CAPWIC 2026 Technical Program

CAPWIC is an ACM Capital Region Celebration of Women in Computing that provides a low-cost, regionally tailored conference for women and minorities in computing. Participants include students, faculty, and professionals, as well as all supporters of women in computing.

Sponsors
CAPWIC 2026 would not be possible without the efforts of the organizers and the generous support of our sponsors. We are grateful for their contributions (see below).

Gold

Silver


Northeastern University
Khoury College of
Computer Sciences


UNIVERSITY of VIRGINIA


LOYOLA
UNIVERSITY MARYLAND

In-kind


aws


WOMEN IN CYBERSECURITY
WiCyS

Steering Committee

Dr. Michael Stewart, James Madison University
Dr. Megan Olsen, Loyola University Maryland
Dr. Swapneel Sheth, University of Pennsylvania
Dr. Nada Basit, University of Virginia

Organizing Committee

Chairs
Dr. Sehrish Basir Nizamani, Virginia Tech
Dr. Mona Rizvi, James Madison University

Local Chairs
Dr. Melissa Cameron, Virginia Tech
Dr. Erika Olimpiew, Virginia Tech

Technical Program Chairs
Dr. Jin-Hee Cho, Virginia Tech
Dr. Mohammed Farghally, Virginia Tech

Web and Registration
Dr. Swapneel Sheth, University of Pennsylvania

Publicity and Outreach
Dr. Zhuojun Duan, James Madison University

Scholarships
Dr. Maha Elouni, Randolph-Macon College

Education Liaison
Deborah Kariuki, University of Maryland, Baltimore County

Data and Birds-of-the-Feather Chair
Allyson Senger, Virginia Tech

Keynote Speakers and Awards Chair
Dr. Megan Olsen, Loyola University Maryland

Technical Workshops
Dr. Nguyen Ho, Loyola University Maryland

Student Posters and Competition
Dr. Jessica Zeitz, University of Mary Washington

Research Shorts and Competition
Dr. Mai Dahshan, University of Virginia

Flash Talks and Competition
Dr. Mona Rizvi, James Madison University

Panels
Dr. Chandani Shrestha, James Madison University

Technical Talks
Dr. Safdar Hussain Bouk, Old Dominion University

Website Management
Dr. Michael Stewart, James Madison University
Dr. Javier Tibau, Unaffiliated

Reviewing Committee (Additional Reviewers)
Prajakta Belsare, James Madison University
Karen Anewalt, University of Mary Washington

Welcome Message from the Chairs

Welcome to CAPWIC 2026, the ACM Capital Region Celebration of Women in Computing. We are delighted to bring together students, faculty, professionals, and allies from across the region for two days of learning, collaboration, and community building.

CAPWIC continues to serve as a platform to support the recruitment, retention, and advancement of women in computing. This year's program highlights a diverse range of technical talks, research shorts, flash talks, posters, workshops, panels and birds-of-a-feather sessions that reflect the breadth of innovation and inquiry in our field. We are especially excited to provide opportunities for students to showcase their work, connect with mentors, and explore academic and industry pathways.

We extend our sincere gratitude to our sponsors, program committee, and organizing team for their dedication and support in making this conference possible. We also thank all participants for contributing to an inclusive and engaging environment.

We hope you find CAPWIC 2026 inspiring, informative, and a valuable opportunity to build lasting connections within the computing community.

Sincerely,
CAPWIC 2026 Chairs

Keynote Speakers

Dr. Laurian Vega
System Engineer, Senior at Booz Allen Hamilton

Tools in Your Toolbox: What I've Learned as a Professional Female Computer Scientist

In today's rapidly evolving **technological** landscape, thriving as a professional in computer science requires not just technical expertise but also adaptability, resilience, and a toolkit of practical strategies for navigating challenges. For women in this field, this journey often involves overcoming unique obstacles while carving out space for innovation, leadership, and growth.

In this keynote address, I'll share the tools I've cultivated throughout my career as a female computer scientist—tools that have helped me break barriers, foster collaboration, and stay inspired in an ever-changing industry. From technical skills and problem-solving approaches to strategies for handling imposter syndrome and advocating for diversity, equity, and inclusion, this talk will offer actionable insights for professionals at every stage of their journey.

Whether you're just building your toolbox or looking to refine it, my goal is to equip you with empowering lessons learned from real-world experiences: navigating workplace dynamics, balancing competing demands, taking risks, and finding opportunities to mentor and uplift others. Join me as I reflect on what it means to innovate and thrive—not just as a computer scientist, but as a leader, a learner, and an advocate for change in the tech world. Together, let's add some powerful new tools to your toolbox.

Candace Aku
Senior Technical Program Manager at Google Public Sector

Goodbye Imposter, Hello Winner: Overcoming Perceptual Expectations to Reclaim Excellence

What happens when your identity is shaped by the "perceptual expectations" of everyone around you? Others often develop their own thoughts about who you are and those thoughts become the very chains that limit your professional potential. For many women in tech, external narratives become a "groomed perfection" that can lead to identity crises, spiraling, and the heavy weight of imposter syndrome.

In this keynote, Candace Aku shares a transparent look at her journey from living under the shadow of others' expectations and traces her unconventional path through the tech industry. This session explores the process of identifying which parts of your professional identity are yours and which are merely the result of perceptual expectations.

Join us to learn how to stop meeting the expectations of others and start exceeding your own.

Dr. Meg Dickey-Kurdziolek
UX Lead / Sr. Staff UX Researcher, Intrinsic.ai

Human-Centered Automation: A Journey through HCI, AI, and the Future of Robotics

As AI shifts from digital chatbots to physical entities that move, build, and interact in our world, the role of the UX professional is undergoing a radical transformation. In this keynote, Dr. Meg Kurdziolek, UX Lead at Intrinsic.ai, explores the fascinating intersection of Human-Computer Interaction and industrial robotics.

Drawing from her career journey, starting from a Ph.D. at Virginia Tech to leading UX at a Google-born robotics startup, Meg will share how her perspective on "human-centeredness" has evolved. She will dive into the unique challenges of designing for complex physical systems, the critical role of Explainable AI (XAI) in building trust with machines, and how AI tools are reshaping the very nature of UX research. Whether you are a student just starting out or a seasoned professional, this talk offers a roadmap for navigating the "wild west" of robotics and automation.

Meg is the UX lead for Intrinsic.ai, where she focuses on making it easier to create automation solutions with industrial robotics. She is a “Xoogler” and previously worked on Vertex AI and Explainable AI services for Google Cloud. Meg has had a varied career working for start-ups and large corporations alike, and she has published on topics such as information visualization, educational-technology design, voice user interface (VUI) design, explainable AI (XAI), and human-robot interaction (HRI). Meg is a proud alumnus of Virginia Tech, where she received her Ph.D. in Human-Computer Interaction.

Technical Program Summary

Panels: 2 abstracts

Technical Workshops: 5 abstracts

Technical Talks: 8 abstracts

Flash Talks: 12 abstracts

Research Shorts: 22 abstracts

Posters: 43 abstracts

Birds of a Feather: 1 abstract

Panels

Total abstracts in this session: 2

Navigating the Path to Grad School: Discuss, Reflect, and Make an Informed Decision

Mohammed Seyam (Virginia Tech), Madison Barton (Northeastern University), Mohammed Farghally (Virginia Tech), Promise Owa (Northeastern University)

Bringing together faculty and current graduate students, this session offers a reflective look at what it's actually like to pursue an advanced degree and how to decide whether it's the right step for you. Designed to make the process feel less overwhelming and more grounded, this interactive panel creates space for real stories, practical advice, and open conversation. We'll start with reflection. What do you hope grad school will change, strengthen, or open up in your life? Panelists will share their own journeys, like their uncertainties, turning points, and lessons learned, while encouraging participants to think intentionally about their goals, interests, and readiness. We'll also explore the application process in a straightforward and approachable way. We'll talk about how to craft a personal statement that sounds like you, how to request strong letters of recommendation, and how to navigate common challenges along the way. Current graduate students will share what they wish they had known before applying and before committing to a program. Finally, we'll focus on making the decision. Together, we'll consider factors, what truly shapes your graduate experience, such as financial support, mentorship, community, well-being, and long-term career goals. Throughout the session, participants will have time to ask questions and engage in open dialogue, creating a supportive environment where uncertainty is normal and thoughtful exploration is encouraged.

Reimagining Classroom Engagement: Innovative Approaches to Student Discussion

Chandani Shrestha, Mona Rizvi, Alexander Heckel, Michael Stewart (James Madison University)

Both students and instructors face barriers to successful classroom discourse. Students may have apprehensions when uncertain how others in the class will respond to their viewpoint, while instructors may experience similar uncertainties along with challenges in guiding discussions, particularly when strong emotions arise. ThoughtSwap (TS) is a classroom discussion facilitation tool designed to foster Conscientious Discourse. Rather than positioning itself as a solution, it functions as a supportive layer that reshapes how discussion occurs. Instructors share a prompt and students submit responses with authorial privacy, creating a brave space for sincere expression while maintaining moderation. Responses can be redistributed so students engage with a peer’s idea, providing an accessible entry point and enabling instructors to design a range of discussion activities. TS can help reduce some barriers to classroom discourse while also drawing attention to underlying emotional and pedagogical challenges that may otherwise remain unseen. One of the instructors who used TS in her class observed that students often have varied and unexpected reactions to disagreement, and balancing sincere expression with guidance that keeps focus on ideas rather than emotions is challenging. Challenges may vary, from the initiation of discussions to their progression and potential effects on students. This CAPWIC session offers an opportunity to bring together instructors and students with diverse experiences to discuss these concerns, better understand multiple perspectives, share strategies, and inform future iterations of the tool and associated discussion activities.

Technical Workshops

Total abstracts in this session: 5

Designing Safer AI Agents: Secure-by-Design Patterns for AI Systems

Pavan Reddy (The George Washington University)

AI agents can sound confident while doing the wrong thing, especially when they can read private data or trigger real business actions. This beginner-friendly, 120-minute hands-on workshop teaches practical LLM security using two concrete examples: a car dealership chatbot that can view inventory and internal notes, and a sandboxed EchoLeak incident involving prompt injection and unauthorized data disclosure in Microsoft 365 Copilot. This workshop will start with a simple threat-modeling introduction (what the agent should and shouldn't do, what "untrusted text" means, and why safety filters aren't the same as security), then move into a guided sandbox where participants try harmless "boundary-pushing" prompts and see how small wording changes can cause oversharing or unintended tool use (such as getting a 99% discount on a car dealership website). Then, it will move onto an EchoLeak case study to explain indirect prompt injection and why mixing instructions with untrusted content is risky in RAG-style systems. Defenses will be added progressively to understand how they impact security, building up to a secure system

Debugging Your Resume

Aubrey Baker (Red Van Workshop), Holly Wilsey (Purple Basil Games)

Led by an eCommerce Web Developer and a Video Game Engineer, this hands-on workshop guides students and young professionals in crafting a clear, effective resume customized to their academic work, internships, jobs, and volunteer experiences. Learn how Applicant Tracking Systems (ATS) screen resumes long before a human ever sees them and how to avoid common pitfalls that can hold your resume back. Participants will work in breakout groups to review, edit, and strengthen their resumes, while receiving tips and guidance in a supportive environment. Bring a copy of your resume and something to write with, and be ready to leave with a stronger, more compelling resume!

Intended Audience: Students and young professionals

Power/AV equipment/space need: Ideal room setup would have a projector, so that we can display some examples. Having round tables or rows of tables setup to facilitate writing on resumes and forming breakout groups would be helpful.

Augmenting Datasets with Generated Data: Methods and Evaluation

Rebecca Ansell (Georgetown University), Lisa Singh (Georgetown University)

As machine learning models demand larger training datasets, synthetic data generation offers a powerful solution for augmenting limited real-world data. This hands-on workshop teaches participants to generate augmented data using LLM-based techniques and rigorously evaluate its effectiveness.

Through interactive coding exercises, attendees will implement data generation methods including prompting strategies, and controlled generation. We'll focus on the critical evaluation challenge: determining whether augmented data improves model performance while maintaining distributional alignment and avoiding artifacts. Participants will explore evaluation metrics including distribution similarity measures, downstream task benchmarking, and human evaluation protocols.

Attendees will leave with practical experience building augmentation pipelines, implementing evaluation frameworks, and making informed decisions about incorporating generated data into ML workflows.

AWS 101 Workshop

Srinija Desiraju (AWS)

This foundational workshop provides a comprehensive introduction to Amazon Web Services for participants at all skill levels. Attendees will gain hands-on experience with core AWS services including compute (EC2), storage (S3), databases (RDS), and networking fundamentals. The workshop covers essential cloud concepts, the AWS shared responsibility model, and basic architectural patterns, enabling participants to start building on AWS with confidence.

All workshops include interactive demonstrations, hands-on labs, and practical examples that participants can apply immediately. These sessions are designed to accommodate various experience levels and provide valuable skills for cloud computing and AI innovation.

Best Practices for Using Gen AI to Manage Cloud Compliance

Srinija Desiraju (AWS)

This workshop explores how generative AI can revolutionize cloud compliance management. Participants will learn to leverage AI-powered tools to automate compliance monitoring, detect security vulnerabilities in real-time, and streamline audit processes. The session covers best practices for implementing AI-driven governance frameworks, ensuring data protection, and maintaining regulatory compliance across cloud environments.

Technical Talks

Total abstracts in this session: 8

Still Here, Still Learning: A Woman's Journey Back to Graduate Computing

Shareka Robinson (Morgan State University)

After more than 20 years working in information technology, I returned to graduate school in computing at Morgan State University confident in my professional experience, yet unexpectedly confronted with self-doubt and identity shifts that can accompany re-entering academia later in one's career. I entered graduate study expecting technical rigor---but not the personal transformation that comes with being both highly experienced and still learning as a woman in computing.

In this talk, I reflect on moments that tested my confidence, from sitting in classrooms alongside much younger peers to navigating research environments that do not always recognize industry expertise. Returning to graduate school has reshaped how I think about leadership, mentorship, and what it means to model lifelong learning in computing. I share strategies that helped me reconnect with my sense of belonging, including reframing experience as an asset, building community, and choosing to show up authentically rather than invisibly.

This talk is for students and professionals who wonder if it is "too late," "too different," or "too risky" to pursue advanced computing education. I close with a call to action for women and allies to value non-linear paths, amplify lived experience, and help create computing spaces where curiosity and experience are equally celebrated.

Designing the AI-Ready Workforce, A Systems Approach to AI–Human Teaming in Roles, Hiring, and Performance

Denise D'Angelo (Referral Reka Rob and Dr. Nizamani)

AI adoption is accelerating across computing education, research, and industry, yet the structures that shape work, evaluation, and advancement have not evolved at the same pace. As AI becomes embedded in everyday work, long-standing assumptions about roles and performance no longer hold, particularly in AI–human teaming environments. As artificial intelligence becomes embedded across computing workflows, a shift is underway in role design and in how performance is approached. An AI-ready workforce extends beyond tool proficiency to include the human and system-level conditions that shape judgment, evaluation, and execution. A systems perspective helps establish clarity around expectations, responsibility, and trust as AI becomes part of everyday work. Drawing on Denise D'Angelo's experience leading large-scale AI and workforce transformation initiatives across industry and public-sector environments, this talk applies a systems approach to building an AI-ready workforce. The systems approach includes the impact of AI-assisted recruiting and filtering tools, particularly where they replace early human interaction and shape opportunity and trust. The session situates hiring, performance interpretation, and trust within a unified framework for AI–human teaming, emphasizing alignment across role design, evaluation practices, and execution. This talk offers a perspective-driven contribution to practice and community direction, focusing on how performance is interpreted, how talent is assessed, and how trust is built when AI is part of everyday work. Attendees will leave prepared to approach AI-enabled work with clearer

performance expectations, stronger hiring and interviewing practices, and a systems-level understanding of how trust and accountability support an AI-ready workforce.

QBTrain: Zero-Cost, Accessible Hands-On Educational Labs for AI Agent Security

Pavan Reddy (The George Washington University)

Hands-on learning in AI is still out of reach for many people: resources are often theory-heavy, paywalled, or require expensive setups and instructor time. Learners and educators need a practical, low-friction way to build real intuition, both for core AI concepts and for the security risks that appear once AI is connected to data and actions. QBTrain is my response: a free, source-available, browser-based training platform built around guided, self-serve interactive labs that can be used independently or in classrooms and community cohorts. In this 20-minute talk (including a live demo), I'll explain the problem QBTrain addresses, the learning model behind it (repeatable practice and safe failure), and how it improves access to modern AI education. I'll close with ways instructors and groups can adopt it quickly and the impact it can have on expanding who gets to learn AI and AI security through hands-on work.

Audience Level: Beginner-Intermediate

Natural Language as the User Interface: Designing Predictive AI That Actually Gets Used

Emily Smith (N/A)

Human interaction with software is dominated by rigid forms, menus, and structured inputs --- yet in daily life, people communicate through natural language. In this talk, I present a production-tested approach to designing systems where natural language is the user interface. By combining schema-driven validation, modular prompt components, and staged processing pipelines, we can translate freeform user input into reliable, structured data that powers real products.

Drawing on multiple real-world deployments --- from volunteer tracking and benefits Q&A tools to scalable lead capture systems --- I'll outline patterns for:

Schema-first design that guides conversation without breaking natural flow

Repair loops and staged validation that gracefully handle ambiguous or partial text

Composable prompt components that generalize across domains

Serverless, multi-tenant architecture that scales without brittle custom code

Attendees will gain practical insights into how predictive AI systems are built for resilience, maintainability, and human trust --- not just accuracy on benchmarks. This talk bridges theory and practice, offering engineers, researchers, and practitioners concrete strategies to bring natural-language interfaces into production with confidence.

AGRI-Fidelity: Evaluating the Reliability of Listenable Explanations for Animal Disease Detection

Sindhuja Madabushi (Virginia Tech), Arda Dogan (Virginia Tech), Jonathan Liu (Virginia Tech), Dian Chen (Virginia Tech), Dong S Ha (Virginia Tech), Sook Shin (Virginia Tech), Jin-Hee Cho (Virginia Tech)

Explainable AI (XAI) methods are increasingly used in audio-based animal disease detection. However, existing XAI evaluation metrics largely assess faithfulness with respect to a single model and ignore explanation instability across near-optimal models and alignment with domain-specific physiological knowledge. In noisy farm environments, models with similar predictive performance may rely on different acoustic cues, yielding explanations that are faithful yet model-dependent or driven by spurious correlations. As a result, faithfulness alone does not ensure explanation reliability. We propose AGRI-Fidelity, a reliability-oriented metric for evaluating listenable explanations in farm animal disease detection. AGRI-Fidelity combines three components: (i) a consensus-based stability score measuring cross-model agreement, (ii) a fidelity score quantifying the impact of explanation-identified regions on model predictions, and (iii) a domain alignment score assessing consistency with established bioacoustic markers of poultry health. Experiments on real-world poultry vocalization data show that AGRI-Fidelity effectively distinguishes reliable, physiologically meaningful explanations from spurious yet faithful ones, highlighting the need to incorporate stability and domain knowledge into XAI evaluation for farm animal disease detection.

X-MAP: eXplainable Misclassification Analysis and Profiling for Spam and Phishing Detection

Qi Zhang (Virginia Tech), Dian Chen (Virginia Tech), Lance Kaplan (U.S. Army DEVCOM Army Research Laboratory), Audun Josang (University of Oslo), Dong Hyun Jeong (University of the District of Columbia), Feng Chen (University of Texas at Dallas), Jin-Hee Cho (Virginia Tech)

Misclassifications in spam and phishing detection are very harmful, as false negatives expose users to attacks while false positives degrade trust. Existing uncertainty-based detectors can flag potential errors, but possibly be deceived and offer limited interpretability. This paper presents X-MAP, an eXplainable Misclassification Analysis and Profilling framework that reveals topic-level semantic patterns behind model failures. X-MAP combines SHAP-based feature attributions with non-negative matrix factorization to build interpretable topic profiles for reliably classified spam/phishing and legitimate messages, and measures each message's deviation from these profiles using Jensen–Shannon divergence. Experiments on SMS and phishing datasets show that misclassified messages exhibit at least two times larger divergence than correctly classified ones. As a detector, X-MAP achieves up to 0.98 AUROC and lowers the false-rejection rate at 95% TRR to 0.089 on positive predictions. When used as a repair layer on base detectors, it recovers up to 97% of falsely rejected correct predictions with moderate leakage. These results demonstrate X-MAP's effectiveness and interpretability for improving spam and phishing detection.

Houston, We Have a Model: Performance Assessment for Mission-Critical AI

Cheryl Howard (IBM), Peyton Cooper (IBM)

As AI and ML systems move from experimentation into mission-critical environments, rigorous performance assessment becomes essential to ensuring reliability, safety, fairness, and business continuity. In high impact settings where model failures can result in financial loss, regulatory violations, reputational damage, or safety risks, accuracy metrics alone are insufficient.

We present a framework for evaluating both classical ML and large language models across multiple dimensions aligned to risk and system impact. We begin with core evaluation

approaches for classification and regression tasks, then extend to LLM-specific assessment techniques, including perplexity, BLEU, ROUGE, and structured human-in-the-loop evaluation.

Emphasis is on challenges unique to mission-critical generative AI, such as hallucination detection, factual accuracy verification, prompt robustness under adversarial or ambiguous inputs, and consistency across executions. We discuss how failure modes manifest in production systems and how to measure them before they cause downstream harm.

We highlight the foundational role of data quality as a prerequisite for trustworthy AI, covering data validation, bias detection in training and evaluation datasets, and the impact of data drift on model performance. We also examine how use-case appropriate metrics enable teams to align model evaluation with operational goals and risk tolerance.

Attendees will learn techniques for detecting model degradation, measuring fairness, and implementing continuous monitoring pipelines that provide real time visibility into model health. Real world examples illustrate how performance assessment frameworks grounded in high quality data and continuous evaluation help prevent costly deployment failures and ensure AI systems deliver reliable, accountable, and sustained value in mission-critical applications.

Beyond the Weakest Link

Deborah Kariuki (University of Maryland Baltimore County), Hidare Debar (University of Maryland Baltimore County)

The phrase "humans are the weakest link" has long shaped cybersecurity practice, influencing training, policy, and system design. In the age of AI-enabled security systems, this framing is no longer merely incomplete; it is technically misleading. As security decisions increasingly rely on automation, adaptive interfaces, and machine-generated recommendations, human behavior is inseparable from system design, model outputs, and organizational incentives. This technical talk argues for moving beyond the weakest-link narrative by examining how contemporary cybersecurity failures emerge from human AI interaction breakdowns, rather than isolated user error. Drawing on examples from phishing detection, alert fatigue, authentication workflows, and AI-assisted security tools, the talk illustrates how automation bias, opaque model behavior, and poorly calibrated trust systematically shape human actions and error patterns.

Grounded in principles from Human-Centered Computing (HCC), the talk reframes humans not as liabilities to be controlled, but as integral components of sociotechnical security systems. From an HCC perspective, security outcomes depend on how well systems support human judgment through meaningful feedback, risk visibility, explainability, and alignment between system intent and user understanding. The session introduces design-oriented technical principles for AI-enabled security systems, including trust calibration, interaction transparency, and failure-aware interface design. These principles emphasize designing with human capabilities and limitations, rather than designing around them.

This talk is intended for cybersecurity practitioners, system designers, and researchers seeking more effective and realistic approaches to security in AI-driven environments approaches where improving system design, not blaming users, is central to improving security outcomes.

Research Shorts

Total abstracts in this session: 22

A Hybrid Multi-Agent Prompting Approach for Simplifying Complex Sentences

Pratibha Zunjare (VT), Michael Hsiao (VT)

This paper addresses the challenge of transforming complex sentences into sequences of logical, simplified sentences while preserving semantic and logical integrity with the help of Large Language Models. We propose a hybrid approach that combines advanced prompting with multi-agent architectures to enhance the sentence simplification process. Experimental results show that our approach was able to successfully simplify 70% of the complex sentences written for video game design application. In comparison, a single-agent approach attained a 48% success rate on the same task.

Do you feel the burn? Exploring lightweight haptic feedback and how it can improve exercise performance.

Julia Larson (James Madison University), Jason Forsyth (James Madison University)

Automated, computerized exercise feedback systems have been proposed to monitor and provide live feedback to users while completing unsupervised exercise training at home. Positive impacts such as reducing errors in motor learning, detecting missed repetitions, and enhancing exercise performance have been observed. Haptic feedback has shown to be an effective method of feedback for these systems and has been explored via various metaphors. Forceful haptic feedback is effective in guiding motion, but has also been reported to be difficult to interpret. In contrast, tapping haptic feedback has shown to be more distinguishable than forceful, while remaining effective in motion trajectory correction. However, there is another feedback method used in practice that has not yet been explored. In addition to forcefully guiding a patient through an exercise, and gently tapping their arms or legs to move into a certain position, they'll also tap on certain muscle regions of the body where a patient should be feeling activation in certain muscles. We propose a new haptic feedback metaphor, MyoTap, that utilizes vibrotactile cues similar to that of the tapping metaphor, but instead of placing an actuator on an end effector that is being instructed to move, the actuator is placed on a muscle region that is supposed to feel. We propose this feedback method will result in a more positive user experience in terms of usability while remaining comparable in motion trajectory correctness.

RESFL: Uncertainty-Aware Federated Learning for Joint Privacy, Fairness, and Utility

Dawood Wasif, Terrence Moore, Jin-Hee Cho

Federated learning (FL) enables collaborative model training without centralizing data, but privacy mechanisms can undermine fairness by obscuring sensitive signals needed to correct demographic performance gaps. We present RESFL, an uncertainty-aware framework that jointly balances privacy, fairness, and utility for FL-based object detection in high-stakes settings. RESFL integrates two components. First, adversarial privacy disentanglement uses a gradient reversal objective to suppress sensitive-attribute information in learned

representations, reducing privacy risks while preserving task-relevant features. Second, uncertainty-guided fairness-aware aggregation uses an evidential neural network to estimate client uncertainty and adaptively weight updates, prioritizing contributions that are both high-confidence and exhibit lower fairness disparities. This aggregation reduces the influence of unreliable or biased updates and improves stability under heterogeneous client conditions. Experiments on the FACET and CARLA benchmarks demonstrate that RESFL achieves strong accuracy while improving equity and privacy relative to FedAvg, including reduced membership-inference attack success and smaller equality-of-opportunity gaps, with improved robustness to adversarial perturbations. RESFL is model- and domain-agnostic, making it applicable beyond autonomous driving to other federated applications where privacy, fairness, and reliability must be optimized together.

Interpretable AI for Animal Health: Bayesian Networks with Uncertainty-Aware Decision Making

Dian Chen (Virginia Tech), Sindhuja Madabushi (Virginia Tech), Raj Kashikar (Virginia Tech), Angelie Nguyen (Virginia Tech), Dong Ha (Virginia Tech), Sook Shin (Virginia Tech), Jin-Hee Cho (Virginia Tech)

Timely and trustworthy animal disease detection is essential for animal welfare, food security, and farm profitability, yet existing AI systems struggle with noisy sensors, class imbalance, and limited interpretability. We present U-DeepBN, an Uncertainty-Aware Bayesian Network with Deep Learning, which integrates deep feature learning, uncertainty quantification, and Bayesian reasoning to enable robust and explainable disease prediction. By modeling feature-level uncertainty and explicit feature dependencies, U-DeepBN provides intrinsic interpretability and transparent inference under sensor noise and adversarial perturbations. Experiments on three real-world animal disease datasets show that U-DeepBN improves accuracy and robustness, achieves higher explanation fidelity and stability, and reduces runtime by orders of magnitude compared to strong baselines. Ablation studies further demonstrate that optimal uncertainty types and feature representations are dataset-dependent, with adaptive uncertainty modeling and hybrid features yielding the most reliable performance. The results highlight U-DeepBN as a practical, deployable decision-support tool for precision livestock monitoring.

OSMoSIS: Enabling Metacognitive Feedback Reflection in LLMs

Isaiah Freeman (Graduate Student at Virginia State University), Joseph Shelton (Faculty at Virginia State University)

In higher education, there are large volumes of open-text educational feedback, such as comments from an end-of-course survey, that are difficult to analyze systematically. Moreover, as educational institutions begin to experiment with using LLMs to cultivate direct feedback conversation for end-of-course surveys, there is a need for automatic assessment of whether the student's responses address the survey's objectives. We propose OSMoSIS, a five-label hierarchical classification system. The system classifies responses through chain-of-thought: first determining if an objective is addressed, then assessing whether addressed responses are direct or partial, and whether unaddressed responses are on-topic or tangential. We propose that by interpreting the logic of OSMoSIS as a chain-of-thought questionnaire, LLMs can provide efficient and explainable zero-shot classification between the five labels. In initial validation using synthetic conversation data, we hypothesize that there should be moderate agreement between human and LLM ratings. Future work aims to directly survey students at the end of a

semester, recompute ratings for human-written data, and support faculty and administration by measuring improvement of feedback from student surveys.

Investigating the Neural Correspondence of Speech-Based Neural Networks with Human Brain Responses

Ivy Brundege (Virginia Tech- Computer Science), Gasser Elbanna (MIT), Josh McDermott (MIT)

Speech recognition, while integral to human communication, remains poorly understood as a neural process. Prior work identified brain regions that are responsive and selective to speech, however, the underlying neural mechanisms that enable humans to recognize speech remain unclear. One bottleneck is the limitations of studying speech responses in animals, as well as poor temporal resolution of human brain imaging that is critical for speech. Recently, artificial neural networks (ANNs) have emerged as powerful models of human perception due to their ability to map complex relationships from sensory signals to behavior, often achieving human-level performance. These models learn representations that exhibit correspondence with patterns of brain responses, motivating the utilization of ANNs to generate hypotheses for sensory system function and organization. While previous work has demonstrated neural correspondence in general auditory ANNs, speech-specific ANNs have not been thoroughly explored. In this work, we investigate different types of speech recognition ANNs comprising a variety of architectures, training datasets, and tasks, for their correspondence to human brain responses. We examine stage-wise internal representations of each model, revealing speech-selective units across layers analogous to speech-selective regions in the human brain. We further compare model and brain responses to known speech phenomena, including temporal sensitivity, noise invariance, and tuning to non-spectrotemporal features. We found that models optimized on sub-word units such as phonemes reproduce properties of speech-selective regions in the brain. This work provides a foundation for leveraging ANNs to study how speech is encoded in the brain and minds.

Automatic identification of relevant scholarly work for initial screening in systematic reviews

Minji Kang (Virginia Tech, Department of Mathematics), Bipasha Banerjee (Virginia Tech, University Libraries)

Systematic reviews are rigorous, transparent, reproducible, and replicable practices to find all relevant literature pertaining to a research question. The screening of literature currently is predominantly a manual process that involves discovering pertinent evidence from available literature and critically appraising it. Manual screening of thousands of published literature makes the process long and hard to keep pace with the rapidly growing scientific literature. This study investigates whether large language models (LLMs) can assist in the initial screening stage of systematic reviews while maintaining high recall and preserving the core principles of systematic review. We develop an LLM-integrated pipeline to automatically classify literature (Relevant, Non-relevant, or Unsure) and record a justification for each classification. We evaluated its performance against ground-truth labels from completed systematic reviews and calculated the Precision, Recall, and F-1 scores. The results show that the model achieved very high recall across both case studies. Literature labelled as Relevant by human reviewers was successfully identified by the model during screening. However, we noticed that the precision was relatively low. Many non-relevant studies were classified as Relevant or Unsure. The results indicate that LLM-based screening can be useful as a first filtering step in systematic reviews.

The model can reduce the number of papers that require manual review by identifying the most relevant studies early. Including an Unsure category also helps make the screening process clearer and reduces the risks of excluding relevant studies too early. Future work will focus on improving results: precision while still maintaining high recall.

From Deficits to Strengths: A Scoping Review on ADHD Adults & Productivity Technology

Suzanna Yaukey (University of Maryland, Baltimore County), Anita Komlodi (University of Maryland, Baltimore County), Ravi Kuber (University of Maryland, Baltimore County), Andrea Kleinsmith (University of Maryland, Baltimore County)

In the past few years, more HCI venues have begun to feature research on neurodivergent users and their experiences. Much of the research focuses on neurodivergent children with a large body of work on autistic and/or ADHD (attention-deficit/hyperactivity disorder) children and interventions. Studies considering the user needs of ADHD individuals are represented in HCI literature, but infrequently with adults as the study population. This scoping review provides an overview of recent literature on technology, tools, and assistive interventions used or proposed for use by adults with ADHD, or neurodivergence generally where ADHD is included as a subpopulation. Three core themes emerge from this review. The first theme of tactics and challenges to achieve productivity includes ways to increase concentration; train on social skills; technologies for managing time; and body doubling (working alongside someone whether in-person or virtually). A second theme focuses on technologies that create deficits for those who are neurodiverse, including technologies developed to "fix" their issues so that they can behave more neurotypically; and deficits created by broadly available technology that is not designed universally. The last theme is the importance of finding strength-based approaches when designing technology for neurodivergent users. The studies in this section prioritize participatory design, feedback from neurodiverse users, and the importance of designing for all.

When Jokes Cross the Line: A Dataset for Distinguishing Humor, Dark Humor, and Harm in Social Media Short-Form Videos

Sydney Johns (Virginia Polytechnic Institute and State University)

Short-form videos on platforms such as TikTok, Instagram Reels, and YouTube Shorts are among the most consumed media worldwide. However, their rapid and multimodal nature makes it difficult to automatically detect harmful content. Social media content is commonly flagged as harmful because of a trigger word it contains but in the context of the entire video there is no harm. Many times, the differentiation between harmful content and humorous content is unclear and nuanced. This paper introduces a new dataset and framework for understanding harmful short-form videos. Our contribution is twofold: (1) a 1,500-video multimodal dataset sampled from Instagram, and YouTube Shorts, and (2) a definitional framework and annotation protocol that distinguish between harmful, dark humor, and humorous content. Each video is fully hand-annotated by trained participants. The objective of this study is to construct a dataset that will advance future multimodal systems in discerning humor, dark humor, and sarcasm within short-form video content.

Urban Giants Under Threat: Unveiling Climate Vulnerabilities and Adaptive Strategies in Megacities

Dhanyasri Bolla (George Washington University)

As rapid urbanization converges with climate change, megacities face unprecedented risks that require robust adaptation strategies. This study presents a quantitative framework and an AI-powered analytics dashboard designed to assess climate resilience across 43 global megacities. We employed a mixed-methods approach, combining statistical analysis---including Chi-square tests, Cramer's V, and regression modeling---with a novel Large Language Model (LLM) pipeline to parse and analyze complex policy documents. The system utilizes GPT-4 to extract structured insights on infrastructure risks, financing mechanisms, and stakeholder engagement from unstructured climate action plans. These insights are visualized in a full-stack web application built with Flask and Leaflet.js, allowing for dynamic multi-city comparison. Our analysis reveals critical gaps in current adaptation planning and identifies correlations between specific urban attributes and resilience outcomes. This work demonstrates how NLP and interactive data visualization can democratize access to policy insights, supporting data-driven decision-making for urban sustainability.

StagePilot: A Deep Reinforcement Learning Agent for Stage-Controlled Cybergrooming Simulation

Heajun An (Virginia Tech), Qi Zhang (Virginia Tech), Minqian Liu (Virginia Tech), Xinyi Zhang (Virginia Tech), Sang Won Lee (Virginia Tech), Lifu Huang (University of California, Davis), Pamela Wisniewski (International Computer Science Institute), Jin-Hee Cho (Virginia Tech)

Cybergrooming is an evolving threat to youth, necessitating proactive educational interventions. We propose StagePilot, an offline RL-based dialogue agent that simulates the stage-wise progression of grooming behaviors for prevention training. StagePilot selects conversational stages using a composite reward that balances user sentiment and goal proximity, with transitions constrained to adjacent stages for realism and interpretability. We evaluate StagePilot through LLM-based simulations, measuring stage completion, dialogue efficiency, and emotional engagement. Results show that StagePilot generates realistic and coherent conversations aligned with grooming dynamics. Among tested methods, the IQL+AWAC agent achieves the best balance between strategic planning and emotional coherence, reaching the final stage up to 43% more frequently than baselines while maintaining over 70% sentiment alignment.

Persona-Sensitive Design Evaluation with Humans and Large Language Models

Khyati Goyal (Virginia Tech), Zannah Zeiw (Virginia Tech), Svastik Arora (Virginia Tech), Sehrish Basir Nizamani (Virginia Tech)

Design evaluation in HCI depends on whether evaluators adapt their judgments to different persona contexts. This research investigates whether design judgments change based on persona-specific priorities and whether LLMs demonstrate context sensitivity comparable to humans. Human participants and LLMs evaluated three dashboard design variants while adopting distinct personas (Productivity App User, Patient, and Educator), each associated with different evaluative criteria. We analyze metric prioritization, design rankings, and qualitative reasoning within each persona context. Results show clear persona-dependent patterns: evaluators emphasized different metrics aligned with persona goals, changed their preferred

designs across personas, and provided persona-appropriate qualitative explanations and feature suggestions. Importantly, LLMs exhibited context-sensitive behavior comparable to human evaluators across all analyses. These findings demonstrate that design evaluation is inherently context-dependent and that LLMs can function as persona-sensitive evaluators. This work supports the use of LLMs as a complementary tool for early-stage, multi-persona design assessment rather than as a replacement for human evaluation.

CREA: A Collaborative Multi-Agent Framework for Creative Image Editing and Generation

Kavana Venkatesh (Virginia Tech), Connor Dunlop (Virginia Tech), Pinar Yanardag (Virginia Tech)

Creativity in AI imagery remains a fundamental challenge, requiring not only the generation of visually compelling content but also the capacity to add novel, expressive, and artistically rich transformations to images. Unlike conventional editing tasks that rely on direct prompt-based modifications, creative image editing requires an autonomous, iterative approach that balances originality, coherence, and artistic intent. To address this, we introduce CREA, a novel multi-agent collaborative framework that mimics the human creative process. Our framework leverages a team of specialized AI agents who dynamically collaborate to conceptualize, generate, critique, and enhance images. Through extensive qualitative and quantitative evaluations, we demonstrate that CREA significantly outperforms state-of-the-art methods in diversity, semantic alignment, and creative transformation. To the best of our knowledge, this is the first work to introduce the task of creative editing.

Reason Impossible: Can LLMs Forecast Wartime Returns?

Farzana Yasmin Ahmad (University of Virginia), Kazi Noshin (University of Virginia), Zakaria Mehrab (University of Virginia)

Background: The 2022 Russian invasion of Ukraine displaced millions of people, yet return migration remains underexplored in current research. This work provides one of the first systematic evaluations of Large Language Models (LLMs) for forecasting wartime return migration. We ask: How well can LLMs forecast crisis-induced temporal return migration patterns using real-world data, how do they reason about their forecasts, and how robust are these forecasting and reasoning processes? Methodology: We evaluate LLM-based forecasts using a PromptCast strategy that provides varying levels of migration, conflict, and historical context. Three models, ChatGPT-5, Gemini 2.5 Flash, and Claude Sonnet 4.5, are assessed using real-world data from the Humanitarian Data Exchange (HDX), ACLED, and UNHCR. Performance is measured using Normalized Root Mean Square Error (NRMSE), Pearson Correlation Coefficient (PCC), and Normalized Continuous Ranked Probability Score (NACRPS). Results: ChatGPT-5 and Gemini-2.5 demonstrate competitive agreement with ground-truth return data, though performance varies across evaluated cases. In quantitative evaluation, Gemini-2.5 achieves strong accuracy and uncertainty quality, while ChatGPT-5 more effectively captures overall trends. Qualitative analysis reveals consistent differences in reasoning behavior. ChatGPT-5 incorporates all provided contextual features and favors linear autoregressive models, whereas Claude Sonnet 4.5 emphasizes non-linear, multi-phase approaches and often omits explicit migration context. ChatGPT-5 also most clearly acknowledges modeling limitations. Conclusion: LLMs show promise for trend-level analysis of crisis-induced return migration but remain unreliable for independent use in high-stakes

humanitarian settings. These findings highlight the importance of human-in-the-loop workflows, where LLM-based forecasts support analysis and situational awareness rather than replace human judgment.

Multimodal Spatio-Temporal Representation Learning for Edge-Based Urban Traffic Control

Seyma Simsek (Virginia Tech), Betül Nur Güner (Odak R&D Center, Türkiye), Chang-Tien Lu (Virginia Tech)

This work extends a previously published and field-validated edge-based traffic signal control system by reframing it as a multimodal representation of intersection dynamics. The existing system simultaneously produces visual observations (vehicle detections), temporal information (tracked trajectories), spatial semantics (lane-level ROI representations), and control states (signal phase information). In this study, these heterogeneous components are explicitly formalized as a synchronized multimodal state representation that is suitable for learning-oriented traffic analysis.

Based on the existing deployment results, the generated multimodal representations exhibit temporal consistency and structural coherence, indicating their suitability for downstream learning tasks such as short-term traffic state prediction, learning-assisted signal control, and digital twin calibration. Importantly, this extension does not require additional sensors or hardware modifications; all multimodal information is derived from the existing edge infrastructure.

Future work will focus on evaluating lightweight temporal prediction models and learning-informed control strategies using the proposed multimodal state representation. However, the primary contribution of this study is to demonstrate, from a methodological perspective, that the deployed edge-based system already provides a learning-ready multimodal foundation for scalable urban traffic modeling and optimization.

Understanding Developer Acceptance of AI-Assisted Code Commenting: An Undergraduate Study Using the AI Code Comment Acceptance Model (AICCAM)

Yoonje Lee (Virginia Tech), Anurag Pokala (Virginia Tech), Jasser Darguech (Virginia Tech), Parth Mehta (Virginia Tech), Sehrish Basir Nizamani (Virginia Tech)

This study explores how developers perceive and accept the use of Artificial Intelligence (AI) tools for generating code comments. As AI becomes increasingly integrated into software development, understanding its impact on coding practices and user attitudes is essential. The proposed AI Code Comment Acceptance Model (AICCAM) is an adaptation of the Technology Acceptance Model (TAM). It includes four main constructs: Perceived Accuracy & Trust (PAT), Ease of Use & Perceived Learning Benefit (EoU–PLB), Perceived Usefulness (PU), and Attitude & Intention to Use (AT/IU). Data were collected through a Likert-scale survey and analyzed to measure reliability, relationships among constructs, and overall acceptance patterns. Results show that ease of use and learning benefits strongly influence perceived usefulness, and both usefulness and trust significantly shape users' positive attitudes toward AI-assisted commenting tools. The findings highlight that while developers trust and find these tools helpful, their intention to continue using them depends mainly on how useful and reliable they perceive the AI-generated comments to be.

Ethics and AI in Education: Understanding Public Discourse on Twitter (X)

Akriti Bagale (George Mason University), Nafisa Mehjabin (George Mason University), Ali Ünlü (George Mason University), Aditya Johri (George Mason University)

The use of artificial intelligence (AI) and generative AI (GenAI) in education is rapidly expanding, creating opportunities but also raising ethical concerns. By capturing large-scale, real-time discussions and reactions to key events, social media platforms provide a valuable lens for examining public debates around AI ethics in education. In this paper, we analyze five years (2019-2024) of discourse on Twitter (now X) to capture the evolving conversations around the ethics of AI-based educational technologies, with particular attention to two pivotal moments: the COVID-19 pandemic, during which online and AI-based education increased significantly, and the release of ChatGPT. Using BERT-based topic modeling and sentiment analysis with TweetNLP, we identify dominant themes and shifts in public perspectives. Although the discourse engages with ethics, risk governance, and responsible deployment, the dominant framing is constructive and forward-looking emphasizing human-in-the-loop adoption, applied ML/AI in practice, and tool-centric workflows. By providing a longitudinal understanding of the ethics discourse on AI, this study contributes a broader public perspective to ongoing debates about the responsible and equitable integration of AI in education.

Temperature and Interface Effects on Carrier Transport in Hydrogen-Terminated Diamond MOSFETs

Nuwayyir Alshammari (George Mason University)

Hydrogen-terminated diamond metal–oxide semiconductor field-effect transistors (HD-MOSFETs) are promising candidates for high-power and high-temperature electronic applications due to diamond's ultra-wide bandgap and excellent thermal properties. In these devices, hydrogen surface termination enables the formation of a two-dimensional hole gas (2DHG), which governs channel transport and overall device performance. Understanding how temperature and interface effects influence carrier mobility is essential for reliable device design and modeling.

In this work, we use TCAD Sentaurus simulations to investigate the temperature-dependent transport behavior of HD-MOSFETs over the range of 300–600 K. The study focuses on the evolution of channel mobility, threshold voltage, and on-state resistance under varying thermal conditions. In addition, the role of interface traps is examined to assess how trapping effects modify the apparent mobility and electrostatic control of the channel.

The results reveal clear temperature-dependent trends in transport characteristics, highlighting the combined influence of thermal effects and interface traps on device operation. These findings provide modeling insight into the performance limitations of hydrogen-terminated diamond MOSFETs at elevated temperatures and demonstrate the importance of accounting for interface effects in predictive TCAD simulations. The presented work contributes to improved understanding of transport mechanisms in diamond-based devices and supports their optimization for high-temperature and high-power applications.

Framework for Finding Attribution of Social Media Screenshots

Tarannum Zaki (Graduate Student, Old Dominion University), Michael Nelson (Faculty, Old Dominion University), Michele Weigle (Faculty, Old Dominion University)

Screenshots of social media posts are widely used for sharing information. Sharing screenshots serve many legitimate purposes, such as keeping evidence of deleted posts, enabling cross-platform sharing, and supporting humor/satire. However, this also creates opportunities for misleading users through the spread of fabricated content. In practice, users rarely examine the attribution of a social media post before sharing a screenshot. We are exploring methods to determine attribution of a social media post shown in a screenshot using resources found on the live web and in web archives. Our ultimate goal is to establish an automated framework that helps users determine whether the screenshot content was really posted by the alleged author. To achieve this, we first collect a data set of screenshots of tweets shared on Twitter/X. We then introduce and evaluate methods for extracting key information such as the Twitter handle, timestamp, and tweet text from screenshot images. While attributing non-deleted posts can often be accomplished using the live web, finding attribution of deleted posts presents greater challenges. Web archives therefore play a crucial role in finding attribution of deleted posts. We introduce an automated approach that uses the Internet Archive's Wayback Machine CDX API to determine screenshot's attribution of deleted posts. We are currently investigating predictive factors that can help estimate the likelihood of a screenshot's attribution when evidence cannot be found on either the live web or in web archives. Overall, our work helps promote accurate information sharing on social media platforms.

MARCO: Multi-Agent Code Optimization for High-Performance Computing

Ariana Nafisi (Virginia Tech), Veljko Cvetkovic (Virginia Tech), Claire Shin (Virginia Tech), Kathleen Reece (Virginia Tech), Krishnateja Reddy (Virginia Tech), Bryan Torres (Virginia Tech), Sina Heidari (Virginia Tech), Margaret Ellis (Virginia Tech), Dimitrios Nikolopoulos (Virginia Tech)

Large language models (LLMs) have transformed software development through code generation capabilities, yet their effectiveness for high-performance computing (HPC) remains limited. The complexity of optimizing high-performance computing (HPC) kernels, particularly the manual effort required to integrate and configure vendor-specific libraries like CUTLASS, poses a significant barrier to maximizing modern GPU performance. We introduce MARCO (Multi-Agent Reactive Code Optimizer), an agentic AI framework designed for automated, reliable kernel generation and optimization. MARCO leverages in-context learning (ICL) to incorporate provided context and exemplars from structured prompts without altering parameters, eliminating the need for resource-intensive fine-tuning and making the system scalable and accessible. Our agentic system implements a two-phase optimization loop, code generation followed by validation, and leverages structured guide resources to overcome the limitations of general-purpose LLM knowledge. The framework supports both the CUTLASS 2.x and 3.x APIs, enabling automatic detection of linear algebra operations, selecting correct device-specific configurations, and performing complex epilogue fusions. In a head-to-head evaluation, MARCO achieves substantial performance gains, generating kernels that run up to faster than naive CUDA baselines. MARCO represents a new paradigm for high-performance code generation, proving that LLMs, when equipped with specialized domain guidance, can reliably automate the development of complex GPU kernels with both high efficiency and optimal performance.

From Vulnerable to Resilient: Examining Parent and Teen Perceptions on How to Respond to Unwanted Cybergrooming Advances

Xinyi Zhang (Virginia Tech), Mamtaj Akter (New York Institute of Technology), Heajun An (Virginia Tech), Qi Zhang (Virginia Tech), Lifu Huang (University of California, Davis), Jin-Hee Cho (Virginia Tech), Pamela Wisniewski (International Computer Science Institute), Sang Won Lee (Virginia Tech)

Cybergrooming is a form of online abuse that threatens teens' mental health and physical safety. Yet, most prior work has focused on detecting perpetrators' behaviors, leaving a limited understanding of how teens might respond to such unwanted advances. To address this gap, we conducted an online survey with 74 participants---51 parents and 23 teens---who responded to simulated cybergrooming scenarios in two ways: responses that they think would make teens more vulnerable or resilient to unwanted sexual advances. Through a mixed-methods analysis, we identified four types of vulnerable responses (encouraging escalation, accepting an advance, displaying vulnerability, and negating risk concern) and four types of protective strategies (setting boundaries, directly declining, signaling risk awareness, and leveraging avoidance techniques). As the cybergrooming risk escalated, both vulnerable responses and protective strategies showed a corresponding progression. This study contributes a teen-centered understanding of cybergrooming, a labeled dataset, and a stage-based taxonomy of perceived protective strategies, while offering implications for educational programs and sociotechnical interventions.

Beyond Participation: Rethinking Classroom Discussion with ThoughtSwap

Chandani Shrestha (James Madison University), Mona Rizvi (James Madison University), Alexander Heckel (James Madison University), Michael Stewart (James Madison University)

Both students and instructors face barriers to successful classroom discourse. Students may have apprehensions when uncertain how others in the class will respond to their viewpoint, while instructors may experience similar uncertainties along with challenges in guiding discussions, particularly when strong emotions arise. ThoughtSwap (TS) is a classroom discussion facilitation tool designed to foster Conscientious Discourse. Rather than positioning itself as a solution, it functions as a supportive layer that reshapes how discussion occurs. Instructors share a prompt and students submit responses with authorial privacy, creating a brave space for sincere expression while maintaining moderation. Responses can be redistributed so students engage with a peer's idea, providing an accessible entry point and enabling instructors to design a range of discussion activities.

TS can help reduce some barriers to classroom discourse while also drawing attention to underlying emotional and pedagogical challenges that may otherwise remain unseen. One of the instructors who used TS in her class observed that students often have varied and unexpected reactions to disagreement, and balancing sincere expression with guidance that keeps focus on ideas rather than emotions is challenging. Challenges may vary, from the initiation of discussions to their progression and potential effects on students. This CAPWIC session offers an opportunity to bring together instructors and students with diverse experiences to discuss these concerns, better understand multiple perspectives, share strategies, and inform future iterations of the tool and associated discussion activities.

Flash Talks

Total abstracts in this session: 12

Methods and Challenges in LLM-Based Vulnerability Detection and Repair

Fatemeh Vares (George Mason University), Brittany Johnson (George Mason University)

Large Language Models (LLMs) are increasingly used for software vulnerability detection and automated repair, evolving from standalone prompting to tool-augmented and agentic workflows that integrate compilers, unit tests, static analyzers, fuzzers, and retrieval from codebases. This poster offers a structured synthesis of recent methods and distills open challenges that motivate new research directions.

We organize the literature along three dimensions: (1) task scope (vulnerability detection, localization, patch generation, and end-to-end pipelines), (2) system architecture (single-LLM approaches, retrieval-augmented generation, and multi-step or multi-agent systems with external tools), and (3) adaptation strategy (prompt engineering, fine-tuning of code models, and feedback-driven iterative refinement). Using this taxonomy, we compare common design patterns and discuss where they succeed and where they break down.

Across studies, recurring limitations include incomplete project context (dependencies, build configurations, and environment assumptions), non-compiling or non-functional patches, semantic regressions, fixes that mask symptoms, and repairs that inadvertently introduce new vulnerabilities. We highlight evaluation gaps such as limited reproducibility of build contexts and overreliance on benchmark pass rates rather than security-relevant outcomes. The poster concludes with a concise research agenda and recommendations for building more reliable and security-grounded evaluation and validation pipelines.

AI Assisted Cross-border Cybersecurity Literacy

Norah Ondus, Sehrish Basir Nizamani (Virginia Tech)

Migrants and international students often face challenges in maintaining online security due to unfamiliar digital environments, language barriers, and cultural differences in technology use. Despite the availability of cybersecurity education tools, most are not designed to address these specific challenges.

This study investigates the use of an AI-assisted cybersecurity literacy tool to support migrants and international students in developing safer online practices and greater confidence in navigating digital systems. Using a mixed-methods approach, participants will complete pre- and post-surveys to measure changes in cybersecurity knowledge and confidence and maintain short daily diary entries to document their experiences with the tool.

The study also examines how cultural background, language proficiency, and trust in digital technologies influence participants' engagement with the tool and their learning outcomes. Findings from this research are expected to inform the design of culturally responsive, accessible, and effective cybersecurity education tools that enhance the ability of diverse users to protect themselves online.

LLM-Guided Input Generation for Causal Fairness Testing

Sadia Afrin Mim (George Mason University)

Software fairness testing is essential for identifying discriminatory behaviors in automated decision-making systems. Traditional causal testing frameworks, such as Themis, rely on generating matched test cases that differ only in sensitive attributes (e.g., race, gender). While subsequent approaches like Aequitas and KOSEI improved efficiency through two-phase search strategies---combining global sampling with local, deterministic perturbations---they remain limited by a "semantic gap." These methods typically treat input attributes as independent, often leading to the generation of implausible or redundant test cases that ignore real-world domain constraints and complex feature dependencies.

This paper proposes a novel framework that integrates Large Language Models (LLMs) with causal fairness testing to transition from blind, syntactic exploration to semantic-guided input generation. By leveraging the inherent world knowledge and reasoning capabilities of LLMs, our approach automates the identification context-aware input ranges, and enforces domain-specific constraints. This integration allows for the prioritization of meaningful regions in the search space, reducing the computational overhead of exhaustive search while improving the interpretability of detected biases. Our preliminary results suggest that LLM-assisted fairness testing not only enhances the plausibility of generated test cases but also significantly lowers the manual burden on developers, making rigorous discrimination testing more accessible for large-scale, ML-based systems.

Pebble

James Mercer (Student at Virginia Tech), Rhea Parekh (Student at Virginia Tech), Nikhita Kuninti (Student at Virginia Tech), Aleysa Sweeney (Student at Virginia Tech)

Pebble is a gamified learning platform that turns studying into a fun feedback loop instead of a chore. Not "notes + quizzes" (that's table stakes). It's a system that forces interaction, repetition, and progress tracking in a way normal note apps refuse to do.

What it actually does - Structured notes → not walls of text. Notes are chunked so they can become questions. - Auto-generated quizzes & multiplayer games → every concept gets turned into recall and quizzes, not rereading. - XP, levels, streaks, rewards → studying has consequences. Skip work, you stagnate.

Who it's for: - Students who wished they studied more but spend too much time on the phone. - ADHD / neurodivergent learners who need stimulation + structure. - Competitive students who respond to numbers, progress bars, and "one more round".

Gender, Culture, and Privacy: Social Media Privacy Concerns in the Saudi Context

Khoulood Alharthi (Virginia Tech), Sehrish Basir Nizamani (Virginia Tech)

Privacy on social media is shaped not only by platform features but also by users' cultural and social contexts. Although prior research has examined privacy concerns extensively, much of this work is grounded in Western settings, leaving non-Western cultural contexts underexplored. This research examines how gender and cultural norms shape privacy concerns among social media users in Saudi Arabia.

Grounded in established privacy frameworks from Human–Computer Interaction and information privacy research, the study focuses on key dimensions of social media privacy concern, including perceptions of data collection, information control, and appropriate information sharing. Particular attention is given to how culturally embedded values---such as modesty, reputation, and social boundaries---intersect with gendered experiences of online visibility.

By centering a Saudi context, this work aims to extend current privacy theories beyond Western assumptions and contribute culturally grounded insights to privacy research. The study seeks to inform the design of social media platforms and privacy controls that better align with users' lived expectations and cultural norms. More broadly, this research highlights the importance of considering cultural and gendered contexts when evaluating privacy concerns and developing socially responsive technologies.

Benchmarking and Advancing Generative Models for Calorimeter Shower Simulation

Farzana Yasmin Ahmad (University of Virginia), Geoffrey Fox (University of Virginia)

Calorimeter detectors play a central role in high-energy physics experiments by enabling particle energy reconstruction through detailed shower simulations. High-fidelity Monte Carlo tools such as Geant4 provide accurate simulations but are computationally expensive, with calorimeter showers accounting for a dominant fraction of total simulation cost. This challenge will intensify with the increased data demands of the High-Luminosity LHC, motivating the development of fast yet physically accurate simulation methods.

Recent generative machine learning models offer promising alternatives to traditional simulation, but evaluating their physical reliability remains nontrivial. Existing metrics largely emphasize high-level observables or distributional similarity and often fail to capture the complex correlation structure across calorimeter layers and spatial voxels.

In this work, we introduce Correlation Frobenius Distance (CFD), a physics-informed metric that quantifies discrepancies in layer-wise and voxel-wise energy correlations between generated samples and Geant4 reference simulations. Using CFD within a unified benchmarking framework, we conduct a systematic comparison of GAN-, VAE-, flow-, and diffusion-based generative models across multiple calorimeter datasets. Our analysis reveals distinct trade-offs between fidelity, sparsity modeling, scalability, and computational efficiency, while exposing failure modes that are not evident under conventional evaluation metrics.

Motivated by these findings, we propose Lantern, a two-stage autoregressive diffusion model that factorizes layer-level energy prediction from voxel-level shower generation. This design improves correlation fidelity while remaining parameter-efficient and flexible across detector configurations.

Together, this work demonstrates the importance of correlation-aware evaluation and model design for advancing trustworthy, scalable generative simulation in scientific machine learning.

Effects of Prompt Writing Training Using the AMERH Framework on the Students' Prompt Quality

Eman A. M. Amer (Virginia Tech), Jeannine Eddleton (Virginia Tech)

The rapid advancement of Generative AI (GenAI) presents significant challenges, and without a foundation in writing prompts and in evaluating AI models' responses, students may not recognize ethical concerns, such as security and privacy issues, which may contribute to

irresponsible use of technology in personal and professional settings. This study investigates the effects of training students in prompt-writing using the AMERH framework and perceptions across three classrooms. Employing a pre-test, intervention, and post-test design, data were collected from 400 students via Canvas.

Initial findings indicate that prompt-writing skills significantly improved following the intervention compared to baseline attempts. Interestingly, the study revealed that many students initially avoided GenAI tools due to specific ethical and environmental concerns. Furthermore, qualitative reflections revealed a shift in student mindset; participants noted that interacting with AI is a conscious process that requires critical thinking and rigorous evaluation of outputs. These results suggest that formal training is a necessary prerequisite for integrating AI tools into educational settings to ensure meaningful and responsible use.

Understanding State-Level AI Readiness Policy

Saanvi Shashikiran (Georgetown University), Henry Deng (Georgetown University), Lisa Singh (Georgetown University), Andrea Headley (Georgetown University), Aditi Bhardwaj (Georgetown University), Julie Dang (Georgetown University), Vinuri Dissanayake (Georgetown University)

Artificial Intelligence (AI) readiness varies significantly across U.S. states, reflecting differences in policy, infrastructure, and institutional support. We examine AI readiness through a policy lens, focusing on the incentives and protections shaping state-level AI adoption and governance. Our analysis centers on five pilot states---New York, California, Texas, Illinois, and Florida---and operationalizes AI readiness by developing an approach that systematically analyzes legislative and executive documents using text-based similarity search. Using Elasticsearch, we compare multiple text-based similarity approaches to determine the relevance of documents to policies that support AI readiness policy. We find that using normalized BM25-based keyword relevance scoring performs better than other strategies. These methods enable us to track changes in AI readiness over time and lay the groundwork for a future tool that will allow policymakers and businesses to compare AI readiness across states.

Data-Driven Exploration of Physiological Factors Perpetuating Bias in Pulse Oximetry Readings for At-Home Use

Arshnoor Bhutani (University of Maryland, College Park), Mahi Sanghavi (University of Maryland, College Park)

Pulse oximeters estimate blood oxygen saturation (SpO_2) non-invasively and are increasingly used in at-home settings. Prior research has shown systematic overestimation of oxygen levels in individuals with darker skin tones due to melanin-related differences in light absorption, raising the risk of undetected hypoxemia. While hospitalized patients may receive confirmatory arterial blood gas (ABG) testing, at-home users lack access to gold-standard measurements, disproportionately affecting underserved communities.

This work analyzes paired SpO_2-SaO_2 measurements from the BOLD dataset to characterize bias patterns in pulse oximetry. The analysis employs the Monk Skin Tone Scale, which offers improved representation of darker skin tones compared to the Fitzpatrick scale, originally developed for dermatological applications. Descriptive statistics and cluster-robust linear regression models (for repeated measures within encounters) were used to examine how bias varies across skin tones while controlling for oxygen saturation, perfusion index, carboxyhaemoglobin, and methoxyhemoglobin.

Preliminary findings show an increasing gradient bias from light to darker skin tones. Clinically significant errors (|bias| > 3 mmHg) rise from 17.4% in the lightest tones to 58.5% in the darkest tones, with an overall rate of 30.4% across the test set. Regression modeling confirms that skin tone remains associated with greater positive bias.

This ongoing project focuses on characterizing these patterns to enable skin tone-specific correction algorithms. By providing quantitative evidence of bias magnitude and identifying when errors are most pronounced, this research aims to support the development of equitable guidance for home pulse oximeter users, ultimately reducing health disparities in respiratory monitoring across diverse populations.

A Structured Dataset for Analyzing Human Group Motion Across Spatial Formations

Meklet Berihun (University of Richmond)

Understanding human group motion patterns is a key step toward enabling autonomous robots to navigate, coordinate, and learn behaviors in shared human environments. In this project, we collected motion data from groups of participants forming structured spatial patterns, including circular, diagonal, horizontal, vertical, and infinity-shaped trajectories. Using low-cost onboard sensors, participant motion trajectories were recorded and organized into a structured CSV dataset capturing spatiotemporal features suitable for analysis.

This dataset enables initial investigations into how autonomous systems can recognize and interpret collective human movement using lightweight sensing and computation. Through basic statistical feature extraction and dimensionality-reduction techniques, we demonstrate that several group formations are clearly separable in feature space, indicating that even simple models can distinguish human motion patterns reliably. These results suggest that robots can learn to identify, anticipate, and ultimately replicate human group behaviors without relying on expensive sensing infrastructure.

By providing a reusable dataset and an analytical framework, this work lays the groundwork for future machine learning models that allow robots to copy and adapt human motion patterns for collaborative tasks. The findings inform the design of perception and imitation-learning algorithms that improve autonomous navigation and coordination in human-rich environments.

A Trust-Aware, Biometrically-Secured Social Network Using Decentralized Identity Protocols and the Analytic Hierarchical Process for Collaboration

Christopher Parham (Virginia State University), Sai Sindhu Tirumalasetty (Virginia State University), Joseph Shelton (Virginia State University)

Human factors remain a critical challenge in cybersecurity, particularly in collaborative digital environments where identity management, trust, and usability intersect. As web technologies evolve, users must manage increasing numbers of credentials, exposing them to risks such as interception and brute-force attacks. WebID protocols address these challenges by enabling decentralized Web 3.0 authentication through user-controlled identity servers. However, compromise of a personal identity server may expose authentication credentials and trust relationships. This research proposes a secure, trust-aware social network architecture integrating WebID-based decentralized identity with biometric authentication using a novel dual-factor method called Permuted Disposable Feature Vectors (PDFV). PDFV generates permuted biometric feature representations designed for one-time authentication, significantly expanding biometric encoding variability and reducing replay attack vulnerability. The system

models a decentralized social network where users maintain personal WebID servers containing RDF-based FOAF profiles that form a distributed web of trust. Participants are evaluated using multidimensional suitability scoring that incorporates structural network metrics, embeddedness, task-performance attributes, and assortativity. These metrics are combined using the Analytic Hierarchy Process to rank job-role suitability and determine task feasibility through ego-centric network collaboration. Implementation includes a one-time facial recognition authentication module using Local Binary Patterns combined with evolutionary optimization to generate disposable feature extractors aligned with PDFV. Network simulations across Erdős–Rényi, Barabási–Albert, and Watts–Strogatz models demonstrate improved resistance to spoofing, replay attacks, and centralized compromise while enhancing usability and trust-based collaboration. This framework advances human-centered cybersecurity by combining decentralized identity, network science, and privacy-preserving biometrics for secure distributed collaboration.

Relationship Extraction using Retrieval Augmented Generation for Biomedical Dataset

Jannat Lnu (Virginia Commonwealth University), Bridget McInnes (Virginia Commonwealth University), Tomasz Arodz (Virginia Commonwealth University), Charlie Dil (Virginia Commonwealth University)

With the increasing number of structured and unstructured data, obtaining reliable information effectively has become crucial. In the biomedical domain, extracting information from the scientific papers is crucial in order to stay up-to-date with accurate information, given the increased pace by which new research studies are published. This work focuses on identifying relationships between entities that are extracted from the abstracts and titles of biomedical research papers.

In this work, we developed a Retrieval Augmented Generation (RAG) based system to automatically identify relations between biomedical entities. We evaluate multiple open source Large Language Models (LLMs) and the number of examples (shots) required to improve the LLM's results. We evaluate our methods using precision, recall and F-1 scores and compare our approach to traditional deep learning methods using DeBERTa with a Convolutional Neural Network (CNN). Our results indicate that Qwen models using the RAG approach with 10-shot examples achieved the highest macro F1 score compared to the baseline and other LLMs under the same setting. At 35 shots, Qwen reasoning and Qwen non-reasoning model performed best, exhibiting the fewest hallucinated labels and maintaining high macro F1 scores.

Posters

Total abstracts in this session: 43

DBWorkout: An Interactive SQL Learning Platform for Students and Professors

Deepika Devaraj et al. (Virginia Tech)

Learning SQL in classroom settings requires a safe environment where students can execute queries without affecting shared data or risking schema corruption. Many existing platforms provide limited isolation or lack the operational workflows instructors need to manage schemas, tasks, and student progress at scale. DBWorkout addresses these challenges by offering a gamified SQL learning platform that combines isolated execution, automated validation, and structured instructor tools.

DBWorkout provides each student with a dedicated PostgreSQL sandbox. When instructors create tasks, the system generates expected outputs and stores them as hashed JSON snapshots. Each student submission is executed inside a fresh PostgreSQL clone, ensuring clean and reproducible execution. The backend compares the student's output against the stored snapshot to determine correctness, enabling real-time feedback while preventing unintended schema modification. To assist with task creation, the platform can optionally leverage the ChatGPT API to generate example problems, refine prompts, or suggest schemas.

The platform includes tools for schema and task management, session organization, and basic progress tracking. DBWorkout is deployed on Virginia Tech's CS Launch Kubernetes cluster using containerized frontend, backend, and database services, enabling secure internal networking, persistence, and scalability for large classes. By combining safe SQL sandboxes, reproducible validation, and integrated instructor workflows, DBWorkout enhances SQL learning by providing students with a low-risk practice environment and instructors with a stable, scalable system for managing coursework. Future work includes analytics dashboards, collaborative group tasks, and richer schema visualization.

Identifying Human Factors in Red Teams for Cyber Exercises

Stephanie Travis, Denis Gracanin (Virginia Tech)

Cybersecurity practitioners are facing an increasingly complex world of attacks from increasingly technical adversaries. One of the primary means of training for these practitioners is through Cyber Defense Exercises (CDX) wherein participants are put into scenarios to defend networks against an attacking red team who emulates the tactics and techniques of real-world threat actors. After conducting a systematic literature review to identify existing human factors of cyber attackers documented in the literature, we undertook a multi-round DELPHI survey with cyber exercise planning and red team experts in order to refine a human factors taxonomy for CDX red team planning, scenarios, and execution with the goal to improve realism in red team behavioral execution. We developed a taxonomy that is a combination of human factors identified in the literature as well as by the expert participants with a view to combine this human factors taxonomy with existing CDX taxonomies in future work.

Experiences in Transition: Survey Development, Validation, and Analysis of Computer Science Transfer Student Pathways

Nawar Wali (Virginia Tech)

Transfer students are a vital yet understudied population in computer science (CS). Unlike fields such as engineering with ABET-standardized curricula, CS programs vary widely across institutions, creating barriers in transferability, credit recognition, and persistence. This project aims to address these gaps by systematically investigating the supports, challenges, and intentions to persist reported by CS transfer students before and after transfer.

Methods and Challenges in LLM-Based Vulnerability Detection and Repair

Fatemeh Vares, Brittany Johnson (George Mason University)

Large Language Models (LLMs) are increasingly used for software vulnerability detection and automated repair, evolving from standalone prompting to tool-augmented and agentic workflows that integrate compilers, unit tests, static analyzers, fuzzers, and retrieval from codebases. This poster offers a structured synthesis of recent methods and distills open challenges that motivate new research directions.

We organize the literature along three dimensions: (1) task scope (vulnerability detection, localization, patch generation, and end-to-end pipelines), (2) system architecture (single-LLM approaches, retrieval-augmented generation, and multi-step or multi-agent systems with external tools), and (3) adaptation strategy (prompt engineering, fine-tuning of code models, and feedback-driven iterative refinement). Using this taxonomy, we compare common design patterns and discuss where they succeed and where they break down.

Across studies, recurring limitations include incomplete project context (dependencies, build configurations, and environment assumptions), non-compiling or non-functional patches, semantic regressions, fixes that mask symptoms, and repairs that inadvertently introduce new vulnerabilities. We highlight evaluation gaps such as limited reproducibility of build contexts and overreliance on benchmark pass rates rather than security-relevant outcomes. The poster concludes with a concise research agenda and recommendations for building more reliable and security-grounded evaluation and validation pipelines.

Towards Practical Discrimination Testing for Software Systems

Sadia Afrin Mim, Brittany Johnson (George Mason University)

Software systems naturally make decisions, but software discrimination, where outcomes vary based on protected attributes like race or gender---poses a significant challenge for developers. Detecting these biases is difficult when they are buried in complex, implicit logic. While automated test generation tools aim to expose these failures, their actual effectiveness in helping practitioners reason about discrimination is not well understood.

In this work, we present a controlled user study evaluating how practitioners perceive and utilize discrimination testing support. We compared two automated tools: Themis, a fairness-specific testing tool, and Pynguin, a general-purpose test generator. Additionally, we investigated the potential for Large Language Models (LLMs) to augment this process by interpreting test outputs, explaining disparate behaviors, and reasoning about inclusivity.

Our findings reveal that Themis significantly outperforms general tools by using a causal pair comparison mechanism that signals bias through fairness metrics and sensitive attribute dependencies. These insights provide practitioners with actionable data to mitigate bias.

Furthermore, our results suggest that LLMs serve as a vital complementary aid, facilitating the generation of explainable, discrimination-aware test cases. This research establishes an empirical foundation of user centric study towards discrimination testing.

Comparative Analysis of Knowledge-Guided Few-Shot Brain Tumor Detection...

Okib Islam et al. (Morgan State University)

Abstract---Brain tumor detection from MRI scans remains challenging due to morphological heterogeneity and limited annotated data, with manual interpretation being time-consuming and prone to inter-observer variability. While vision-language models (VLMs) and single-shot detectors have shown promise individually, no systematic comparison exists evaluating their performance under knowledge-guided few-shot learning constraints. This study presents the first head-to-head comparison of two state-of-the-art VLMs (PaliGemma-2 and SmolVLM) against a single-shot detector (YOLOv12), each augmented with a structured medical knowledge graph (KG) encoding ICD-10 and SNOMED-CT ontologies. We construct a multimodal KG linking brain MR image embeddings with clinical reports and inject semantic priors through graph-aware attention modules for VLMs and feature wise linear modulation for YOLOv12. Experiments are conducted under strict 1 shot, 3 shot, and 5 shot scenarios using a publicly available brain MRI dataset with glioma and meningioma annotations. Our comparative analysis reveals distinct advantages: PaliGemma-2 with KG achieves the largest performance gain (55.1% mAP, 18.9% F1), demonstrating superior ability to leverage structured knowledge. SmolVLM + KG shows notable efficiency gains (31.3% mAP) with minimal parameter overhead, while YOLOv12 + KG excels in precise localization with sustained spatial accuracy (5.3% mAP) and real-time inference capability. These findings establish that medical knowledge graph augmentation is a model-agnostic enhancement strategy, with each architecture offering unique clinical deployment advantages.

Beyond the Percentage: An LTI Proxy to Support Mastery-Based Grading in CS

Ilse van Duijl (Virginia Tech)

Traditional points-based grading often acts as a barrier to student success in Computer Science, frequently rewarding grade-chasing over actual learning. As educators move toward Equitable Grading Practices (EGP), such as mastery-based assessment and flexible resubmissions, they generally hit a technical wall. Most Learning Management Systems (LMS) like Canvas are hard-coded for percentages and points. This project introduces the EGP Broker, a custom middleware solution, designed to bridge this gap. Built using the MERN stack and the LTI 1.3 protocol, the broker acts as a translator between external learning tools and the LMS. By intercepting and transforming data through the LTI 1.3 protocol, the broker allows instructors to use rich, categorical feedback while automatically mapping mastery to the numerical requirements of the university gradebook. Ultimately, the EGP Broker shifts the focus from points to proficiency, ensuring that the software we use to manage learning supports the success of a diverse student body.

Uncertainty-Aware Interfaces Between Neural Predictors and Large Language Models

Xinyue Zeng (Virginia Tech)

Integrating large language models (LLMs) into high-stakes machine learning pipelines remains nontrivial due to limited numerical grounding and brittle failure modes, especially when LLMs are coupled with neural predictors trained on structured data. This work proposes uncertainty-aware latent-variable models as principled interfaces between neural predictors and LLMs, enabling faithful post-hoc reasoning while preserving predictive reliability.

We instantiate this paradigm in DiagnoLLM, a hybrid framework that unifies Bayesian deconvolution, regulatory feature modeling, and structured language-based explanation. At its core, GP-unmix is a Gaussian Process-based hierarchical model that infers cell-type-specific gene expression from bulk RNA-seq and produces calibrated uncertainty estimates. These uncertainty-aware representations, augmented with eQTL-derived regulatory priors, are used to train a neural classifier for Alzheimer's disease detection, achieving up to 88.0% accuracy under low-data regimes.

Crucially, the LLM is not used as an end-to-end predictor, but as a post-hoc reasoning module that translates model outputs, uncertainty signals, and attributions into audience-specific explanations. Through divergence analysis, interface ablations, and human evaluation, we demonstrate that neural models and LLMs exhibit complementary and structurally predictable failure patterns, and that uncertainty-aware interfaces substantially improve explanation faithfulness, robustness, and usability.

Virginia Transit Deserts: A Geo-Demographic Study of Public Transportation Access

Jayla Hall (Undergraduate)

Public transportation access is crucial to economic growth and social equity. However, as the distance from metropolitan regions increases, public transportation often thins out, creating transit deserts. Because these areas frequently lack proper infrastructure, such as sidewalks or frequent routes, they can intensify historical inequities that disproportionately affect low-income, marginalized communities. This project uses spatial visualization to reveal geographic patterns and make disparities visible. By combining U.S. Census demographic data with route and stop data from the Virginia Department of Rail and Public Transportation, this analysis examines how transit access shifts across Virginia counties and census tracts. Initial data exploration was conducted using Microsoft Excel, followed by the use of Python libraries such as Folium and Geopandas to develop an interactive visualization that illustrates regional patterns of transit access related to racial demographics and median household income.

Information Quality Assessment Framework to Design Human-Centered AI Systems for Assistive Technology

Anuridhi Gupta (GMU), Hemant Purohit (George Mason University)

Assistive Technology (AT) is an umbrella term for products and services used to increase, maintain, or improve the functional capabilities of people with disabilities. With the advent of Artificial Intelligence (AI), there has been a noticeable shift in the development of AT. Given the diverse applications of AI and the unique needs of people with disabilities, AI enabled AT provides a practical approach of facilitating informed decision-making for all stakeholders while supporting freedom and control for people with disabilities. However, while AI may offer these

opportunities for people with disabilities, the rapid evolution of technology presents challenges in implementing AI-enabled AT, including issues of data quality, data value, and reliability. Moreover, the evaluation processes and metrics for AI-enabled AT remain underexplored, and serious concerns such as bias and explainability are often overlooked. Currently, there are no transparent, quantifiable, and robust metrics to systematically measure the functional outcomes of AI-enabled AT. To overcome these challenges, our work presents a set of robust information quality metrics that one can check for, as the first step of a data science pipeline when creating AI-enabled AT. These metrics comprise the proposed information quality framework, and they not only help quantify the different principles of AI enabled AT but also are comprehensive enough for domain specific contexts such as assistive technology. Further, we also provide a set of remediation techniques that can be employed on datasets to prioritize the creation of assistive technology. These contributions provide a foundation for developing reliable, and user-centered assistive technologies.

SETRL: Verifier-Guided Reinforcement Learning Under Query Budgets for Cross-Mode Generalization in SET

Michelle Lin (Thomas Jefferson High School for Science and Technology)

Verification is often the limiting resource in learning systems: a model can propose many candidate actions, but only a few can be checked for correctness, especially when the test distribution differs from training. This project developed SETRL, a controlled benchmark for verifier guided reinforcement learning using the game SET, where correctness is exact but access to checking can be deliberately restricted. Each card was encoded as a four dimensional vector over the finite field with three elements, so that validity of a proposed three card triple is determined by a deterministic algebraic rule. I generated boards under two distinct generation modes to induce a controlled mode shift between training and testing. Using this environment, I trained and evaluated three policies under identical conditions: RU as a baseline, BC as behavior cloning from demonstrations, and PL or RL as verifier guided reinforcement learning from rollouts with limited verifier access. Policies were evaluated by sweeping the maximum number of verifier calls allowed per episode and measuring both in mode and cross mode accuracy. With a fixed 20 call allowance, baseline cross mode accuracy ranged from 52 to 73 percent, while PL or RL reached 75 to 81 percent. For a 90 percent target accuracy, learned policies typically required about 10 verifier calls compared with about 14 for baselines. These results indicate that verifier guided reinforcement learning improves query efficiency and robustness under distribution shift, and that SETRL provides a clean testbed for studying selective verification when checking is scarce.

Roots for STEM: Promoting STEM Accessibility

Isabella Divietro (NHREC Governor's School for Science and Technology), Giada Grant (NHREC Governor's School for Science and Technology), Evan Sturgill (NHREC Governor's School for Science and Technology)

Roots for STEM is a student-led, minority-oriented organization located at the Governor's School for Science and Technology (GSST). Roots for STEM aims to empower underprivileged youth across seven school divisions in the Hampton Roads community to pursue their interests in various STEM fields. Roots for STEM's student leadership organizes and runs free, monthly STEM camps for students in grades 1-8 to actively participate in opportunities in a hands-on environment surrounded by like-minded peers on GSST's campus. In addition to gaining awareness of STEM-related opportunities, children explore different aspects of STEM; for

instance, camps include computer programming in Python, building/programming robots, constructing bridges, or engaging in other engineering and science focused events. To accommodate the hundreds of students participating in these camps, juniors and seniors from GSST sign up to volunteer after school. Whether answering questions, helping piece together robots, or walking students through challenging coding prompts, volunteers help students while promoting a mindset powered by excitement and curiosity. In addition to the benefits that Roots for STEM provides to students, volunteers gain opportunities to practice motivating others, collaborating, supporting students, and effectively sharing ideas. In STEM, publishing and working with other researchers are critical to success, making the volunteer-student relationship mutually beneficial. Roots for STEM prioritizes the promotion of educational opportunities in a fun, interactive format, creating an invested and interconnected community that will continue inspiring new generations of computer scientists, engineers, chemists, and mentors.

Octree-Based Spatial Partitioning for Multi-Agent UAV Navigation

Emma Faith (Christopher Newport University Student)

This work focuses on developing an Octree-based spatial partitioning framework for efficient multi-agent UAV navigation using the Crazyflie 2.1+ nano aerial platform. Drone systems that rely on centralized architectures for 3D navigation often suffer from high computational overhead, making them less ideal in nano aerial vehicle (NAV) swarm environments with limited onboard resources. This is handled by defining three-dimensional "exclusion zones" where the Crazyflie is or is not allowed to operate, serving as a form of lightweight air traffic control. By reducing computational overhead while maintaining safe flight planning, this research aims to contribute to the field of robotics a lightweight spatial partitioning method for NAV swarms to operate in complex 3D environments.

The framework uses a preloaded static map of the environment and an Octree data structure to partition 3D space, recursively dividing into 8 equal-sided cubic regions. Navigating the Octree map is done using the Crazyflie's Lighthouse positioning system, a vision-based tracking system that uses infrared laser sweeps to calculate its position in world space. Navigation decisions are generated using the A* search algorithm, which finds an optimal path by balancing path cost and a heuristic estimate, all while avoiding restricted areas. The system is implemented within the Robot Operating System (ROS 2) ecosystem and integrates Crazyflie drones, RViz, Python, C++, and the Flexible Behavior Engine (FlexBE).

Expert Systems as Accessible Pedagogical Infrastructure: Developing a Cross-Disciplinary General Education AI Literacy Course

Mackenzie Swain (University of Mary Washington), Eric McGowan (University of Mary Washington), Karen Anewalt (University of Mary Washington), Jennifer Polack (University of Mary Washington)

The growing influence of artificial intelligence across all academic disciplines underscores the importance of AI literacy across liberal arts and STEM education. However, most technical AI courses exclude non-technical students by assuming prior computer science knowledge, creating a substantial gap in preparing all students to be informed, critical citizens in an AI-shaped world. The development of educational materials and hands-on activities that make fundamental AI concepts accessible to all students regardless of computer science background is essential in ensuring all students have a foundation in AI literacy.

This approach, inspired by an assignment from Professor Yang Wang of La Salle University, is focused on researching interactive tools, assessing current AI-related activities, and forming educational materials based off of the findings of the research conducted. The combined experience and diverse perspectives of two computer science faculty and two undergraduate student researchers will culminate in a framework that encompasses technical and ethical considerations in AI pedagogy.

This project concentrates on educational materials and hands-on assignments for the topic of expert systems to create a template for technical foundations across key AI topics. These materials will be evaluated for accessibility, conceptual effectiveness, and engagement levels to ensure the development of an educational model to serve as the basis of a general education course that fosters AI literacy in students from all backgrounds and academic disciplines.

Hope or Hype? Understanding Vibe Coding through Software Practitioner Discussions

Fairuz Nawer Meem (George Mason University), Fatema Tuz Zohra (George Mason University), Justin Smith (Lafayette College), Brittany Johnson (George Mason University)

Recent advances in AI-assisted programming have sparked new forms of creative practice that blur the boundaries between development, authorship, and emotion. Among these, 'vibe coding', a term introduced in February of 2025 and later popularized by developers to describe an intuitive, conversational mode of coding with large language models, has become both a technical workflow and a cultural discourse. Our work presents a longitudinal analysis of 896 Hacker News conversations (155 valid posts and 7,223 comments) to examine how discussions around vibe coding evolved from its introduction to the present. Through thematic and sentiment analysis, we trace a clear progression with early months of optimism and play, mid-year skepticism and fatigue, and late-year consolidation around responsibility and augmentation. Community sentiment mirrored this trajectory, with positive affect rising from ≈50% to ≈64% while negative tone declined from ≈33% to ≈20%. Together, these findings reveal that developers' perception for vibe coding follows an emotional arc of adaptation, from excitement to reflection to mature integration. We argue that studying such affective trajectories offers a window into how software practitioners collectively negotiate creativity, identity, and well-being in the age of vibe coding.

Well-Being in AI-Assisted Software Development

Fairuz Nawer Meem (George Mason University), Justin Smith (Lafayette College), Brittany Johnson (George Mason University)

Well-being in software development has traditionally been examined through lenses such as productivity, burnout, and workplace stress. However, the growing adoption of AI-assisted programming tools introduces new cognitive, emotional, and experiential dimensions of software work that remain insufficiently understood. In particular, little is known about how large language models (LLMs) shape practitioners' moment-to-moment experiences of stress, frustration, confidence, and overall well-being during development tasks.

Our work investigates how well-being is defined and experienced in software development practice, and how AI-assisted programming influences these experiences across professional and academic contexts. We propose a controlled, within-subject experimental study conducted under a broader umbrella of Well-Being in Software Development, with parallel tracks for

software professionals and computer science students. Participants complete two coding sessions involving common software development tasks, such as code writing, debugging, and code comprehension, one without AI assistance and one with access to an LLM (e.g., ChatGPT or Gemini).

Each session is accompanied by think-aloud protocols and pre- and post-task psychological assessments measuring perceived stress, frustration, affect, and well-being. By comparing experiences across AI-assisted and non-assisted conditions, this study aims to identify both the benefits and risks associated with AI-assisted software development. The anticipated contributions include empirical insights into how LLMs alter developers' emotional and cognitive experiences, and design implications for building AI tools that support healthy, sustainable, and human-centered software practice.

Drone On and Code Along: Inclusive Robotics for Non-Tech Majors

Warren Emmell (Christopher Newport University), Evangelina Grimes (Christopher Newport University), Abhishek Phadke (Christopher Newport University)

This research looks into the efficacy of block-based educational coding platforms in making robotics-centered coding accessible to students outside of the STEM demographic. Hands-on automated drone flight is paired with guided instruction with the aim of building confidence and digital fluency. Results will help inform the evaluation and development of interdisciplinary educational initiatives by measuring self-assessed proficiency in technology, engagement, and familiarity with fundamental programming skills.

Flexible Swarm: ROS2 Action-based Coordination of Heterogeneous Drone Swarms

Eva Grimes (Christopher Newport University Undergraduate), Aubrie Kooiker (Christopher Newport University Undergraduate), Emma Faith (Christopher Newport University Graduate)

This research presents Flexible Swarm, an open source framework for heterogeneous drone swarm coordination using ROS2 action-based interfaces. The proposed approach introduces a modular swarm server architecture that provides structured feedback, which can be dynamically acted upon during execution. This framework addresses the limitations of traditional command-and-response drone control interfaces, provides greater insight into task progress, and enables behavior response to runtime feedback. Building on current tools for Crazyflie nanocopters, such as CFLib and Crazyswarm2, Flexible Swarm generalizes swarm coordination to support heterogeneous aerial platforms. The research demonstrates high-level coordination in simulation, real-world Crazyflie nanocopters, and PiHawk quadcopters.

A Comparative Study of Spatio-Temporal Deep Learning Methods for Multimodal Urban Mobility Prediction

Seyma Simsek (Virginia Tech), Melike Yildiz Aktas (Virginia Tech), Chang-Tien Lu (Virginia Tech)

Urban mobility prediction plays a critical role in designing efficient, sustainable, and responsive cities. With the increasing availability of open urban data, researchers now have the opportunity to study mobility patterns across multiple transportation modes. However, existing studies often focus on specific model architectures or isolated data sources, limiting the understanding of how

different spatio-temporal modeling strategies perform when applied to multimodal urban mobility data. This work presents a comparative study of spatio-temporal deep learning approaches for urban mobility prediction using publicly available datasets. The study leverages open-access data sources, including taxi trip records, bike-sharing data, and road network information from platforms such as OpenStreetMap. These heterogeneous data sources are integrated into a unified spatio-temporal representation to capture both temporal dynamics and spatial interactions inherent in urban transportation systems. Rather than proposing a new model architecture, this research focuses on evaluating and comparing existing spatio-temporal learning strategies, including recurrent neural networks, spatially informed deep learning models, and ensemble-based approaches. All models are evaluated under consistent experimental settings to ensure fair comparison and reproducibility.

Adaptive Knowledge Retrieval with Graph-Enhanced and Document-Centric RAG

Melike Yildiz Aktas (Virginia Tech), Min Zhang (Virginia Tech), Meenu Ravi (Virginia Tech), Chang-Tien Lu (Virginia Tech)

This research focuses on a hybrid Retrieval-Augmented Generation (RAG) framework that dynamically integrates graph-based retrieval with document-based retrieval to improve answer quality across queries of varying complexity. Traditional document-centric RAG systems excel at surface-level semantic matching but often struggle with multi-hop reasoning and relational queries. Conversely, while graph-based RAG approaches provide stronger structural reasoning, they can be inefficient or redundant for simpler information needs. Our approach introduces a query complexity assessment layer that automatically selects the relevant retrieval strategy. It leverages document-based RAG for single-hop, factual, or contextual queries, and graph-based RAG for multi-entity, relational, or reasoning-intensive queries. The system combines semantic embeddings, structural signals, and query analysis features to guide routing decisions. This hybrid design demonstrates that retrieval strategy awareness is a key factor in advancing robust, context-sensitive generation systems, potentially reducing response time while delivering more precise reasoning and making next-generation RAG pipelines significantly more efficient and transformative.

PrivacyR1: Privacy-Preserving Collaborative Reasoning in Multi-Agent Systems

Min Zhang (Virginia Tech), Chang-Tien Lu (Virginia Tech)

Complex reasoning, which demands both contextual understanding and logical inference, is challenging for low-capacity local models deployed on computation-constrained devices. Although such complex reasoning queries could be routed to powerful remote models like GPT-4, exposing local data raises significant data leakage concerns. Existing privacy-preservation methods generate problem descriptions or examples for remote assistance via prompting. However, the inherent complexity of these tasks hinders local models from generating effective auxiliary queries and accurately inferring answers from remote guidance. In this work, we propose teaching local and remote agents flexible interactions, enabling the local agent to propose a privacy-preserving query to obtain remote guidance for local reasoning. The model is optimized by rewarding with task accuracy and privacy preservation metrics via reinforcement learning.

Drone Trajectory Tools with KKT Conditions and CasADi

Aubrie Kooiker (Christopher Newport University)

This research presents a variety of tools to create and translate optimized trajectories for quadrotor drones. The tools are scripted in Python and can be run entirely on their own. Each trajectory is generated using KKT conditions to enforce exactness in straight lines and small curves, or using the CasADi Python library and penalty constraints to allow for flexibility with inexact trajectories and obstacle avoidance. This framework allows quadcopter drones to fly through hoops and perform complex courses. Though currently structured to build and deliver trajectories in a format that suits Crazyflie nanocopters, these tools are being expanded and generalized to be compatible with other autopilots that can be used in drones like PiHawk quadcopters. To align with other work on drone swarms, a ROS2 action server has been created to allow all trajectory tools to be accessed from an external behavioral control system, such as a FlexBE state machine.

Between Places and Pasts: Enhancing Cultural Heritage Experiences through Location-Based Systems

Yuanqi Zhen (University of Pennsylvania, School of Engineering and Applied Science)

New ways to get, organize, and share information about cultural heritage have come about due to improvements in web-based mapping services and location-aware technologies. Location-based systems can link cultural content to specific spatial contexts, enabling more flexible, context-specific heritage mapping. This is different from traditional digital heritage platforms that use static lists or one-way narratives.

This project focuses on cultural heritage settings in China and aims to create a database-driven, location-based interactive system for finding historical sites and planning trips. China provides a representative setting due to its extensive and diverse historical resources, and there is a growing public demand for digital tools that help people explore cultural heritage information independently.

We introduce an ongoing web-based prototype system that integrates an interactive map interface with a structured cultural heritage database. The system is designed to help people find and explore heritage sites based on historical period, related figures, topics of interest, and all spots and regions. With user accounts, bookmarking locations, and creating preliminary travel plans are allowed. In addition, users may submit heritage locations that are not yet included in the database, which will then be reviewed by a moderation process.

The poster discusses the system architecture, database organization, and primary interaction mechanisms. It also outlines key design challenges related to data integration, location accuracy, content organization, and user-contributed information management.

Evaluating the Expressibility of Data Encoding Methods for Quantum Machine Learning

Natalie Lee (The Catholic University of America)

In recent years, there has been a focus on the development of Quantum Machine Learning (QML), an interdisciplinary field that combines classical machine learning methods and the distinct hardware capabilities of QCs. However, despite rapid advancements in QML, one fundamental challenge that remains is determining which method of classical data encoding should be used in a quantum algorithm.

In hybrid quantum-classical algorithms, classical data must be transformed into a quantum representation before processing. This step, known as data encoding, plays an integral role in determining algorithm performance by shaping the structure and complexity of the underlying quantum circuit. While many encoding strategies exist, selecting an appropriate encoding method is largely a trial-and-error process. This lack of systematic guidance introduces developmental and computational overhead in the QML workflow.

This project investigates whether expressibility, a measure of how richly a quantum circuit can explore the space of quantum states, can serve as a metric for comparing and selecting data encoding strategies. Specifically, this work examines how the expressibility of quantum encoding circuits influences learning performance across QML models. To accomplish this, encoding methods are evaluated to assess their accessible state spaces and corresponding impact on classification performance using tailored synthetic datasets.

By establishing a relationship between data encoding expressibility and model performance, this project aims to provide a framework for informed encoding selection in quantum algorithms. Improving this foundational step in QML could reduce development costs, enhance model performance, and help identify classes of data that are naturally suited to quantum approaches.

Beyond the Android Manifest: Analyzing Native Libraries and Eye-Tracking Use in Virtual Reality Applications

Kimberly Giordano (Virginia Tech), Brendan David-John (Virginia Tech), Bo Ji (Virginia Tech), Margaret Ellis (Virginia Tech)

Eye-tracking sensors in virtual reality (VR) systems enable advanced interaction techniques and performance optimizations, but they also introduce significant privacy risks due to the sensitivity of gaze data. Prior analyses of VR applications have shown inconsistencies between disclosed sensor usage and in-app consent mechanisms, particularly when relying solely on Android Manifest declarations. This work builds on that foundation by extending eye-tracking audits beyond the manifest layer to examine native library behavior in VR applications.

In earlier stages of this project, I analyzed the Android Manifests from VR applications available on the Meta Store to identify declared permissions and eye-tracking-related capabilities. While this approach surfaced gaps in transparency, it provided limited insight into how eye-tracking data is actually accessed and what it is used for within applications. To address this limitation, my current work investigates the native .so APK files, where core eye-tracking functionality is implemented.

I am currently developing a multi-layer static analysis methodology that combines Android Manifest evaluation with native code inspection to identify eye-tracking use cases such as gaze-based interaction and foveated rendering. Early findings suggest that some eye-tracking functionality may be implemented at the native library level without clear or consistent user-facing disclosure. This ongoing work highlights the limitations of permission-based auditing alone and motivates the need for more comprehensive analysis tools to improve transparency and accountability in VR systems.

Emotion-Aware Speech Analysis for Mental Health Support Using Whisper

Claire Gillaspie (VCU), Cyd Oldham (VCU), Nigel Uson (VCU), Jae Skeete (VCU)

Emotional Self Awareness (ESA) refers to a person's ability to understand and identify their own emotions. Low ESA is associated with depressive and anxiety disorders, while increasing ESA

is associated with a reduction in symptoms. Depression and anxiety are widely experienced, debilitating conditions, and methods to reduce the symptoms of these disorders have critical impacts on public health. Self-monitoring approaches for emotional states are often used with the intention of increasing ESA. However, many people struggle to pinpoint exactly how they feel, when those feelings started, and why. Thus, manual self-monitoring imposes a daily obligation onto the user and is limited by the individual's own self-awareness and ability to reflect honestly. Providing continuous, automatic, and objective emotion tracking would address these shortcomings, yet no current methods to do so exist. In this work, we developed a system to solve this problem using automatic, AI/ML powered emotion monitoring through speech and textual emotion classification. The privacy module transcribes a user's speech while removing all other speakers from the analysis. Within the classification module, OpenAI's Whisper is used to perform a multi-modal analysis using both audio and text to classify the user's emotions. Automatic tracking removes the individual's responsibility of logging entries, creating accurate and consistent mood trends. This project introduces a novel and cost-effective approach to improving ESA by transforming everyday speech into data driven insights.

Bridging the AI Education Gap: A Self-Funded AI Awareness Initiative in Cocoa-Farming Villages of Ghana

Zahra Rizvi (William and Mary), Sophia Rizvi (Grafton High School), Derek Atiawu (William and Mary)

Access to Artificial Intelligence education remains limited in many rural communities in the Global South, particularly where infrastructure and digital resources are scarce. This work describes a self-funded AI awareness initiative founded and led by Derek Worlasi Atiawu, a William & Mary graduate and healthcare data professional with experience in business analysis and software development, focused on AI for developing countries. The initiative introduces foundational AI concepts to middle school students in cocoa-farming villages in Ghana, communities that face economic hardship, limited technology access, and environmental challenges that threaten local livelihoods. The program delivers remote, interactive training sessions to build basic understanding of data, algorithms, machine learning, neural networks, image recognition, and robotics. In the first training cycle in late 2025, 60 students participated, including 35 girls and 25 boys. Recently donated projectors and robotics materials will soon support more engaging, hands-on instruction. Major challenges affecting the initiative include poor internet access for online training, and limited availability of laptops, tablets, projectors, AI books, and robotics resources. Following training, students will apply AI to practical community projects that support cocoa farming and the broader chocolate production value chain. Planned projects include an AI-powered weather application to recommend optimal conditions for drying cocoa seeds, and an AI tool to distinguish good cocoa beans from poor-quality ones to improve product quality and farmer income. The long-term goal is to reach 100,000 rural students by 2035, with at least 60 percent participation from girls, helping close the global AI education gap.

Bridging the AI Education Gap: Understanding How Students Learn to Maintain AI Systems

Fatema Tuz Zohra (George Mason University), Brittany Johnson (George Mason University)

While the technology industry is rapidly advancing the deployment of artificial intelligence (AI) systems, academic curricula are still struggling to address how these systems should be maintained over time. Despite the widespread adoption of AI, model maintenance, including

testing, debugging, and repairing, remains underrepresented in formal education. Findings from our recent survey of 301 academic stakeholders, including both educators and students, reveal a significant gap between AI development training and maintenance preparedness. Although 84.1% of participants reported incidental exposure to maintenance-related concepts within general machine learning or software engineering courses, 95.4% indicated a critical absence of dedicated instructional modules focused on AI model maintenance. Participants further identified systemic barriers to effective learning, including the inherent opacity of model errors, the lack of standardized curricula, and limited exposure to industry-standard maintenance tools. These findings suggest that current educational practices disproportionately emphasize model development while overlooking the long-term sustainability and reliability of AI systems.

Securing Vehicular Ad Hoc Networks (VANETs) Against Cyber Threats

Susan Zehra (Old Dominion University), Syed Rizvi (Old Dominion University), Stephan Olariu (Old Dominion University)

Vehicular Ad Hoc Networks (VANETs) enable real-time communication among vehicles and infrastructure, improving traffic management and road safety. However, their open and dynamic nature makes them vulnerable to cyber threats such as record-and-relay attacks, man-in-the-middle attacks, and deception by pranksters. This work proposes a security framework that operates without a major central authority, relying instead on a decentralized authentication mechanism for verifying vehicles. The mobility model considers two-way lanes, varying vehicular densities, and unplanned entry and exit on highways, ensuring a realistic traffic environment. The network model integrates multi-hop packet relays, vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication, encryption and decryption of packets, data chunking, and multi-source/multi-destination routing for enhanced security and resilience. The proposed security methodology is grounded in probability theory, modeling attack likelihood and the effectiveness of countermeasures using stochastic analysis. Simulation experiments evaluate system robustness under different attack scenarios, measuring key metrics such as message integrity, delay, and attack detection rates. Access Points (APs) serve as communication hubs, facilitating secure data transmission. Our results demonstrate that the proposed decentralized security mechanism mitigates cyber threats effectively, ensuring secure and resilient VANET communication in dynamic and adversarial environments.

Heterogeneous Paging in PagedAttention: Adaptive Memory Management for LLM Inference Workloads

Nishat Sultana (William and Mary)

Efficient key-value (KV) cache memory management is critical for serving large language models (LLMs) at scale. While PagedAttention uses fixed-size memory blocks to reduce fragmentation, the dynamic nature of LLM workloads suggests potential benefits from adaptive paging strategies. Motivated by heterogeneous paging techniques in operating systems, we extend PagedAttention with a MultiPool Architecture that supports variable block sizes (16 and 128 tokens) and dynamically allocates memory based on workload characteristics. We implement adaptive allocation policies, modify CUDA attention kernels for cross-pool operations, and integrate workload-aware scheduling to optimize memory utilization. Through comprehensive evaluation on NVIDIA A100 GPUs using real-world workloads (ShareGPT) and multiple LLM architectures, we investigate whether heterogeneous paging benefits observed in CPU systems transfer to GPU memory management. Our findings reveal that heterogeneous paging introduces significant performance overhead due to increased kernel complexity,

memory coalescing constraints, and the memory-bound nature of attention operations. Contrary to expectations from OS literature, uniform small blocks (16 tokens) consistently outperform heterogeneous strategies across diverse workloads. Through detailed profiling, we identify fundamental architectural differences between CPU virtual memory and GPU KV cache management that explain this performance gap. Our work provides the first systematic study of heterogeneous paging for LLM serving, offering quantitative analysis of the fragmentation-versus-complexity trade-off and design guidelines for future systems. This demonstrates that architectural insights from one domain do not automatically transfer to another, highlighting the importance of empirical validation in AI system design.

Moodoodle: Generative Soundscapes for Reflective Emotional Exploration

Mindy Zheng (University of Virginia)

Emotions are often communicated through language-based systems that rely on predefined labels, written reflection, or corrective feedback. While these approaches can support emotional awareness, they can also feel limiting when emotions are vague, mixed, or difficult to articulate. This project, Moodoodle, explores sound as an alternative medium for engaging with emotional experience in a more open-ended and reflective way. Rather than asking users to label or optimize their emotions, Moodoodle allows users to externalize how they feel through short, generative soundscapes.

Moodoodle is a web-based application in which users select a color and write a short freeform text describing their current emotional state. These inputs are mapped onto a valence-arousal model and used as generative controls to produce a one-minute ambient soundscape. Valence influences harmonic mode and timbral brightness, while arousal affects tempo and rhythmic density. The system does not aim to detect emotions accurately or guide users toward a desired emotional state. Instead, it treats emotional input as subjective and interpretive, embracing ambiguity and variation as core features of emotional experience.

The design of Moodoodle is informed by constructionist theories of emotion, which view emotions as constructed through context and interpretation, as well as principles from reflective design in HCI that emphasize supporting reflection rather than behavioral optimization. By generating sound as a discrete artifact that users can listen back to and interpret on their own terms, Moodoodle positions sound as a reflective medium rather than a regulatory or diagnostic tool.

Generative Models for Reconstructing Cosmological Initial Conditions from Realistic Observations

Giti Doolabi (University of Virginia)

Recent advances in generative models have demonstrated strong performance in reconstruction tasks across scientific domains, including the reconstruction of cosmological initial conditions from late-time matter density fields. While these approaches have shown promising results, most existing studies rely on idealized observations, such as full dark matter density fields, which are not directly accessible in real astronomical surveys. This gap limits the practical applicability of current reconstruction methods to realistic cosmological data.

In this work, we investigate the impact of stochastic generative models, Denoising Diffusion Probabilistic Models (DDPMs), Denoising Diffusion Implicit Models (DDIMs), and Conditional Flow Matching (CFM), for reconstructing the initial conditions of the universe under more realistic observational settings. Compared to diffusion-based models, CFM offers faster

sampling and deterministic inference, making it well suited for large-scale scientific inverse problems.

Our experiments focus on conditioning the generative models on halo-based observations, which more closely reflect real galaxy survey data than idealized density fields. Preliminary results suggest that while reconstruction from halo fields is significantly more challenging, both conditional flow-based models and diffusion models outperform existing deterministic reconstruction approaches. We further explore how incorporating richer datasets and additional observational information can enhance reconstruction quality.

This poster presents ongoing PhD research and aims to stimulate discussion on moving beyond idealized assumptions in cosmological reconstruction and the role of generative modeling choices in enabling practical, uncertainty-aware AI tools for scientific discovery.

Exploring Socioeconomic Status Narratives of Computer Science Students

Jennifer Alexandra Thompson (Virginia Tech), Sara Hooshangi (Virginia Tech)

Socioeconomic status (SES) has been an underexplored identity factor in computer science education. This gap is attributable to many factors, such as the difficulties with defining and measuring SES, the sensitivity of the data being collected, and the lack of readily available SES data collected by large-scale datasets. Despite these challenges, socioeconomic status plays a large role in determining a student's access to and familiarity with technology. In this in-development study, I seek to understand the relationship between a student's socioeconomic identity and the impact it has on their performance in computer science (CS) courses. I will develop a mixed-methods approach to understand what aspects of socioeconomic status most impact a CS student's prior access to technology, prior programming experience, and academic achievement in computer science. I plan to understand these perspectives by looking at populations of college-level computer science students from both a rural community college and a public research-focused university. This method consists of using short surveys and interviews to understand the perceived impact that a student's socioeconomic status has had on their ability to succeed in computing. Interviews will further understand what aspects of socioeconomic inequality were most relevant to a student's choice to apply to, to enroll in, and to persist in computer science. I will utilize narrative analysis methods to identify key themes that emerge throughout this study. The goal of this study is to further understand the relationship between a computer science student's SES and their ability to succeed in computer science courses.

Deep Metric Learning for Ancient Coin Identification

Dhanshree Atre (James Madison University), Nathan Sprague (James Madison University), Jason Forsyth (James Madison University), Trevor Schonbrun (James Madison University)

The identification of specific ancient coins within catalog records is a challenging computer vision task due to surface wear, faded details, inconsistent lighting, and the limited availability of labeled images per coin. In the absence of large numbers of examples per coin, our approach relies on image augmentations to simulate images taken under different conditions. This project explores the use of deep metric learning for ancient coin recognition by evaluating convolutional neural network backbones using ArcFace loss. ResNet50 and InceptionV3 models were trained on augmented images from the RRC-60 and RRCD datasets, incorporating transformations such as rotation, brightness adjustment, blur, and perspective distortion to improve robustness

and generalization. Experimental results demonstrate strong performance on augmented images, achieving approximately 99-100% classification accuracy on a holdout test set, with only 25 misclassifications out of 2,954 images. While both architectures perform competitively, further analysis is needed to determine their relative accuracy on a real-world dataset consisting of unaugmented images of individual coins captured under varying conditions. Future work will focus on extracting explainability from the trained models, enabling visualization and interpretation of the features and regions that influence coin identification decisions.

MITransfer: Towards Model-level Generalization of Mechanistic Interpretability

Wendy Zheng (University of Virginia), Yinhan He (University of Virginia), Jundong Li (University of Virginia)

Mechanistic interpretability (MI) has made significant progress in reverse-engineering the internal algorithms of large language models, uncovering circuits responsible for specific capabilities. However, current MI research predominantly focuses on analyzing individual models in isolation, requiring researchers to repeat computationally expensive interpretability analyses from scratch for each new model. This approach becomes increasingly impractical as model architectures proliferate and diversify, motivating the need for mechanistic interpretations that generalize across models. Achieving such generalizability faces two major challenges: (1) Models vary in layer count, hidden dimensions, and attention mechanisms, preventing straightforward one-to-one component mapping. (2) Functionality can be distributed across multiple components, or single components can serve multiple functions, obscuring how to identify equivalent circuits across models.

To address these challenges, we propose MITransfer, a cross-model circuit transfer framework that aligns functionally equivalent components across architectures without requiring structural identity. Our framework introduces a flexible alignment mechanism that matches components based on functional similarity rather than positional correspondence, enabling the transfer of discovered circuits to unseen models with minimal additional analysis. Through extensive experiments spanning model families with different scales and architectures, we demonstrate that MITransfer successfully transfers circuits while preserving their functional roles, establishing that mechanistic insights can generalize beyond individual models.

Spatial Seer: Exploiting Telemetry to Expose XR User Environment

Gayatri Kamtala (Virginia Tech), Allie Craddock (Virginia Tech), Bo Ji (Virginia Tech), Brendan David-John (Virginia Tech)

The applications of extended reality (XR) are rapidly expanding across healthcare, retail, education, and entertainment. As the XR market is projected to reach an $84.86 billion valuation by 2029, concerns surrounding security and user privacy have grown accordingly. This research examines how system-level performance data generated by XR headsets can be exploited to infer sensitive details about a user's physical environment without direct access to camera data, highlighting the need for stronger security protections in XR platforms.

Using developer-facing scanning tools on the Magic Leap 2 and Meta Quest 3, we study patterns in memory usage and power consumption to infer environmental features. Data were collected and cross-validated using Magic Leap's Power Profiler, Perfetto Traces, Unity Profiler, and Unity Memory Profiler. These metrics proved highly informative, enabling the classification of room type using supervised machine learning models. We achieved cross-validated accuracies of 91% with a support vector machine trained on Magic Leap 2 data, 98.67% with a

Random Forest model using Meta Quest 3 Perfetto data, and 100% with a Random Forest model trained on Unity profiling data. Across devices and profilers, these results demonstrate that performance metrics can reliably fingerprint physical environments, underscoring significant side-channel risks in XR systems.

Future work aims to automate data collection to evaluate whether memory and power profiling can be triggered through in-application object interactions and leveraged in a realistic end-to-end cyberattack, moving beyond theoretical risk to assess the feasibility of unauthorized access and remote data exfiltration within Unity-based XR applications.

Exploring the Factors Contributing to the Gender Disparity in Computing Education

Sara Hailemariam (Roanoke College)

Despite the growing demand for computing professionals, women remain underrepresented in computer science and data science programs. This project explores the factors contributing to the gender disparity in computing education, with a focus on student experiences and enrollment trends at a small liberal arts college. This study uses a mixed-methods approach. Quantitative data are analyzed to examine patterns in course enrollment and progression within computing-related majors. In addition, qualitative interviews with students provide insight into their motivations, experiences, and perceived barriers in computing courses. The goal is to better understand how classroom environments, prior exposure, confidence, and sense of belonging may influence students' decisions to pursue or leave computing fields. Preliminary findings suggest that factors such as limited early exposure to computing, lack of representation, and classroom dynamics may affect students' interest and persistence. By identifying these influences, this research aims to highlight opportunities for institutions to create more inclusive and supportive computing pathways. This project contributes to ongoing conversations about equity in computing education and seeks to inform strategies that improve recruitment, retention, and success for women in computing programs.

Towards Large Language Model-powered Automation of Detecting Metadata Related Bugs

Yeana Bond (Virginia Tech)

Misuse of metadata can introduce severe issues in Enterprise Applications written in Java, including misconfiguration, incorrect runtime behavior, developer misunderstanding, and security vulnerabilities. Existing tools provide limited support for analyzing and debugging metadata bugs. MeCheck is a metadata checker that offers partial automation for detecting metadata issues in Java applications using annotations and XML. Inspired by the engine capable of analyzing both annotations and XML files, we developed an automated pipeline that generates rules written in Rule Specification Language (RSL) from metadata-related bug descriptions extracted from JetBrains inspections. These rules analyze Java codebases that use XML configuration and metadata annotations.

The pipeline leverages Large Language Models (LLMs) to generate rules, extract relevant metadata annotations from bug descriptions, augment existing example rules, locate relevant web resources related to metadata usage, and validate generated rules. We evaluate the pipeline using GPT-4 and GPT-5 across these capabilities as an initial step toward fully automated detection of metadata bugs. Experimental results show that rules generated by GPT-5 demonstrate higher semantic correctness and implementation completeness than those

generated by GPT-4, although incomplete or partially correct rules remain for both models. GPT-5 is also less likely to violate pipeline constraints. However, extending MeCheck with additional built-in functions and manually refining generated rules remain necessary for automation. Overall, this present work demonstrates the potential of LLMs for generating and validating rules and for advancing automated metadata debugging tools.

PhysioAI

Amulya Chinnala (Virginia Tech), Aleysa Sweeney (Virginia Tech), James Mercer (Virginia Tech), Nico Parong (Virginia Tech), Rhea Parekh (Virginia Tech), Vinicius Rezanejad (Virginia Tech), Sehrish Nizamani (Virginia Tech)

Our project addresses the critical gap in patient care that occurs after leaving the hospital. While in-hospital care is strictly monitored, patients often struggle to maintain their rehabilitation regiments once discharged, leading to improper exercise form, diminished motivation, and ultimately, stalled recovery. To solve this, we have developed an AI-powered physical therapy app that provides patients with immediate, personalized feedback. Our project also has a doctor's side that allows healthcare professionals to verify their patient's recovery is on track, identify patients who are falling behind, and receive automated alerts for urgent issues. This app is targeted towards hospitals who want to invest in our services to improve remote patient care.

Cyber Hygiene That Sticks: Research in the K-12 Space on Cybersecurity

Deborah Kariuki (University of Maryland Baltimore County), Hidare Debar (University of Maryland Baltimore County)

As cybersecurity concepts expand into K-12 education, cyber hygiene is often taught through warnings, rules, and one-time awareness lessons. While these approaches increase exposure, they rarely support long-term reasoning or transfer to unfamiliar digital contexts. This research examines how hands-on, human-centered instructional design can strengthen cybersecurity learning in K-12 classrooms. Drawing on classroom-informed implementations and educator professional development, the work explores how interactive activities focused on phishing detection, password practices, and data-sharing decisions promote student engagement and visible reasoning. Rather than framing students as the "weakest link," this study positions cyber hygiene as a learnable practice shaped by instructional design and sociotechnical context. Early findings suggest that repeated, scenario-based practice supports deeper understanding and improved recognition of new cyber threats. This work contributes evidence-based strategies for computing educators seeking scalable, developmentally appropriate approaches to cybersecurity education that build confidence, judgment, and durable digital safety habits.

Benchmarking DAOS Filesystem on Aurora

Rebecca George (William & Mary), Xinxin Mei (Jefferson Lab), Jie Ren (William & Mary)

The growing scale and complexity of scientific computing workloads, including large-scale simulations, AI/ML training, and real-time data acquisition, have exposed fundamental I/O bottlenecks in traditional parallel file systems. The Distributed Asynchronous Object Storage (DAOS) system is an open-source, scale-out object store that addresses these bottlenecks by bypassing the OS kernel and employing a distributed key-value architecture that eliminates centralized metadata limitations, delivering high throughput, low latency, and high IOPS through interfaces including POSIX, MPI-IO, HDF5, and DFS.

In this work, we evaluate the performance of DAOS on Aurora, an exascale supercomputer at Argonne National Laboratory (the third largest supercomputer on the Top500 list). DAOS on Aurora currently achieves the highest storage performance on the IO500 Production list. Using standard HPC I/O benchmarks (fio, IOR, mdtest, and IO500), we systematically explore throughput, IOPS, and latency scaling across varying block sizes, read/write ratios, access patterns, and concurrency levels. Key findings include that write-dominant workloads achieve higher bandwidth than read-dominant ones, that random and sequential access patterns exhibit minimal performance differentiation, and that peak throughput is attained at block sizes exceeding 1 MiB. These results offer practical guidance for optimizing I/O performance on DAOS-based exascale storage systems.

SongSmith: An AI Lyric Study

Andre Daniels (Virginia State University), Denise Daniels (Virginia State University), Joon Lee (Virginia State University)

Fueled by the rapid growth of Artificial Intelligence (AI) tools built on large language models (LLMs), AI is transforming creative industries at an unprecedented pace. Tasks once considered uniquely human---such as composing music, writing poetry, and crafting emotionally powerful lyrics---are increasingly being performed by AI systems. This shift challenges long-held assumptions about creativity, originality, and artistic ownership. SongSmith: An AI Lyric Tool builds on these concerns by exploring two key questions. First, can creative activities traditionally dependent on human imagination be replicated by artificial agents? Second, how do computer science students---future engineers who may shape the next generation of AI---perceive the ethical, cultural, and intellectual property issues surrounding machine-generated creative work? Copyright law offers limited guidance in this area. While some creative fields receive minimal protection and certain reuse may qualify as fair use, the boundaries become unclear when AI generates lyrics or melodies that closely resemble a recognizable artist's style. A major concern may also lie in the datasets used to train these systems, which often contain copyrighted works without users realizing it. SongSmith serves as both a creative system and a research tool. It presents the user with original generated lyrics based on prompts such as "in the style of Adele" or "inspired by Kendrick Lamar." In a controlled study, participants compare unlabeled human-written lyrics and AI-generated lyrics to evaluate emotional authenticity and artistic engagement. Interviews further examine how students interpret ethical and social implications of AI-assisted creativity.

Knowledge Triple Filtration for Explainable Question-Answering

Joanne Varughese (Virginia Commonwealth University), Nicole Yazbeck (Virginia Commonwealth University)

A significant application of Natural Language Processing (NLP) is the development of question answering (QA) systems, which are instrumental in enabling information retrieval from extensive knowledge bases. This research explores Graph Retrieval-Augmented Generation (GraphRAG) and PathRAG, techniques that enhance data retrieval and processing capabilities in NLP. GraphRAG leverages graph structures to improve data connectivity, while PathRAG optimizes retrieval paths, both aiming to improve the accuracy, transparency, and reliability of Medical QA systems. In a typical QA system, users find it difficult to trace answers back to the source. However, our system can enhance answer traceability, allowing users to query vast collections of medical studies and receive verifiable, evidence-based answers, thereby supporting healthcare professionals to make better-informed decisions and helping patients gain more

reliable insights into their health. To address the challenge of balancing retrieval quality with explainability, we developed a novel GraphRAG system. This system employs a graph filtration strategy using the Linking Term Count metrics to reduce noise and redundancy before query-time retrieval. From our preliminary tests, we expect our system to enhance efficiency and explainability compared to traditional RAG systems. To evaluate performance, we measure the accuracy, and efficiency is assessed by comparing the retrieval time between GraphRAG and RAG systems. Our work improves the traceability and accuracy of medical QA systems, this research contributes to developing AI tools that make medical knowledge more transparent, verifiable, and accessible to both healthcare professionals and patients.

Improving Automated Coding of News Data: A Machine Learning Approach to News Framing

Emily Lugos (Villanova University), Mauricio Gruppi (Villanova University)

The modern news cycle has been fundamentally reshaped by the rapid exchange of information online, producing media framing that evolves as new information, political responses, and social reactions emerge. Understanding how these narratives form and shift over time is essential for interpreting public discourse. While large-scale event datasets such as GDELT provide broad coverage and standardized coding of political and social events, preliminary results suggest that automated coding approaches are often insufficient for identifying complex event relationships and evolving narratives. Moreover, as the volume of online reporting expands, coding frameworks must become more sophisticated to accurately identify event subjects and topics and to filter the substantial informational noise that accompanies real-time news production. As a result, existing event data often fails to capture the nuances of framing and context that shape how events are communicated and understood. To address these limitations, we adopt an empirical approach to develop a representation learning framework that encodes news articles according to event similarity. Our goal is to make event data more robust for large-scale analysis while preserving the nuanced distinctions necessary for interpretive research. Using a combination of BART models for zero-shot and few-shot classification and Sentence Transformers embeddings for semantic similarity, we extract article-level information to train article embeddings that capture relations between events and stories. In doing so, this study aims to bridge the gap between automated event coding systems and human-centered analyses of news discourse, while improving the ability of computational methods to track media framing of events.

Praxly: An Online IDE for the Praxis CS Test Pseudocode

Ellona Macmillan (James Madison University), Chris Mayfield (James Madison University), Chris Johnson (James Madison University)

Praxly (praxly.cs.jmu.edu) is a web-based IDE that empowers users to read, write, and run the pseudocode on the Praxis CS test. The development of Praxly has a significant impact on CodeVA's Praxis Prep course by providing a hands-on tool to make learning more interactive. The Praxly team is eager to extend this resource to educators across the country who are aiming to pass the CS Praxis test to obtain their state's computer science endorsement. Teachers have also asked if Praxly could be used for student instruction. We are currently developing Praxly2, a rewrite of Praxly that supports additional languages including CSP pseudocode, Java, and Python. We are also adding language features for the Praxis pseudocode not currently supported by Praxly, such as multidimensional arrays, classes, and object-oriented features. In addition to expanded language support, Praxly2 introduces a code

explainer feature that translates pseudocode into plain-English descriptions. This feature is designed to help novice programmers understand the meaning and intent of pseudocode fragments commonly used on the Praxis exam, supporting conceptual understanding and code-tracing skills. Praxly2 is built on a robust, object-oriented abstract syntax tree architecture that enables program execution, code generation, and future language extensions, making the platform easier to maintain and extend as instructional needs evolve.

Birds of a Feather

Total abstracts in this session: 1

Thriving in Computing Across Career Stages: A Values-Based Framework for Retention, Resilience, and Leadership

Adalena Oliver (Amtrak)

Women in computing often navigate careers that are non-linear, fast-changing, and shaped by both opportunity and exclusion. While technical skills are essential, they are rarely sufficient for long-term retention, resilience, and leadership growth. This Birds-of-a-Feather session invites students, early-career professionals, and allies into an open, facilitated conversation about how core values influence career decisions, persistence, and identity across different stages of a computing career.

Drawing from industry leadership, organizational change management, and mentoring experience, I will introduce a values-based developmental lens to help participants reflect on where they are in their journey, what pressures they are experiencing, and how misalignment between personal values and professional environments often contributes to burnout, disengagement, or departure from the field. Rather than presenting a prescriptive solution, the session emphasizes shared experience, peer learning, and reflection.

Participants will explore questions such as why some women thrive in computing while others quietly leave, how values evolve across career stages, and what success can look like beyond titles and technical mastery. Attendees will leave with practical reflection prompts and a clearer understanding of how values awareness can support intentional decision-making, resilience, and leadership development in computing fields.

Acknowledgments

We extend our sincere gratitude to all those who made CAPWIC 2026 possible. We thank our organizers and volunteers for their dedication and hard work, our program committee and reviewers for their thoughtful evaluations and contributions, and our sponsors for their generous support.

Your collective efforts have helped create a welcoming and impactful conference experience for our community.