The validation did more than test the schema; it changed how we represent culture.

The talk connects findings to the remaining phases, including studies examining power distance and horizontal individualism-collectivism [7], and closes with a challenge: how can we operationalise culture without reducing it to rigid categories? By moving from identifying dark patterns to asking when, why, and for whom they become manipulative, this work aims to make deceptive-design research more culturally responsive.

Keywords: dark patterns, deceptive design, cross-cultural HCI, user vulnerability, persuasive technology ethics, interface auditing

References

[1] A. Mathur *et al.*, 'Dark patterns at scale: Findings from a crawl of 11K shopping websites', *Proceedings of the ACM on human-computer interaction*, vol. 3, no. CSCW, pp. 1–32, 2019.

[2] C. M. Gray, Y. Kou, B. Battles, J. Hoggatt, and A. L. Toombs, 'The dark (patterns) side of UX design', in *Proceedings of the 2018 CHI conference on human factors in computing systems*, 2018, pp. 1–14.

[3] C. M. Gray, C. T. Santos, N. Bielova, and T. Mildner, 'An ontology of dark patterns knowledge: Foundations, definitions, and a pathway for shared knowledge-building', in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–22.

[4] K. Seaborn *et al.*, 'Global and Transdisciplinary Perspectives on Dark Patterns and Deceptive Design Practice: ダークパターンとひとをだますデザインに関する国際的かつ学際的な交流', in *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 2025, pp. 1–6.

[5] R. Orji and Others, 'Persuasion and Culture: Individualism-Collectivism and Susceptibility to Influence Strategies', *Ppt@ Persuasive*, vol. 1582, pp. 30–39, 2016.

[6] G. Hofstede, 'Culture's consequences: Comparing values, behaviors, institutions and organizations across nations', *Ppt@ Persuasive*, vol. 1582, pp. 30–39, 2001.

[7] T. M. Singelis, H. C. Triandis, D. P. S. Bhawuk, and M. J. Gelfand, 'Horizontal and vertical dimensions of individualism and collectivism: A theoretical and measurement refinement', *Cross-cultural research*, vol. 29, no. 3, pp. 240–275, 1995.

The Digital Death Directory: Rating How Platforms Handle Death

Lara De Leon, University of Manitoba, deleonl@myumanitoba.ca
Dr. Celine Latulipe, University of Manitoba, celine.latulipe@umanitoba.ca

Format: Lightning Talk

ABSTRACT

What happens to our digital lives after we die? Can our families access our Facebook accounts or recover the years of photos on our Instagram? Our digital accounts can become both a source of connection and a practical challenge for the people we leave behind. Many of us have not planned for this, and it can be difficult to tell how carefully platforms have planned for it either. Companies spend enormous effort designing the onboarding experience, but far less attention goes toward what happens when the user never comes back. Instead, this responsibility is left to help-center articles that are hard to find, hard to parse, and inconsistent across companies. Grieving families encounter this gap at the worst possible time.

HCI has studied this gap for over a decade. *Thanatosensitivity*, introduced by Michael Massimi and Andrea Charise, describes an approach that integrates mortality, dying, and death into design [1]. It names the assumption underneath most account systems: that the user will always be alive to manage them. When that assumption breaks, grief becomes, in effect, a usability failure.

I developed the Digital Death Directory to make that failure visible. It is a consumer-facing website that grades major platforms, including Apple, Google, Facebook, and Instagram, on how well they handle posthumous account management. It functions like a Better Business Bureau for a person's digital life after death. Can a grieving family gain access to an account? Is there a clear process to memorialize or delete an account? Did the platform make any of this information easy to find in the first place? Answering those questions required deciding, in a defensible way, what "handling this well" means. So, we convened an expert panel spanning law, sociology, privacy and security, philosophy, and business to brainstorm and jointly develop a grading rubric. Their answers became the rubric underneath the Directory's grades.

In this talk, I'll discuss how the panel's expertise became a working rubric, what it measures, and what grades it produces for the platforms that hold our digital lives. The Directory does more than rank posthumous account policies. It turns thanatosensitive design into a public standard of evaluation and asks whether platforms anticipate death as part of the user lifecycle, support the people who navigate its aftermath, and make those commitments visible enough to be held accountable.

Keywords: digital legacy, digital death, thanatosensitivity, posthumous account management, human-computer interaction

References

[1] Massimi, M., and Charise, A. (2009). Dying, death, and mortality: Towards thanatosensitivity in HCI. *In CHI '09 Extended Abstracts on Human Factors in Computing Systems (CHI EA '09).* Association for Computing Machinery, New York, NY, USA, 2459–2468. https://doi.org/10.1145/1520340.1520349

Understanding the Privacy Experiences of Assistive Technology Users with Physical Disabilities

Negar Abbasi, Master of Arts in Information Systems and Technology, ngr1380@yorku.ca

Format: Lightning Talk

ABSTRACT

Assistive technologies (ATs) play a vital role in supporting the participation and independence of people with physical disabilities. These technologies may include augmentative and alternative communication (AAC) devices, alternative keyboards and switches, speech-to-text technologies, smart-home technologies, and other tools which enable communication and interaction with the world.

While assistive technologies can provide significant benefits, their use may also introduce privacy challenges for users. For example, ATs may collect, store, or share users' personal information [3]. In some cases, physical accessibility barriers may make it difficult for users to independently manage authentication or access to their devices and personal information [4]. At times, they may rely on caregivers, family members, or other support persons to assist with technology use due to accessibility barriers [4], [5]. Thus, ATs which are intended to increase users' independence may also result in them having less control over their privacy.

Although previous work has examined the privacy and accessibility of ATs, the perspectives of people with different types of physical impairments remain limited [1], [2]. This research addresses that gap by exploring how adults with physical disabilities who use assistive technologies understand, experience, and manage privacy in their everyday lives. Specifically, we explore:

- **RQ1**: How do assistive technology users with physical disabilities understand privacy?
- **RQ2**: What are their experiences with privacy?
- **RQ3**: What strategies do they use to manage privacy challenges?
- **RQ4**: How does caregiver or support-person involvement influence their privacy experiences?

To address these research questions, we conducted a user study with AT users with physical disabilities. Participants used a range of assistive technologies, including augmentative and alternative communication (AAC) devices, alternative switches, built-in accessibility features, and smart-home technologies. In the study, participants completed a *demographic questionnaire* followed by a *semi-structured interview* with the researcher. We are analyzing interview data using thematic analysis to identify common patterns across participants' experiences. In this lightning talk, we will share emerging themes from the study, focusing on participants' privacy experiences, the challenges they encounter, and the strategies they use to manage their privacy.

Keywords: assistive technology, privacy, accessibility, physical disability, human-computer interaction, usable privacy

References

[1] K. A. Mack, Y.-J. Chen, L. Zhang, D. Gurari, T. Sharma, Y. Wang, and L. Findlater, "The Accessibility, Security, and Privacy Nexus: Trends and Opportunities," in *Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility*, 2025, pp. 1–5, doi: 10.1145/3663547.3759688.

[2] A. Zezulak, F. Tazi, and S. Das, "SoK: Evaluating Privacy and Security Concerns of Using Web Services for the Disabled Population," *arXiv preprint arXiv:2302.13261*, 2023, doi: 10.48550/arXiv.2302.13261.

[3] K. Crawford, Y. X. Khoo, A. Kumar, H. Mentis, and F. Hamidi, "Decoding the Privacy Policies of Assistive Technologies," in *Proceedings of the 21st International Web for All Conference*, 2024, pp. 87–95, doi: 10.1145/3677846.3677850.

[4] B. Lewis and K. Venkatasubramanian, "'I...Got my Nose-Print. But it Wasn't Accurate': How People with Upper Extremity Impairment Authenticate on their Personal Computing Devices," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–14, doi: 10.1145/3411764.3445070.

[5] S. K. Kane, A. Guo, and M. R. Morris, "Sense and Accessibility: Understanding People with Physical Disabilities' Experiences with Sensing Systems," in *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, 2020, pp. 1–14, doi: 10.1145/3373625.3416990.

Can Search Interfaces Reduce Memory Demands for Students with Dyslexia?

Anita Seihoon, University of Regina, ast562@uregina.ca

Format: Lightning Talk

ABSTRACT

Exploratory academic search often requires students to formulate and reformulate queries, examine multiple results, and keep track of useful information as their understanding of a topic develops [1]. Prior research with university students with dyslexia has identified memory as an important factor in information seeking and documented persistent uncertainty during the search process [2,3]. This raises a design question: how can search interfaces externalize parts of the search process while balancing memory support with additional interface complexity?

Motivated by this question, SearchPath was designed to externalize search history and resource organization so that previous search activity does not have to be reconstructed entirely from memory. It provides a visual timeline of previous search activity, allows users to attach personal tags to saved resources, and enables those resources to be filtered using the tags.

This approach introduces an important accessibility design trade-off. Adding interface features can provide external support for remembering and organizing information, but those same features can also introduce additional information and interactions that users must process. SearchPath is currently being evaluated in an ongoing within-subjects study with a target sample of 16 undergraduate students with dyslexia. Participants complete exploratory academic search tasks using both SearchPath and a baseline interface with the same underlying academic search mechanism and result presentation. Interaction logs capture how participants search and organize information, while questionnaires examine their experiences with each interface.

This lightning talk will demonstrate SearchPath and discuss its ongoing empirical evaluation, focusing on the broader accessibility trade-off between providing external support for memory and introducing additional interface complexity.

Keywords: exploratory search, dyslexia, accessibility, academic search, human-computer interaction, inclusive design

References

[1] Marchionini, G. (2006). Exploratory search: From finding to understanding. *Communications of the ACM, 49*(4), 41–46. DOI: 10.1145/1121949.1121979.

[2] Cole, L., MacFarlane, A., & Makri, S. (2020). More than words: The impact of memory on how undergraduates with dyslexia interact with information. *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, 353–357. DOI: 10.1145/3343413.3378005.

[3] Beveridge, L., Makri, S., & MacFarlane, A. (2022). ‘I’m just not sure.’ The persistence of uncertainty in the information seeking of undergraduate students with dyslexia. *Information Research, 27*, Article isic2207. DOI: 10.47989/irisic2207.

Assistive, Inclusive, and Accessible Furniture using FDM 3D Printing

Nuo (Bella) Xu, Queen's University, bella.xu@queensu.ca
Sara Nabil, Queen's University, sara.nabil@queensu.ca

Format: Lightning Talk

ABSTRACT

Why do we often assume that accessible design has to look purely functional? In assistive products, functionality is often the first priority, yet appearance, personal preferences, and how an object fits into a living space also shape people's experiences. This project asks whether assistive furniture can support people with physical disabilities while remaining visually appealing, personal, and enjoyable to live with.

This ongoing project explores filament-based Fused Deposition Modeling (FDM) 3D printing as a medium for personalized and inclusive furniture for people with physical disability, focusing on limited hand dexterity. Access to FDM 3D printers through universities, libraries, and makerspaces, together with customizable digital models, creates opportunities for more people to participate in fabrication. Building on research in digital fabrication and personal fabrication [2, 3], we are developing a prototype of a drawer cabinet as a case study of how accessibility, aesthetics, and customization can be considered together.

Everyday interactions, such as grasping small handles or opening drawers, can be difficult for people with limited hand dexterity. The prototype design therefore explores alternative ways of interacting with furniture, including a push-to-open approach that can reduce reliance on precise gripping movements. In this work, we examine design considerations that can guide future evaluation.

Producing furniture-sized objects with FDM 3D printing also presents practical challenges. Previous work has examined how printing parameters affect accuracy in FDM-fabrication [1]. For example, printer size constraints, unsuccessful prints, CAD revisions, and material-specific settings require repeated adjustments. The prototype is still being fabricated, and these experiences inform our understanding of the opportunities and limitations of customized furniture fabrication.

Looking ahead, we aim to explore participatory and co-design approaches with people with physical disabilities, informed by prior work on mixed-ability collaboration in assistive technology design [4]. We also hope to investigate how modularity and material reuse could support adaptation as people's needs change. Ultimately, this project considers how digital

fabrication can enable furniture that is accessible, expressive, and customizable to individual preferences.

Keywords: 3D Printing, Inclusive Design, Furniture, Personalization, Material Reuse, Accessibility, Assistive Technology

References

[1] Zhan, Sheng, Seonyoung Youn, and Kavita Mathur. “Effect of Fused Deposition Modeling Printing Parameters on the Accuracy of Virtual Drape Using 3D Printed Textiles.” *International Conference on Human-Computer Interaction*. Cham: Springer Nature Switzerland, 2025.

[2] Abdullah, Muhammad, Romeo Sommerfeld, Bjarne Sievers, Leonard Geier, Jonas Noack, Marcus Ding, Christoph Thieme, Laurenz Seidel, Lukas Fritzsche, Erik Langenhan, Oliver Adameck, Moritz Dzingel, Thomas Kern, Martin Taraz, Conrad Lempert, Shohei Katakura, Hany Mohsen Elhassany, Thijs Roumen, and Patrick Baudisch. “HingeCore: Laser-Cut Foamcore for Fast Assembly.” *The 35th Annual ACM Symposium on User Interface Software and Technology (UIST ’22)*. ACM, 2022, 13 pages. DOI: 10.1145/3526113.3545618.

[3] Abdullah, Muhammad, Laurenz Seidel, Ben Wernicke, Mehdi Gouasmi, Anton Hackl, Thomas Kern, Conrad Lempert, Clara Lempert, David Bizer, Wieland Storch, Chiao Fang, and Patrick Baudisch. “PopCore: Personal Fabrication of 3D Foamcore Models for Professional High-Quality Applications in Design and Architecture.” *ACM Symposium on Computational Fabrication (SCF ’24)*. ACM, 2024, 14 pages. DOI: 10.1145/3639473.3665787.

[4] Baldwin, Mark S., Sen H. Hirano, Jennifer Mankoff, and Gillian R. Hayes. “Design in the Public Square: Supporting Assistive Technology Design Through Public Mixed-Ability Cooperation.” *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, CSCW, Article 155, 2019, 22 pages. DOI: 10.1145/3359257.

The Landmark Puzzle: How Landmark Quantity and Complexity Shape Spatial Learning of Commands in Graphical Interfaces

Tahsin Aziz, University of Regina, taj541@uregina.ca
Md Sami Uddin, University of Regina, sami.uddin@uregina.ca

Format: Lightning Talk

ABSTRACT

Large graphical user interfaces often contain many commands, making it difficult for users to learn and remember where each command is located. Visual landmarks can provide spatial reference points that support this learning [1]. However, an important interface design question remains unresolved: how many landmarks are useful, and how does their visual complexity affect spatial learning? Adding landmarks may make locations easier to distinguish, but too many or overly complex landmarks could also create visual interference.

This proposed PhD research investigates the combined effects of landmark quantity and visual complexity on learning and recalling command locations. The central goal is to determine whether increasing the number of landmarks consistently helps users or whether its benefits depend on how visually complex those landmarks are. Studying these factors together is important because practical interface designs must balance informative spatial cues against visual clutter.

We propose a controlled experiment in a spatially stable graphical interface containing 160 icons. Participants will perform repeated command-selection tasks as the number and design complexity of coloured landmark regions vary, with a no-landmark condition for comparison. The experiment will examine learning and recall performance through selection time and errors. By comparing performance across conditions, we aim to determine whether additional landmarks improve spatial memory, yield diminishing returns, or interfere with learning, and whether these patterns vary with landmark complexity.

The expected contribution is evidence-based guidance for designing spatial cues in command-dense interfaces. This work also builds on research into intrinsic landmarks, including icon-size variation [2], and may inform future studies of how different types of visual landmarks can support efficient interaction.

Keywords: spatial memory, graphical user interfaces, artificial landmarks, landmark quantity, visual complexity

References

[1] Md. Sami Uddin, Carl Gutwin, and Andy Cockburn. 2017. The Effects of Artificial Landmarks on Learning and Performance in Spatial-Memory Interfaces. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17). Association for Computing Machinery, New York, NY, USA, 3843–3855. https://doi.org/10.1145/3025453.3025497

[2] Tahsin Aziz and Md. Sami Uddin. 2026. Size Does Matter: Effects of Icon Size Variation on Spatial Learning in Interfaces with Large Command Sets. In Proceedings of the 52nd Graphics Interface Conference (GI '26). Association for Computing Machinery, New York, NY, USA, Article 33, 1–12. https://doi.org/10.1145/3831423.3831446

CAN-CWIC 2026

University of Waterloo

Abstracts for Presentations and Lightening Talks

Session: Journeys in Tech

Building Your Brand from the Inside Out: Making Your Strengths Visible in Computing

Dilara J. Leamen, Kyndryl / Ontario Tech University, dj.leamen@ontariotechu.net

Format: Lightning Talk

ABSTRACT

Technical ability is only one part of building a career in computing. For students and early-career professionals, the challenge is often recognizing what makes them valuable, communicating those strengths clearly, and creating opportunities for others to see them. This lightning talk introduces an inside-out approach to personal branding that begins with self-awareness rather than self-promotion.

Drawing on practical experience as a student and early-career AI and machine learning developer, the session explores why capable people can remain professionally invisible. Imposter syndrome [1], comparison with peers, and the tendency to minimize skills can make it difficult to articulate what we bring to a team or community. Rather than treating personal branding as a polished online persona, the inside-out approach treats it as an intentional extension of the skills, interests, relationships, and contributions that already exist.

The session presents a three-stage framework. First, participants identify their strengths through pattern hunting, an energy test, and the "compliment clue": noticing what others repeatedly ask for help with, what work gives them energy, and what positive feedback they tend to dismiss. Second, they incorporate those strengths into their professional brand by deliberately choosing communities, leadership opportunities, research, projects, and public-facing activities that reinforce what they want to be known for. Finally, they translate those experiences into professional impact by communicating achievements consistently and authentically.

Participants will leave with practical exercises for identifying their own strengths and strategies for making those strengths visible through campus involvement, professional communities, networking, and online presence. The goal is not to become an influencer or manufacture a persona. It is to help emerging computing professionals recognize the value they already possess, communicate it with confidence, and build a professional identity that is both intentional and authentic.

Keywords: personal branding, career development, early-career computing professionals, professional identity, women in computing, confidence, communication

References *(if any)*

[1] Paulise, L. 2023. 75% Of Women Executives Experience Imposter Syndrome In The Workplace. Forbes. https://www.forbes.com/sites/lucianapaulise/2023/03/08/75-of-women-executives-experience-imposter-syndrome-in-the-workplace/

Title: Beyond the Numbers: Lessons from My First Analyst Internship

Zhiye (Zoe) Su
University of Toronto Faculty of information
Format: Lightning Talk

ABSTRACT

During my internship as a Quantitative Research Analyst at Worldpanel by Numerator, I had the opportunity to work with large-scale consumer panel data and support projects related to market performance, consumer behaviour, and business decision-making. Drawing on this experience, this lightning talk will reflect on three lessons that became especially meaningful to me: looking beyond individual tasks to understand the system and workflow, learning to respond constructively to mistakes and data quality issues, and recognizing the importance of human side like communication, customer needs, and team culture alongside technical skills.

One reflection from the internship was that data analysis rarely exists in isolation. Becoming curious about where data comes from, how it is collected and processed, and how analytical outputs eventually support business decisions helped me understand the wider context behind my own tasks. Observing how technologies such as AI, OCR, and automation were incorporated into existing workflows also encouraged me to think more broadly about how technology changes the way people work.

The experience also changed how I viewed mistakes. Data-related work requires close attention to detail, but errors and unexpected issues can still occur. Instead of seeing mistakes only as failures, I gradually learned the importance of identifying where problems originated, asking questions when necessary, seeking practical solutions, and developing stronger validation habits. This helped me become more comfortable taking responsibility while also treating mistakes as opportunities to improve my working process.

Finally, the internship made me appreciate the human side of professional growth. Communication with colleagues and clients, exposure to different working styles, and supportive mentorship all shaped how much I was able to learn. It also made me realize that internships can help us understand not only what type of work interests us, but also what kind of team and working environment allows us to contribute and grow.

Through this talk, I hope to share an early-career perspective on internships as spaces not only for developing technical skills, but also for building broader professional awareness and discovering how we work, learn, and grow.

Keywords: Consumer Insights; Data Analytics; Systems Thinking; Early-career perspective

CAN-CWIC 2026

University of Waterloo

Abstracts for Presentations and Lightening Talks

Session: Theory and Scientific Computing

Numerical Approximation of Exercise Boundary of American Options

Yifan Qu, Department of Computer Science, University of Toronto, quyf@cs.toronto.edu
Christina Christara, Department of Computer Science, University of Toronto, ccc@cs.toronto.edu

Format: Lightning Talk

ABSTRACT

This work falls in the general area of computational methods for which the accuracy of the computed quantities as well as the efficiency of the resulting calculations are equally important. The work focuses on the computation of American option prices and of the so-called exercise boundary (also known as free boundary) associated with them.

American options are among the most widely traded financial contracts. An American option gives its holder the right, but not the obligation, to buy or sell a stock at a fixed price at any time before the option expires. There is no explicit formula that gives the price of an American option, therefore, computational methods are used for its approximation. These methods often involve the well-known Black–Scholes model, which is a parabolic partial differential equation (PDE). Many computational methods exist for calculating the value of an American option, and their accuracy has been well studied. Far fewer methods address an equally important question: when is the best time to exercise the option? The answer is described by the "exercise boundary," the stock price at which exercising becomes the optimal choice. This boundary is hard to compute because no formula describes it directly, and it shifts as time proceeds and the option approaches expiry. Simple numerical techniques produce a staircase-like boundary instead of a smooth curve, making the exercise decision less reliable.

To find the exercise boundary, we first solve the Black-Scholes PDE for American options by using the penalized formulation and standard finite difference methods to compute both the option prices and their sensitivities to the model inputs (given by first and second derivatives of the option price with respect to the stock price), commonly known as the Greeks, ensuring second-order convergence for both.

We then investigate several techniques for estimating the free boundary based on different local interpolants of the values or derivative values in the PDE region, extrapolated towards the exercise region. From theory, we know that these interpolants must satisfy some continuity or continuity of derivative conditions. We compute the free boundary as a solution to these conditions.

Our goals are a boundary that moves steadily in one direction as the option approaches expiry, matching the theory and avoiding the staircase effect, and accuracy that improves predictably as the computational method uses more data points. Furthermore, the computational complexity is linear (optimal) in the number of data points.

Joint work with Christina Christara.

Keywords:Partial Differential Equations, Finite Difference Method, Free Boundary Estimation, American Options

References

[1] Christara, C. and Leung, N. C.-H. (2018). Analysis of quantization error in financial pricing

via finite difference methods. SIAM Journal on Numerical Analysis, 56(3):1731–1757.
[2] Forsyth, P. A. and Vetzal, K. R. (2002). Quadratic convergence for valuing American options using a penalty method. SIAM Journal on Scientific Computing, 23(6):2095–2122.
[3] Kim, B. J., Ahn, C., and Choe, H. J. (2013). Direct computation for American put option and free boundary using finite difference method. Japan journal of industrial and applied mathematics, 30(1):21–37.
[4] Kreiss, H.-O., Thom´ee, V., and Widlund, O. (1970). Smoothing of initial data and rates of convergence for parabolic difference equations. Communications on Pure and Applied Mathematics, 23(2):241–259.
[5] Wang, D., Serkh, K., and Christara, C. (2023). A high-order deferred correction method for the solution of free boundary problems using penalty iteration, with an application to American option pricing. Journal of Computational and Applied Mathematics, 432:115272.
[6] Wu, L. and Kwok, Y.-K. (1997). A front-fixing finite difference method for the valuation of American options. Journal of Financial Engineering, 6(4):83–97.
[7] Zhu, S.-P. (2006). An exact and explicit solution for the valuation of American put options. Quantitative Finance, 6(3):229–242.

The Information Complexity of Decision Trees

Avantika Agarwal, University of Waterloo, a243agar@uwaterloo.ca
Shalev Ben-David, University of Waterloo, shalev.b@uwaterloo.ca
Eric Blais, University of Waterloo, eric.blais@uwaterloo.ca

Format: Lightning Talk

ABSTRACT

We define and study a measure of information complexity for randomized decision trees. We show that the information complexity of randomized decision tree is equal to the logarithm of the amortized worst-case randomized tree size complexity of computing a function f. That is, when computing f on n inputs, the logarithm of the randomized tree size is exactly equal to the amount of information needed to compute the function.

The talk will be self-contained; no background is assumed in information theory or query complexity. Based on joint work with Shalev Ben-David and Eric Blais [1].

Keywords: Query Complexity, Information Theory, Decision Trees

References

[1] Avantika Agarwal, Shalev Ben-David, Eric Blais (2026). The Information Complexity of Decision Trees. arXiv preprint.

Vertex Coloring in Sublinear Time and Space

Sepehr Assadi, University of Waterloo, sepehr@assadi.info
Helia Yazdanyar, University of Waterloo, hyazdanyar@uwaterloo.ca

Format: Lightning Talk

ABSTRACT

A basic graph coloring problem asks us to assign colors to vertices so that adjacent vertices receive different colors. For an n-vertex graph G with maximum degree Δ, a greedy algorithm finds a coloring using $\Delta+1$ colors. Achieving this guarantee with limited memory, time, or parallel resources is more challenging.

The **palette sparsification theorem (PST)** of Assadi, Chen, and Khanna (SODA 2019) [1] states that sampling a list of $O(\log n)$ colors from $\{1,\ldots,\Delta+1\}$ for every vertex independently and uniformly, with high probability, allows for finding a $(\Delta+1)$ vertex coloring of G by coloring each vertex only from its sampled list. PST naturally leads to a host of **sublinear algorithms** for $(\Delta+1)$ vertex coloring, including in semi-streaming, sublinear time, and MPC models, which are all proven to be nearly optimal and in the case of the former two are the only known sublinear algorithms for this problem.

While being a quite natural and simple-to-state theorem, PST suffers from two drawbacks. Firstly, all its known proofs require technical arguments that rely on sophisticated graph decompositions and probabilistic arguments. Secondly, finding the coloring of the graph from the sampled lists in an efficient manner requires a considerably complicated algorithm.

In this talk, I will present our **asymmetric palette sparsification theorem (APST)**, which allows different vertices to receive lists of different sizes and only bounds the *average* size of these lists. The benefit of this weaker requirement is that we can now easily show the graph can be $(\Delta+1)$ colored from the sampled lists using the standard greedy coloring algorithm. This way, we can recover nearly-optimal bounds for $(\Delta+1)$ vertex coloring in all the aforementioned models using algorithms that are much simpler to implement and analyze.

I will focus on the intuition behind asymmetric sampling and why it makes greedy coloring possible. This talk is based on joint work with Sepehr Assadi, published at SOSA 2025.

Keywords: graph coloring, palette sparsification, sublinear algorithms

References

[1] Sepehr Assadi, Yu Chen, and Sanjeev Khanna. Sublinear algorithms for $(\Delta+1)$ vertex coloring. In Timothy M. Chan, editor, *Proceedings of ACM-SIAM Symposium on Discrete Algorithms, SODA 2019*, pages 767–786. SIAM, 2019.

Bounds for Hardness Condensation in the Query Model

Chandrima Kayal, Université Paris Cités, kayal@irif.fr
Rajat Mittal, Indian Institute of Technology Kanpur, rmittal@cse.iitk.ac.in
Sai Soumya Nalli, University of Waterloo, ssnalli@uwaterloo.ca
Manaswi Paraashar, Indian Institute of Technology Hyderabad, manaswi@cse.iith.ac.in
Karthikeya Polisetty, Google Research India, CS22B026@smail.iitm.ac.in
Jayalal Sarma, Indian Institute of Technology Madras, jayalal@cse.iitm.ac.in
Nitin Saurabh, Indian Institute of Technology Hyderabad, nitin@cse.iith.ac.in

Format: Lightning Talk

ABSTRACT

For any Boolean function $f: \{0,1\}^n \to \{0,1\}$ with a complexity measure having value $k \ll n$, is it possible to restrict the function f to $\Theta(k)$ variables while keeping the complexity preserved at $\Theta(k)$? In other words, can a Boolean function be reduced to a much smaller number of relevant variables without significantly reducing its computational complexity? This question, in the context of query complexity, was recently studied by Göös, Newman, Riazanov and Sokolov [1]. They showed, among other results, that query complexity cannot be condensed losslessly. They asked if complexity measures like block sensitivity or unambiguous certificate complexity can be condensed losslessly?

In this work[2], we show that decision tree measures like block sensitivity and certificate complexity cannot be condensed losslessly. That is, there exists a Boolean function f such that any restriction of f to O(M(f)) variables has M(·)-complexity at most $\tilde{O}(M(f)2/3)$, where $M \in \{bs, fbs, C, D\}$. This also improves upon a result of Göös, Newman, Riazanov and Sokolov [1].

We also complement the negative results on lossless condensation with positive results about lossy condensation. In particular, we show that for every Boolean function f there exists a restriction of f to O(M(f)) variables such that its M(·)-complexity is at least $\Omega(M(f)1/2)$, where $M \in \{bs, fbs, C, UCmin, UC1, UC, D, de\tilde{g}, \lambda\}$. We also show a slightly weaker positive result for randomized and quantum query complexity.

Keywords: computational complexity, query complexity, Boolean functions

References

[1] Mika Göös, Ilan Newman, Artur Riazanov, and Dmitry Sokolov. Hardness condensation by restriction. In Bojan Mohar, Igor Shinkar, and Ryan O'Donnell, editors, Proceedings of the 56th Annual ACM Symposium on Theory of Computing, STOC 2024, Vancouver, BC, Canada, June 24-28, 2024, pages 2016–2027. ACM, 2024.
[2] Chandrima Kayal, Rajat Mittal, Sai Soumya Nalli, Manaswi Paraashar, Karthikeya Polisetty, Jayalal Sarma, and Nitin Saurabh. *Bounds for Hardness Condensation in the Query Model*. In 41st Computational Complexity Conference (CCC 2026), LIPIcs, Vol. 383, pp. 30:1–30:20, 2026. Schloss Dagstuhl–Leibniz-Zentrum für Informatik. DOI: 10.4230/LIPIcs.CCC.2026.30.

CAN-CWIC West 2026

Simon Fraser University & University of British Columbia

Abstracts for Presentations

Session 1A

The WEST Effect: How Female-Only Clubs Benefit Women in STEM

Rahma Ismail, Women in Engineering, Science & Technology Club at the University of Victoria, rahmaismail@uvic.ca
Kariman Hamad, Women in Engineering, Science & Technology Club at the University of Victoria, karimanhamad@uvic.ca
Presentation

ABSTRACT

Background

Women remain underrepresented in engineering and computing, and many students report isolation and limited access to hands-on, industry-facing experience early in their degrees. Women in Engineering, Science & Technology (WEST), founded at the University of Victoria in 2022, is a student-run, female-only club that addresses this. This work asks: can such a club create a repeatable pathway from first project to leadership? An earlier version was presented at CCWESTT 2026 in Calgary this past May.

Approach

WEST's core model is the semester-long, student-led design project, run with industry and campus partners, where students solve real problems for real stakeholders. Partners include FortisBC, Schneider Electric, Introba, UVic Robotics, Workday, Victoria Hand Project, Fraser Health, and Shift (Redbrick), with several more upcoming. We analyze participation records across nine projects, collect partner feedback, and track what participants do afterward: taking leadership roles, joining the WEST executive, or joining other Engineering and Computer Science clubs. Several participants and executives have also received co-op opportunities with partner companies and others.

Results

Across nine projects, 165 students participated (over 300 across all WEST initiatives). Of these, 89 (54%) went on to take leadership roles or join other clubs, and over 20% joined the WEST executive. Conversion varied: UVic Robotics had the highest rate (73%, 8 of 11 students), while Introba 2024 and one FortisBC project each reached 57%. Cohorts grew from roughly 10 students in early projects to 20-25 in later ones, and the four FortisBC projects together engaged 61 students. The organizing team grew from 5 to 30 members (500%), and partners have continued working with WEST across multiple years.

Discussion and Contribution

The project itself appears to be the on-ramp: a structured, low-barrier first experience in a women-only setting that leads many participants into leadership and the wider ECS community. Our contribution is a replicable model for student organizations with three parts: (1) multi-semester projects co-designed with industry partners; (2) inclusive practices, including

accessibility, psychological safety, and intersectionality; and (3) an explicit leadership pipeline from participant to executive. Challenges include sustaining mentorship quality as cohorts grow, coordinating timelines with external partners, and maintaining continuity as leaders graduate. Because these are descriptive data from one institution, they show association, not causation; future work could compare participants with non-participants.

Conclusion

Small, student-led, community-rooted initiatives can produce measurable leadership outcomes for women in STEM. By sharing WEST's results and lessons, we hope to help other institutions build similar pathways. More information: https://onlineacademiccommunity.uvic.ca/uvicwest/, uvicwest.official@gmail.com

Traceable Question Answering over SEC 10-K Filings: Every Number Cited, Unanswerable Questions Refused

Shuo Yan, Simon Fraser University, sya236@sfu.ca

Format: Presentation

ABSTRACT

Financial question-answering systems, especially LLM-based ones, often return fluent answers with numbers that cannot be traced to a source, or answer questions the document does not cover. I present a traceable question-answering system over SEC 10-K filings that answers only from the filing, cites the exact source location of every answer, and refuses questions the filing cannot answer. The system uses no generative model. It routes each question to one of three paths: a structured lookup using the filing's machine-readable tags (XBRL) for standard metrics, structure-aware hybrid retrieval with verbatim cell or span extraction for everything else, or an explicit refusal. I study two questions: (RQ1) whether chunking by 10-K Item, table and note improves answer and citation accuracy over fixed-length chunking, and (RQ2) whether a confidence- and scope-based abstention rule refuses out-of-filing questions without refusing answerable ones. Chunks carry a deterministic address line — filing, Item, note and table number — taken from the filing's own structure, so retrieval can match a chunk to where it sits; when a retrieved chunk cites a note or appendix, retrieval follows that reference one step to pull the target chunk in. Evaluation uses a human-verified gold set of 100 questions from Microsoft's 10-K (30 dev / 70 held-out test), reporting answer accuracy, citation accuracy, retrieval recall, refusal precision and recall, coverage and selective accuracy. Planned extensions add Alphabet and Meta, an 80-item red-team suite spanning eight attack types, and a standard retrieval-augmented generation baseline.

Keywords: traceable question answering, financial filings, citation, refusal, XBRL

CAN-CWIC West 2026

Simon Fraser University & University of British Columbia

Abstracts for Presentations

Session 1B

Clinical User-Centered Explainable AI for Medical Image Analysis: A Critical Technical Practice

Weina Jin, School of Computing Science, Simon Fraser University. weinaj@sfu.ca
Ghassan Hamarneh (PhD supervisor), School of Computing Science, Simon Fraser University.

Format: Presentation

ABSTRACT

This presentation is based on my PhD thesis research. My thesis explores the overarching research question: how do we-both as the collective technical community and as the individual technical developers and researchers-develop explainable artificial intelligence (XAI) and medical image analysis (MIA) techniques responsibly to genuinely benefit patients' healthcare and society? XAI aims to explain AI model predictions in human-understandable ways, and it has the potential to empower clinical users to verify and challenge AI predictions and hold technical developers accountable. My thesis first conducts a computational evaluation and a neurosurgeon user study, which identify that the existing XAI development agenda fails to fulfill clinical requirements, and the evaluated XAI technique cannot effectively improve clinicians' task performance; furthermore, a critical examination of the dominant XAI evaluation criterion of plausibility shows that this criterion is unscientific and unethical to use. These symptoms expose systemic issues in the technical community. Drawn from relational ethics and feminist epistemology, I diagnose the root causes as the power imbalance between for-profit sectors and end users/the public.

Based on the diagnosis, my thesis then provides a systematic intervention denoted as critical technical practice (CTP), analogous to strengthening the technical immune system. My thesis demonstrates CTP across different levels of technical practice, defending against the influence of unjust power on the values, technical standards, and specific decisions in technical development. At the high level, by critically examining the existing general vision of AI development, my thesis provides an alternative imagination of AI, "AI for just work", that outlines an alternative worldview, values, foundational assumptions, and technical properties of AI grounded in the rights and power of stakeholders and the public.

At the lower level, this new imagination is further developed as technical development guidelines for two tasks: 1) For the medical image synthesis (MISyn) task, my thesis derives two intrinsic limits of MISyn based on the new foundational assumption, and provides the ESIM checklist (ethical synthesis of images for medicine) to account for the intrinsic limits in problem formulation, design, and evaluation of MISyn techniques. 2) For the XAI task, my thesis develops three XAI frameworks; the EUCA framework supports end-user-centered participatory design and evaluation of XAI techniques from the initial development phase; the clinical XAI guidelines encode user-centered clinical requirements as specific XAI design and evaluation criteria; the responsible XAI framework provides systematic support to encode justice assumptions in XAI design and evaluation.

Keywords: Medical image analysis; Explainable AI; Human-centered AI; AI ethics; Critical technical practice

Towards Clinical Implementation of Artificial Intelligence in Cancer Care: Concept Mapping Analysis of Provincial Workshop Findings

Hong Hao Xu, University of Pittsburgh School of Medicine; Faculty of Medicine, University of British Columbia
Chhavi Nayyar (presenting author), Faculty of Science, University of British Columbia, chhavi09nayyar@gmail.com
Daniel Hilbers, Faculty of Medicine, University of British Columbia
Jonathan Avery, Supportive Care Service, BC Cancer
Srinivas Raman, Faculty of Medicine, University of British Columbia; BC Cancer – Vancouver
Alan T. Bates, Supportive Care Service, BC Cancer; Faculty of Medicine, University of British Columbia
Cristina Conati, Department of Computer Science, University of British Columbia
Ahmad Fayaz–Bakhsh, Tehran University of Medical Sciences; University of British Columbia
John–Jose Nunez, Supportive Care Service, BC Cancer; Faculty of Medicine, University of British Columbia

Format: Presentation

ABSTRACT

Artificial intelligence (AI) is increasingly discussed as a tool for cancer care, but little is known about what patients, clinicians, and other professionals see as the main concerns, benefits, and priorities for implementing it. We ran a concept mapping study, a mixed-methods approach that turns stakeholder ideas into a structured visual map, at a provincial Cancer Summit in Vancouver, Canada. Forty-eight workshop participants generated 265 statements about AI in cancer care, which we consolidated to 100. A purposively sampled expert panel of 13 members sorted the statements by similarity and rated each on importance and feasibility. We applied nonmetric multidimensional scaling and hierarchical clustering (Ward's method), compared clusters with paired t-tests and effect sizes, and used a go-zone analysis to place statements into priority quadrants.

The analysis produced two clusters, each with four subclusters: "Challenges and Safeguards for AI Implementation" (62 statements) and "Clinical Benefits and Efficiency Gains" (38 statements). The benefits cluster was rated higher than the safeguards cluster on both importance (3.50 vs. 3.14; Cohen's d = 0.94) and feasibility (3.62 vs. 2.99; d = 1.59). Clinical and nonclinical panelists rated most clusters similarly; the one significant difference was that clinical panelists rated the benefits cluster as more feasible (P = .029, Hedges' g = 0.465).

In the go-zone analysis, 36% of statements were both high in importance and high in feasibility, 13% were important but hard to implement, 16% were less important but easy, and 35% were less important and hard. The high-priority quadrant was dominated by workflow-embedded applications, such as automating routine tasks and improving documentation efficiency. Agreement among panelists was moderate at the group level (ICC(2,k) 0.67 for importance and 0.77 for feasibility) but low for single raters (ICC(2,1) 0.08 and 0.13).

Participants saw the largest near-term opportunity in AI that fits existing clinical workflows and reduces routine burden, while safeguards and governance were rated as important but harder to achieve. These findings can help health-system leaders sequence AI projects, starting with workflow-aligned use cases while planning governance in parallel. Limitations include a single

summit, a small panel, and that concept mapping captures perceptions, not outcomes. This work combines computing, data science, and health research, and is published in Implementation Science Communications [1].

Keywords: AI in healthcare, cancer care, concept mapping, implementation science, stakeholder engagement

References *(if any)*

[1] Xu, H. H., Nayyar, C., Hilbers, D., Avery, J., Raman, S., Bates, A. T., Conati, C., Fayaz–Bakhsh, A., & Nunez, J.–J. (2026). Towards clinical implementation of artificial intelligence in cancer care: concept mapping analysis of provincial workshop findings. Implementation Science Communications. https://doi.org/10.1186/s43058-026-01075-x

CAN-CWIC West 2026

Simon Fraser University & University of British Columbia

Abstracts for Presentations

Session 2A

Mapping Sustainability Education in Canadian Computer Engineering

Esther Roorda, University of British Columbia, estherr@ece.ubc.ca
Sathish Gopalakrishnan, University of British Columbia, sathish@ece.ubc.ca

Format: Presentation

ABSTRACT

Given the socio-environmental impacts of their work, sustainability should be a critical concern for Electrical and Computer Engineering (ECE) students. eWaste is the world's fastest growing waste stream, with 62 million tonnes produced each year [1]. AI datacenter energy utilization is increasing dramatically, causing a surge in fossil fuel-based power generation [2,3,4]. The supply chains that support this industry impact communities and ecosystems around the world. To meaningfully address these environmental concerns, a paradigm shift is urgently required, and we argue that this shift begins with education. Our work [5] shows that across Canada (and in the US), ECE students receive the least sustainability education of all major engineering disciplines. Through analysis of course descriptions and curricula and interviews with students and educators, we have explored the existing landscape of sustainability education in Canada, and in this presentation, we propose avenues for change.

We performed a qualitative analysis of all Canadian engineering course descriptions and program documents, categorizing them by sustainability coverage amount and type. Our analysis shows that ECE students have fewer sustainability-related electives and core courses than any other major engineering discipline. Accreditation requires that all engineering students take sustainability-related courses, but ECE students typically take these courses from outside of their departments. Many students then report not understanding how sustainability relates to computing, or not seeing how they could engage in climate action through their work. Students and educators also report a lack of engagement in some existing courses, with non-technical sustainability courses sometimes perceived as 'box-ticking requirements', or not in-line with dominant conceptualizations of engineering identity.

Based on our findings, we will discuss for five avenues for change:

- *Increasing sustainability coverage in technical electives:* There are many opportunities across technical courses to integrate concrete discussions of environmental impact.
- *ECE-specific core sustainability courses:* Fulfilling sustainability related accreditation requirements through discipline specific courses links student work more directly to sustainability.
- *Climate action as praxis*: Our prior work uses electronic repair as an example of experiential education, teaching by engaging students in concrete climate action.
- *Changing the 'hidden curriculum'*: Reevaluating how ECE and sustainability are presented in program descriptions could change student perceptions and bring more sustainability-oriented students into the discipline.
- *Community and student-oriented education*: Students have a role to play in changing the departmental culture, through practical climate action and climate leadership.

In our talk, we outline a roadmap of sustainability education in ECE based on our previous mixed methods analysis, and briefly discuss how we are enacting these changes through ongoing work.

Keywords: engineering education, sustainability education,

References

[1] Bald´e, C. P., Kuehr, R., Yamamoto, T., McDonald, R., D'Angelo, E., Althaf, S., Bel, G., Deubzer, O., Fernandez-Cubillo, E., Forti, V., et al. (2024). Global e-waste monitor 2024.
[2] International Energy Agency (2025). Energy and ai. Technical report, International Energy Agency, Paris
[3] Kollar, J. (2026). The fossil-ai nexus: petrostate capitalism, computing power, and the production of powered land. New Political Economy, pages 1–22.
[4] Martos, J. (2026). Proposed gas-fired power plants in the united states rise due to ai energy demand speculation, but remain largely in early development stage. Technical report, Global Energy Monitor.
[5] Roorda, E., & Gopalakrishnan, S. (2026). Structuring Sustainability Education: The Canadian Engineering Landscape. Proceedings of the Canadian Engineering Education Association (CEEA).

From Models to Megawatts: Efficiency and Environmental Impact of AI

Ananyaa Chopra, University of British Columbia, reachananyaa@gmail.com
Giulia Toti, University of British Columbia, giulia.toti@ubc.ca

Format: Presentation

ABSTRACT

Artificial intelligence is commonly evaluated by predictive accuracy, while the electricity, hardware, and infrastructure required to achieve that performance remain less visible. Green AI argues that computational efficiency should be considered alongside model quality [1]. Translating this principle into undergraduate education, however, requires more than reporting an emissions number. We present From Models to Megawatts, a self-contained Jupyter notebook assignment designed to teach students to quantify, critically interpret, and use environmental estimates in machine-learning decisions.

The assignment connects three activities. First, students calculate power, energy, and carbon-dioxide-equivalent emissions, including how electricity-grid carbon intensity changes the estimated emissions of identical computations. Second, they examine AI infrastructure and reporting boundaries, questioning what published inference estimates include and omit. Third, they conduct a nonlinear classification experiment using 10,000 synthetic observations and six model configurations: logistic regression, random forests with 100 and 500 trees, gradient boosting with 100 estimators, and multilayer perceptrons with 32 and 128 hidden units. Accuracy is assessed using stratified five-fold cross-validation repeated three times. Separately, CodeCarbon [2] estimates energy and emissions for standardized fit-and-predict workloads. Students construct an accuracy-emissions Pareto frontier and justify a model-selection decision.

One illustrative execution demonstrates the trade-off: increasing a random forest from 100 to 500 trees raised mean accuracy from 91.620% to 91.623% (0.003 percentage points), while estimated energy per standardized workload increased from 1.558 to 6.328 mWh (approximately 4.1-fold). Four configurations were Pareto-optimal; RF500 and the wider multilayer perceptron were dominated. These results describe one workload and execution environment, not universal model rankings or statistically established performance improvements.

Our contribution is the educational sequence: students progress from calculating resource costs to questioning measurement assumptions and defending context-sensitive technical choices. The assignment complements efforts to integrate sustainability into computing education [3], while acknowledging that local resource estimates do not capture full system or lifecycle impacts. Future work will evaluate student learning using a pre/post study.

Keywords: Green AI, computing education, energy literacy, CodeCarbon, Pareto frontier

References

[1] Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. Communications of the ACM, 63(12), 54–63.

[2] CodeCarbon (2026). CodeCarbon 3.3.1. github.com/mlco2/codecarbon.

[3] Peters, A.-K., et al. (2024). Sustainability in Computing Education: A Systematic Literature Review. ACM TOCE, 24(1), 13.

Investigating Holos Calibration for Climate and Sustainability Research

Fariba Choudhury, Athabasca University, fchoudhury1@learn.athabascau.ca

Jason Bernard, Athabasca University, cbernard@athabascau.ca

ABSTRACT

Holos estimates greenhouse-gas emissions and soil-carbon changes in Canadian farming systems [1], supporting sustainability research in an agricultural sector facing climate and food-security pressures [2,3]. Interpreting these estimates requires understanding how well model inputs reflect observed farm conditions. Calibration adjusts uncertain inputs to match observations, but different input combinations can reproduce similar outputs [4]. Input-recovery tests use known generating conditions to determine whether a close fit also reconstructs the correct inputs; this distinction is central to calibration because matching outputs alone need not identify the conditions that produced them [5].

An earlier global sensitivity analysis (GSA), which identifies inputs that contribute to variation in model outputs [6], measured how farm inputs contributed to variation in Holos's estimated gross greenhouse-gas emissions. Building on this work, we tested recovery of fertilizer nitrogen, starting soil-carbon stock and an offset to the historical temperature series from synthetic Holos outputs for one crop-farm case. Starting carbon initialized the 1985-2024 history, whereas the targets described 2024.

A genetic algorithm, simulated annealing and Nelder–Mead completed 135 noiseless trials in total. Each of three output sets had 45 trials: three methods, three known input combinations and five repetitions. We fitted total gross emissions, separate gases, or gases plus 2024 soil-carbon stock. These output sets tested whether additional observations could distinguish farm inputs that a total-emissions estimate alone could not resolve. Each noiseless trial evaluated 200 candidate input combinations against the targets, counting repeated combinations toward this limit. A further 45 trials at one input combination used 1% observation noise and 100 evaluations each. Successful input recovery required absolute errors within 5% of each input's allowed range. Output agreement required each absolute output error divided by its fixed reference scale to be at most 0.001.

Nitrogen was recovered in all 180 trials, with a mean absolute error of 0.55 kg N/ha. Including the 2024 soil-carbon stock improved recovery of the starting stock from 4/45 noiseless trials using total emissions and 5/45 using separate gases to 44/45. Local checks at the tested inputs found no gas response to starting soil carbon, explaining why emissions alone provided little information about this input.

With gases plus carbon, Nelder–Mead achieved output agreement in all 15 of its noiseless trials and recovered all three inputs in nine of those 15. Continuing these same searches from 100 to 200 evaluations increased joint input recovery from 0/15 to 9/15, although all had met the output-fit criterion at 100. Temperature could still be misidentified: one acceptable fit used a −0.69°C offset when the known offset was +0.50°C.

Assessing both output agreement and input recovery can guide data collection for Holos calibration: soil-carbon observations reduced ambiguity in starting carbon, while temperature remained uncertain despite acceptable output fits. The budget comparison shows that further search can improve input recovery even after the target outputs have been matched. Field validation across farms and years is needed before applying these findings to emissions-mitigation or soil-carbon management decisions. This approach provides a basis for testing whether observations adequately constrain inputs in other climate and environmental models, with the potential to strengthen the evidence behind sustainability assessments.

Keywords: Holos, Optimization, Input Recovery, Climate Modelling, Agricultural Sustainability

References

[1] Agriculture and Agri-Food Canada. n.d. Holos. https://agriculture.canada.ca/en/agricultural-production/holos

[2] Rahman, T. U., Faraz, A., Nawaz, T., Saud, S., Fahad, S., & Harrison, M. T. (2025). Towards sustainable solutions: Climate change and food security in a globalized world. *Food and Energy Security*, *14*(5), e70126.

[3] IPCC. (2023). Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC. https://doi.org/10.59327/IPCC/AR6-9789291691647

[4] Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, *249*(1-4), 11-29.

[5] Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society Series B Statistical Methodology*, *63*(3), 425-464. https://doi.org/10.1111/1467-9868.00294

[6] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). *Global Sensitivity Analysis The Primer*. Wiley.

CAN-CWIC West 2026

Simon Fraser University & University of British Columbia

Abstracts for Presentations

Session 2B

Who Governs Decentralized AI? Accountability, Agency, and Ethical Governance

Sandhya Joshi, Vancouver Island University, Nanaimo, BC, Canada, sandhya.joshi@viu.ca
Yulia Bobkova, Vancouver Island University, Nanaimo, BC, Canada, yulia.bobkova@viu.ca
Saksham Sohal, Vancouver Island University, Nanaimo, BC, Canada, saksham.sohal@viu.ca
Roy Jonuk, Vancouver Island University, Nanaimo, BC, Canada, roy.jonuk@viu.ca
Ajay K. Shrestha, Vancouver Island University, Nanaimo, BC, Canada, ajay.shrestha@viu.ca

Format: Presentation

ABSTRACT

Keeping data on devices does not necessarily give people greater control over artificial intelligence (AI). Federated learning and edge intelligence distribute aspects of AI training and processing across devices and organizations. However, important decisions about how these systems are designed, managed, and used may still be controlled by a small number of organizations. This raises an important question: when AI is decentralized, who is accountable for its decisions and their consequences?

This study draws on an ongoing literature review of research in computing, law, and related fields to examine how accountability is understood and addressed in federated learning and edge intelligence. Our initial examination of the literature has helped identify key questions about who makes decisions, how systems are monitored, who is responsible for explaining outcomes, and how affected individuals can challenge decisions or seek remedies. As the review progresses, we will examine these questions across the AI lifecycle, considering the roles of developers, organizations, users, and other affected groups.

Building on this work, we are developing an Authority–Accountability–Agency (A3) matrix to examine who holds decision-making power and how responsibilities are divided. Preliminary findings suggest that keeping data local does not always give individuals meaningful control, and technical monitoring may not necessarily lead to accountability. They also suggest that people who participate in or are affected by AI systems may have limited opportunities to influence decisions and outcomes. We will build on these initial findings as we continue reviewing the literature.

Ultimately, this research aims to identify gaps in existing approaches to accountability and develop a practical framework to support more responsible, transparent, and inclusive decentralized AI systems.

Keywords: Decentralized Artificial Intelligence, Federated Learning, Edge Intelligence, AI Accountability, Ethical AI Governance, Human Agency

LLMs redraw argumentative patterns across languages

Gurneet Gill, Simon Fraser University, gkg36@sfu.ca
Jasleen Kaur, Simon Fraser University, jka349@sfu.ca
Stella Maltcheva, Simon Fraser University, sma345@sfu.ca
Khoa Ha Minh Duong, Simon Fraser University, hmd6@sfu.ca
Rodrigo Armando Anasco Santillan, Simon Fraser University, raa69@sfu.ca
Maria Kapitanenko, Simon Fraser University, mka260@sfu.ca
Ankita Gupta, University of Massachusetts Amherst, ankitagupta@cs.umass.edu
Marzena Karpinska, Simon Fraser University, karpinsk@sfu.ca

Format: Presentation

ABSTRACT

Large language models (LLMs) are increasingly used to draft opinion articles across languages, yet research on homogenization in argumentative writing has largely focused on English. In this work, we compare 4,000 human-written opinion articles with 43,750 articles generated by five LLMs across eight languages, using three complementary methods: (1) discourse structure analysis, (2) Toulmin-based argument analysis, and (3) data-driven feature discovery. Using only the 45 discovered features, a classifier identifies the language of a human opinion article 42% of the time (chance 12.5%), consistent with previous findings in contrastive rhetoric [1]. LLMs retain some cross-linguistic contrasts while flattening others. For example, Japanese still uses more rhetorical questions than English, but the human hedging gap between English and Hindi nearly disappears in LLM articles (shrinking from 14.2 to 1.3 percentage points). Across all eight languages, LLM articles use less irony, colloquial language, and personal discrediting, but more concession, and more often end by restating the thesis. To support their claims, LLM articles appeal more to general principles, whereas human articles rely more on testimony and statistics. In experiments with five prompting variations, certain features move closer to human writing (e.g., providing the author's stance restores the writer's voice), but no prompt eliminates this less conversational, less confrontational style. Some prompts even widen other gaps: providing the author's stance or requesting diverse articles in a single prompt increases thesis-first openings from 70% to 91% and 95%, respectively. We release our code and data to support research on multilingual homogenization in LLMs.

Keywords: Argument structure, argumentative writing, multilingual LLM evaluation, homogenization, opinion articles, machine-generated text, cross-lingual analysis

References

[1] U. M. Connor, *Contrastive Rhetoric: Cross-Cultural Aspects of Second-Language Writing*. Cambridge, UK: Cambridge University Press, 1996.

CAN-CWIC West 2026

Simon Fraser University & University of British Columbia

Abstracts for Presentations

Session 3C

Tool-Aware Ergodic Stroke Planning and Voxel Simulation for Robotic Sgraffito

Chloe Yip, SFU School of Computing Science, chloe_yip@sfu.ca
Gaurika Sharma, SFU School of Computing Science, gaurika_sharma@sfu.ca
Siqi Zhou, SFU School of Computing Science, zhou_siqi@sfu.ca

In collaboration with Lidia Atanasova, Kathrin Dörfler, and Angela P. Schoellig @ TU Munich

Format: Presentation

ABSTRACT

Sgraffito is a decorative technique used in Renaissance architecture and ceramics, in which carving through a top layer of clay reveals a contrasting coloured layer underneath. The process is slow and repetitive, and a single mistake can ruin a slab that took hours to prepare. We present an end-to-end system that turns an artist's drawing into a robot carving plan and lets the artist preview the result before any irreversible cut is made.

Through a web interface, an artist uploads a drawing, selects a slab, and chooses a stylization (regions, outlines, hatching, or stipple), along with carving rules and constraints. Our pipeline then works in three stages: task decomposition, ergodic stroke placement under tool-footprint constraints, and voxel-simulated verification. Each shape is classified as one of three action primitives: traces for thin marks under 1.1 mm, dot presses for small round shapes under 80 mm^2, and scrapes for clearing larger areas. Regions are filled using ergodic coverage [1], where a greedy planner repeatedly selects the stroke that most reduces the gap between target and achieved coverage. The planner is tool-aware: by slicing the tool mesh at the carve depth, it keeps the tool from cutting outside the lines. We also found that tool tilt changes only the approach angle, not the cut width. The full plan is simulated at 0.25 mm voxel resolution [2, 3] and rendered in Blender [4].

We evaluated six designs. The share of each design that was cut ranged from 62.9% (Branch) to 92.4% (Pyramid), and four designs showed no over-cutting. The Mixed design, with many small dots and thin lines, had the most missed area. Overall, large shapes are carved more accurately, while details smaller than the tool remain the hardest. On the Sun design, coverage rose quickly and then levelled off, reaching 28%, 44%, 61%, and 76% after 10, 20, 40, and 210 strokes. The first 40 strokes delivered about 80% of the final coverage, suggesting that a short plan could provide a fast rough preview.

The work is currently simulation-only, and greedy placement stalls near 95% coverage, leaving small gaps. Next steps include executing planned strokes on a robot arm with real clay, comparing the results to simulation, and gathering feedback from artists using the interface.

Keywords: robotic sgraffito, ergodic coverage, stroke planning, tool-aware planning, voxel simulation, computational fabrication, digital art tools

References

[1] Mathew, G., & Mezić, I. 2011. Metrics for ergodicity and design of ergodic dynamics for multi-agent systems. *Physica D*, 240(4–5), 432–442.

[2] Heiden, E., Macklin, M., Narang, Y., Fox, D., Garg, A., & Ramos, F. 2021. DiSECt: A

differentiable simulation engine for autonomous robotic cutting. *Robotics: Science and Systems (RSS)*.

[3] Museth, K. 2013. VDB: High-resolution sparse volumes with dynamic topology. *ACM Transactions on Graphics*, 32(3).

[4] Blender Foundation. Blender. `https://www.blender.org`

CAN-CWIC West 2026

Simon Fraser University & University of British Columbia

Abstracts for Posters

Automated Detection, Correction, and Surface Fairing for Skin-to-Muscle Registration Artifacts

Ilian Khankhalaev, SFU, florykhan@gmail.com
Zhiyun Peng, SFU, zpa18@sfu.ca
Feifei Yu, SFU, fya39@sfu.ca
Nikolay Deinego, SFU, nikolaydeinego@gmail.com
Khoa Ha Minh Duong, SFU, hmd6@sfu.ca
Qiu Shi, SFU, lajiyoujian14@gmail.com
Eugene Fiume, SFU, eugene_fiume@sfu.ca

Format: Lightning Talk

ABSTRACT

Digital characters in film, games and virtual reality increasingly use biomechanical facial models, which make faces move realistically by simulating the muscles, fat and bone beneath the skin. Before a character can be simulated, a generic skin surface must be fitted, or registered, onto that character's own anatomy. This fitting is imperfect. The skin can end up with small bumps and ridges, and in some places the fat or muscle underneath pokes through it. Today, artists fix these problems by hand, which is slow and hard to repeat for every new character.

We study how to clean up the registered skin automatically. The difficulty is a trade-off. Smoothing removes bumps, but it can also pull the skin inward until the anatomy shows through. We therefore ask how much the skin can be smoothed while the anatomy stays covered, except at the eye and mouth openings, where it is meant to be visible.

Working in Autodesk Maya on a skin mesh of 15,905 vertices, we built detectors that locate rough regions, and an exact triangle intersection test that counts where the skin crosses the anatomy. Methods that enforced safety everywhere removed every crossing but stopped improving smoothness. An experiment with plain smoothing and no safety showed that the smoothing itself worked well, and that the global safety rules were what held it back. This led to our main method, a blocked smoother. It smooths freely, and whenever a step would push the skin through the anatomy, it undoes the move only for the vertices involved and freezes them.

Before cleanup, 93 skin triangles crossed the anatomy. Plain smoothing reduced roughness in the roughest regions by about 59%, but it created 143 crossing triangles. A prototype of the blocked smoother kept 92% of that improvement with no crossings, while freezing fewer than 4% of the smoothed vertices. The saved final mesh, which combines milder smoothing with about 130 manual edits, is 28% smoother overall and has 3 crossing triangles instead of 93.

All results come from one character, and the detectors have not been checked against labelled data. Next, we plan to automate the remaining problem regions, run the blocked smoother inside the full Maya pipeline, and extend the method from the face to the head, neck and shoulders to test how well it generalizes.

Keywords: digital humans, facial simulation, geometry processing, mesh smoothing, artifact detection

Imagining Technological Futures for Perinatal Mental Health and Well-being

Xueying Zhang, Simon Fraser University, xueying_zhang@sfu.ca
Nardos Solomon, Simon Fraser University, nms15@sfu.ca
Felicia Jiang, Simon Fraser University, felicia_jiang_2@sfu.ca
Yucong Cai, Simon Fraser University, yucong_cai@sfu.ca
Lawrence Kim, Simon Fraser University, lawkim@sfu.ca

Format: Presentation

ABSTRACT

Perinatal mental health (PMH) challenges occur within a period of substantial emotional, social, and embodied change, while technologies are increasingly positioned as potential sources of support. Yet designing technologies for PMH requires understanding not only what forms of support perinatal people may want, but also what roles they consider appropriate for technology within broader ecologies of care. We conducted participatory design workshops with 12 pregnant and postpartum participants, who first articulated their experiences and support needs before engaging with emerging technological possibilities. Our findings show how participants imagined technology taking on multiple roles in PMH support while simultaneously negotiating its boundaries, limitations, and relationships with human, material, and institutional care. We contribute an empirically grounded account of technological imaginaries in PMH and derive implications for designing technologies that respond to perinatal people's desired forms of care while preserving space for non-technological support.

Keywords: Perinatal Mental Health, Perinatal Technology, Participatory Design

Exploring LLMs' Translation Strategies in Literary Translation Under Multiple Generations

Hala Obeid*, Saemi Park*, Sophia Don Tranho*, Amandeep Kaur Manan,
Aniyah Bohnen, Emmy Fong, Weidi Li, Marzena Karpinska
Simon Fraser University

Motivation. Machine translation is already being used in the industry, Amazon's Kindle offers AI translation of books, yet literary text is full of expressions without a single 'correct' translation. In Agatha Christie's *The Man in the Brown Suit*, Suzanne answers a compliment with "I love butter," meaning she loves flattery. A translator can keep the butter, swap in an expression for flattery, or find a local idiom (Table 1). Previous work, using one output per model, finds that LLMs often translate such expressions literally, but LLM output varies between runs. We ask which strategies LLMs use for creative expressions, whether this changes across prompts and repeated generations, and whether an acceptable translation keeps the literary effect.

Table 1: Renderings of *"I love butter," purred Suzanne*, back-translated. Flattery and sweet talk keep the meaning but lose the food image, the Punjabi idiom keeps both.

Approach	Language	Back-translation
Literal	Arabic	"I love butter."
Flattery	Arabic	"I adore flattery."
Sweet talk	Korean	"I really do like sweet talk."
Butter idiom	Punjabi	"I like buttering up."

The TARTE dataset. We selected 357 creative expressions, mostly wordplay (39%) and cultural references (15%), from 51 English books published 1869–2026, 18 are from 2025–2026, since these are less likely to appear in training data. Six LLMs (Gemini 3.1 Pro, Gemini 3.8 Flash, GPT-5.6 Terra, GPT-5.6 Sol, Claude Sonnet 5, Qwen 3.6 27B) translated each ∼300-word passage into Arabic, Dutch, Korean, Punjabi and Vietnamese, with a simple and a literary prompt, ten times each. That gave us 214,200 translations.

Labelling. An LLM labeller assigned each rendering one of ten strategies, such as literal, substitution or explicitation, and it matched the annotator-reviewed labels on 83% of 150 test items. A second LLM judge decided whether the intended meaning reaches a target-language reader. On 150 blind human labels its agreement was modest ($\kappa = 0.37$), catching about half of the rejected renderings, so its 83.6% acceptance rate is an estimate.

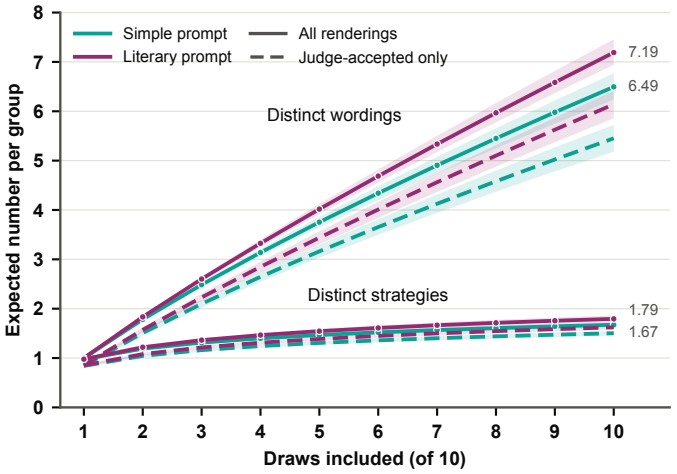


Figure 1: Distinct wordings and strategies among the first k of ten draws of the same expression, language, model and prompt, on the 169,200 renderings where every model has all draws.

Finding 1: expression drives strategy, not the model. The expression accounts for 42% of the variation in strategy and repeated draws 32% (partly label noise), against 14% for the model, 8% for the language and about 1% for the prompt. Only 0.8% reflects a model consistently preferring one strategy. Language still shifts choices: draws of one expression share a strategy about 80% of the time within a language, but 66–72% across languages.

Finding 2: literal translation is the default. Half of all renderings are literal across the six models used with two prompts and ten generations. The judge accepts 86% of them, 94% for metaphors but only 64% for idioms. In idiomatic expressions, models use substitution almost as often as literal.

Finding 3: models rarely recreate an expression. Recreation was only 2.5% of renderings with the simple prompt, and 4.1% with the literary prompt. Compensation appeared in only 58 out of 213,000 renderings.

Takeaway. Judging a model by one output can mistake sampling variation for a real difference, so creative translation should be evaluated across multiple generations. Acceptable is also not the same as creative: in a 300 translation pilot, 33 of the 195 translations that two LLM judges both accepted still lost a literary feature.

Mathematical Conditions for Human-AI Complementarity: Comparing AI-Initiated and Human-Initiated Deferral

Anita Hashemi[†], Mohid Rashid[†], Weina Jin[*], Ghassan Hamarneh[*]

Fall 2026

In healthcare, AI is increasingly being integrated into medical image interpretation workflows to reduce clinicians' workloads. However, AI integration alone does not necessarily guarantee improved performance. Therefore, human-AI complementarity becomes an important objective measured by the achievement of a lower expected loss in hybrid model performance rather than either the human or AI agent alone. Additionally, several deferral-based collaboration workflows have been proposed where a case may be passed from one decision-maker to another. Existing papers have explored conditions for complementary performance and various deferral mechanisms, but how the conditions vary between deferral systems remains undetermined.

Our goal in this study is to examine whether changing which decision-maker controls deferral affects the mathematical conditions required to achieve complementarity i.e., what constraints and relationships specific parameters of the system must fulfill. To achieve this, we will set up a theoretical framework to model and analyze simple human-AI deferral systems. Our approach will investigate how complementarity relates to the deferral function and pattern in each workflow along with measured uncertainty and agent accuracy. We will also empirically validate our results with a full simulation study. Doing so, we will address the following questions: (i) Does changing who controls deferral change the conditions required to achieve complementarity? (ii) Can both workflows achieve complementarity under shared assumptions?

Currently, this project is in very early stages, but we anticipate that conditions found between deferral patterns will be symmetric holding all other variables constant, only swapping the positions of the AI and human parameters. We hope to prove this and expand on our work by considering scenarios where this is not the case. This analysis may help deepen our understanding of how to design systems that integrate AI as a supportive tool, particularly in high-stakes clinical scenarios.

$$x_i \rightarrow A(x_i) \rightarrow D_A(x_i) \rightarrow \begin{cases} 0 \rightarrow \text{AI decision} \\ 1 \rightarrow \text{Human decision} \end{cases}$$

$$x_i \rightarrow H(x_i) \rightarrow D_H(x_i) \rightarrow \begin{cases} 0 \rightarrow \text{Human decision} \\ 1 \rightarrow \text{AI decision} \end{cases}$$

Figure 1: Illustrative AI-Initiated (top) and Human-Initiated (bottom) workflows. High uncertainty leads to deferral, while low uncertainty leads to retaining the case.

$$L_C < \min(L_H, L_A)$$

L_A:Expected loss of standalone AI . L_H:Expected loss of standalone human

Figure 2: Formal definition of complementarity. L_C represents the loss of the combined human-AI system. When this inequality is true, complementarity is achieved

Auto-Graded Assessments for Mastery-Based Learning

Layan Barrieshee, Simon Fraser University, lab22@sfu.ca
Ouldooz Baghan Karimi, Simon Fraser University, ouldooz@sfu.ca

Format: Poster

ABSTRACT

The majority of computer science course offerings adopt fixed timelines for when a student must complete a course element. This structure can limit a student's learning potential and understanding of vital course concepts. Mastery-based learning aims to improve student learning through flexibility. Allowing students to retake assessments and providing feedback can improve their learning of course material and their performance. Moreover, mastery-based learning has been shown to reduce stress levels for students and support meaningful learning [1]. Artificial intelligence can assist with the integration of mastery-based learning in courses. It has helped professors in grading and managing student work, providing tailored feedback to help students identify gaps in their understanding, as well as providing practice and resources for students to strengthen their understanding [2]. By introducing mastery-based learning, students are given the opportunity to grasp course concepts before moving onto the next topic. Moreover, it allows students to fix any learning gaps they have early in their learning, making it easier to tackle future topics. Furthermore, it gives students more flexibility in their learning pace and redirects their focus from getting high grades to acquiring mastery. In our study, we introduce mastery-based learning through one of the course components: interactive sessions. Every two weeks, students complete a trivial activity related to the previous weeks' material. After doing so, they answer a set of multiple-choice questions and one theoretical question. Variations of these automatically graded questions can be completed numerous times until students have achieved proficiency. On each try, students receive a different set of questions to prevent memorization and ensure that their knowledge is being tested. We are trying to understand the patterns of use and possible connections between mastery-based engagement and course performance in an automated-grading setup.

References:

[1] Manan Bhargava, Mehul Gandhi, Eemon Qayumi, Victoria Phelps, Zixuan Zhuang, Naveen Nathan, Connor Bernard, and Dan Garcia. 2025. GradeSync: A Tool for Automating Incomplete Processing to Support Mastery Learning. ACM SIGCSE TS 2025, 1387–1388. doi:10.1145/3641555.3705192

[2] Nikhil Kapasi, Derek Kirschbaum, Aryaman Singh, and Diba Mirza. 2026. Caliber:AI-Assisted Infrastructure for Mastery-Based Computer Science Education atScale. ACM ITiCSE 2026. Association for Computing Machinery, New York, NY, USA, 805–806. doi:10.1145/3803401.3811974

Green TTA: Benchmarking the Energy Consumption of Test-Time Adaptation

Vivian White, UBC + Vector Institute, vivwhite@cs.ubc.ca
Michika Hiratsuka, UBC, micieume@student.ubc.ca
Evan Shelhamer, UBC + Vector Institute, evan.shelhamer@vectorinstitute.ai

Poster Presentation

ABSTRACT

Deep learning achieves remarkable performance when the train and test data come from the same distributions, yet degrade significantly under distribution shift. Test-time adaptation (TTA) offers a realistic setting for deployment where models encounter data shifts at inference time, such as blur or sensor noise, and adapt on-the-fly without retraining. For TTA to be practical at deployment scale, it must be accurate and efficient. Green AI research to quantify and reduce the environmental impact of AI systems. A 2026 systematic literature survey by Lambert and Luccioni of 61 Green AI papers across eight AI life cycle stages reveals a blind spot: while over 50% of surveyed papers studied the energy consumption of static deployment and inference, zero papers evaluate models that update after training [2]. Existing TTA benchmarks measure efficiency using computation time as a proxy, but this ignores the effects of hardware usage on the environment. To fill this gap, we provide the first systematic benchmark of energy consumption during TTA. Our research questions are: (1) How do TTA methods differ in their energy-accuracy tradeoff compared to each other and to standard inference and finetuning baselines? (2) Which computational phases of TTA (setup, forward, backward, optimizer, or method-specific phases) drive energy cost, and how does this vary across methods and architectures? and (3) How do TTA methods perform in an energy-bounded setting?

We use the CodeCarbon and Zeus software packages to create the first systematic benchmark of energy consumption during TTA on ImageNet-C, a large-scale dataset composed of 15 visual corruptions (such as snow and motion blur) applied on top of natural images. We study 7 different TTA methods, including Tent and SAR, and four architectures, including ResNet50 and ViT. Our benchmark reports the accuracy on corrupted images, the energy consumption in kWh, the wall-clock time, and the number of adapted parameters. Our results show that TTA methods cover a broad range of energy-accuracy tradeoffs. We find that lightweight methods such as ETA can outperform more energy-consuming methods, such as SAR. Our work fills an important gap in both the TTA and Green AI literature. By studying the energy-accuracy tradeoff across methods and architectures, we aim to support practitioners choosing TTA methods to deploy under energy constraints, and establish a foundation for Green TTA research.

Keywords: test-time adaptation, green AI, energy efficiency

References

[1] Katherine Lambert and Sasha Luccioni. 2026. From Cradle to Cloud: A Life Cycle Review of AI's Environmental Footprint. Proceedings of the 2026 ACM Conference on Fairness, Accountability, and Transparency. Association for Computing Machinery, New York, NY, USA, 518–540. https://doi.org/10.1145/3805689.3812401

Exploring Human-Centric Multimedia Networking

Gizem Kurekci Polat, Simon Fraser University, gizem_kurekci_polat@sfu.ca
Ouldooz Baghban Karimi, Simon Fraser University, ouldooz@sfu.ca

Format: Poster

ABSTRACT

Multimedia applications are an important part of how people communicate, share information, and interact with digital services. As these applications evolve and new multimedia applications are adopted (e.g., Augmented and Extended Reality), communication networks need to support not only increasing volumes of data and their demands on the network, but also diverse human needs and expectations that come with more complex patterns of use, such as movement and interactions, and dependence on networked technologies.

Traditional networking approaches mainly focus on Quality of Service (QoS) metrics, such as bandwidth, latency, and throughput. However, good network performance does not always guarantee a satisfying user experience [1]. Therefore, recent works has explored Quality of Experience (QoE) aware approaches to multimedia delivery that consider both network conditions and users' perceived quality [2]. There is an increasing need for incorporating human-centric approaches to multimedia delivery to improve the end-user experience. These approaches take into account network performance in addition to user behaviour, mobility patterns, preferences, and contextual factors [3].

This work in progress explores the main concepts, challenges, and developments in human-centric multimedia networking through a review of existing research. The study focuses on understanding how multimedia networks are designed to better respond to user requirements, and what requirements need to be met for a guarantee of human-centricity. It also examines the supporting technologies and the implications of adopting human-centric approaches in connected applications, as well as access and core networks.

We review existing literature, identify common challenges and limitations, and highlight potential directions for future research. We aim to provide a clearer understanding of the challenges in developing multimedia networks that better meet user needs and identify factors influencing the adoption and integration of emerging technologies.

Keywords: Human-Centric Networking, Multimedia Applications, Quality of Experience

References

[1] M. Alreshoodi and J. Woods, “Survey on QoE/QoS Correlation Models for Multimedia Services,” *International Journal of Distributed and Parallel Systems*, vol. 4, no. 3, pp. 53–72, 2013. doi: 10.5121/ijdps.2013.4305.

[2] R. Kalan and I. Dulger, "A Survey on QoE Management Schemes for HTTP Adaptive Video Streaming: Challenges, Solutions, and Opportunities," in IEEE Access, vol. 12, pp. 170803-170839, 2024, doi: 10.1109/ACCESS.2024.3491613.

[3] M. T. Sultan and H. El Sayed, "QoE-Aware Analysis and Management of Multimedia Services in 5G and Beyond Heterogeneous Networks," in IEEE Access, vol. 11, pp. 77679-77688, 2023, doi: 10.1109/ACCESS.2023.3298556.

LLMs redraw argumentative patterns across languages

Gurneet Gill, Simon Fraser University, gkg36@sfu.ca
Jasleen Kaur, Simon Fraser University, jka349@sfu.ca
Stella Maltcheva, Simon Fraser University, sma345@sfu.ca
Khoa Ha Minh Duong, Simon Fraser University, hmd6@sfu.ca
Rodrigo Armando Anasco Santillan, Simon Fraser University, raa69@sfu.ca
Maria Kapitanenko, Simon Fraser University, mka260@sfu.ca
Ankita Gupta, University of Massachusetts Amherst, ankitagupta@cs.umass.edu
Marzena Karpinska, Simon Fraser University, karpinsk@sfu.ca

Format: Presentation

ABSTRACT

Large language models (LLMs) are increasingly used to draft opinion articles across languages, yet research on homogenization in argumentative writing has largely focused on English. In this work, we compare 4,000 human-written opinion articles with 43,750 articles generated by five LLMs across eight languages, using three complementary methods: (1) discourse structure analysis, (2) Toulmin-based argument analysis, and (3) data-driven feature discovery. Using only the 45 discovered features, a classifier identifies the language of a human opinion article 42% of the time (chance 12.5%), consistent with previous findings in contrastive rhetoric [1]. LLMs retain some cross-linguistic contrasts while flattening others. For example, Japanese still uses more rhetorical questions than English, but the human hedging gap between English and Hindi nearly disappears in LLM articles (shrinking from 14.2 to 1.3 percentage points). Across all eight languages, LLM articles use less irony, colloquial language, and personal discrediting, but more concession, and more often end by restating the thesis. To support their claims, LLM articles appeal more to general principles, whereas human articles rely more on testimony and statistics. In experiments with five prompting variations, certain features move closer to human writing (e.g., providing the author's stance restores the writer's voice), but no prompt eliminates this less conversational, less confrontational style. Some prompts even widen other gaps: providing the author's stance or requesting diverse articles in a single prompt increases thesis-first openings from 70% to 91% and 95%, respectively. We release our code and data to support research on multilingual homogenization in LLMs.

Keywords: Argument structure, argumentative writing, multilingual LLM evaluation, homogenization, opinion articles, machine-generated text, cross-lingual analysis

References

[1] U. M. Connor, *Contrastive Rhetoric: Cross-Cultural Aspects of Second-Language Writing*. Cambridge, UK: Cambridge University Press, 1996.

SpeakSwarm: Generating and Executing Co-Speech Gestures with Swarm Robots

Saarang Anand*, Simon Fraser University, saarang_anand@sfu.c
Minh Duc Dang*, Simon Fraser University, duc_dang@sfu.ca
Dou Gwon*, Simon Fraser University, dou_gwon@sfu.ca
Dora Ciobanu, Simon Fraser University, dora_ciobanu@sfu.ca
James Cheuong, Simon Fraser University, james_chuong@sfu.ca
Lawrence Kim, Simon Fraser University, lawkim@sfu.ca

Format: Presentation

ABSTRACT

Co-speech gestures convey meaning alongside speech, but systems that generate them are mostly built for humanoid robots and virtual avatars. Tabletop swarm robots offer a non-anthropomorphic alternative, though existing swarm gesture vocabularies are bounded by the elicitation studies that produce them. We present SpeakSwarm, a system that generates co-speech formations for arbitrary keywords, and executes them on 25 robots in synchrony with speech, through a five-stage generative pipeline separating planning, rendering, correction, evaluation, and extraction. In a formation validation study (N=30, 40 keywords), 92.5% of keywords yielded a formation that most participants judged semantically aligned. A comparative evaluation study (N=20) compared the running system against a projected simulation of the same formations and against an existing zero-shot language-to-coordinate approach. Participants judged SpeakSwarm's formations better aligned with the keywords than the baseline's in 93% of trials, and rated the physically execution higher than the projection on semantic accuracy, coherence, and rapport.

Keywords: Co-speech Gestures, Swarm Robotics, Swarm User Interface,
Embodied Conversational Agents, Generative AI

Beneath the Skin:
Interactive Biomechanical Facial Simulations via Composable Co-Expressions

Zhiyun Peng, Simon Fraser University, zhiyun_peng@sfu.ca
Michael Tao, Augmenta, michael.w.tao@gmail.com
William Huang, Simon Fraser University, william_huang_4@sfu.ca
Sophia Don Tranho, Simon Fraser University, syd3@sfu.ca
Eugene Fiume, Simon Fraser University, eugene_fiume@sfu.ca

Format: Lightning Talk

ABSTRACT

Facial animation usually trades physical accuracy for artistic control. Kinematic rigs offer direct control but lack biological constraints, so they lose volume and need many hand-sculpted corrective shapes. Musculoskeletal and finite-element (FEM) models capture anatomy but are too slow for interactive digital twins or VR. We ask whether a biomechanical face can be both physically grounded and interactive.

Beneath the Skin drives an epidermal skin layer from internal muscle dynamics. Given a rest pose and a sculpted target (e.g., the phoneme N or a raised brow), we solve for energy-optimized muscle trajectories. The optimizer tracks 300 sampled landmarks rather than the full non-linear deformation, balancing landmark distance, elastic strain energy, and Hill-based muscle work. The model includes a muscle proxy mesh with fat compartments, a rotational-joint mandible, and a skin layer registered once per model. Expressions are composable: combining facial action units (AUs) AU12 and AU2 on the left yields an asymmetric expression. It also reproduces volume preservation, passive head motion, soft-tissue contact, and pathologies such as burn scars, and responds to clicks, text prompts, or keyboard input.

During optimization, landmark distance drops steadily while strain energy and accumulated muscle work rise smoothly to convergence. We benchmark against FaceDiffuser, a state-of-the-art learned speech-animation model, using lip aperture, per-landmark travel, stretch ratio, and lip contact area. Our model articulates 2.2-2.4 times further and keeps the lips apart where the baseline collapses toward closed. Volume change, rigid-body validation, and jaw angular velocity cannot be measured for the baseline. We also compare the phoneme AA against film footage and Media2Face.

For micro-expressions, we isolated muscle activations for four AUs: AU1 (surprise), AU12 (happiness), AU15 (repression), and AU20 (fear), matching target poses to CASME II footage. These last about 500 ms and are barely visible. Our model reproduces them but not their subtlety: simulated movements had to be scaled up to be visible, while real ones are far smaller. All results use a single lab facial model.

To replace manual setup, we are prototyping a smartphone pipeline combining MediaPipe landmarks with iPhone TrueDepth data (MRI is a planned second pathway) to infer skull dimensions, muscle anchors, and skin mesh boundaries before the solver takes over. Next, we will validate anatomical accuracy with electromyography and MRI, add a high-fidelity skin layer, and extend the benchmark to the four AUs, where volume and stretch changes are largest.

Keywords: facial animation, biomechanical simulation, muscle-driven animation, physics-based animation, micro-expressions, benchmark evaluation, digital human, digital puppeteering

Is It the Data? Testing Target-Language Supervision for Fairer Prompt Compression Across Languages

Dezhao Feng, Jiehuan Zhao, Jonathan Yang
Simon Fraser University · {dfa54, jza396, sya171}@sfu.ca

Motivation

Large language models are billed and limited by tokens (units of text), and speakers of many languages already need more tokens than English speakers to say the same thing [1]. Prompt compressors such as LLMLingua-2 [2] shorten long inputs by deleting unimportant words, so they could help these users most. Instead, a recent cross-lingual audit [3] found that at one-third token retention, LLMLingua-2 preserves substantially less question-answering benefit in Chinese and Lithuanian than in English. The audit attributes this to the English-only data used to teach the compressor which words to keep, but it does not isolate supervision language experimentally.

Research question and design

We plan a controlled test. Following the LLMLingua-2 recipe, a teacher language model shortens Wikipedia passages by deleting words only, and its keep/drop decisions become training labels. We will continue training the released compressor on Chinese labels, on Swahili labels (a lower-resource language), and, as a control, on an equal amount of English labels from the same teacher. If a target language gains more from its own labels than from the English control, the language of the supervision matters, not just more data. We will evaluate on Belebele [4], a parallel reading-comprehension benchmark, at three compression rates.

Work so far

We built and tested the labelling pipeline: reproducible Wikipedia sampling with a document-level train/dev split; one word segmentation shared by labelling, training and inference (Chinese has no spaces, so what counts as a "word" must be fixed first); and an aligner that accepts a teacher output only if it is an order-preserving selection of the original words. Without API access, we use open models on a 24 GB GPU. Unconstrained, a 3B teacher produced no legal deletions on our pilot passages: it paraphrased, reformatted and added punctuation. We therefore added constrained decoding, which lets the teacher emit only original words in their original order, so every output is a legal label by construction.

Preliminary findings: legal is not the same as good

All three teachers in Table 1 produce 9/9 legal outputs, yet they fail in different ways. The small model deletes too much and loses key facts; the 13B model copies the passage or keeps only its opening; Qwen3-14B, given an explicit length target, produces some useful English and Chinese compressions but remains inconsistent: it barely compresses Swahili, and its outputs flip between near-copies and real compression under small prompt changes. These results come from nine pilot passages, and the Swahili outputs have not yet been checked by a native speaker.

Teacher (constrained decoding)	en	zh	sw	Main failure
Tiny Aya Global (3B)	0.13	0.22	0.08	over-deletes, loses key facts
Aya-101 (13B)	0.74	0.62	0.73	copies, or keeps only the opening
Qwen3-14B + length target	0.45	0.41	0.93	barely compresses Swahili

Table 1: Mean share of words kept on three pilot passages per language (lower = more compression). Only Qwen receives an explicit one-third length target.

Takeaway and next steps

For a lower-resource language, the first bottleneck may be obtaining good supervision, before any question about the compressor itself. Next we will validate the teacher on unseen passages, choose a Swahili label source (for example, a small human-annotated set), then train the compressors and run the Belebele evaluation. Either outcome shows where effort should go to make compression fair across languages.

[1] Petrov et al. Language Model Tokenizers Introduce Unfairness Between Languages. NeurIPS 2023.
[2] Pan et al. LLMLingua-2: Data Distillation for Efficient and Faithful Task-Agnostic Prompt Compression. Findings of ACL 2024.
[3] Lukauskas. Lost in Compression: A Controlled Cross-Lingual Audit of Extractive Prompt Compressors. arXiv 2026.
[4] Bandarkar et al. The Belebele Benchmark: a Parallel Reading Comprehension Dataset in 122 Language Variants. ACL 2024.

Where Should Domain Safety Policies Live? Policy Placement in LLM Safety Pipelines

Dezhao Feng, Tracy Cui
Simon Fraser University · contact: dfa54@sfu.ca

Motivation

Assistants built on large language models (LLMs) for regulated domains such as finance and medicine are wrapped in safety pipelines: an input guard screens the request, the answering model may refuse, and an output guard screens the answer. Any stage can block a benign request. Explaining how to spot fraudulent financial statements should be allowed, while helping to fabricate them should not. A deployer can write a domain policy that draws such boundaries, but it is unclear where in the pipeline the policy should be applied, and improving one stage may change little in what the user finally receives.

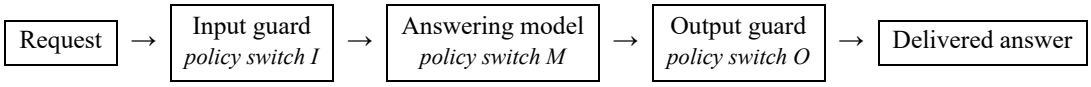


Figure 1: The pipeline we study. Each switch either supplies the domain policy or keeps the default.

Research question and design

When the same domain policy is given to the answering model, the input guard or the output guard, which placement yields more compliant, useful answers while limiting harmful assistance? Three switches (Figure 1) give eight configurations in which both guards are always present, plus two model-only references. We use Qwen2.5-7B-Instruct and Llama-3.1-8B-Instruct as answering models and Llama-Guard-3-8B [1] as both guards, medical requests for the pilot, and planned benign/harmful financial request pairs with similar wording but different intent. Both outcomes are measured end to end: the share of benign requests that receive a compliant, adequate answer, and the share of harmful requests that receive harmful help. Custom-policy guards and per-stage attribution have prior work; our focus is a controlled comparison of where the same policy is placed.

Implementation findings

Building the harness surfaced three pitfalls that can silently invalidate guard experiments. (1) The Hugging Face chat template of Llama-Guard-3-8B hardcodes its fourteen default categories and ignores a custom categories argument without warning, so a "policy-adapted" guard can quietly run the default policy. Our harness builds guard prompts itself, and its default arm reproduces the official template token for token. (2) Re-tokenizing a rendered chat template added a duplicate start-of-text token for Llama-3.1-8B-Instruct, so earlier Llama results must be re-checked. (3) Guard outputs began with a blank line before "safe" or "unsafe", so verdict scores must be read at the verdict token.

Pilot observations (hypotheses, not results)

In a 15-pair medical pilot, adding the policy to the guard changed no input-guard verdicts (0/60 across both models) and 2 of 120 output-guard verdicts. Both changes were Llama answers to harmful seed requests that used educational or simulation framing: the default output guard blocked them and the policy guard passed them. Neither has been judged yet, and both reached the 200-token cap, which truncated 42/60 Qwen and 27/60 Llama answers. On 25 boundary requests compared with one annotator's input labels, the policy input guard unblocked none of the 11 requests that the annotator allowed but the default guard blocked. These are small, unjudged, component-level observations; they do not show that any placement is safer or more useful.

Next steps and takeaway

We will add a second annotator, a frozen judging rubric for refusal, policy compliance and minimum answer adequacy, higher token caps, and the financial evaluation set, and compare configurations with paired statistics and confidence intervals. Our early lesson: where a policy lives, and whether it is applied at all, has to be verified end to end.

[1] Llama Team, AI @ Meta. The Llama 3 Herd of Models. arXiv:2407.21783, 2024.

CAN-CWIC East 2026

Dalhousie University

Academic Initiatives Panel: "Building Pathways for Women in Computing Across Atlantic Canada"
Dr. Sazia Mahfuz (Acadia University), Dr. Dania Tamayo-Vera (UPEI), Laurie Ricker (Mount Allison University), Paula Crouse (Mount Saint Vincent University), Dr. FE Bordeleau (Dalhousie University)

Faculty from five Atlantic universities discuss the programs, mentorship and outreach that help women enter and stay in computing, and how institutions across the region can work together.

Workshop: "An Overview of Technology Risk: The Intersection between IT, Risk and Business"
Jamie Wadden, Shreya Rai and Rashae Hart, EY

An introduction to technology risk and how IT, risk management and business decisions connect, with insight into careers in the field.

Workshop: "Step Into Consulting: A Day of Real AI Problems and Real Career Advice"
Rose Boudreau and Sana Kashgouli, CGI

Participants work through real-world AI problems the way consultants do and receive practical career advice from the CGI team.

Workshop: Women in AI
"Confidently Wrong: Evaluating AI for Transparency and Accountability in High-Stakes Decisions," Moriya Dechtiar, Chapter Lead, Research & Content, Women in AI
"AI Research: Factors Affecting AI Adoption," Shweta Pai, Senior Technical Product Owner, GeoSpectrum Technologies

Two talks on evaluating AI systems when the stakes are high, and on what drives or slows AI adoption in practice.

Industry Panel
Anita Bianchini (RBC), Chido Mbavarira (AI-First Consulting), Claire Quirion (Digital Nova Scotia), Mikaela Malson (IBM), Moriya Dechtiar (Women in AI)
Industry leaders share their career paths, how AI is changing their work, and advice for students entering the tech industry.

Networking and Poster Showcase
Students connect with industry partners at networking tables and view research posters from Acadia University.