VECTOR INSTITUTE | INSTITUT VECTEUR

REMARKABLE 2026

Research Poster Presentation Guide

February 20, 2026

VECTOR INSTITUTE | INSTITUT VECTEUR

TABLE OF POSTERS

Poster #	Poster Title	Presenter and Affiliation	Poster Session Type
1.	An Offline Agentic AI Approach to Voice-Controlled Surgical Navigation	Colton Barr, Queen's University	Agentic AI
2.	Adaptive Sustainability Policy: Reinforcement Learning–Driven Environmental Governance in Crisis-Informed Agent-Based Economies	Azin Sharifi, University of Toronto	Agentic AI
3.	Agent-Guided Relational Concept Discovery: Toward Explainable Surgical Margin Assessment	Nooshin Maghsoodi, Queen's University	Agentic AI
4.	Designing Ethical and Secure Agentic AI Pipelines Using MCP, RAG, and GitHub Actions	Aryan Gulati, University of Ottawa	Agentic AI
5.	Repurposing Optical Mice for Acoustic Eavesdropping	Zhimin Mei, Western University	Data, Challenges, Implementation & Software
6.	Mini Amusement Parks (MAPs): A Testbed for Modelling Business Decisions	Ian Berlot-Attwell, University of Toronto	Data, Challenges, Implementation & Software
7.	Open-PMC-18M: A High-Fidelity Large Scale Medical Dataset for Multimodal Representation Learning	Negin Baghbanzadeh, York University	Data, Challenges, Implementation & Software
8.	Accelerating AI via Learning Optimal Datatypes	Andreas Moshovos; Ali Hadizadeh, University of Toronto	Data, Challenges, Implementation & Software
9.	PFNSyn: Pretrained Diffusion Model for Tabular Data	Rozhan Akhound-Sadegh, University of Waterloo	Deep Learning
10.	Interpreting Vision Model Latent Spaces via Spatial Probing	Amoon Jamzad, Queen's University	Deep Learning
11.	Predicting Glucose Test Ordering in Hospitalized Patients Using Temporal Models of Clinical Context Embeddings	Joud El-Shawa, Western University	Deep Learning

12.	Self-Distillation of Hidden Layers for Self-Supervised Representation Learning	Scott Lowe, Vector Institute	Deep Learning
13.	Sensors Framework for Remote Detection of Functional Decline	Safea Altef, Vector Institute	Deep Learning
14.	Efficient Subsampling for GNN Downstream Tasks	Hirad Daneshvar, Toronto Metropolitan University	Deep Learning
15.	The Dataset Matters: Linking Image Memorability to Adversarial Robustness	Ehsan Ur Rahman Mohammed, Western University	Deep Learning
16.	Exploring Visual Prompt Tuning for Demographic Adaptation in Foundation Models for Medical Imaging	Matin Tavakoli, York University	Deep Learning
17.	Structure-Aware Graph Hypernetworks for Neural Program Synthesis	Wenhao Li, University of Toronto	Deep Learning
18.	Adaptive Latent-Space Constraints in Personalized Federated Learning	David Emerson, Vector Institute	Deep Learning
19.	VQ-Transplant: Efficient VQ-Module Integration for Pre-trained Visual Tokenizers	Xianghong Fang, University of Toronto	Generative Models
20.	Self-Supervised Transformers as Iterative Solution Improvers for Constraint Satisfaction	Yudong Will Xu, University of Toronto	Generative Models
21.	Catalyst GFlowNet for Electrocatalyst Design	Lena Podina, University of Waterloo	Generative Models
22.	Interactive Generative Image Model Merging via Bayesian Optimization	Chenxi Liu, University of Toronto	Generative Models
23.	On the Gradient Complexity of Private Optimization with Private Oracles: DP-SGD is Provably Slow	Michael Menart, University of Toronto	ML Privacy & Security
24.	MaskSQL: Safeguarding Privacy for LLM-Based Text-to-SQL via Abstraction	Sepideh Abedini, University of Waterloo	ML Privacy & Security
25.	Cascading Robustness Verification: Toward Efficient Model-Agnostic Certification	Mohammadreza Maleki, Toronto Metropolitan University	ML Privacy & Security

26.	Secure and Confidential Certificates of Online Fairness	Olive Franzese-McLaughlin, University of Toronto	ML Privacy & Security
27.	Majority of the Bests: Improving Best-of-N via Bootstrapping	Amin Rakhsha, University of Toronto	NLP/LLMs
28.	Attention Sinks: A 'Catch, Tag, Release' Mechanism for Embeddings	Stephen Zhang, University of Toronto	NLP/LLMs
29.	Predicting and Improving Test-Time Scaling Laws via Reward Tail-Guided Search	Muheng Li, University of Toronto	NLP/LLMs
30.	LLMs Uncertainty Quantification via Adaptive Conformal Semantic Entropy	Hamed Karimi, Toronto Metropolitan University	NLP/LLMs
31.	Say It Another Way: Auditing LLMs with a User-Grounded Automated Paraphrasing Framework	Rebecca Ma, University of Waterloo	NLP/LLMs
32.	Dialect Preference Bias in Large Language Models: Evidence from African American English	Huan Wu, York University	NLP/LLMs
33.	MedPerturbing LLMs: A Comparative Study of Toxicity, Prompt Tuning, and Jailbreaks in Medical QA	Arash Asgari, York University	NLP/LLMs
34.	Attention Sinks in the First Layers Enable Jailbreaking	Jonah Mackey, University of Toronto	NLP/LLMs
35.	Entropy-Gated Branching for Efficient Test-Time Reasoning	Xianzhi Li, Queen’s University	NLP/LLMs
36.	Deception Under Public Commitments in N-Player Games	Jerick Shi, Carnegie Mellon University	NLP/LLMs
37.	Bias in the Picture: Benchmarking VLMs with Social-Cue News Images and LLM-as-Judge Assessment	Aravind Narayanan, Vector Institute	NLP/LLMs
38.	LinguaMark: Do Multimodal Models Speak Fairly? A Benchmark-Based Evaluation	Ananya Raval, Vector Institute	NLP/LLMs

39.	Human-Centric Framework for Large Multimodal Models Evaluation	Vahid Reza Khazaie, Vector Institute	NLP/LLMs
40.	Auto-Regressive Masked Diffusion Models	Mahdi Karami, University of Waterloo	NLP/LLMs
41.	Neural Operators Can Play Dynamic Stackelberg Games	Xuwei Yang, McMaster University	Optimization
42.	Capacity-Constrained Online Learning with Delays: Scheduling Frameworks and Regret Trade-offs	Alexander Ryabchenko, University of Toronto	Optimization
43.	Efficient Bilevel Optimization with KFAC-Based Hypergradient	Disen Liao, University of Waterloo	Optimization
44.	Cross-Tokenizer Likelihood Scoring Algorithms for Language Model Distillation	Truong Buu Phan, University of Toronto	Probabilistic Methods
45.	First-Explore, then Exploit: Meta-Learning to Solve Hard Exploration-Exploitation Trade-Offs	Benjamin Norman, University of British Columbia	Reinforcement Learning & Planning
46.	Learning to Negotiate via Voluntary Commitment	Shuhui Zhu, University of Waterloo	Reinforcement Learning & Planning
47.	EM-DQN: Expectation-Maximization for Gaussian-Mixture Value Distributions in Distributional Reinforcement Learning	Michal Lisicki, University of Guelph	Reinforcement Learning & Planning
48.	SocialHarmBench: Revealing LLM Vulnerabilities to Socially Harmful Requests	Punya Syon Pandey, University of Toronto	Societal Considerations
49.	A Unified Theory of Attention and State Space Dynamics	Aref Jafari, University of Waterloo	Theory
50.	Fairness and Optimization	Mojtaba Kolahdouzi, York University	Theory
51.	Understanding and Improving Shampoo and SOAP via Kullback–Leibler Minimization	Wu Lin, Vector Institute	Other AI Applications
52.	Chrysalis: Towards Scalable AI-Enabled Peer Tutoring & Assessment	Prashanth Arun, University of Waterloo	Other AI Applications

53.	Enigma: An Efficient Model for Deciphering Regulatory Genomics	Andrew Jung, University of Toronto	Other AI Applications
54.	From Individual to Multi-Agent Algorithmic Recourse: Minimizing the Welfare Gap via Capacitated Bipartite Matching	Zahra Khotanlou, University of Waterloo	Other AI Applications
55.	Benchmarking Histology Foundation Models for Glioblastoma Molecular Prediction with Spatial Transcriptomic Validation of Attention	Dilakshan Srikanthan, Queen's University	Other AI Applications
56.	Reimagining LLMs as Ethical and Adaptive Co-Creators in Mental Health Care	Abeer Badawi, York University	Other AI Applications
57.	BarcodeMamba+: Advancing State-Space Models for Fungal Biodiversity Research	Tiancheng Gao, University of Guelph	Other AI Applications
58.	Vision-Language Models Learn Clinical Concepts for Breast Cancer Detection	Mohamed Harmanani, Queen's University	Other AI Applications
59.	Hash Collisions in Molecular Fingerprints: Effects on Property Prediction and Bayesian Optimization	Walter Virany, University of Toronto	Other AI Applications
60.	ProstNFound+: A Prospective Study using Medical Foundation Models for Prostate Cancer Detection	Paul Wilson, Queen's University	Other AI Applications

Agentic AI

Poster #1: An Offline Agentic AI Approach to Voice-Controlled Surgical Navigation

Presenter: Colton Barr, Queen's University
Google Scholar: https://scholar.google.com/citations?hl=en&user=-Xi-UPgAAAAJ
Collaborators: Colin Galvin (Brigham and Women's Hospital, Harvard Medical School), Amirali Azimi (Brigham and Women's Hospital, Harvard Medical School), Sarah Frisken (Brigham and Women's Hospital, Harvard Medical School), Steve Pieper (Isomics Inc.), Gabor Fichtinger (Queen's University), Alexandra Golby (Brigham and Women's Hospital, Harvard Medical School), Parvin Mousavi (Queen's University)

Inexpensive surgical navigation systems aim to improve patient outcomes in low-resource settings but often require additional personnel and technical expertise to operate. We introduce NOVA, a hands-free voice assistant for low-cost surgical navigation systems that runs large language model (LLM) agents locally on a consumer-grade laptop. NOVA employs an agentic architecture with distinct LLMs for question answering, user-interface control, and agent delegation. Two semi-synthetic benchmark datasets were created to optimize individual agents, informing the integration of a complete NOVA prototype into NousNav, an existing low-cost neuronavigation platform. A user study evaluated system latency and command success, with nine participants performing hands-free neuronavigation tasks on a simulated patient. Offline, open-source LLM agents performed comparably to commercial online models across both benchmarks, and all participants successfully completed the workflow hands-free. These results demonstrate the feasibility of offline LLMs for enabling low-cost, voice-controlled surgical navigation with clinically viable latency and command success, highlighting their potential for deployment in resource-constrained settings.

Poster #2: Adaptive Sustainability Policy: Reinforcement Learning–Driven Environmental Governance in Crisis-Informed Agent-Based Economies

Presenter: Azin Sharifi, University of Toronto
Google Scholar: https://scholar.google.com/citations?user=abYgy1QAAAAJ&hl=en
Collaborators: Luis Seco (University of Toronto), Shiva Zamani (Sharif University of Technology)

Sustainability policy design requires adaptive strategies that can respond to environmental degradation, economic heterogeneity, and systemic crises. Traditional equilibrium-based models often fail to capture the dynamic and shock-driven nature of real-world green transitions. We propose Adaptive Sustainability Policy (ASP), a computational framework that integrates agent-based modeling with multi-agent reinforcement learning to design resilient environmental governance mechanisms. Building on the AI Economist paradigm, we introduce a Gather–Trade–Build–Pollute (GTBP) environment in which pollution arises endogenously from economic activity and feeds back into long-term economic and environmental outcomes. Within this simulated economy, heterogeneous agents adapt their behavior while an AI policymaker dynamically optimizes sustainability interventions such as emission taxes and green subsidies. We further evaluate policy robustness under crisis-informed scenarios calibrated to real-world events, including the 2023 Canadian wildfires. Results show that adaptive reinforcement learning policies consistently outperform static benchmarks, achieving higher welfare, greater economic stability, and faster pollution reduction under disruptive conditions. ASP provides a scalable testbed for designing crisis-resilient sustainability policies in complex eco-economic systems.

Poster #3: Agent-Guided Relational Concept Discovery: Toward Explainable Surgical Margin Assessment

Presenter: Nooshin Maghsoodi, Queen’s University
Google Scholar: https://scholar.google.ca/citations?hl=en&user=M_pFT_IAAAAJ
Collaborators: Robert Policeli, Amoon Jamzad, Mohammad Farahmand, Martin Kaufmann, Kevin Y. M. Ren, John F. Rudan, Doug McKay, Gabor Fichtinger, Parvin Mousavi

Accurate identification of cancerous tissue during surgery is critical for complete tumor removal and reducing recurrence risk. Rapid Evaporative Ionization Mass Spectrometry (REIMS) enables real-time analysis of surgical smoke, providing rich molecular information, but developing reliable and interpretable models remains challenging. We present Agent-Guided Relational Concept Discovery, an explainable framework for surgical margin assessment that integrates concept learning with biochemical reasoning. The model learns discriminative representations from REIMS spectra while automatically discovering latent metabolic concepts. During training, a reasoning agent analyzes concept activations, identifies informative m/z regions, assigns concise semantic descriptions, and adaptively adjusts concept importance without manual concept annotations. The discovered concepts are grounded in known metabolites and metabolic pathways, and their relationships are organized into a concept-level knowledge structure that supports interpretability. After training, the agent is removed, enabling efficient inference while preserving concept-level explanations. We evaluate the approach on a REIMS dataset of 693 spectra from 91 patients, including basal cell carcinoma and multiple healthy tissue types, demonstrating clinically relevant and interpretable surgical margin assessment.

Poster #4: Designing Ethical and Secure Agentic AI Pipelines Using MCP, RAG, and GitHub Actions

Presenter: Aryan Gulati, University of Ottawa

Agentic AI systems are increasingly used to automate decision-making, data processing, and security workflows. However, deploying such systems responsibly raises challenges around ethics, cybersecurity, transparency, and governance. This poster presents the design of a new, modular agentic AI architecture that integrates Model Context Protocol (MCP) servers, Retrieval-Augmented Generation (RAG), and GitHub Actions to enable secure and auditable AI automation.

The proposed system leverages AI agents capable of tool use, contextual reasoning, and policy-aware execution while enforcing ethical safeguards such as least-privilege access, data minimization, and human-in-the-loop controls. GitHub Actions are used to orchestrate CI/CD-style AI workflows, enabling reproducibility, versioning, and traceability, while MCP standardizes agent-tool communication. RAG is incorporated to ground agent responses in curated knowledge sources, reducing hallucinations and improving trustworthiness.

This work focuses on architectural design rather than model novelty and is intended to be accessible to a general AI audience. The poster highlights practical design trade-offs, security considerations, and ethical implications for deploying agentic AI systems in real-world environments such as cybersecurity automation and enterprise workflows.

Data, Challenges, Implementation, and Software

Poster #5: Repurposing Optical Mice for Acoustic Eavesdropping

Presenter: Zhimin Mei, Western University

Google Scholar: https://scholar.google.com/citations?user=5YgNO8IAAAAJ&hl=zh-CN

Paper: https://ieeexplore.ieee.org/abstract/document/11044584

Collaborators: Donghui Dai (The Hong Kong Polytechnic University)

Acoustic eavesdropping presents a longstanding challenge in personal information security and privacy. We introduce JerryAttack, a method that repurposes an optical mouse as a covert eavesdropping device. The mouse's integrated low-resolution, high-frame-rate image sensor is transformed into a high-speed camera for visual vibrometry, capable of capturing acoustic vibrations from nearby loudspeakers. Our contributions are threefold: (1) using the ‘pixel grabber’ register as a backdoor to extract the pixel stream, (2) establishing an acoustic-optical side channel for effective eavesdropping, and (3) exploring two attack scenarios: voice profiling and speech reconstruction. Sound recovered through this side channel achieves a mean SNR of 7.3 dB, comparable to standard microphones in noisy environments. Combined with a classification neural network, JerryAttack identifies individuals with 83.27% accuracy across six languages, and achieves good intelligibility with a median STOI score exceeding 0.7 when integrated with joint channel information.

Poster #6: Mini Amusement Parks (MAPs): A Testbed for Modelling Business Decisions

Presenter: Ian Berlot-Attwell, University of Toronto
Google Scholar: https://scholar.google.ca/citations?user=jVdqanEAAAAJ
Paper: https://arxiv.org/abs/2511.15830
Collaborators: Stéphane Aroca-Ouellette (Skyfall AI), Panagiotis Lymperopoulos (Skyfall AI, Tufts University), Abhiramon Rajasekharan (Skyfall AI, University of Texas at Dallas), Tongqi Zhu (Skyfall AI), Herin Kang (Skyfall AI), Kaheer Suleman (Skyfall AI), Sam Pasupalak (Skyfall AI)

Despite recent breakthroughs in artificial intelligence, current systems struggle with complex, uncertain real-world decision making. Many domains, such as business management, require efficient domain learning, long-horizon planning in non-deterministic environments, multi-objective optimization, and multi-modal reasoning. Existing human-AI benchmarks fail to capture these intersections. We introduce Mini Amusement Park (MAP), an amusement park simulator that evaluates a player’s ability to anticipate long-term effects of actions under uncertainty and plan strategically to maximize profit. Human performance baselines and systematic evaluations of leading AI models show humans outperform state-of-the-art LLM agents by 2.2× on easy mode and 15.7× on medium mode. MAP highlights challenges in optimization, sample-efficient learning, and the practicality of LLM world models, providing a foundation for developing robust AI capable of complex real-world reasoning.

Poster #7: Open-PMC-18M: A High-Fidelity Large Scale Medical Dataset for Multimodal Representation Learning

Presenter: Negin Baghbanzadeh, York University
Google Scholar: https://scholar.google.com/citations?hl=en&user=5gOFl4gAAAAJ
Collaborators: Vector Institute, York University

We present Open-PMC-18M, a high-fidelity, large-scale medical dataset designed for multimodal representation learning. The dataset facilitates the development of AI models capable of integrating diverse data modalities, supporting research in medical AI applications that require high-quality, large-scale, multimodal data.

Poster #8: Accelerating AI via Learning Optimal Datatypes

Presenter: Ali Hadizadeh, University of Toronto
Google Scholar: https://scholar.google.com/citations?hl=en&user=TIU-YmgAAAAJ
Paper:
https://proceedings.mlsys.org/paper_files/paper/2024/hash/185087ea328b4f03ea8fd0c8aa96f747-Abstract-Conference.html
Collaborators: Milos Nikolic (University of Toronto & ByteShape Inc.), Enrique Torres Sanchez (University of Toronto & ByteShape Inc.), Ali Hadi Zadeh (University of Toronto & ByteShape Inc.)

AI inference and training costs are rising rapidly, despite coordinated efforts across industry and academia to reduce them by optimizing hardware and software throughout the training and serving stack. At the hardware level, vendors are introducing AI-optimized datatypes such as FP8, FP4, and microscaling; however, effectively exploiting these capabilities is largely left to end users. We present a fully automated method that learns “optimal” datatype/quantization assignments for groups of values (e.g., tensors). The method can targets any model/task and reduces both training and inference cost. We report preliminary results on accelerating model training, along with real-system measurements demonstrating inference speedups on recent LLMs. Beyond cost reduction, our approach enables deployments that would otherwise be infeasible due to memory constraints. We demonstrate real-time inference on a Raspberry Pi 5 running a 30B MoE model. This technology is being commercialized by ByteShape (byteshape.com) to maximize ROI while meeting QoS requirements.

Deep learning

Poster #9: PFNSyn: Pretrained Diffusion Model for Tabular Data

Presenter: Rozhan Akhound-Sadegh, University of Waterloo
Collaborators: Wei Pang, Yaoliang Yu, Xi He

Deep learning for tabular data is challenged by limited data availability and heterogeneous structures. While TabPFN demonstrated that pretraining on synthetic data from Structural Causal Models (SCMs) enables strong in-context classification, large-scale pretraining for tabular data generation remains underexplored. We introduce PFNSyn, a pretrained diffusion model for tabular data, trained on large-scale synthetic datasets generated from SCMs. PFNSyn can be applied directly to unseen datasets without further training, making it well suited for low-data or privacy-sensitive settings. The model supports class-conditional synthesis and row-level conditioning, providing a flexible foundation for privacy-preserving and controllable tabular data generation, bringing the benefits of pretrained diffusion models in vision to tabular domains.

Poster #10: Interpreting Vision Model Latent Spaces via Spatial Probing

Presenter: Amoon Jamzad, Queen’s University
Google scholar: https://scholar.google.ca/citations?user=StQV08sAAAAJ&hl=en
Collaborators: Nooshin Maghsoodi, Parvin Mousavi (School of Computing, Queen's University)

Vision foundation models produce dense, high-dimensional latent feature maps with detailed spatial structure. Traditional extrinsic evaluation via downstream tasks offers only indirect insight into latent representations, and global dimensionality reduction methods often obscure localized variation. We introduce an interactive spatial probing framework that localizes dimensionality reduction to user-defined spatial regions, analyzing corresponding latent features to identify dominant directions of variation. This approach reveals coherent local patterns often hidden in global embeddings. An interactive software tool supports exploratory analysis of vision embeddings, providing a transparent and interpretable way to analyze model latent spaces.

Poster #11: Predicting Glucose Test Ordering in Hospitalized Patients Using Temporal Models of Clinical Context Embeddings

Presenter: Joud El-Shawa, Western University
Google scholar: https://scholar.google.com/citations?user=bwfx2yMAAAAJ&hl=en&oi=ao
Paper: https://ojs.aaai.org/index.php/AAAI-SS/article/view/36924
Collaborators: Elham Bagheri (Vector Institute & Western University), Amol Verma (Unity Health Toronto), Yalda Mohsenzadeh (Western University & Vector Institute)

Laboratory test overuse drives costs, patient discomfort, and low-value care, with glucose testing as a prominent example. We present a deep learning framework integrating structured and unstructured electronic medical record data to predict whether a glucose test will be ordered in the next AM/PM time bin. Using multi-hospital GEMINI data, Long Short-Term Memory models combined with Clinical BioBERT embeddings capture timing and clinical context. On held-out test data, our best model achieved ROC-AUC of 0.92 and PR-AUC of 0.67, generalizing across sites (ROC-AUC 0.84). Temporal recency cues improved performance, and exploratory regression for glucose values highlighted the need for diverse, frequent data. This framework lays the foundation for real-time decision support to reduce unnecessary laboratory testing.

Poster #12: Self-Distillation of Hidden Layers for Self-Supervised Representation Learning

Presenter: Scott Lowe, Vector Institute
Google Scholar: https://scholar.google.com/citations?user=ZFPhxuAAAAAJ&hl=en
Collaborators: Anthony Fuller (Vector Institute & Carleton University), Sageev Oore (Vector Institute & Dalhousie University), Evan Shelhamer (Vector Institute & University of British Columbia), Graham W. Taylor (Vector Institute & University of Guelph)

Self-supervised learning (SSL) leverages unlabelled data, yet current methods like DINOv3 rely on large batches and hand-crafted augmentations. We introduce Bootleg, a self-distillation method predicting mid-level embeddings across hidden layers from masked inputs, bridging the gap between masked auto-encoders and joint predictive embedding methods. Bootleg improves performance on ImageNet-1K (+7% frozen probing over I-JEPA) and ADE20K (+10% frozen probing), providing an augmentation-free, batch-size independent approach that extracts richer intermediate representations for self-supervised learning.

Poster #13: Sensors Framework for Remote Detection of Functional Decline

Presenter: Safea Altef, Vector Institute
Google Scholar: https://scholar.google.ca/citations?hl=en&user=estyJnkAAAAJ
Collaborators: Trent University

Kinera is a sensor framework for monitoring functional health decline in older adults through continuous, non-intrusive observation of home behavioral patterns. Combining sensor data with statistical models, deep learning-based sequence analysis, and rule-based reasoning, Kinera detects subtle deviations in Activities of Daily Living (ADLs), such as bathing, sleeping, and toileting. By identifying spatial-temporal irregularities in behavior, the framework enables early detection of mild cognitive decline.

Poster #14: Efficient Subsampling for GNN Downstream Tasks

Presenter: Hirad Daneshvar, Toronto Metropolitan University
Google Scholar: https://scholar.google.com/citations?user=qHFgIJMAAAAJ&hl=en&oi=ao
Paper: https://openreview.net/forum?id=xE2dvWKpHx
Collaborators: Reza Samavi (Toronto Metropolitan University & Vector Institute)

Graph Neural Networks (GNNs) are effective for data integration using graphs, yet subsampling methods lag behind. We propose an importance-based data subsampling framework that identifies graph inputs based on predictive uncertainty and clusters them using learned graph representations. Subsampling favors high-impact points while maintaining diversity. Evaluation on multi-source child and youth mental health datasets demonstrates statistically significant improvements (average 10.13% across metrics) over random sampling, supporting better predictive performance in downstream tasks.

Poster #15: The Dataset Matters: Linking Image Memorability to Adversarial Robustness

Presenter: Ehsan Ur Rahman Mohammed, Western University
Google Scholar: https://scholar.google.ca/citations?user=nrIjRoMAAAAJ&hl=en
Collaborators: Elham Bagheri (Vector Institute), Apurva Narayan (Western University), Yalda Mohsenzadeh (Western University & Vector Institute)

Adversarial attacks exploit imperceptible perturbations to mislead image classifiers. While defenses often focus on model architecture, we investigate the role of dataset properties, particularly image memorability, on adversarial robustness. Using MemCat and THINGS datasets with known memorability scores, we analyzed multiple architectures (ViT, EfficientNet, ResNet-50) under PGD and CW attacks. High- and low-memorability images exhibited significant differences in adversarial accuracy. Our findings highlight that human-salient attributes influence vulnerability and suggest dataset-aware strategies for robust AI, aligning training data properties with human perception to improve resistance to attacks.

Poster #16: Exploring Visual Prompt Tuning for Demographic Adaptation in Foundation Models for Medical Imaging

Presenter: Matin Tavakoli, York University
Google Scholar: https://scholar.google.ca/citations?hl=en&user=dSaBcsEAAAAJ
Paper: https://neurips.cc/virtual/2024/104969
Collaborators: Artur Parkhimchyk (York University), Amirreza Naziri (York University), Laleh Seyyed-Kalantari (York University)

Pretrained medical foundation models require extensive computation for adaptation. Visual Prompt Tuning (VPT) allows efficient adaptation to new tasks with minimal architecture changes. We explore demographic (race) adaptation of MAE and MoCoV3 models for disease classification on imbalanced medical imaging data. Comparing linear probing, full fine-tuning, and VPT, we find that VPT boosts performance even with small demographic subsets, achieving results comparable to full fine-tuning while requiring fewer resources. This demonstrates VPT's utility for efficiently adapting foundation models in low-resource medical settings.

Poster #17: Structure-Aware Graph Hypernetworks for Neural Program Synthesis

Presenter: Wenhao Li, University of Toronto
Google Scholar: https://scholar.google.com/citations?user=rmaFOGEAAAAJ
Paper: https://openreview.net/forum?id=x7zOzUwtR7
Collaborators: Yudong Xu (University of Toronto), Scott Sanner (University of Toronto), Elias Boutros Khalil (University of Toronto)

A structure-aware meta-learner can directly generate full neural network weights as a continuous program modality. This approach demonstrates strong zero-shot generalization across unseen tasks and outperforms baseline methods in neural program synthesis.

Poster #18: Adaptive Latent-Space Constraints in Personalized Federated Learning

Presenter: David Emerson, Vector Institute
Collaborators: Sana Ayromlou (Vector Institute and Google)

Federated learning (FL) is an effective and widely used approach to training deep learning models on decentralized datasets held by distinct clients. FL also strengthens both security and privacy protections for training data. Common challenges associated with statistical heterogeneity between distributed datasets have spurred significant interest in personalized FL (pFL) methods, where models combine aspects of global learning with local modeling specific to each client's unique characteristics. This work investigates the efficacy of theoretically supported, adaptive MMD measures in pFL, primarily focusing on the Ditto framework, a state-of-the-art technique for distributed data heterogeneity. The use of such measures significantly improves model performance across a variety of tasks, especially those with pronounced feature heterogeneity. Additional experiments demonstrate that such measures are directly applicable to other pFL techniques and yield similar improvements across a number of datasets. Finally, the results motivate the use of constraints tailored to the various kinds of heterogeneity expected in FL systems.

Generative Models

Poster #19: VQ-Transplant: Efficient VQ-Module Integration for Pre-trained Visual Tokenizers

Presenter: Xianghong Fang, University of Toronto
Google Scholar: https://scholar.google.com.hk/citations?user=hQfxe5QAAAAJ
Paper: https://openreview.net/pdf?id=eETr3lrOQB
Collaborators: Xianghong Fang (University of Toronto), Yuan Yuan (Boston College), Dehan Kong (University of Toronto), Tim G. J. Rudner (University of Toronto)

Vector Quantization (VQ) underpins modern discrete visual tokenization. However, training quantization modules for state-of-the-art VQ-based models requires significant computational resources, which in practice prevents the development of novel VQ techniques under resource constraints. To address this limitation, we propose VQ-Transplant, a framework that enables plug-and-play integration of new VQ modules into frozen, pre-trained tokenizers by replacing their native VQ modules. The transplantation process preserves all encoder-decoder parameters, eliminating the need for costly end-to-end retraining.

To mitigate decoder-quantization mismatch, we introduce a lightweight decoder adaptation strategy trained for only 5 epochs on ImageNet-1k. Empirical evaluation shows that VQ-Transplant achieves near state-of-the-art reconstruction fidelity for industry-level models like VAR while reducing training cost by 95%. This approach democratizes quantization research by enabling resource-efficient integration of novel VQ techniques while matching industry-level reconstruction performance.

Poster #20: Self-Supervised Transformers as Iterative Solution Improvers for Constraint Satisfaction

Presenter: Yudong Will Xu, University of Toronto
Google Scholar: https://scholar.google.ca/citations?user=aiBPHn0AAAAJ
Paper: https://icml.cc/virtual/2025/poster/45737
Collaborators: Wenhao Li, Scott Sanner, Elias B. Khalil (Department of Mechanical & Industrial Engineering, University of Toronto)

We present a Transformer-based framework for Constraint Satisfaction Problems (CSPs), which are widely used across applications and are NP-complete. Most existing approaches rely on supervised learning from feasible solutions or reinforcement learning, which require either known solutions or large training budgets with complex expert-designed rewards. We propose ConsFormer, a self-supervised framework that leverages a Transformer as a solution refiner. ConsFormer constructs solutions iteratively, mimicking local search, and uses differentiable approximations to discrete CSP constraints to guide training. The model is trained to improve random assignments for a single step but deployed iteratively at test time, allowing it to tackle out-of-distribution CSPs without requiring large labeled datasets or expert reward functions. Experiments on Sudoku, Graph Coloring, Nurse Rostering, and MAXCUT demonstrate its effectiveness.

Poster #21: Catalyst GFlowNet for Electrocatalyst Design

Presenter: Lena Podina, University of Waterloo
Google Scholar: https://scholar.google.ca/citations?hl=en&user=0sSUAEQAAAAJ
Paper: https://arxiv.org/abs/2510.02142
Collaborators: Christina Humer (ETH Zurich), Alexandre Duval (Entalpic), Ali Ramlaoui (Entalpic), Victor Schmidt (Entalpic), Shahana Chatterjee (Mila), Yoshua Bengio (Mila, Université de Montréal), Alex Hernandez-Garcia (Mila, Université de Montréal), David Rolnick (Mila, McGill University), Felix Therrien (Mila)

Efficient and inexpensive energy storage is critical for renewable energy adoption and grid stability. Electrocatalysts enable energy storage as hydrogen, but discovering high-performance, low-cost catalysts is challenging. We introduce Catalyst GFlowNet, a generative model that uses machine learning predictors of formation and adsorption energy to design crystal surfaces as efficient catalysts. We demonstrate its performance on the hydrogen evolution reaction, successfully identifying platinum as the most efficient known catalyst. Future work will extend the approach to the oxygen evolution reaction and open the search space to discover new catalysts. This generative modeling framework provides a promising pathway to accelerate the search for novel, efficient catalysts.

Poster #22: Interactive Generative Image Model Merging via Bayesian Optimization

Presenter: Chenxi Liu, University of Toronto
Collaborators: Selena Ling, Alec Jacobson

Fine-tuning-based adaptation is widely used to customize diffusion-based image generation, leading to large collections of community-created adapters that capture diverse subjects and styles. Adapters derived from the same base model can be merged with weights, enabling the synthesis of new visual results within a vast and continuous design space. To explore this space, current workflows rely on manual slider-based tuning, an approach that scales poorly and makes weight selection difficult, even when the candidate set is limited to 20–30 adapters. We propose GimmBO to support interactive exploration of adapter merging for image generation through Preferential Bayesian Optimization (PBO). Motivated by observations from real-world usage, including sparsity and constrained weight ranges, we introduce a two-stage BO backend that improves sampling efficiency and convergence in high-dimensional spaces. We evaluate our approach with simulated users and a user study, demonstrating improved convergence, high success rates, and consistent gains over BO and line-search baselines, and further show the flexibility of the framework through several extensions.

ML Privacy & Security

Poster #23: On the Gradient Complexity of Private Optimization with Private Oracles: DP-SGD is Provably Slow

Presenter: Michael Menart, University of Toronto
Google Scholar: https://scholar.google.com/citations?view_op=list_works&hl=en&user=U3Vwd3YAAAAJ
Paper: https://arxiv.org/pdf/2511.13999
Collaborators: Aleksandar Nikolov (University of Toronto)

We present some of the first runtime lower bounds for differentially private (DP) optimization. We show that a large class of DP optimizers, including the ubiquitous DP-SGD algorithm, necessarily suffers a dimension-dependent slowdown compared to their non-private counterparts. Our lower bounds are tight in high dimensions.

Poster #24: MaskSQL: Safeguarding Privacy for LLM-Based Text-to-SQL via Abstraction

Presenter: Sepideh Abedini, University of Waterloo
Google Scholar: https://scholar.google.ca/citations?hl=en&user=ZeCvVAMAAAAJ

MaskSQL is a framework for privacy-preserving text-to-SQL tasks using abstraction to mask sensitive information in LLM prompts. Unlike redaction or generalization, abstraction retains essential information while discarding unnecessary details, balancing privacy and utility. MaskSQL outperforms leading small language models and approaches the performance of state-of-the-art LLMs while enabling local deployment and compliance with privacy regulations, making it practical for sensitive systems.

Poster #25: Cascading Robustness Verification: Toward Efficient Model-Agnostic Certification

Presenter: Mohammadreza Maleki, Faculty Affiliate Researcher (supervised by Vector Faculty Affiliate), Toronto Metropolitan University
Google Scholar: https://scholar.google.com/citations?hl=en&user=CsD1-QwAAAAJ
Paper: https://arxiv.org/pdf/2602.04236

Certifying the robustness of neural networks (NNs) against adversarial examples remains a major challenge in trustworthy machine learning. Providing formal guarantees that inputs remain robust against all adversarial attacks within a perturbation budget often requires solving non-convex optimization problems. Hence, incomplete verifiers, such as those based on linear programming (LP) or semidefinite programming (SDP), are widely used because they scale efficiently and substantially reduce the cost of robustness verification compared to complete methods. We identify the limitations of relying on a single incomplete verifier, which can underestimate robustness due to false negatives arising from loose approximations or misalignment between training and verification methods. In this work, we propose Cascading Robustness Verification (CRV), which goes beyond an engineering improvement by exposing fundamental limitations of existing robustness metrics and introducing a framework that enhances both reliability and efficiency in verification. CRV is a model-agnostic verifier, meaning that its robustness guarantees are independent of the model's training process. The key insight behind the CRV framework is that when using multiple verification methods, an input is certifiably robust as long as one method verifies the input as robust. Rather than relying solely on a single verifier with a fixed constraint set, CRV progressively applies multiple verifiers to balance the tightness of the bound and computational cost. Starting with the least expensive method, CRV halts as soon as an input is certified as robust; otherwise, it proceeds to more expensive methods. For each computationally expensive method, we introduce a Stepwise Relaxation Algorithm (SR) that incrementally adds constraints and checks for certification at each step, thereby avoiding unnecessary computation. Our theoretical analysis demonstrates that CRV consistently achieves equal or higher verified accuracy across all settings compared to powerful but computationally expensive incomplete verifiers in the cascade, such as SDP-based methods, while significantly reducing verification overhead. Empirical results confirm that CRV certifies at least as many inputs as benchmark approaches, while improving runtime efficiency by up to ~90%.

Poster #26: Secure and Confidential Certificates of Online Fairness

Presenter: Olive Franzese-McLaughlin, University of Toronto
Google Scholar: https://scholar.google.com/citations?user=V0918CIAAAAJ&hl=en
Collaborators: Ali Shahin Shamsabadi, Carter Luck, Hamed Haddadi

The black-box service model enables ML service providers to serve clients while keeping their intellectual property and client data confidential. Confidentiality is critical for delivering ML services legally and responsibly, but makes it difficult for outside parties to verify important model properties such as fairness. Existing methods that assess model fairness confidentially lack either (i) reliability because they certify fairness with respect to a static set of data and therefore fail to guarantee fairness in the presence

of distribution shift or service provider malfeasance; and/or (ii) scalability due to the computational overhead of confidentiality-preserving cryptographic primitives.
We address these problems by introducing online fairness certificates, which verify that a model is fair with respect to data received by the service provider online during deployment. We then present OATH, a deployably efficient and scalable zero-knowledge proof protocol for confidential online group fairness certification. OATH exploits statistical properties of group fairness via a cut-and-choose style protocol, enabling scalability improvements over baselines.

NLP/LLMs

Poster #27: Majority of the Bests: Improving Best-of-N via Bootstrapping

Presenter: Amin Rakhsha, University of Toronto
Google Scholar: https://scholar.google.com/citations?user=Uqpl3zwAAAAJ&hl=en
Collaborators: Tianyu Zhang (Autodesk), Kanika Madan (Autodesk), Amir-massoud Farahmand (UofT, MILA, Polytechnique Montréal), Amir Khasahmadi (Autodesk)

Sampling multiple outputs from a Large Language Model (LLM) and selecting the most frequent (Self-consistency) or highest-scoring (Best-of-N) candidate is a popular approach to achieve higher accuracy in tasks with discrete final answers. Best-of-N (BoN) selects the output with the highest reward, and with perfect rewards, it often achieves near-perfect accuracy. With imperfect rewards from reward models, however, BoN fails to reliably find the correct answer and its performance degrades drastically. We consider the distribution of BoN's outputs and highlight that, although the correct answer does not usually have a probability close to one under imperfect rewards, it is often the most likely outcome. This suggests that the mode of this distribution can be more reliably correct than a sample from it. Based on this idea, we propose Majority-of-the-Bests (MoB), a novel selection mechanism that estimates the output distribution of BoN via bootstrapping and selects its mode. Experimental results across five benchmarks, three different base LLMs, and two reward models demonstrate consistent improvements over BoN in 25 out of 30 setups. We also provide theoretical results for the consistency of the bootstrapping. MoB serves as a simple, yet strong alternative to BoN and self-consistency, and more broadly, motivates further research in more nuanced selection mechanisms.

Poster #28: Attention Sinks: A 'Catch, Tag, Release' Mechanism for Embeddings

Presenter: Stephen Zhang, University of Toronto
Google Scholar: https://scholar.google.ca/citations?user=rrMHOLYAAAAJ&hl=en
Paper: https://arxiv.org/abs/2502.00919
Collaborators: Mustafa Khan (Peripheral Labs), Vardan Papyan (University of Toronto)

Large language models (LLMs) often concentrate their attention on a few specific tokens referred to as attention sinks. Common examples include the first token, a prompt-independent sink, and punctuation tokens, which are prompt-dependent. While the tokens causing the sinks often lack direct semantic meaning, the presence of the sinks is critical for model performance, particularly under model compression and KV-caching. Despite their ubiquity, the function, semantic role, and origin of attention sinks—especially those beyond the first token—remain poorly understood. In this work, we conduct a comprehensive investigation demonstrating that attention sinks: catch a sequence of tokens, tag them using a common direction in embedding space, and release them back into the residual stream, where tokens are later retrieved based on the tags they have acquired. Probing experiments reveal these tags carry semantically meaningful information, such as the truth of a statement. These findings extend to reasoning models, where the mechanism spans more heads and explains greater variance in embeddings, or recent models with query-key normalization, where sinks remain just as prevalent. To encourage future theoretical analysis, we introduce a minimal problem which can be solved through the 'catch, tag, release' mechanism, and where it emerges through training.

Poster #29: Predicting and improving test-time scaling laws via reward tail-guided search

Presenter: Muheng Li, University of Toronto
Google Scholar: https://scholar.google.ca/citations?user=wkwU1r8AAAAJ&hl=en
Collaborators: Jian Qian (University of Hong Kong), Wenlong Mou (University of Toronto)

Test-time scaling has emerged as a critical avenue for enhancing the reasoning capabilities of Large Language Models (LLMs). Though the straightforward best-of-N (BoN) strategy has already demonstrated significant improvements in performance, it lacks principled guidance on the choice of N, budget allocation, and multi-stage decision-making, leaving substantial room for optimization. In this work, we propose new methodologies to predict and improve scaling properties via tail-guided search. By estimating the tail distribution of rewards, our method predicts the scaling law of LLMs without the need for exhaustive evaluations. Leveraging this prediction tool, we introduce Scaling-Law Guided (SLG) Search, a new test-time algorithm that dynamically allocates compute to identify and exploit intermediate states with the highest predicted potential. We theoretically prove that SLG achieves vanishing regret compared to perfect-information oracles and achieves expected rewards that would otherwise require a polynomially larger compute budget when using BoN. Empirically, we validate our framework across different LLMs and reward models, confirming that tail-guided allocation consistently achieves higher reward yields than Best-of-N under identical compute budgets.

Poster #30: LLMs Uncertainty Quantification via Adaptive Conformal Semantic Entropy

Presenter: Hamed Karimi, Toronto Metropolitan University
Google Scholar: https://scholar.google.com/citations?user=QOIx27EAAAAJ&hl=en
Collaborators: Vaishali Meyappan (Toronto Metropolitan University), Reza Samavi (Toronto Metropolitan University; Vector Faculty Affiliate)

LLMs' overconfidence, particularly when hallucinating, poses a significant challenge for deployment in safety-critical settings and makes reliable estimation of uncertainty necessary. Existing approaches for uncertainty quantification typically prioritize lexical or probabilistic measures; however, these techniques often ignore the semantic variance of different responses with similar meaning. In this paper, we propose Adaptive Conformal Semantic Entropy (ACSE), a method for estimating prompt-level uncertainty by adaptively measuring semantic dispersion in LLM outputs. Our uncertainty scoring function is based on clustering semantic entropy of multiple diverse responses to the same prompt. The function adaptively adjusts the uncertainty score based on semantic features of each cluster. To ensure statistical reliability of our score, we use conformal calibration to apply a decision rule to accept/abstain the prompts, providing a finite-sample, distribution-free guarantee such that the error rate among the accepted responses remains bounded by a user-specified tolerance. Our extensive experimental evaluations using different LLMs and datasets demonstrate that our approach consistently outperforms state-of-the-art uncertainty quantification baselines using discriminative performance, conformal guarantees, and probabilistic calibration indicators. For example, for TriviaQA, AUROC of our approach is 0.88 compared to 0.65 produced by the token entropy approach.

Poster #31: Say It Another Way: Auditing LLMs with a User-Grounded Automated Paraphrasing Framework

Presenter: Rebecca Ma, University of Waterloo

Collaborators: Cléa Chataigner (McGill University & Mila), Rebecca Ma (University of Waterloo), Prakhar Ganesh (McGill University & Mila), Yuhao Chen (University of Waterloo), Afaf Taïk (Université de Sherbrooke), Elliot Creager (Elliot Creager), Golnoosh Farnadi (McGill University & Mila)

Large language models (LLMs) are highly sensitive to subtle changes in prompt phrasing, posing challenges for reliable auditing. Prior methods often apply unconstrained prompt paraphrasing, which risk missing linguistic and demographic factors that shape authentic user interactions. We introduce AUGMENT (Automated User-Grounded Modeling and Evaluation of Natural Language Transformations), a framework for generating controlled paraphrases grounded in user behaviors. AUGMENT leverages linguistically informed rules and enforces quality through checks on instruction adherence, semantic similarity, and realism, ensuring paraphrases are both reliable and meaningful for auditing. Through case studies on the BBQ and MMLU datasets, we show that controlled paraphrases uncover systematic weaknesses that remain obscured under unconstrained variation. These results highlight the value of the AUGMENT framework for reliable auditing.

Poster #32: Dialect Preference Bias in Large Language Models: Evidence from African American English

Presenter: Huan Wu, York University

Collaborators: Muhammad Furquan Hassan (York University), Ali Emami (Brock University), Faiza Khan Khattak (Monark Health), Laleh Seyyed-Kalantari (York University)

African American English (AAE) is spoken by over 30 million people, yet language technologies routinely misinterpret and penalize this dialect. We present a comprehensive empirical study of dialect preference bias in large language models (LLMs) using the first large-scale corpus of authentic AAE data paired with Standard American English (SAE) equivalents. Unlike prior benchmarks that synthesize AAE from SAE using rules or LLMs, our corpus of 16,000+ sentence pairs originates from real AAE tweets, preserving natural dialectal features. Using this resource, we show that state-of-the-art LLMs consistently favor SAE: models assign inconsistent sentiment labels to equivalent AAE and SAE inputs, and prefer SAE continuations even when prompted with AAE context. Critically, our feature-level analysis identifies the linguistic triggers of this bias, revealing that preverbal markers (e.g., finna, done) act as universal bias triggers across all models tested, while syntactic features show model-specific effects. Finally, we demonstrate that dialect bias can be mitigated at test time without retraining: a lightweight steering method that injects learned dialect directions into model activations improves dialect invariance (e.g., 0.81 to 0.89 on Phi-4) while preserving fluency.

Poster #33: MedPerturbing LLMs: A Comparative Study of Toxicity, Prompt Tuning, and Jailbreaks in Medical QA

Presenter: Arash Asgari, York University
Google Scholar: https://scholar.google.ca/citations?hl=en&user=W0mt3IwAAAAJ
Collaborators: Amirreza Naziri, Laleh Seyyed-Kalantari (York University)

Bias evaluation in large language models (LLMs) uses many metrics and benchmarks, but lacks a systematic way to measure agreement across bias metrics and models. As a result, improvements observed under one metric may contradict another, and model rankings may reflect benchmark-specific artifacts rather than stable bias profiles. In this work, we introduce Metric Agreement Score (MeAS) and Model Agreement Score (MoAS), which quantify cross-metric and cross-model agreement in bias rankings, respectively. We apply these measures to ten LLMs, seven bias metrics, and nine corpora. Our results reveal disagreement among both metrics and models: contrary to expectations, we find that metrics within the same category (generation-based and probabilistic) often behave independently of each other. For instance, HONEST shows independence with toxicity metrics, and the CAT score shows no correlation with Language Modeling Bias metric. At the model level, DeepSeek-family models invert bias rankings relative to most others, indicating that the model family strongly shapes specific bias profiles. These findings challenge the assumption that bias mitigation is universally transferable and highlight the need for agreement-aware evaluation.

Poster #34: Attention Sinks in the First Layers Enable Jailbreaking

Presenter: Jonah Mackey, University of Toronto
Google Scholar: https://scholar.google.ca/citations?hl=en&user=M_pFT_IAAAAJ
Collaborators: Stephen Zhang (University of Toronto, Vector Institute), Vardan Papyan (University of Toronto, Vector Institute)

Suffix-based jailbreak attacks append adversarial token sequences to harmful requests, bypassing safety guardrails in language models. Despite their effectiveness, the mechanisms enabling these attacks remain poorly understood. We find that tokens in adversarial suffixes are prone to inducing attention sinks—a phenomenon where certain tokens receive disproportionately high attention from subsequent tokens. We demonstrate that effective adversarial suffixes induce pronounced attention sinks, especially in the early model layers. Ablating these early-layer sinks significantly impairs attack performance, reducing success rates by up to 56%. Building on this mechanistic insight, we introduce SinkGCG, an attack that explicitly promotes sink formation during optimization. SinkGCG consistently outperforms baseline GCG, improving attack success rates by up to 38% across diverse open-weights models. Our results expose a fundamental structural vulnerability in transformer attention mechanisms that can be exploited to bypass safety alignment.

Poster #35: Entropy-Gated Branching for Efficient Test-Time Reasoning

Presenter: Xianzhi Li, Queen's University
Google Scholar: https://scholar.google.com/citations?user=F7B1QQsAAAAJ&hl=en
Paper: https://arxiv.org/abs/2503.21961
Collaborators: Xianzhi Li, Ethan Callanan, Abdellah Ghassel, Xiaodan Zhu (Queen's University)

Test-time compute methods can significantly improve the reasoning capabilities and problem-solving accuracy of large language models (LLMs). However, these approaches require substantially more computational resources, with most compute wasted on exploring low-diversity branches where the model already exhibits high confidence. We observe that a small subset of uncertain reasoning steps has a disproportionately large impact on final prediction accuracy, and branching at these critical junctures tends to yield more diverse and higher-quality candidate reasoning steps. We propose Entropy-Gated Branching (EGB), which branches only at high-uncertainty steps and prunes expansions with a lightweight verifier. On mathematical and financial reasoning benchmarks, EGB improves accuracy by 22.6% over standard inference while operating 31%-75% faster across math benchmarks than test-time beam search with higher performance. Our results show that dynamic resource allocation during inference can substantially improve both efficiency and effectiveness, offering a more scalable pathway to enhanced LLM reasoning capabilities.

Poster #36: Deception Under Public Commitments in N-Player Games

Presenter: Jerick Shi, Carnegie Mellon University
Google Scholar: https://scholar.google.ca/citations?hl=en&user=6wj2mTQAAAAJ
Collaborators: Zhijing Jin (University of Toronto), Terry Zhang

Large language models are increasingly deployed as strategic agents that publicly commit to actions while retaining the ability to privately deviate. Whether such agents break commitments for profit, and whether they do so optimally, remains unclear. We study deception as a unilateral deviation from a publicly announced action that strictly increases payoff in a known one-shot normal-form game. By exhaustively enumerating public announcement profiles across a diverse set of canonical games, we identify all profitable deception opportunities and evaluate both the frequency and strategic quality of agents' deviations. Across multiple state-of-the-art models, we find that agents reliably engage in profit-driven deception. When deviation is profitable, models lie in 80 to 100% of cases, while deception rates drop to near zero when deviation is unprofitable in simple games. However, deception is not always strategically optimal. While agents select optimal deviations in simple binary-action settings, substantial optimality gaps emerge in more complex games such as auctions and two-thirds guessing. These results show that LLM agents readily exploit profitable deception opportunities, but often fail to select payoff-maximizing lies as strategic complexity increases, a gap we trace to computational limits rather than strategic reasoning about others' behavior.

Poster #37: Bias in the Picture: Benchmarking VLMs with Social-Cue News Images and LLM-as-Judge Assessment

Presenter: Aravind Narayanan, Vector Institute
Collaborators: Vahid Reza Khazaie, Shaina Raza (Vector Institute)

Large vision–language models (VLMs) can jointly interpret images and text, but they are also prone to absorbing and reproducing harmful social stereotypes when visual cues such as age, gender, race, clothing, or occupation are present. To investigate these risks, we introduce a news-image benchmark consisting of 1,343 image–question pairs drawn from diverse outlets, annotated with ground-truth answers and demographic attributes (age, gender, race, occupation, and sports). We evaluate a range of state-of-the-art VLMs and employ a large language model (LLM) as judge, with human verification. Our findings show that: (i) visual context systematically shifts model outputs in open-ended settings; (ii) bias prevalence varies across attributes and models, with particularly high risk for gender and occupation; and (iii) higher faithfulness does not necessarily correspond to lower bias.

Poster #38: LinguaMark: Do Multimodal Models Speak Fairly? A Benchmark-Based Evaluation

Presenter: Ananya Raval, Vector Institute
Google Scholar: https://scholar.google.com/citations?user=JfJueNMAAAAJ&hl=en
Paper: https://arxiv.org/abs/2507.07274

Large Multimodal Models (LMMs) are typically trained on vast corpora of image-text data but are often limited in linguistic coverage, leading to biased and unfair outputs across languages. While prior work has explored multimodal evaluation, less emphasis has been placed on assessing multilingual capabilities. In this work, we introduce LinguaMark, a benchmark designed to evaluate state-of-the-art LMMs on a multilingual Visual Question Answering (VQA) task. The dataset comprises 6,875 image-text pairs spanning 11 languages and five social attributes. We evaluate models using three key metrics: Bias, Answer Relevancy, and Faithfulness. Findings reveal that closed-source models generally achieve the highest overall performance. Both closed-source (GPT-4o and Gemini 2.5) and open-source models (Gemma 3, Qwen 2.5) perform competitively across social attributes, and Qwen 2.5 demonstrates strong generalization across multiple languages. The benchmark and evaluation code are released to encourage reproducibility and further research.

Poster #39: Human-Centric Framework for Large Multimodal Models Evaluation

Presenter: Vahid Reza Khazaie, Vector Institute
Google Scholar: https://scholar.google.com/citations?hl=en&user=lEWvRbIAAAAJ
Collaborators: Shaina Raza, Aravind Narayanan, Ashmal Vayani, Mukund Sayeeganesh Chettiar, Deval Pandya (Vector Institute)

Large multimodal models (LMMs) have been widely tested on tasks like visual question answering (VQA), image captioning, and grounding, but lack rigorous evaluation for alignment with human-centered values such as fairness, ethics, and inclusivity. To address this gap, we introduce HumaniBench, a benchmark of 32,000 real-world image-question pairs and an evaluation suite. Labels are generated via an AI-assisted pipeline and validated by experts. HumaniBench assesses LMMs across seven key alignment principles: fairness, ethics, empathy, inclusivity, reasoning, robustness, and multilinguality, through open-ended and closed-ended VQA tasks. Grounded in AI ethics and real-world needs, these principles provide a holistic lens for societal impact. Benchmarking results show that proprietary models generally lead in reasoning, fairness, and multilinguality, while open-source models excel in robustness and grounding. Most models struggle to balance accuracy with ethical and inclusive behavior. HumaniBench offers a rigorous testbed to diagnose limitations and promote responsible LMM development.

Poster #40: Auto-Regressive Masked Diffusion Models

Presenter: Mahdi Karami, University of Waterloo
Collaborators: Ali Ghodsi
Google Scholar: https://scholar.google.com/citations?user=Ul0H6rMAAAAJ&hl=en

Masked diffusion models (MDMs) have emerged as a promising approach for language modeling, yet they face a performance gap compared to autoregressive models (ARMs) and require more training iterations. In this work, we present the Auto-Regressive Masked Diffusion (ARMD) model, an architecture designed to bridge this gap by unifying the training efficiency of autoregressive models with the strengths of diffusion-based learning. Our key insight is to interpret the masked diffusion process as a block-wise causal model. This allows us to design a strictly causal, permutation-equivariant, attention-based architecture that computes all conditional probabilities across multiple denoising steps in a single, parallel forward pass. The resulting architecture supports efficient, autoregressive-style decoding and a progressive permutation training scheme, allowing the model to learn both canonical left-to-right and random token orderings. On standard language modeling benchmarks, ARMD achieves state-of-the-art performance, outperforming established diffusion-based methods while requiring significantly fewer training steps.

Optimization

Poster #41: Neural Operators Can Play Dynamic Stackelberg Games

Presenter: Xuwei Yang, McMaster University
Google Scholar: https://scholar.google.com/citations?user=IfLPuCYAAAAJ&hl=en
Paper: https://arxiv.org/abs/2411.09644
Collaborators: Guillermo Alvarez (University of Michigan), Ibrahim Ekren (University of Michigan), Anastasis Kratsios (McMaster University), Xuwei Yang (McMaster University)

Dynamic Stackelberg games are a broad class of two-player games in which the leader acts first, and the follower chooses a response strategy to the leader's strategy. Unfortunately, only stylized Stackelberg games are explicitly solvable since the follower's best-response operator (as a function of the control of the leader) is typically analytically intractable. This paper addresses this issue by showing that the follower's best-response operator can be approximately implemented by an attention-based neural operator, uniformly on compact subsets of adapted open-loop controls for the leader. We further show that the value of the Stackelberg game where the follower uses the approximate best-response operator approximates the value of the original Stackelberg game. Our main result is obtained using our universal approximation theorem for attention-based neural operators between spaces of square-integrable adapted stochastic processes, as well as stability results for a general class of Stackelberg games.

Poster #42: Capacity-Constrained Online Learning with Delays: Scheduling Frameworks and Regret Trade-offs

Presenter: Alexander Ryabchenko, University of Toronto
Google Scholar: https://scholar.google.com/citations?user=hQ6XzGUAAAAJ&hl=en
Paper: https://proceedings.mlr.press/v291/ryabchenko25a.html
Collaborators: Daniel Roy (University of Toronto and Vector Institute), Idan Attias (IDEAL, hosted by UIC and TTIC)

Our paper studies online learning with delayed feedback under capacity constraints that limit how many outstanding rounds the learner can actively track, capturing real resource bottlenecks that arise in practice. For instance, in healthcare, a limited pool of physicians can constrain how many patients can be followed while outcomes arrive with latency (e.g., side effects or recovery measured days later). We design algorithms that jointly learn and schedule: they decide which delayed rounds to keep in the tracking set and which to ignore or defer, using tools such as delay batching and proxy delays to compress long, irregular delay patterns into a tractable surrogate. The resulting methods achieve minimax-optimal regret across all capacity regimes, and degrade gracefully as capacity shrinks, recovering classical delayed-learning guarantees at high capacity while remaining provably robust when tracking resources are scarce.

Poster #43: Efficient Bilevel Optimization with KFAC-Based Hypergradient

Presenter: Disen Liao, University of Waterloo
Collaborators: Felix Dangel, Yaoliang Yu

Bilevel optimization (BO) is widely applicable to many machine learning problems. However, to scale BO, practitioners often adopt crude approximations like one-step gradient unrolling or identity/short-Neumann surrogates, which discard curvature information. We build on implicit function theorem-based algorithms and propose incorporating Kronecker-factored approximate curvature (KFAC), yielding curvature-aware hypergradients with a better performance–efficiency trade-off than Conjugate Gradient (CG)/Neumann methods and consistently outperforming unrolling. We evaluate our method across diverse tasks, including meta-learning and AI safety-related problems. On models up to BERT, we show that curvature information is valuable at scale, and KFAC can provide it with only modest memory and runtime overhead.

Probabilistic Methods

Poster #44: Cross-Tokenizer Likelihood Scoring Algorithms for Language Model Distillation

Presenter: Truong Buu Phan, University of Toronto
Collaborators: Karen Ullrich (Meta AI), Ashish Khisti (University of Toronto)

Computing next-token likelihood ratios between two language models (LMs) is a standard task in training paradigms such as knowledge distillation. Since this requires both models to share the same probability space, it becomes challenging when the teacher and student LMs use different tokenizers, for instance, when edge-device deployment necessitates a smaller vocabulary size to lower memory overhead. In this work, we address this vocabulary misalignment problem by uncovering an implicit recursive structure in the commonly deployed Byte-Pair Encoding (BPE) algorithm and utilizing it to create a probabilistic framework for cross-tokenizer likelihood scoring. Our method enables sequence likelihood evaluation for vocabularies different from the teacher model native tokenizer, addressing two specific scenarios: when the student vocabulary is a subset of the teacher vocabulary, and the general case where it is arbitrary. In the subset regime, our framework computes exact likelihoods and provides next-token probabilities for sequential sampling with only O(1) model evaluations per token. When used for distillation, this yields up to a 12% reduction in memory footprint for the Qwen2.5-1.5B model while also improving baseline performance up to 4% on the evaluated tasks. For the general case, we introduce a rigorous lossless procedure that leverages BPE recursive structure, complemented by a fast approximation that keeps large-vocabulary settings practical. Applied to distillation for mathematical reasoning, our approach improves GSM8K accuracy by more than 2% over the current state of the art.

Reinforcement Learning & Planning

Poster #45: First-Explore, then Exploit: Meta-Learning to Solve Hard Exploration-Exploitation Trade-Offs

Presenter: Benjamin Norman, University of British Columbia
Collaborators: Jeff Clune (Vector Institute and Canada CIFAR AI Chair)

Standard reinforcement learning (RL) agents never intelligently explore like a human (i.e., taking into account complex domain priors and adapting quickly based on previous exploration). Across episodes, RL agents struggle to perform even simple exploration strategies, such as systematic search that avoids revisiting the same location multiple times. This poor exploration limits performance on challenging domains. Meta-RL is a potential solution, as unlike standard RL, meta-RL can learn to explore and potentially learn highly complex strategies beyond those of standard RL, including experimenting in early episodes to learn new skills or conducting experiments to understand the current environment. Traditional meta-RL focuses on optimally balancing exploration and exploitation to maximize cumulative reward. We identify a limitation of state-of-the-art cumulative-reward meta-RL methods: when optimal behavior requires early sacrifice of reward to enable higher subsequent reward, existing methods become stuck on a local optimum. Our method, First-Explore, overcomes this limitation by learning two separate policies: one to explore and one to exploit. When exploration requires forgoing early-episode reward, First-Explore significantly outperforms existing cumulative meta-RL methods, representing a step toward human-like exploration in meta-RL.

Poster #46: Learning to Negotiate via Voluntary Commitment

Presenter: Shuhui Zhu, University of Waterloo
Google scholar is: https://scholar.google.ca/citations?user=mKti-YAAAAAJ&hl=en&oi=ao.
Paper: https://proceedings.mlr.press/v258/zhu25b.html
Collaborators: Baoxiang Wang (The Chinese University of Hong Kong, Shenzhen), Sriram Ganapathi Subramanian (Carleton University), Pascal Poupart (University of Waterloo & Vector Institute)

Partial alignment and conflict among autonomous agents lead to mixed-motive scenarios in many real-world applications. Agents may fail to cooperate even when cooperation yields better outcomes, often due to non-credible commitments. To facilitate cooperation, we define Markov Commitment Games (MCGs), where agents can voluntarily commit to their proposed future plans. Based on MCGs, we propose a learnable commitment protocol using policy gradients, coupled with incentive-compatible learning to accelerate convergence to equilibria with higher social welfare. Experiments in challenging mixed-motive tasks demonstrate faster empirical convergence and higher returns compared with existing methods.

Poster #47: EM-DQN: Expectation-Maximization for Gaussian-Mixture Value Distributions in Distributional Reinforcement Learning

Presenter: Michal Lisicki, University of Guelph
Google Scholar: https://scholar.google.ca/citations?hl=en&user=ePgzZ18AAAAJ
Collaborators: Graham Taylor (University of Guelph), Mihai Nica (University of Guelph)

Distributional reinforcement learning (DRL) shows that modeling the full return distribution can improve stability and performance; however, existing approaches often rely on computationally expensive objectives or restrictive parametric families. We revisit Gaussian mixture models (GMMs) as value-distribution approximators and propose EM-DQN, a distributional DQN variant that trains a GMM head using an online Expectation–Maximization update. Across synthetic benchmarks and Atari 2600 environments, EM-DQN accurately tracks evolving return distributions and achieves performance competitive with state-of-the-art Wasserstein-based approaches. Importantly, EM-DQN attains this performance with substantially lower computational overhead, scaling linearly as O(BMK) compared to the quadratic O(BTK^2) complexity of Wasserstein-based methods, providing a simple and scalable alternative for efficient multimodal value estimation in DRL.

Societal Considerations

Poster #48: SocialHarmBench: Revealing LLM Vulnerabilities to Socially Harmful Requests

Presenter: Punya Syon Pandey, University of Toronto
Google Scholar: https://scholar.google.com/citations?user=jPxoCxIAAAAJ&hl=en
Paper: https://arxiv.org/pdf/2510.04891

Collaborators: Hai Son Le (Toronto Metropolitan University), Devansh Bhardwaj (Indian Institute of Technology Roorkee), Rada Mihalcea (University of Michigan), Zhijing Jin (University of Toronto, Max Planck Institute for Intelligent Systems, Tubingen, Germany)

Large language models (LLMs) are increasingly deployed in contexts where failures can carry sociopolitical consequences. Existing safety benchmarks sparsely test vulnerabilities in domains such as political manipulation, propaganda generation, or surveillance and information control. To address this, we introduce SocialHarmBench, a dataset of 585 prompts across 7 sociopolitical categories and 34 countries with real-world events, designed to evaluate LLM vulnerabilities to sociopolitical harms. SocialHarmBench enables: (1) adversarial evaluation of high-risk domains including authoritarian surveillance, disinformation campaigns, erosion of democratic processes, and crimes against humanity; (2) evaluation across open-source models, establishing baseline robustness and measuring attack efficiency in politically charged settings; and (3) insights into domain-specific vulnerability comparisons, temporal investigations to trace vulnerable periods, and region-specific vulnerabilities. Findings show that existing safeguards fail to generalize effectively to sociopolitical contexts, exposing partisan biases and limitations in preserving human rights and democratic values.

Theory

Poster #49: A Unified Theory of Attention and State Space Dynamics

Presenter: Aref Jafari, University of Waterloo
Google scholar: https://scholar.google.com/citations?user=HlKxEOEAAAAJ&hl=en
Collaborators: Ali Ghodsi (University of Waterloo), Aref Jafari (University of Waterloo)

Sequence modeling has produced diverse architectures—from classical recurrent neural networks to modern Transformers and state space models (SSMs)—yet a unified theoretical understanding of their expressivity and trainability trade-offs remains limited. We introduce a unified framework representing sequence maps via an input-dependent effective interaction tensor W(X)W(X)W(X), distinguishing between: (i) Explicit Factorization (Attention), where weights are dynamic scalar coefficients applied to shared values, and (ii) Implicit Dynamics (SSMs), where weights are induced by a latent recurrence. Using this framework, we derive three theoretical results: first, we quantify the Interaction Rank Gap, proving that single-head attention is algebraically insufficient to represent high-rank Linear Time-Invariant (LTI) dynamics; second, we establish Head-Count Equivalence, linking the number of attention heads $H \geq k$ to the interaction rank k of an LTI system; third, we characterize the Stability-Trainability Trade-off, showing that while attention mechanisms admit distance-independent “gradient highways,” strictly stable LTI dynamics enforce exponential gradient decay. These results provide a theoretical foundation for hybrid architectures that balance high interaction rank with efficient state compression.

Poster #50: Fairness and Optimization

Presenter: Mojtaba Kolahdouzi, York University
Google Scholar: https://scholar.google.com/citations?user=s94QG7sAAAAJ&hl=en
Paper: https://neurips.cc/virtual/2025/loc/san-diego/poster/115722
Collaborators: Laleh Seyyed-Kalantari (York University), Elham Dolatabadi (York University), Ali Etemad (Queen’s University)

First, we study whether and how the choice of optimization algorithm can impact group fairness in deep neural networks. Through stochastic differential equation analysis of optimization dynamics in an analytically tractable setup, we demonstrate that the choice of optimization algorithm indeed influences fairness outcomes, particularly under severe imbalance. Furthermore, when comparing two categories of optimizers, adaptive methods and stochastic methods, RMSProp (from the adaptive category) has a higher likelihood of converging to fairer minima than SGD (from the stochastic category). Building on this insight, we derive two new theoretical guarantees showing that, under appropriate conditions, RMSProp exhibits fairer parameter updates and improved fairness in a single optimization step compared to SGD. We then validate these findings through extensive experiments on three publicly available datasets, namely CelebA, FairFace, and MS-COCO, across different tasks as facial expression recognition, gender classification, and multi-label classification, using various backbones. Second, we continue this work by analyzing the effects of optimization hyperparameters on the fairness. Using functional Anova, we find that some hyperparameters like learning rate and momentum can largely affect the fairness. We are building on this finding to propose a new class of fair optimizers.

Other AI Applications

Poster #51: Understanding and Improving Shampoo and SOAP via Kullback–Leibler Minimization

Presenter: Wu Lin, Vector Institute
Paper: https://arxiv.org/abs/2509.03378
Collaborators: Wu Lin, Scott C. Lowe, Felix Dangel, Runa Eschenhagen, Zikun Xu, Roger B. Grosse

Shampoo and its efficient variant, SOAP, employ structured second-moment estimations and have shown strong performance for training neural networks. However, Shampoo typically requires step-size grafting with Adam, and SOAP mitigates this at the cost of extra memory overhead. This work recasts these methods as covariance estimation under KL divergence minimization, revealing theoretical limitations and motivating principled redesigns. KL-Shampoo and KL-SOAP match or exceed Shampoo and SOAP performance while achieving SOAP-level runtime. Notably, KL-Shampoo avoids Adam entirely, reducing memory overhead, and consistently outperforms prior methods, establishing KL-based optimization as a promising approach for neural network training.

Poster #52: Chrysalis: Towards Scalable AI-Enabled Peer Tutoring & Assessment

Presenter: Prashanth Arun, University of Waterloo
Collaborators: Pascal Poupart, Igor Grossmann, Kyle Scholz, Vinita Vader, Ana Crisan, Haolin Yu, Arezoo Alipanah, Robert Cai

Chrysalis is a learning companion and assessment platform enabling active learning through AI interactions. Using a “learning by teaching” paradigm, learners teach an AI peer that can “forget” information, prompting learners to reflect critically. Novel assessment techniques leverage human-AI collaboration, showing that teaching AI improves engagement and conversation density. Pilot studies in Computer Science, Psychology, and Environment demonstrate Chrysalis’s versatility and potential to bridge student engagement with institutional adoption of AI learning tools.

Poster #53: Enigma: An Efficient Model for Deciphering Regulatory Genomics

Presenter: Andrew Jung, University of Toronto
Google Scholar: https://scholar.google.com/citations?hl=en&user=tXOXUr8AAAA
Collaborators: Helen Zhu, Roujia Li, Alice J. Gao, Tammy T.Y. Lau, Vivian S. Chu, Declan Lim, Christopher B. Cole, Leo J. Lee, Albi Celaj, Brendan J. Frey

Enigma is a genomics sequence-to-function model achieving competitive performance with state-of-the-art transformer-based models while drastically reducing computational cost. Training on a smaller curated genome track set, Enigma improves zero-shot variant effect prediction and fine-tunes effectively on ChIP-seq, RNA half-life, and translation efficiency tasks. This efficient architecture facilitates broader development and deployment of regulatory genomics models.

Poster #54: From Individual to Multi-Agent Algorithmic Recourse: Minimizing the Welfare Gap via Capacitated Bipartite Matching

Presenter: Zahra Khotanlou, University of Waterloo
Paper: https://arxiv.org/abs/2508.11070
Collaborators: Kate Larson, Amir-Hossein Karimi

This work extends algorithmic recourse from individual recommendations to multi-agent scenarios. It models interactions between multiple seekers and providers as a capacitated weighted bipartite matching problem, optimizing social welfare while minimizing the gap between individual and collective outcomes. A three-layer optimization framework addresses capacity redistribution, cost-aware welfare maximization, and inequality-averse objectives. Experiments show near-optimal welfare with minimal system modification, providing a tractable path to fair multi-agent AI systems.

Poster #55: Benchmarking Histology Foundation Models for Glioblastoma Molecular Prediction with Spatial Transcriptomic Validation of Attention

Presenter: Dilakshan Srikanthan, Queen's University
Google Scholar: https://scholar.google.com/citations?user=VND1sioAAAAJ&hl=en
Collaborators: Amoon Jamzad, Nooshin Maghsoodi, John Rudan, Parvin Mousavi

Evaluating pathology foundation models (CONCH, UNI, Virchow2, GigaPath, H-Optimus-1) on predicting nine molecular alterations in GBM, this study validates attention patterns against spatial transcriptomics. Foundation models improve prediction for specific mutations (IDH1, TP53, CDKN2A) and reveal morphologically silent alterations, offering interpretable and biologically coherent insights for clinical translation.

Poster #56: Reimagining LLMs as Ethical and Adaptive Co-Creators in Mental Health Care Through Trustworthy Evaluation and Alignment

Presenter: Abeer Badawi, York University
Google Scholar: https://scholar.google.com/citations?user=Se-GA04AAAAJ&hl=en
Paper: https://arxiv.org/pdf/2510.19032
Position Paper: https://icml.cc/virtual/2025/poster/40113
Collaborators: Md Tahmid Rahman Laskar, Jimmy Xiangji Huang, Shaina Raza, Elham Dolatabadi, Elahe Rahimi, Sheri Grach, Lindsay Bertrand, Lames Danok, Frank Rudzicz

This work studies responsible deployment of LLMs in mental health. SAFE-i implementation guidelines and HAAS-e evaluation framework operationalize ethical, adaptive, human-centered AI use. Two benchmarks (MentalBench-100k, MentalAlign-70k) evaluate cognitive and affective support. Findings reveal strong LLM performance in cognitive tasks but lower reliability in affective dimensions, highlighting boundaries for safe LLM use and the need for human oversight.

Poster #57: BarcodeMamba+: Advancing State-Space Models for Fungal Biodiversity Research

Presenter: Tiancheng Gao, University of Guelph
Google Scholar: https://scholar.google.com/citations?user=Zf_2nxsAAAAJ
Paper: https://arxiv.org/abs/2512.15931
Collaborators: Scott C. Lowe, Brendan Furneaux, Angel X. Chang, Graham W. Taylor

BarcodeMamba+ is a foundation model for fungal DNA barcode classification, addressing sparse labeling and long-tailed taxa distributions. Using a pretrain and fine-tune paradigm with hierarchical label smoothing, weighted loss, and multi-head output layers, it outperforms traditional supervised approaches across all taxonomic levels. This scalable model advances genomics-based biodiversity research.

Poster #58: Vision-Language Models Learn Clinical Concepts for Breast Cancer Detection

Presenter: Mohamed Harmanani, Queen’s University
Google Scholar: https://scholar.google.ca/citations?hl=en&user=dNDG5CgAAAAJ
Collaborators: Bining Long, Zhuoxin Guo, Paul F.R. Wilson, Amirhossein Sabour, Minh Nguyen Nhat To, Gabor Fichtinger, Purang Abolmaesumi, Parvin Mousavi

Foundation models improve breast cancer mutation prediction using H&E slides. Compared to baseline and existing models, vision-language models achieve higher AUROC for key mutations and identify morphologically silent alterations. Attention patterns validated with spatial transcriptomics provide interpretable insights for clinical translation.

Poster #59: Hash Collisions in Molecular Fingerprints: Effects on Property Prediction and Bayesian Optimization

Presenter: Walter Virany, University of Toronto
Google Scholar: https://scholar.google.com/citations?user=S2GA1h8AAAAJ&hl=en
Paper: https://arxiv.org/abs/2511.17078
Collaborators: Austin Tripp, Valence Labs

This work investigates the impact of hash collisions in molecular fingerprints on property prediction and Bayesian optimization. Using exact fingerprints improves predictive accuracy on molecular property benchmarks but does not significantly affect Bayesian optimization outcomes.

Poster #60: ProstNFound+: A Prospective Study using Medical Foundation Models for Prostate Cancer Detection

Presenter: Paul Wilson, Vector Researcher (supervised by Vector Faculty Member), Queen’s University
Linkedin: https://www.linkedin.com/in/paul-wilson-832313179/
Poster: https://arxiv.org/abs/2510.26703

Collaborators: Mohamed Harmanani, Minh Nguyen Nhat To, Amoon Jamzad, Tarek Elghareb, Zhuoxin Guo, Adam Kinnaird, Brian Wodlinger, Purang Abolmaesumi, Parvin Mousavi

Purpose: Medical foundation models (FMs) offer a path to build high-performance diagnostic systems. However, their application to prostate cancer (PCa) detection from micro-ultrasound (µUS) remains untested in clinical settings. We present ProstNFound+, an adaptation of FMs for PCa detection from µUS, along with its first prospective validation.
Methods: ProstNFound+ incorporates a medical FM, adapter tuning, and a custom prompt encoder that embeds PCa-specific clinical biomarkers. The model generates a cancer heatmap and a risk score for clinically significant PCa. Following training on multi-center retrospective data, the model is prospectively evaluated on data acquired five years later from a new clinical site. Model predictions are benchmarked against standard clinical scoring protocols (PRI-MUS and PI-RADS).
Results: ProstNFound+ shows strong generalization to the prospective data, with no performance degradation compared to retrospective evaluation. It aligns closely with clinical scores and produces interpretable heatmaps consistent with biopsy-confirmed lesions.
Conclusion: The results highlight its potential for clinical deployment, offering a scalable and interpretable alternative to expert-driven protocols.

VECTOR INSTITUTE | INSTITUT VECTEUR

Thank you for attending **Remarkable 2026**

This publication is for informational purposes only and Vector Institute is not, by means of this publication, rendering professional advice or services or providing an opinion of any kind on any subject. No one should act upon or refrain from acting upon the information contained in this publication without obtaining the advice of a qualified professional advisor. You are urged to contact a qualified professional advisor for guidance. Vector Institute does not warrant or guarantee the accuracy, currency, usefulness, or completeness of the information contained in this publication, which may include information obtained from third party sources. Vector Institute shall not have any responsibility or owe any duty to any person in respect of this publication or be responsible for any loss whatsoever sustained by any person who relies on this publication.

About Vector Institute
Launched in 2017, the Vector Institute works with industry, institutions, startups, and governments to build AI talent and drive research excellence in AI to develop and sustain AI-based innovation to foster economic growth and improve the lives of Canadians. Vector aims to advance AI research, increase adoption in industry and health through programs for talent, commercialization, and application, and lead Canada towards the responsible use of AI. Programs for industry, led by top AI practitioners, offer foundations for applications in products and processes, company-specific guidance, training for professionals, and connections to workforce-ready talent.

Vector is funded by the Province of Ontario, the Government of Canada through the Pan-Canadian AI Strategy, and leading industry sponsors from across multiple sectors of Canadian Industry.

For further information or media enquiries, please contact: media@vectorinstitute.ai