THE UNIVERSITY OF BRITISH COLUMBIA
Department of Computer Science, Mathematics, Physics and Statistics
Okanagan Campus

CMPS Undergraduate Research Conference

April 2026

General Information

- Date: Friday, April 24, 2026
- Time: 8:30 am – 5:00 pm
- Discipline: COSC, MATH, PHYS, and STAT/DATA
- Location: SCI 247, SCI 333, and SCI 337
- Organizers: Alex Hill, Ifeoma Adaji, Paul Tsopméné, and Yves Lucet
- Contact:
  - ⋆ COSC: Yves Lucet (yves.lucet@ubc.ca)
  - ⋆ MATH/PHYS/STAT/DATA: Paul Tsopméné (paul.tsopmene@ubc.ca)
- Website: The up to date schedule will be kept at

  http://cmps.ok.ubc.ca/undergraduate/conference

Contents

1 Parallel Session I (Computer Science) – SCI 337

1.1 Names of Speakers

1. Peter Okpoga Idoko
2. Jimi Ademola
3. Shreya Saxena
4. Lucas Xu
5. Om Mistry
6. Alexander Argatoff
7. Adara Putri
8. Kia Khademi
9. Huynh Ngo Vu
10. Prina Metha
11. Aliasgar Sakarwala
12. Mohamed Sakr
13. Tahsin Jawwad
14. Samira Almuallim
15. Abdullah Munir

1.2 Location

- SCI 337

1.3 Zoom Link

- https://ubc.zoom.us/j/65806999712?pwd=btM11PXcQDHABhw8Q22xvaCgdsqiAf.1

1.4 Length of Each Talk

- 16 minutes followed by a 4-minute question period.

1.5 Schedule and Titles

Welcome (SCI 333): 8:30 am – 9:00 am

Session 1 (SCI 337): 9:00 am – 10:00 am
Chair: Gema Rodriguez-Perez

- 9:00 am – 9:20 am: **Peter Okpoga Idoko**
  - ⋆ Title: Designing a Recommendation System to Optimize Mentor-Mentee Pairs in OSS Programs
  - ⋆ Supervisor: Gema Rodriguez-Perez
- 9:20 am – 9:40 am: **Jimi Ademola**
  - ⋆ Title: Beyond the Pull Request: Examining Agentic Code Quality in Mature Open Source Projects
  - ⋆ Supervisor: Gema Rodriguez-Perez
- 9:40 am – 10:00 am: **Shreya Saxena**
  - ⋆ Title: Dynamic Scheduling for the Satellite Range Scheduling Problem Using Heuristic, Optimization, and AI-Based Methods
  - ⋆ Supervisor: Gema Rodriguez-Perez

Break: 10:00 am – 10:15 am

Session 2 (SCI 337): 10:15 am – 11:15 am

Chair: Ramon Lawrence

- 10:15 am – 10:35 am: **Lucas Xu**
  - ⋆ Title: Bridging the Gap Between NL2SQL Models and Users: A VSCode Plugin with Verified SQL Generation
  - ⋆ Supervisor: Ramon Lawrence
- 10:35 am – 10:55 am: **Om Mistry**
  - ⋆ Title: Image Processing System for Apple Trait Analysis
  - ⋆ Supervisor: Ramon Lawrence
- 10:55 am – 11:15 am: **Alexander Argatoff**
  - ⋆ Title: Translating Research Algorithms into Production Systems: Adding Sorting to EmbedDB
  - ⋆ Supervisor: Ramon Lawrence

Break: 11:15 am – 11:30 am

Session 3 (SCI 337): 11:30 am – 12:30 pm

Chair: Bowen Hui

- 11:30 am – 11:50 am: **Adara Putri**
  - ⋆ Title: Feasibility of EEG-Based Confusion Detection for Student Reading Using a Low-Cost BCI Headset
  - ⋆ Supervisor: Bowen Hui
- 11:50 am – 12:10 pm: **Kia Khademi**
  - ⋆ Title: A Data-Driven Analysis of Assessment Item – Quality in Gamified Computer Science Learning Environment
  - ⋆ Supervisor: Bowen Hui
- 12:10 pm – 12:30 pm: **Huynh Ngo Vu**
  - ⋆ Title: All You Need is Density: A Clustering Approach to Polyline Extraction from Survey Points
  - ⋆ Supervisor: Yves Lucet

Lunch (SCI 337): 12:30 pm – 1:30 pm, we will offer the lunch (pizzas)

Session 4 (SCI 337): 1:30 pm – 2:30 pm

Chair: Fatemeh Fard

- 1:30 pm – 1:50 pm: **Prina Metha**
  - ⋆ Title: Evaluating Language Models Across Bloom's Taxonomy for Educational Contexts
  - ⋆ Supervisors: Mostafa Mohamed and Fatemeh Fard
- 1:50 pm – 2:10 pm: **Aliasgar Sakarwala**
  - ⋆ Title: Retirement-Oriented Robo-Advising
  - ⋆ Supervisors: Fatemeh Fard and John Thompson
- 2:10 pm – 2:30 pm: **Mohamed Sakr**
  - ⋆ Title: Sina: A Multi-Agent System For Continuous Health Monitoring and Personalized Coaching
  - ⋆ Supervisor: Fatemeh Fard

Break: 2:30 pm – 2:45 pm

Session 5 (SCI 337): 2:45 pm – 3:45 pm

Chair: Patricia Lasserre

- 2:45 pm – 3:05 pm: **Tahsin Jawwad**
  - ⋆ Title: Extending a Rule-Based SQL-to-MongoDB Translator for a Two-Stage Text-to-NoSQL Pipeline
  - ⋆ Supervisor: Ramon Lawrence
- 3:05 pm – 3:25 pm: **Samira Almuallim**
  - ⋆ Title: Rewriting Dialogue, Not Meaning: Controlling LLM Behavior with Prompt Engineering
  - ⋆ Supervisor: Patricia Lasserre
- 3:25 pm – 3:45 pm: **Abdullah Munir**
  - ⋆ Title: Digitizing Plains Cree: Adapting CRAFT Character-Level Detection for Historical Syllabics Documents
  - ⋆ Supervisor: Patricia Lasserre

Break: 3:45 pm – 4:00 pm

Prize Draw and Closing/Party (SCI 337):
4:00 pm – 5:00 pm

1.6 Abstracts

1. **Peter Okpoga Idoko:** Open-source software mentorship programs are instrumental in attracting and retaining contributors to projects. These programs cultivate community building, skill development, and exposure to ongoing projects. However, challenges during mentorship programs have resulted in reduced retention rates among new contributors. Existing literature defines the roles of mentors and mentees and explores the factors that contribute to successful mentorship pairings and the retention of mentees. Whereas the existing tools focus on expert knowledge, sociability, and activity level, this study proposes a general-purpose recommendation system that predicts successful pairings based on psychological profiles, capacity, behavior, and user demographics. The system uses a random forest model trained on Google Summer of Code (GSoC) data collected in the OpenMent dataset to provide a ranking of the ideal mentorship pairings.

2. **Jimi Ademola:** The increasing adoption of AI-authored coding tools raises questions about how these tools affect software quality and integration practices in open-source software development. In this study, we empirically compared AI- and human-authored Pull Requests (PRs) using a curated subset of the AIDev dataset comprising 36 mature repositories. We analyzed 1,962 PRs and over 66k file-level changes using delta-based static analysis metrics from SonarQube, capturing changes in maintainability, Technical Debt, reliability, security, and complexity introduced by each PR. Using nonparametric statistical tests, we find that AI-authored PRs introduce significantly smaller per-file increases in complexity, code smells, and Technical Debt indicators than human-authored PRs. Overall, our results suggest that AI-authored code does not disproportionately contribute to immediate Technical Debt accumulation.

3. **Shreya Saxena:** Computational satellite scheduling is a critical challenge in modern space communication systems. This study investigates the Satellite Range Scheduling Problem (SRSP), which involves coordinating limited satellite-ground communication opportunities known as visibility windows. As satellite networks expand and communication demands become increasingly dynamic, static scheduling approaches become less effective. This research develops and evaluates a dynamic scheduling framework that integrates heuristic, optimization, and AI-based methods to handle new communication requests in real time. Using open orbital and ground-station data, the study quantifies measurable criteria such as throughput (MB), runtime (s), and schedule disruption (%) to compare algorithm performance under varying request frequencies. The goal is to establish a reproducible, quantitative framework for analyzing adaptive scheduling strategies in satellite communication systems.

4. **Lucas Xu:** This study addresses three critical issues in existing natural language to SQL (NL2SQL) systems: benchmark bias, LLM hallucination, and lack of real-world deployment. We propose a two-part optimization. First, we correct gold-answer errors in the SPIDER dataset and develop an error-aware module with a three-tier validation mechanism and a two-tier correction strategy, improving DAIL-SQL execution accuracy from 80.9% to 87.5%. Second, we optimize the model pipeline and implement a VSCode plugin to bridge the gap between research models and practical user experience, enabling more reliable and usable NL2SQL systems in real-world applications.

5. **Om Mistry:** This thesis presents an end-to-end image processing system that automates apple phenotyping for agricultural research at the AAFC. The system integrates a YOLOv11 object detector with classical computer vision techniques to detect individual fruit and stems, extract calibrated physical measurements via a reference card, and compute 40 engineered features spanning physical, colour, shape, texture, russet, and contextual categories. A lightweight YOLOv11n model trained on a 333-image, 9-class dataset achieves 99.5% test accuracy and 100% recall. The processing pipeline is exposed through a FastAPI backend and Next.js web frontend supporting single-image analysis, batch hot-folder processing with automated QR-code-based dataset organisation, and persistent storage in PostgreSQL. This talk describes the system design, the hybrid HSV + CIELAB colour classification approach, the rule-based quality grading system, and key evaluation results from work developed in collaboration with AAFC.

6. **Alexander Argatoff:** Embedded systems with constrained random access memory require external sorting algorithms that carefully manage memory across buffer allocation and SD file read/write operations. This thesis advances the EmbedDB sorting implementation to a production-ready state by introducing improved interfaces, robust memory management, and comprehensive testing. The sorting features also allow the implementation of GROUP BY on any column that can be used in conjunction with the existing aggregate methods and queries. Benchmarks between EmbedDB and SQLite sorting on different devices is performed, showing faster speeds for SQLite but improved relative performance for EmbedDB on less performant hardware.

7. **Adara Putri:** This study investigates whether a consumer-grade, low-cost EEG headset can detect reading-induced confusion in undergraduate students, with the long-term goal of informing accessible brain-computer interface (BCI) tools for educational settings. Ten computer science undergraduates wore the Neurosity Crown, an 8-channel dry-electrode headset, while reading 30 paragraphs spanning neutral, partly confusing, and fully confusing difficulty levels. Participants self-labelled each sentence as neutral or confusing via a button click, while continuous EEG was recorded. Offline processing applied bandpass and notch filtering, sliding-window segmentation, and extraction of theta, alpha, and beta band power features alongside time-domain statistics. Random Forest and XGBoost classifiers were trained per participant on the resulting 56-feature windows. A key methodological finding was the identification of a reading-pace confound: models trained with sentence-duration features achieved inflated performance (mean XGBoost ROC AUC = 0.643), which fell substantially when those features were removed (0.567), confirming that apparent signal was partly behavioural rather than neural. On pure EEG features, three of ten participants showed genuine above-chance confusion detection (best case: XGBoost balanced accuracy = 0.791, AUC = 0.763), providing proof-of-concept that consumer EEG can capture confusion-related neural signals. However, performance did not generalise uniformly across individuals, underscoring the necessity of per-user calibration. Following the EEG classification phase, participants completed a simulation of the BCI-assisted reading tool and a post-simulation survey. All participants rated automated confusion-triggered explanations as very helpful, and qualitative note-taking behaviour revealed three recurring study strategies: vocabulary tracking, conceptual unpacking, and system agreement-checking. Together, these findings suggest that low-cost EEG-based confusion detection is feasible for some individuals and that stu-

dents have a clear preference for systems that not only detect confusion but actively support its resolution through contextual explanations and linked resources.

8. **Kia Khademi:** This thesis examines how large-scale interaction data from gamified computer science learning platforms can be used to evaluate and improve assessment quality. Unlike traditional single-attempt exams, gamified systems involve repeated attempts and feedback cycles, which complicate the interpretation of classical item metrics such as difficulty and discrimination. The study addresses the lack of integrated analysis connecting psychometric measures, question design features, and cognitive frameworks such as Bloom's taxonomy. Using data from a gamified educational platform, the research analyzes item difficulty, discrimination index, distractor efficiency, and Bloom-level classification based on the ACM's computing-specific taxonomy. The study is guided by three research questions: (RQ1) What are the general characteristics of the question bank in terms of item difficulty, discrimination, distractor performance, and Bloom classification? (RQ2) How do difficulty, discrimination, distractor efficiency, and inferred Bloom's taxonomy levels correlate with one another? (RQ3) Do different types of questions exhibit differential performance patterns across student subpopulations? The goal is to identify relationships among these measures and assess how well cognitive classifications align with observed student performance. The findings aim to support instructors in identifying ineffective questions, improving distractor design, and making more informed, data-driven assessment decisions in computing education.

9. **Huynh Ngo Vu:** Automated extraction of meaningful road features, such as road centerlines and boundaries, from raw survey point clouds remains a significant challenge in geomatics. Most conventional methods rely on clear geometric cues, such as curbs and height discontinuities, or rigorous assumptions about road structure, often supplemented by intensity data or manual annotations. While effective in controlled environments, these approaches lack robustness when faced with noisy, unlabeled data exhibiting variable point density, occlusions, or ambiguous boundaries—conditions common in real-world survey datasets.

   This thesis proposes a fully unsupervised, density-driven pipeline that derives road features exclusively from the spatial distribution of X, Y, Z coordinates, without requiring labels, intensity values, or predefined geometric models. By leveraging an automated HDBSCAN parameter sweep to establish consensus clusters, the method robustly identifies linear high-density regions corresponding to road surfaces amid outliers and noise. Subsequent K-means clustering, minimum spanning tree construction, and principal point offsetting generate smooth, parallel polylines that closely approximate expert-drawn annotations.

   Experiments on industry-provided survey datasets demonstrate high coverage, effective outlier rejection, and computational efficiency, offering a practical, assumption-light alternative for automated feature extraction in challenging surveying contexts.

10. **Prina Mehta:** Existing LLM benchmarks conflate task difficulty with cognitive complexity, obscuring whether models can support the full range of thinking that programming education demands. We evaluate six LLMs across a 3×6 factorial design crossing three difficulty levels with Bloom's Revised Taxonomy's six cognitive levels, using 4,476 questions spanning mathematics, science, reading comprehension, and computer science (26,856 scored responses, scored 0–10 where higher

scores indicate better performance). A key methodological strength is human validation: judge scores were validated on 378 stratified responses (80.4% within-one-point agreement between human raters and the automated judge; inter-rater $\kappa = 0.719$), and Bloom classifications were separately validated on 354 questions (κ = 0.774, 83.6% human-classifier agreement). All models peak at Apply (L3), but five of six decline sharply at Higher-Order levels, with smaller open-source models showing the most severe drop. Bloom's level explains substantially more performance variance than difficulty (η^2 = 0.189 vs. η^2 = 0.059). Rubric-based scoring aligned to learning objectives is more informative than surface-level lexical metrics, and tool selection for classroom use should be matched to the cognitive demands of each task.

11. **Aliasgar Sakarwala:** Existing robo-advisors rely on questionnaires to assign clients to rigid portfolio templates and evaluate recommendations against proxy risk metrics rather than the outcomes clients actually care about. This thesis devel- ops a retirement-oriented robo-advisor that learns from real client holdings and evaluates recommended transitions against each client's probability of reaching their retirement goal. The robo-advisor was built on proprietary data from a Canadian wealth management firm covering approximately 13,000 clients and 22,000 accounts. First, assets are clustered into risk tiers from historical price data, then a LightFM hybrid recommende sys- tem learns personalized asset rankings from weighted holdings data. Next, a portfolio-aware reranking layer balances relevance against directional risk alignment and a Monte Carlo projection engine evaluates transitions by their improvement in retirement success probability. Our main finding is that 53% of transitions classified as successful under a proxy-risk criterion fail to improve retirement success probability by a meaningful margin. This calls into question the sufficiency of proxy-based evaluation that is standard in the robo-advisory literature. Further, we found that advisor-assigned and holdings-based risk agree exactly in only 40.2% of accounts, which con- firms that neither measure alone is a reliable ground truth. The hybrid recommender achieves Recall@10 of 0.975 and MRR of 0.944, substantially outperforming all baselines and demonstrating that collaborative patterns in real holdings data carry a strong predictive signal. Transition policy struc- ture matters more than reranker tuning: sell-and-reallocate policies produce meaningful retirement improvements that add-only policies cannot.

12. **Mohamed Sakr:** Recent advancements in personalized mobile health (mHealth) span classical machine learning, Large Language Model (LLM) integration, and multi-agent architectures. However, these paradigms exhibit persistent structural gaps, including reliance on self-reported data, isolated pipeline execution without shared physiological state, static user targets, and a lack of daily adaptivity. To address these limitations, we introduce Sina, a novel multi-agent system that synthesizes and extends foundational models from prior mHealth literature into a single, coordinated closed-loop architecture. Sina operates via four integrated components: a Supervisor, a Data Science Agent (DSA), a Nutrition Planning Agent (NPA), and an Exercise Planning Agent (EPA). The Supervisor adapts Google's Personal Health Agent (PHA) framework into a robust orchestration layer featuring deterministic processing pipelines and typed inter-agent data contracts, ensuring users receive seamlessly coordinated, holistic health interventions. Coordinated by this layer, the DSA extends Google's PHIA framework to analyze live wearable data via ReAct and sandboxed Python execution, providing users with dynamically

updated macronutrient targets, daily readiness scores, and continuous physiological state summaries and analysis. The NPA actualizes a previously unrealized Retrieval-Augmented Generation (RAG) architecture over the USDA FoodData Central database, delivering validated, highly personalized daily meal plans based on DSA-computed metabolic targets rather than static user inputs. Concurrently, the EPA enhances the conversational PlanFitting framework by integrating daily, wearable-grounded readiness scores, supplying users with structured, conversational exercise schedules and coping plans that adapt day-to-day based on their recovery state. Unified by a continuous synchronization loop that dynamically adjusts both dietary and exercise artifacts based on live physiological signals, Sina establishes the first production-ready mHealth system to successfully bridge orchestration, health data analysis, nutritional grounding, and conversational exercise planning grounded on user physiological data into a fully adaptive multi-agent ecosystem.

13. **Tahsin Jawwad:** This thesis develops a SQL-to-MongoDB translation algorithm using a two-stage Text-to-NoSQL pipeline. The translator, embedded in the UnityJDBC middleware, converts SQL produced by Text-to-SQL models into MongoDB Query Language (MQL) executable against a MongoDB instance. This work extends the baseline translator, which handled only simple SELECT–WHERE–LIMIT queries, with native support for GROUP BY, HAVING, DISTINCT combined with ORDER BY, set operations (UNION, INTERSECT, EXCEPT), IN/NOT IN with correlated subqueries, scalar subqueries, arithmetic and string-concatenation expressions in projections, and JOIN support with correct field resolution. Cross-cutting fixes address case-insensitive schema lookup, cross-branch table resolution for set operations, and a planner NPE, together with a SQL preprocessing pass that normalises double-quoted string literals produced by the Text-to-SQL model. On the TEND benchmark of 2,775 SQL–MQL pairs derived from Spider and BIRD, using SQL predictions from DAIL-SQL, these changes reduced translation failures from 1021 to 82, a 92% reduction, and improved the translation success rate from 63.2% to 97.0% with zero regressions. Execution accuracy on the same benchmark rose from 18.0% to 40.2%, nearly four times the 10.8% reported for the rule-based Grammar Converter in Lu et al. and within 4.5 percentage points of their LLM-based SQL-to-MQL converter, at a fraction of the inference cost.

14. **Samira Almuallim:** Large Language Models (LLMs) present a promising approach for automating dialogue rewriting in video game systems, enabling scalable content generation and adaptation for different audiences. However, their reliability remains highly sensitive to prompt design and the amount of conversational context provided. This thesis investigates how prompt engineering strategies and context scaling affect the structural reliability, lexical complexity, and overall quality of LLM-generated dialogue rewrites. We introduce a multi-metric evaluation framework tailored to dialogue rewriting tasks, measuring hallucination rate, readability (via Dale-Chall approximation), output size, formatting reliability, and textual uniqueness. Using this framework, we conduct experiments across two model scales, comparing a smaller local model (Phi-2) and a larger model (GPT-4o), under varying prompt schemas and context lengths. Results show that prompt structure has a significant impact on output quality, with single-line prompts combined with explicit output constraints yielding the most reliable rewrites. In contrast, multiline and dialogue level prompts increase hallucination, formatting errors, and verbatim copying. Context scaling further reveals divergent behavior across models:

while the larger model maintains high performance, the smaller model exhibits rapid degradation as context increases, often failing to perform meaningful rewriting beyond minimal context. Additionally, we analyze the extent to which prompt wording can control lexical complexity and observe patterns of unintended lexical drift in neutral prompts. Based on these findings, we provide practical guidelines for designing stable and controllable LLM-based dialogue rewriting pipelines. Overall, this work highlights the limitations of current LLMs in constrained rewriting tasks and emphasizes the importance of prompt and context design for reliable deployment in interactive media systems.

15. **Abdullah Munir:** Many Plains Cree language materials remain available only as scanned archival pages, which limits their use for search, correction, language learning, and future digital tools. Optical Character Recognition (OCR) can help convert these images into machine-readable text, but Plains Cree syllabics presents challenges that differ from high-resource OCR settings: the script is specialized, labelled data are limited, and historical scans often contain noise, degradation, and irregular print quality. This project focuses on the detection stage of the OCR pipeline, where individual syllabic characters must be located before they can be recognized. Building on earlier OCreeR work showing that isolated syllabic recognition performs strongly once characters are available, this project adapts CRAFT, a character-aware text detection model, to Plains Cree syllabics. Manually annotated page images are converted into region and affinity supervision maps so the detector can learn character locations and relationships between neighbouring characters. These results establish a detection baseline for future Plains Cree OCR development.

2 Parallel Session II (Computer Science) – SCI 247

2.1 Names of Speakers

1. Priyansh Mathur
2. Ahmad Memon
3. Ronit Buti
4. Shlok Shah
5. Erem Ozdemir
6. Bill Vo
7. Sammie Scully
8. Salma Vikha Ainindita
9. Anuk Ahangamgoda
10. Syed Saad Ali
11. Shuyu Yan
12. Stavan Shah
13. Luke Yin
14. Regan Nguyen and Aaditya Golash

2.2 Location

- SCI 247

2.3 Zoom Link

- https://ubc.zoom.us/j/68411277870?pwd=Mam7UBra3TLOKbAg3hau7FTbT0viJh.1

2.4 Length of Each Talk

- 16 minutes followed by a 4-minute question period.

2.5 Schedule and Titles

Welcome (SCI 247): 8:30 am – 9:00 am

Session 1 (SCI 247): 9:00 am – 10:00 am
Chair: Abdallah Mohamed

- 9:00 am – 9:20 am: **Priyansh Mathur**
  - ⋆ Title: Can We Trust LLM Answers on Algorithm Problems? Evaluating Algorithmic Reasoning Across Domains Through Disagreement Driven Refinement
  - ⋆ Supervisor: Abdallah Mohamed
- 9:20 am – 9:40 am: **Ahmad Memon**
  - ⋆ Title: From AI Grading to Trustworthy AI-Generated Rubrics in CS1 Assessment
  - ⋆ Supervisors: Abdallah Mohamed
- 9:40 am – 10:00 am: **Ronit Buti**
  - ⋆ Title: Evaluating RAG-Based Course Advising Chatbots Using LLM-as-a-Judge Methodology
  - ⋆ Supervisor: Abdallah Mohamed

Break: 10:00 am – 10:15 am

Session 2 (SCI 247): 10:15 am – 11:15 am

Chair: Shan Du

- 10:15 am – 10:35 am: **Shlok Shah**
  - ⋆ Title: Notional Machines, Not Just Errors: Toward Belief Attribution with Instructor-Facing LLMs
  - ⋆ Supervisor: Abdallah Mohamed
- 10:35 am – 10:55 am: **Erem Ozdemir**
  - ⋆ Title: Digitizing Neuropsychological Memory Assessments with Speech Processing AI Models
  - ⋆ Supervisors: Shan Du and Barrett Ens
- 10:55 am – 11:15 am: **Bill Vo**
  - ⋆ Title: Enhancing RSGPT with Custom Dataset Annotation and Cross-Domain Adaptation for Remote Sensing Tasks
  - ⋆ Supervisor: Shan Du

Break: 11:15 am – 11:30 am

Session 3 (SCI 247): 11:30 am – 12:30 pm

Chair: Ifeoma Adaji

- 11:30 am – 11:50 am: **Sammie Scully**
  - ⋆ Title: Design and Development of a Customized Foster Placement Management System
  - ⋆ Supervisor: Ifeoma Adaji
- 11:50 am – 12:10 pm: **Salma Vikha Ainindita**
  - ⋆ Title: Exploring AI Startup Reviews with Sentiment, Emotion, and Topic Modeling Through an Interactive Dashboard
  - ⋆ Supervisor: Ifeoma Adaji
- 12:10 pm – 12:30 pm: **Anuk Ahangamgoda**
  - ⋆ Title: But The Bot Told Me! Proposing a Benchmark Concept to Evaluate the Reliability of Large-Language Models as a Tool for Day-to-Day Knowledge Gathering
  - ⋆ Supervisor: Ifeoma Adaji

Lunch (SCI 337): 12:30 pm – 1:30 pm, we will offer the lunch (pizzas)

Session 4 (SCI 247): 1:30 pm – 2:30 pm

Chair: Abdallah Mohamed

- 1:30 pm – 1:50 pm: **Syed Saad Ali**
  - ⋆ Title: AI-Assisted Assessment Management for Equivalent Exam Variants
  - ⋆ Supervisors: Abdallah Mohamed and Mostafa Mohamed
- 1:50 pm – 2:10 pm: **Shuyu Yan**
  - ⋆ Title: Identifying Learner Attributes for Cold-Start Personalization via AI-Based Student Simulation
  - ⋆ Supervisors: Abdallah Mohamed and Mostafa Mohamed
- 2:10 pm – 2:30 pm: **Stavan Shah**
  - ⋆ Title: Answer Leakage in Multi-Turn AI Tutoring for Computing Education
  - ⋆ Supervisors: Abdallah Mohamed and Mostafa Mohamed

Break: 2:30 pm – 2:45 pm

Session 5 (SCI 247): 2:45 pm – 3:25 pm
Chair: Ifeoma Adaji

- 2:45 pm – 3:05 pm: **Luke Yin**
  - ⋆ Title: Accessible Deepfake Detection on Everyday Devices
  - ⋆ Supervisor: Ifeoma Adaji
- 3:05 pm – 3:25 pm: **Regan Nguyen and Aaditya Golash**
  - ⋆ Title: Human Eye Tracking For Application Usability
  - ⋆ Supervisor: Ifeoma Adaji

Prize Draw and Closing/Party (SCI 337): 4:00 pm – 5:00 pm

2.6 Abstracts

1. **Priyansh Mathur:** Large language models are increasingly used by students in computer science courses as learning aids for understanding algorithmic concepts, including selecting appropriate algorithms under changing constraints and reasoning about time complexity. For these tools to support learning effectively, their responses must remain reliable not only on familiar textbook problems but also when problem structure changes in ways that alter the correct solution. Despite this need, there is still limited understanding of how consistently LLM reasoning holds across algorithmic settings and whether apparent failures reflect genuine reasoning limitations or ambiguity in problem formulation. This honors thesis addresses three research questions: how LLM reasoning accuracy changes across algorithmic domains, whether models perform differently on algorithm selection and time complexity deduction, and to what extent disagreement driven refinement can help distinguish true reasoning failures from underspecified constraints. The study was conducted in two phases. The first phase evaluated five LLMs, ChatGPT 5, Gemini 2.5 Pro, Claude 4 Sonnet, Grok, and DeepSeek V3, on a benchmark of 20

sorting problems presented in both baseline and structurally varied forms. In this phase, models were generally more reliable at selecting the correct sorting algorithm than at deducing its time complexity, and several showed reduced stabilities when standard constraints were rephrased or altered. These findings suggested that correct answers in sorting did not always reflect robust reasoning, motivating a broader second phase. The second phase extended the analysis beyond sorting by adding searching and graph problems, producing a 33-problem benchmark for cross domain evaluation. Across the expanded study, performance remained relatively strong in linear domains but declined sharply on graph problems, where no model exceeded 63% accuracy. When models disagreed, a disagreement driven refinement process was used to identify ambiguous prompt elements and generate clarifications reviewed by a human with computer science knowledge before reevaluation. Refinement improved performance across all five models, with the largest gain observed for DeepSeek V3, which increased from 73% to 87%, and ChatGPT 5 rising from 82% to 93%. These findings suggest that many apparent reasoning failures in algorithmic settings are partly communication failures caused by underspecified constraints rather than purely computational limitations, highlighting the importance of linguistic precision in the responsible use and evaluation of LLMs in computer science education.

2. **Ahmad Memon:** Assessment in introductory programming (CS1) demands significant instructor effort, from evaluating diverse student code to designing rubrics that support fair and consistent grading. This presentation summarizes a progression of research exploring how generative AI can support these assessment tasks while preserving reliability and instructor oversight. We first investigate the use of large language models for grading CS1 code, comparing AI-generated grades with human evaluation and examining how rubric structure influences grading performance. Building on the finding that rubric quality is central to reliable AI grading, we then explore AI-generated rubric drafting as a way to reduce instructor workload. This leads to our current work on evaluating when AI-generated rubrics can be trusted, using an alignment-based framework that compares generated rubrics with human reference rubrics and applies threshold-based escalation for human review when needed. Together, these studies trace a shift from asking whether AI can assist with grading toward how AI outputs can be deployed responsibly in practice. The work contributes toward reliable human-in-the-loop AI support for CS1 assessment and highlights future directions for scalable educational assessment tools.

3. **Ronit Buti:** Retrieval-Augmented Generation (RAG) chatbots are increasingly deployed for academic advising, yet it remains unclear whether these systems can provide reliable, hallucination-free guidance. We investigate this question by evaluating a production course advising chatbot using LLM-as-a-judge methodology with Gemini 3 Flash. To enable systematic assessment, we develop AdvisingBench, a benchmark comprising 55 expert-evaluated questions across three complexity categories (common, specific and edge cases). We evaluate the system on four metrics: answer relevancy, faithfulness, context relevancy and answer correctness. Our results reveal performance variation across query categories, with edge cases showing 6% higher overall scores than common queries despite lower context relevancy. We find that context relevancy is the metric most sensitive to query complexity, while answer relevancy remains consistently high (¿0.95) across all categories. Notably, hallucination failures were rare (3 of 53 cases), suggesting

that RAG architectures substantially constrain models to retrieved content in academic advising contexts, though they do not eliminate fabrication risk. This work provides initial empirical evidence on whether AI can support course advising, and offers a methodology and benchmark for future validation studies.

4. **Shlok Shah:** Explaining a bug is a straightforward task; however, the true labor of tutoring lies in understanding why a student wrote it. In CS1, the better instructional signal is often a student's mental model of how programs execute, rather than surface correctness alone. Reasoning about these beliefs can reveal where instruction breaks down, but it demands epistemic restraint: separating what the code structurally supports from higher-level inferences about student intent. Moreover, belief attribution is risk-asymmetric: falsely diagnosing a misconception is significantly more harmful than abstaining.

   As large language models (LLMs) enter CS1 classrooms, we argue that their primary opportunity lies not in automated fixing but in supporting instructors in reasoning about plausible student beliefs expressed through code. LLMs should be treated as instructor-facing hypothesis generators that help instructors understand students' mental models, beyond syntax and logic errors.

   To ground this position, we report evidence from TRACER, a controlled testbed for evaluating belief attribution. In 5-fold cross-validation on 1,200 synthetic CS1 Java submissions spanning 18 notional machines, models achieve recall of 0.87 and specificity of0.85, with a majority of false positives driven by over-diagnosis on correct programs. A label-inclusive matching strategy, by comparison, saturates recall (0.98) while degrading specificity (0.77), illustrating how evaluation shortcuts can inflate impressions of capability while worsening safety. Consequently, we argue that belief-oriented LLM support is a worth while research direction, but it requires standards that prioritize diagnostic humility over broad coverage.

5. **Erem Ozdemir:** Neuropsychological assessments such as the Selective Reminding Task (SRT) and the Auditory Consonant Trigram task (CCC) are clinically important for evaluating verbal learning, memory, and attention in the assessment of dementia and related cognitive disorders. However, their paper-based administration is time-intensive and difficult to standardize. This thesis investigates whether speech-processing AI can support their digitization through a local browser-based platform that standardizes stimulus delivery, records spoken responses, and automates scoring. The project was conducted in collaboration with an ongoing normative study and used two main data sources, including a clinical survey dataset with 247 audio recordings from eight participants and a filtered subset of 4,823 recordings drawn from the Mozilla Common Voice corpus. Using these data, this thesis developed a local browser-based platform for standardized SRT and CCC administration, provided a comparative evaluation of automatic speech recognition (ASR), signal processing, and keyword spotting (KWS) approaches for this clinical setting, and implemented an MFCC- and DTW-based quality-assessment pipeline. A KWS proved best suited to SRT and CCC scoring than either large-scale ASR models or standalone signal-processing methods, achieving 94.3% accuracy on participant SRT recordings, compared with 57.5% for the best Whisper model and 25.5 to 31% for Granite Speech. Although the work remains a proof of concept due to the small participant sample and the absence of a KWS-based CCC model, the results show that digitizing the SRT and CCC is both technically feasible and worth further study. More broadly, the thesis demonstrates that successful au-

tomation in this setting depends less on model scale than on choosing methods whose objective closely matches the structure of the neuropsychological task.

6. **Bill Vo:** Remote Sensing Image Captioning (RSIC) addresses the challenge of automatically generating natural language descriptions for aerial and satellite imagery, supporting applications in disaster response, urban planning, and environmental monitoring. Unlike general-domain captioning, RSIC must contend with top-down viewpoints, heterogeneous object distributions, sparse annotations, and domain-specific vocabulary. RSGPT, proposed by Hu et al. (2023, 2025), represents a state-of-the-art approach that adapts InstructBLIP — a vision-language model with instruction tuning — to the remote sensing domain through fine-tuning on the RSICap dataset.

   This thesis develops the work in three phases.

   Phase 1 reproduces RSGPT on the RSICD benchmark, achieving approximately 77.6% mean fidelity across BLEU-1/2/3/4, ROUGE-L, and CIDEr using a single A100 GPU — one-eighth the compute reported in the original work. The effort also fills several reproducibility gaps left open by the original release, most notably the absence of an evaluation pipeline in the published code.

   Phase 2 builds a region-guided extension for grounded paragraph-level captioning on RSITMD. This phase contributes 23,715 manually annotated bounding boxes across the 4,743 images of RSITMD and proposes a novel bounding-box attention-mask mechanism that conditions Q-Former's cross-attention on patch-level masks derived from the bounding boxes. The formulation preserves the full spatial context of the image, requires no modifications to the underlying Q-Former, and processes multiple regions through a single vision-encoder pass. Three new modules are introduced — a SequentialRegionProcessor, a CrossRegionTransformer, and a WeightedFusionModule — producing an interpretable softmax over region importance. An initial five-region design in this phase exposed a limitation for scene-level captioning: because every region saw only a masked subset of the image, no representation ever encoded the full spatial layout, leaving relational descriptions such as "near the highway" or "on the east side" ungroundable.

   Phase 3 addresses this limitation through a six-region architecture developed in collaboration with the project's postdoc. The full image is treated as an additional region with an all-ones mask, so scene context and object-specific detail are captured jointly before being merged by the fusion module.

   Beyond the architecture and dataset, this work contributes a reproducibility audit of design choices left underspecified in the original RSGPT publication

7. **Sammie Scully:** Small volunteer-run organizations are frequently under-examined in computing research, despite operating under significant administrative complexity with limited technical resources. This presentation reports on the design and development of a full-stack foster placement management system built in collaboration with a local dog rescue organization in Kelowna, BC. The work is motivated by a gap in the literature: while structured digital tools have demonstrated measurable operational benefits in large nonprofits and for-profit organizations, purpose-built systems tailored to the specific workflows of small, volunteer-run rescues remain rare and underexplored.

   The system was developed using Flask, SQLAlchemy, and MySQL, deployed on DigitalOcean App Platform, and designed around a five-tier architecture support-

ing four user roles. Core features include foster profile management with capacity and preference tracking, away block and coverage scheduling, a coordinator-facing calendar view, adoption preference matching, and an in-system and email notification system. The data model was structured to generalize beyond the reference organization.

The presentation covers the research context and literature review, system architecture and design decisions, key feature implementation, testing strategy, deployment, and feedback gathered from the partner organization. Limitations of the current system and directions for future formal evaluation are also discussed.

8. **Salma Vikha Ainindita:** Artificial intelligence (AI) startups increasingly market their products as "AI-powered," yet there is limited systematic understanding of how users perceive these products at scale. This thesis addresses that gap by analyzing approximately 33,000 English-language reviews from 46 AI startups across four AI startup clusters and four public review platforms. The study combines sentiment analysis, emotion modelling, and topic modelling in an interactive dashboard that supports accessible exploration of review insights. Results show that deep learning models consistently outperformed traditional baselines. RoBERTa achieved the strongest performance for sentiment analysis (F1 = 0.970). For emotion modelling, DistilBERT performed best in both valence-arousal prediction (R^2 = 0.629–0.824) and discrete emotion prediction (R^2 = 0.748). BERTopic produced the most coherent topics overall (mean cluster-level coherence = 0.676). Beyond overall trends, the study provides cluster-specific insights and implications. It shows how different types of AI startups vary across three dimensions: product use cases, customer-related concerns, and product quality and trust.

9. **Anuk Ahangamgoda:** While there are several comprehensive benchmarks that evaluate the performance of Large Language Models (LLMs) on specific tasks and domains of knowledge, there remains a lack of a benchmark to measure the reliability of LLMs as a tool for knowledge-gathering across non-professional, everyday domains. This paper therefore proposes a conceptual benchmark that can evaluate an LLM's ability to answer objective questions reliably within this mundane context. By crafting a novel dataset of prompts across eight distinct knowledge domains and three knowledge classes, the output of an LLM is quantitatively analyzed with respect to key reliability metrics. A secondary dataset, created by converting the prompts into non-sensical counterparts, is then used for further evaluation. A reliability matrix and overall reliability score is then given to each tested LLM to create a benchmarking leaderboard. The results were able to identify an LLM's overall reliability as a tool for mundane knowledge gathering as well as the reliability within specific knowledge domains and classes to highlight. Further, benchmarking variations of the same model found that a smaller parameter count and edge-deployment variations were less reliable than their counterparts, indicating a positive correlation between model size and reliability.

10. **Syed Saad Ali:** When students take parallel versions of an exam, fairness depends on each form targeting the same constructs at a comparable level of difficulty. In this paper, a question variant is an item intended to substitute for a reference item: different enough to avoid duplication, but aligned in topic, expected reasoning, and instructional purpose. Producing such variants manually is possible for small assessments, yet difficult to maintain consistently at exam scale. We investigate whether structured question metadata and an instructor-mediated AI workflow can

support full-exam variant construction while preserving rubric-level alignment with a reference form. Reference items are annotated with topic, cognitive level, difficulty band, format, and constraints, then used to guide slot-matched LLM generation. A separate LLM-as-judge rubric scores candidate items on construct and difficulty alignment, structural validity, answer correctness, topic alignment, and distinctness. Because automated judging of automated output has known limitations, independent instructor review is required before content is committed or exported. We report a case study using a first-year programming exam in which two full variants were generated and evaluated. LLM-judge results indicate strong alignment with the reference exam on construct and topical dimensions, while instructor review confirms that the most common issues are limited distinctness and occasional minor edits rather than pervasive conceptual drift. Findings are bounded to one course context and text-only items. The workflow offers a practical approach to scaling variant authoring while keeping instructors in control.

11. **Shuyu Yan:** Before cold-start adaptive systems can support interaction and collaboration in engineering and computing education, we need to establish which learner attributes are worth representing. This paper investigates that question using an LLM-based simulation pipeline across ten mathematics topics as a controlled proxy for structured STEM problem solving. Mathematics is chosen for its explicit solution structure, interpretable difficulty progression, and objective correctness — properties that enable rigorous analysis of early learning behaviour, with findings interpreted for computing education, where cold-start challenges arise but ground-truth trajectories are harder to isolate.

    We operationalize four attributes at five levels each: learning rate, slip propensity, and problem-solving strategy — linked to performance differences in prior learner-modelling research — and a demographic descriptor (gender-related expression), included to examine whether demographic attributes should drive personalisation decisions at all. We evaluate simulated behaviour using learning curves, accuracy-by-difficulty trends, reasoning-step distributions, time to mastery, and rank-order correlations, and compare these with a real computer science student dataset as illustrative grounding.

    Results show clear stratification for learning rate, slip propensity, and problem-solving strategy, while the demographic descriptor remains nearly invariant. These findings identify which learner descriptors are plausible early priors for cold-start decisions, indicate which should not drive personalisation, and suggest that simulation can help educators stress-test adaptive assumptions before deployment — offering transferable principles for collaborative and interactive learning under uncertainty.

12. **Stavan Shah:** AI tutors can fail educationally even when their explanations are fluent and technically plausible: under sustained help-seeking pressure, the tutor may reveal enough information that the learner no longer has to make the final inference. We study this failure mode as answer leakage in synthetic, six-turn tutoring conversations over computing multiple-choice questions. Using 8,538 automatedjudge-labeled conversations over 1,423 questions from CSBench and PeerWise, we ask when leakage emerges, whether it is explained by the tutor model's direct-answer accuracy on the same item, how question source changes the observed risk profile, and how much a lightweight review loop reduces student-visible leakage. In single-tutor mode, leakage accumulated over turns rather than

appearing only in the first response. GPT leakage reached 15.8% on CSBench and 22.5% on PeerWise; Gemini leakage was lower, at 6.1% and 6.4%. Direct-answer accuracy was 22.5–23.5 percentage points lower on PeerWise than on CSBench, but correctness-conditioned leakage gaps did not show a stable pattern. Review loops reduced leakage in every matched tutor-source comparison, with absolute reductions from 3.4 to 20.4 percentage points and average revision iterations between 1.02 and 1.22 per supervised turn. Because labels are automated, we report these as automated-judge-labeled patterns, not validated human leakage rates. The contribution is a compact evaluation frame for testing whether computing tutors preserve the learner's final inferential move across interaction.

13. **Luke Yin:** The proliferation of AI-generated synthetic media poses growing risks to information integrity and public trust. Most existing deepfake detectors are evaluated under controlled conditions and fail under real-world distribution shifts-particularly face restoration post-processing, which can suppress the very artifacts detectors rely upon. This work presents a lightweight deepfake detection system deployable on consumer devices, targeting two dominant manipulation categories: face-swap artifacts and diffusion-based AI synthesis. We fine-tune a Self-Blended Images (SBI) model on a restoration-enriched dataset to address catastrophic performance degradation caused by GFPGANv1.3 face restoration, recovering AUC from 0.75 to 0.97 on an in-domain mixed test set. For AI-synthesized image detection, we replace the ResNet-50 backbone in DistilDIRE with a ConvNeXt-Base architecture pre-trained with CLIP, improving accuracy on an in-domain split of the Deepfake-Eval-2024 benchmark from 61% to 87%. GPT-5.4 Vision is evaluated as a zero-shot comparative reference using logprob-based confidence calibration; results indicate inconsistent outputs across images, confirming that lightweight specialized detectors remain more reliable for public-facing deployment. The system is delivered as a web application and Android app, demonstrating practical accessibility without specialized hardware.

14. **Regan Nguyen and Aaditya Golash:** Evaluating mobile applications through surveys can be unreliable because user responses may be affected by bias, incomplete recall, or forgetfulness. Eye-tracking glasses provide a more direct way to study user attention and interaction by capturing where users look in real time. However, eye-tracking recordings generate large amounts of data that can be difficult and overwhelming to interpret. This directed studies project explores how to make Tobii Pro Glasses 3 eye-tracking data easier to analyze and understand for research use. The project improves an interactive dashboard for uploading, processing, and exploring eye-tracking recordings within a single system. The dashboard brings together gaze visualizations, heatmaps, fixation shots, confusion-related metrics, and pupillometry outputs, while also adding features that improve usability and support easier deployment. By making eye-tracking analysis more organized and accessible, this work helps researchers use rich behavioral data more effectively when evaluating user experience.

3 Parallel Session III (Math, Phys, Stat/Data) – SCI 333

3.1 Names of Speakers

1. Michael Burns
2. Emma Hendry
3. Ananda Thomlinson
4. Karim Akhitamov
5. MacKenzie Richards
6. Justin Li
7. Colin Haggarty
8. Margaret Angtuaco
9. Shengbo Lin
10. Medha Naidu
11. Aimee Menard
12. Rachel Brookes
13. Slade Coffman

3.2 Location

- SCI 333

3.3 Zoom Link

- https://ubc.zoom.us/j/61900526719?pwd=c9qcNzbruwtpEeaYzeWrlmTC5Us657.1
- Meeting ID: 619 0052 6719
- Passcode: 0

3.4 Length of Each Talk

- 16 minutes followed by a 4-minute question period.

3.5 Schedule and Titles

Welcome (SCI 333): 8:30 am – 8:40 am

Session 1 (SCI 333): 8:40 am – 10:00 am, PHYS and MATH
Chair: Jake Bobowski

- 8:40 am – 9:00 am: **Michael Burns**
  - ⋆ Title: Custom Fabry-Perot transmission lines: A quantitative testbed for studying RF Plasmas
  - ⋆ Supervisors: Jake Bobowski
- 9:00 am – 9:20 am: **Emma Hendry**
  - ⋆ Title: Data Wrangling for Machine Learning in Personalized Radiation Oncology
  - ⋆ Supervisor: Andrew Jirasek
- 9:20 am – 9:40 am: **Ananda Thomlinson**
  - ⋆ Title: Exploring the Effects of the Higgs Mechanism for New Gauge Forces
  - ⋆ Supervisor: John Hopkinson, Andrey Shkerin and Marco Costa

- 9:40 am – 10:00 am: **Karim Akhitamov**
  - ⋆ Title: Dealing with simulation failures in blackbox optimization
  - ⋆ Supervisor: Warren Hare

Break: 10:00 am – 10:15 am

Session 2 (SCI 333): 10:15 am – 11:15 am, MATH

Chair: Paul Tsopméné

- 10:15 am – 10:35 am: **MacKenzie Richards**
  - ⋆ Title: The Ideal Class Group
  - ⋆ Supervisor: Chad Davis
- 10:35 am – 10:55 am: **Justin Li**
  - ⋆ Title: Manifold Calculus of Functors and Classification of Homogeneous Functors
  - ⋆ Supervisor: Paul Tsopméné
- 10:55 am – 11:15 am: **Colin Haggarty**
  - ⋆ Title: Kähler Manifolds and the Geometry of Projective Varieties
  - ⋆ Supervisor: Sylvie Desjardins

Break: 11:15 am – 11:30 am

Session 3 (SCI 333): 11:30 am – 12:30 pm, STAT/DATA

Chair: Lengyi Han

- 11:30 am – 11:50 am: **Margaret Angtuaco**
  - ⋆ Title: Mixed-Effects Model-Based Control Charts For The Trail-Making Test
  - ⋆ Supervisor: Lengyi Han
- 11:50 am – 12:10 pm: **Shengbo Lin**
  - ⋆ Title: Hidden Seasonal and Nonseasonal Minification Processes
  - ⋆ Supervisor: Lengyi Han
- 12:10 pm – 12:30 pm: **Medha Naidu**
  - ⋆ Title: The Linearity Trap - Curvature Based Detection of Nonlinearity Under Measurement Error in BP Data
  - ⋆ Supervisor: Xiaoping Shi

Lunch (SCI 337): 12:30 pm – 1:30 pm, we will offer the lunch (pizzas)

Session 4 (SCI 333): 1:30 pm – 2:10 pm, PHYS

Chair: Alex Hill

- 1:30 pm – 1:50 pm: **Aimee Menard**
  - ⋆ Title: Arcminute-Resolution View of the Polarization Structure in the Fan Region near the W3/W4/W5 Complex
  - ⋆ Supervisor: Alex Hill
- 1:50 pm – 2:10 pm: **Rachel Brookes**
  - ⋆ Title: Modeling Instrumental Polarization of the John A. Galt Telescope
  - ⋆ Supervisors: Tim Robishaw and Alex Hill
- 2:10 pm – 2:30 pm: **Slade Coffman (mathematical physics)**
  - ⋆ Title: Topological Quantum Field Theory and Partition Functions
  - ⋆ Supervisor: Paul Tsopméné

Prize Draw and Closing/Party (SCI 337): 4:00 pm – 5:00 pm

3.6 Abstracts

1. **Michael Burns:** A custom air-dielectric coaxial transmission line with airtight seals and a transparent copper-mesh outer conductor was designed and constructed to ignite and study radio-frequency (RF) plasma. RF power was capacitively coupled into one end of the coax while the opposite end was terminated by an open circuit. This open termination establishes a region of high electric field strength that can sustain a stable ionized plasma. We measured the plasma ignition power as a function of the air pressure within the transmission line, producing an RF Paschen curve. At high pressure, strong scattering limits the electron mean free path and thus reduces the electrons' ability to ionize additional air particles. At low pressure, on the other hand, scattering events become infrequent and electrons are lost through collisions with the transmission line walls. As a result, the plasma ignition power is minimized at intermediate pressures and increases at both low and high pressures. We also developed a model that allows the plasma electric susceptibility and conductivity to be extracted from reflection-coefficient measurements made with a vector network analyzer (VNA). Our measurements show that both the plasma susceptibility and conductivity increase as the incident RF power is increased and as the air pressure is decreased from 900 to 200 mTorr.

2. **Emma Hendry:**
   Purpose: Radiotherapy (RT) is used to treat up to 50% of cancer patients in Canada, and current dosage and treatment options are based on population-level statistics from previous patient responses. However, immune and biochemical factors affect radiosensitivity and treatment response in both malignant and benign tissues. By analysing RT response through immune cell counts and biochemical abundances, treatment may be tailored to individual patients. Biochemical abundances can be measured using Raman spectroscopy, a label-free optical technique. This project focuses on two objectives: automating Raman spectra preprocessing to improve data quality and efficiency, and developing an image registration program to correlate histology, immunofluorescence (IF), and Raman spectra collection sites.

   Methods: Raman spectra preprocessing was automated using an R-based clustering model (PGMM). A subset of spectra was labeled as "good" or "bad" and used to

cluster all spectra in one-dimensional space. The automated program was compared to the manual approach based on user interaction time, spectra removed, and standard deviation of the final dataset. An image registration program was developed in MATLAB. Optimization-, correlation-, and feature-based methods were tested, and masking and thresholding were used to improve registration.

Results: The automated preprocessing was ∼20% faster and removed 8.54% more spectra. The manual program produced an STD of 3.06×10-4, while the automated program produced 2.85×10-4, indicating reduced noise and less user interaction time. Intensity-based registration produced the best results, as other methods failed. Image masking improved registration, resulting in a simple but effective program. Conclusion: Both objectives were successfully accomplished. Although the automated preprocessing removed more spectra, the 8.54% increase is unlikely to affect dataset utility, and the benefits in noise reduction and efficiency outweigh this drawback. The registration program aligns IF, histological, and Raman images. While quality could be improved with better masking and further automation, the program has fulfilled its basic purpose. Overall, these methods will facilitate data correlation for this project and future work within the lab group.

3. **Ananda Thomlinson:** New long-range forces that are mediated by massive vector bosons may arise from gauging the global symmetries of the Standard Model. Such theories include $U(1)_{B-L}$, $U(1)_{L_\mu - L_\tau}$, and $U(1)_{L_e - L_\mu}$, where $B - L$ is the difference between baryon and lepton numbers, and $L_i - L_j$ are the differences in electron (e), muon (μ), and tauon (τ) lepton numbers. We investigate gauge boson mass generation via the Higgs mechanism using a simple model consisting of a gauge field, a complex scalar Higgs field, and a charged source. The source is taken to be a static or rotating sphere of uniform charge density. The source models astrophysical bodies like the Earth or a neutron star, which have significant charge under $B-L$ or $L_i - L_i$ symmetries. Using the finite element method, we numerically solve the nonlinear, coupled partial differential equations for the gauge and Higgs fields. We analyze how the electric and magnetic components of the gauge field are modified in the presence of the dynamical Higgs field, for different parameters of the charged body. Above a critical charge, the Higgs field is found to exhibit a significant back-reaction, leading to electric field screening. For rotating sources, large charge currents drive Higgs field symmetry restoration, analogous to the destruction of condensate in superconductors. We conclude by pointing out the region where the existing constraints on the new gauge boson mass and coupling from fifth-force experiments must be revisited.

4. **Karim Akhitamov:** Blackbox optimisation (BBO) is the study of optimising a system whose inner workings are unknown. We study a situation where simulation failures occur in a batch of requested evaluations. Several algorithms are developed and compared under this condition. The work is ongoing; we present current progress and outline future directions.

5. **MacKenzie Richards:** Let K be a number field over Q of degree n. It is well known that the ring of integers of K, O_K, is not generally a unique factorization domain (UFD); that is, there is not generally an analogue of the Fundamental Theorem of Arithmetic in O_K. The ideal class group of K, Cl_K, gives a measure of how far O_K is from being a UFD. In this talk, we describe the ideal class group, calculate it explicitly with an example, and demonstrate how to prove, using the theory of lattices, one of the classical results in algebraic number theory: Cl_K is finite.

6. **Justin Li:** Manifold calculus of functors, developed by M. Weiss, aims to study the contravariant functors from the open sets category of a manifold to the category of topological spaces. Similar to the sequences of Taylor approximations of smooth functions, there exist sequences of "Taylor approximations" of functors. A functor that is equivalent to its kth approximation and has a trivial (k-1)th approximation is called homogeneous of degree k. In his foundational work on Manifold Calculus in 1999, Weiss classified homogeneous functors through fibrations over certain spaces. In 2025, Tsopméné and Stanley obtained a classification result of homogeneous functors into an arbitrary category through the construction of a complicated topological space $\hat{A}$. We studied the space $\hat{A}$ and showed that it is equivalent to a well known structure, and used this make a connection between the classification of Weiss and that of Tsopméné and Stanley. In this talk, we will go over the concepts of topology and category theory, explain the idea of manifold calculus of functors, and present our result. A possible use of our result is in the understanding of topological quantum field theory.

7. **Colin Haggarty:** This talk gives an overview of a rich connection between Riemannian geometry and algebraic geometry through the theory of Kähler manifolds. We begin with the language of differential forms on smooth manifolds, which unifies many classical results in vector calculus via the generalized Stokes' theorem. We then introduce Kähler manifolds, spaces that carry mutually compatible Riemannian, complex, and symplectic structures. Focusing on complex projective space, we explain why every smooth complex projective variety naturally inherits a Kähler structure. This correspondence underlies major open problems such as the Hodge conjecture.

8. **Margaret Angtuaco:** Aerial firefighting pilots operate under extreme conditions that induce fatigue, and an earlier attempt to use a wearable physiological monitor proved counterproductive. This motivates the use of the Trail-Making Test (TMT), a non-intrusive, repeated cognitive measure, to monitor a pilot's cognitive performance prior to flight assignments. Constructing a control chart for repeated-measures cognitive test data is complicated by non-normality, autocorrelation, a learning-driven non-stationary mean, and between-subject heterogeneity. To address this, a population-level control chart derived from the nonlinear mixed-effects (NLME) model for Part A and Part B of the TMT was developed. The data consist of 25 pilots with 50 attempts each for Part A and Part B (10 per day over 5 days). An NLME model is fitted to the reciprocal duration, a transformation applied to address skewness, capturing both the across-day and within-day learning while accounting for between-subject heterogeneity. The fitted model provides the centerline, control limits, and a model-driven burn-in period, which is the number of initial attempts during which the learning effect inflates the false-alarm rate (FAR) above a 1% threshold. The asymptotic mean duration is estimated at 9.09 seconds for Part A and 11.8 seconds for Part B, with population control limits at [6.13, 17.6] seconds for Part A and [7.32, 29.9] seconds for Part B. The model-driven burn-in recommendation is 7 attempts for Part A and 12 attempts for Part B at this threshold. Post-burn-in exceedance rates were 0.28% and 0.21%, respectively, which are consistent with the nominal one-sided FAR of 0.135% within Poisson sampling variation. The resulting framework provides an application of NLME-based control charts to neuropsychological data, enabling cognitive-performance monitoring for a new pilot without per-subject model fitting. Detection performance for fatigue-induced shifts is not characterized in this cohort and is identified as future work.

9. **Shengbo Lin:** The minification model is a discrete time, continuous state Markov process which can be applied to nonnegative time series. Properties of the exponential and tailed exponential models are reviewed, including stationarity, moments and autocorrelations. Two ways of extending the model are proposed. One extension is a seasonal model which could be applied to time series that are periodic. The other includes a multiplicative error which makes it a hidden Markov model. This extension can be applied to time series where the minification process is subject to excess noise. It is shown that the hidden minification model has a stationary process, and the moments and autocorrelations are presented. Parameter estimation can be carried out using moment-matching. An application to Fire Weather Index data from Northwestern Ontario is presented, including a comparison between the nonseasonal and seasonal models.

10. **Medha Naidu:** Non-invasive BP monitoring assumes linear relationships between physiological signals (ECG, PPG, PCG) and arterial pressure, but noise from cuff placement and white coat effects creates a 'Linearity Trap,' leading standard tools to underestimate cardiovascular risk in hypertensive patients.

    We propose a doubly-sharpened curvature diagnostic for nonparametric linearity testing in errors-in-variables models, using a Summation of Squared Second Derivatives (SSSD) test statistic with residual bootstrap inference, extended to changepoint detection. Applied to 6,756 NHANES observations, the method detects nonlinearity and a significant structural break between Diastolic Blood Pressure and Maximum Inflation Level, with implications for clinical BP protocols and risk assessment.

11. **Aimee Menard:** The Fan Region is a bright, large-scale feature observed in polarized radio continuum emission in the Milky Way Galaxy. It is thought to trace structures associated with the Perseus Arm or foreground interstellar medium (ISM). Previous studies have explored either large-scale structure at low resolution or fine-scale structure at high resolution. We use combined single-dish and interferometric observations to achieve arcminute resolution results to investigate the magnetized ISM surrounding the W3/W4/W5 complex located within the Fan Region. Our results reveal complex, multi-scale depolarization structures, with distinct behaviour in polarized intensity and polarization angle.

12. **Rachel Brookes:** A planned survey with the John A. Galt 26-m Telescope, located at the Dominion Radio Astronomical Observatory, aims to measure weak circular polarization (Stokes V) in the 21-cm neutral hydrogen (HI) line. A primary source of systematic error arises from polarized beam pickup, where the interaction of Galactic HI emission with the beam response corrupts the on-boresight Stokes V measurement. This contribution depends on the region of sky being observed and varies over time due to the Earth's rotation and orbital motion. To investigate this, we developed a simulation pipeline that receives an all-sky HI data cube in the sky reference frame, converts it to the telescope's reference frame, convolves it with the Stokes V beam pattern, and transforms the result back to the sky reference frame. By comparing the convolved and original data, we can estimate the instrumental contribution to observed signals for a given time and pointing.

13. **Slade Coffman:** Since the axiomatization of Topological Quantum Field Theories (TQFTs) by Michael Atiyah in the late 1980s, these theories have had a profound impact on both mathematics and physics. Because TQFTs depend only on topology

(and not on geometric details such as curvature), they provide some of the simplest and most accessible examples of Quantum Field Theories. For this reason, they offer a natural entry point into the broader framework of QFT.

A central object of interest in any quantum field theory is the partition function, which encodes global physical information about a system, such as its possible states and invariants. In the topological setting, partition functions often reduce to computable quantities that capture deep geometric and algebraic features of the underlying space. Understanding how to compute and interpret these functions is therefore a key motivation for studying TQFTs.

Despite their conceptual appeal, much of the literature on TQFTs relies on advanced mathematical machinery, making the subject difficult to approach at the undergraduate level. In this talk, we aim to bridge this gap by presenting a clear and accessible introduction to 2-dimensional TQFTs. We begin by reviewing the necessary background, including basic ideas from category theory, Frobenius algebras, and cobordisms. We then outline the classification result that the category of 2-dimensional TQFTs is isomorphic to the category of commutative Frobenius algebras. Finally, we show how this correspondence can be used in practice to compute partition functions.

4 CMPS Research Awards and Closing/Party – SCI 337

Location: SCI 337

- 4:00 pm – 4:30 pm: CMPS Research Awards
- 4:30 pm – 5:00 pm: Closing/Party