

Shera Potka

Vancouver, BC | potka.sheraa@gmail.com | (778) 535-3774 | linkedin.com/in/shera-potka
github.com/sherapotka | sherapotka.github.io

Professional Summary

Researcher in Responsible AI, focusing on **fairness in recommender systems**, **bias in large language models**, and **privacy in NLP**. My work combines algorithm design with social impact, producing **award-winning publications** at leading conferences. As an **instructor**, I have designed and taught both **graduate and undergraduate courses** and mentored large cohorts of computer science students. I aim to advance inclusive, accountable AI while training the next generation of researchers and practitioners.

Core Competences

- Research expertise in **Responsible AI**, with focus on fairness in recommender systems, bias auditing in LLMs, and privacy-preserving NLP.
- Proven track record of **scholarly impact**, including award-winning publications at international conferences.
- Experienced in **teaching and mentoring**, having designed and taught graduate-level courses and supported large undergraduate cohorts.
- Skilled in developing **scalable algorithms and data-driven methods**, bridging theoretical advances with practical applications.

Education

- **PhD in Computer Science**, University of Victoria (May 2023 – May 2025)
Thesis: “Three Ethical Dimensions of AI: Fairness in Social Recommenders, Bias in LLMs, and Privacy in NLP”
Focus: Responsible AI, Fairness in Recommendation, Bias Auditing, Privacy-Preserving NLP
- **Master’s in Media and Computer Science**, University of Cologne (Aug 2021 – Apr 2023)
Thesis in AI, machine learning, and computational linguistics
- **Bachelor’s in Media and Computer Science**, University of Cologne (Jun 2019 – Aug 2021)
Thesis on graph-based text analysis for social media
- **Advanced Training:** Data Mining, Systems for Massive Datasets, Data Privacy, Foundations of Disease Analytics

Academic Appointments & Teaching

Postdoctoral Research Fellow, Digital Global Health & AI Group (Dr. Richard Lester), Division of Infectious Diseases, University of British Columbia May 2026 – Present

- Lead AI and NLP research on patient–provider mobile-health communication data, developing topic-classification and clinical-language models that support real-world digital health programmes.
- Advise on AI implementation strategy across the group’s mHealth platforms, weighing in-house and external model approaches against data-governance, ethics, and language-coverage requirements.
- Apply **responsible-AI methods** (bias auditing, privacy-preserving NLP) to global-health datasets, bridging methodological work from my PhD with applied digital-health research.
- Contribute to the Division’s broader AI in Infectious Diseases initiative through strategy work and cross-disciplinary engagement.
- Collaborate on international digital-health research, including a cluster-randomized trial of digital learning for community health workers.
- Mentor graduate trainees on machine-learning and NLP components of ongoing projects.

Instructor, Data Models and Algorithms (CSC 501), University of Victoria Spring 2025 – Spring 2026

- Designed and taught this core graduate course across three offerings (≈ 50 students each), emphasizing active learning and algorithmic problem-solving.
- Created comprehensive teaching materials (slides, notes, and supplementary resources) aligned with current research in data management.
- Mentored students on projects, guiding applications of data models and algorithms to real-world datasets.
- Currently teaching Spring 2026 offering with updated materials and pedagogy based on prior cohort feedback.

Lead Teaching Assistant, Data Mining (SENG 474), University of Victoria Spring, Summer, Fall 2024

- Coordinated teaching support across three academic terms (Spring, Summer, Fall), with class sizes ranging from 50 to 180 students.
- Designed and delivered lab sessions to reinforce core methods in classification, clustering, and pattern mining.
- Provided individualized support through office hours, assignment feedback, and project guidance, contributing to improved student performance.

Lead Teaching Assistant, Web Design (CSC 130), University of Victoria Spring 2024

- Guided ≈ 120 students through web technologies (HTML, CSS, JavaScript), with an emphasis on usability and accessibility.
- Provided detailed project feedback and mentoring, enabling students to deliver functional and well-structured websites.

Researcher, Cologne Center for eHumanities (CCeH), University of Cologne Nov 2022 – Nov 2023

- Built workflows for large-scale text and metadata analysis in digital humanities.
- Advanced *Project RACIR* by developing tools for higher education and cross-disciplinary research.
- Bridged humanities and computer science through data-driven experiments.
- Produced reports, software prototypes, and scholarly dissemination materials.
- **Link:** [CCeH Profile](#)

Research Fellow, Electronic Textual Cultures Lab (ETCL), University of Victoria Oct 2022 – Mar 2023

- Built visualizations and computational tools for digital scholarship and open knowledge.
- Collaborated across disciplines to design research workflows and integrate digital methods.
- Produced scholarly outputs, including presentations, abstracts, and conceptual frameworks.
- Directed the *HSS Commons* initiative, shaping both platform design and implementation.
- **Link:** [ETCL Profile](#)

Selected Projects

ConVisScope, Visual Analytics Workbench for WeTel mHealth Programmes 2026 – Present

- Research workbench converting the *WeTel* two-way SMS stream into population-level intelligence for Rwanda's digital health ecosystem; co-developed with UBC Computer Science and NLP research teams.
- Production deployment on AWS Elastic Beanstalk with authenticated conversation viewer, paginated cohort cards, and GPT-4o-labelled topics carrying confidence scores and verification status.
- Currently extending the platform into a researcher-facing tool with cohort filtering, label review and correction, taxonomy editing, and reproducible cohort export, generating the validated correction stream that will train the next-generation in-house classifier.

NoSpin-EBM, AI Evidence-Based Medicine Search Engine for Clinicians 2026 – Present

- Designed and built an end-to-end clinical-search application that returns a Key Recommendation, a Logic Model, GRADE-rated Evidence Cards with verified DOIs (Crossref/PubMed), and a *Skeptic's Corner* that flags potential bias in the underlying evidence.
- Implemented **fail-closed typed runtime validation**: a field is rendered only when anchored to a retrieved

record, enforcing citation and quotation provenance at the field level. Retrieval spans Crossref, PubMed, openFDA, and the Health Canada Drug Product Database.

- Stack: FastAPI (Python 3.11) + React 19 + MongoDB; Groq (Llama 3.3 70B) generation behind a pluggable OpenAI-compatible provider layer; JWT and Google OIDC auth; Stripe billing; Resend email; deployed via Docker Compose with CI on GitHub Actions.
- Built an evaluation harness covering ablations, baselines, entailment scoring, adversarial prompt-injection, and stress testing, producing the measured provenance–entailment results now written up as a manuscript in preparation with A. Thomo (UVic) and R. Lester (UBC).

PageLink, Auditable Clinical Paging and Consult Platform (Prototype) 2026

- Contributing developer on a physician-to-physician paging and consult concept designed to replace hospital pagers under Canadian privacy constraints (FIPPA/PIPA); collaboration with Dr. Richard Lester (AboutHealth.ai).
- Built a high-fidelity, zero-install React product simulation (≈ 60 components) covering the full clinical workflow: paging and acknowledgement, on-call scheduling and shift coverage, in-app consult chat, structured clinical coding, and analytics.
- Used as the stakeholder artifact for clinical and investor conversations, scoping the requirements for a production MVP.

Publications

UNDER REVIEW

Selective Overconfidence: How Large Language Models Underserve Social Welfare Law Across Jurisdictions AIES 2026

Shera Potka, Alex Thomo, Wulf Loh, Paula Helm

AIES 2026 (AAAI/ACM Conference on AI, Ethics, and Society) — under review

TLDR: Probed 1,200 court decisions across Germany, France, and Canada, showing that Llama 3.3 70B admits uncertainty on 49.8% of tax-law questions but only 6.8% of social-welfare questions ($p < 0.0001$), projecting false confidence precisely where users are most vulnerable. A retrieval intervention closes 87% of the gap, identifying a knowledge deficit rather than a stylistic artifact.

People, Processes, Platforms: A Coding Framework and Comparative Benchmark for Global AI Governance AIES 2026

Shera Potka, Jens Weber

AIES 2026 (AAAI/ACM Conference on AI, Ethics, and Society) — under review

TLDR: Developed a 17-sub-dimension coding framework and applied it to 12 regulatory instruments across 10 jurisdictions, validated against an AI-assisted coding pipeline (87.6% agreement). Identifies formal AI classification and an “institutional creation threshold” (76% vs. 27% provision coverage) as the structural drivers of regulatory divergence, arguing that harmonization barriers are architectural rather than value-based.

MANUSCRIPTS IN PREPARATION

Multilabel Topic Classification of Patient Messages Across Two Clinical Contexts 2026

Riyaz Ur Rehman, Andy Liu, Hyeju Jang, Hassan Mugabo, **Shera Potka**, Richard T. Lester

Targeted at *npj Digital Medicine*

TLDR: A single multilabel pipeline classifying patient SMS across 19 topics in two contrasting WelTel deployments (COVID-19 monitoring in Kinyarwanda in Rwanda; pediatric diabetes care in English in Canada). A BERT+SVM+XGBoost stacking ensemble reaches macro-F1 0.78 (95% CI 0.75–0.81), outperforming fine-tuned BERT (0.71), with explicitly bounded performance on safety-critical labels supporting human-in-the-loop rather than autonomous triage.

**Evaluating Off-the-Shelf Machine Translation Systems for
Kinyarwanda-to-English Clinical Dialogue: A Benchmark and Cautionary
Analysis**

2026

Hassan Mugabo, Riyaz Ur Rehman, Andy Liu, Matthew Manson, Justin Tuyisenge, Muhammed Semakula,
Shera Potka, Richard T. Lester

Targeted at *The Lancet Digital Health*

TLDR: First head-to-head benchmark of six deployed machine translation systems on 600 real Kinyarwanda patient-provider clinical dialogues. All systems reach TER ≈ 80 (four in five tokens require correction), and automatic metrics systematically understate the quality gap clinicians perceive, indicating that no off-the-shelf system is safe for unsupervised clinical use in this language pair.

PEER-REVIEWED PUBLICATIONS

**Gender and Race Bias in Consumer Product Recommendations by Large
Language Models**

AINA-2025

Ke Xu, **Shera Potka**, Alex Thomo

AINA-2025 (39th International Conference on Advanced Information Networking and Applications)

TLDR: Investigated biases in consumer product recommendations, focusing on gender and race bias. Analyzed implications for fairness and diversity in AI-driven recommendation systems.

**CluSanT: Differentially Private and Semantically Coherent Text
Sanitization**

NAACL 2025

Ahmed Musa Awon, Yun Lu, **Shera Potka**, Alex Thomo

NAACL 2025 (Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics)

TLDR: Introduced CluSanT, a novel framework for text sanitization using Metric Local Differential Privacy (MLDP), balancing privacy and semantic coherence through clustering and embedding techniques.

Word Embedding Bias in Large Language Models

I-SPAN 2025

Poomrapee Chuthamsatid, **Shera Potka**, Alex Thomo

I-SPAN 2025 (17th International Symposium on Pervasive Systems, Algorithms, and Networks)

TLDR: Examined gender and race bias in modern large language models, expanding beyond previous research with new insights using SC-WEAT tests and clustering techniques.

**Enhancing Structural Minority Visibility in Link Recommendations
(Best Paper Award)**

MEDES 2024

Shera Potka, Isla Li, Jason Kepler, Alex Thomo

MEDES 2024 (16th International Conference on Management of Digital EcoSystems)

TLDR: Introduced MinWalk, an algorithm to improve visibility of minority groups in social networks, balancing fairness and reducing popularity bias.

**Community Structure and Coherence in Digital Humanities Works
(Best Paper Award)**

IISA 2023

Shera Potka, Alex Thomo

IISA 2023 (14th International Conference on Information, Intelligence, Systems & Applications): 1–8

TLDR: Analyzed a decade of Digital Humanities publications, revealing community structure and cohesion using text similarity networks.

Technical Expertise

Extensive experience with **Python** and its scientific ecosystem (PyTorch, scikit-learn, NumPy, Pandas) for data mining, bias analysis, and privacy-preserving NLP.

Proficient in **databases and large-scale systems**, including PostgreSQL, MySQL, and ElasticSearch, applied to recommender systems and network analysis.

Skilled in **web and platform development** (React, Node.js, WordPress) for building interactive applications and research dissemination tools.

Competent in **knowledge representation and reproducibility**, using Protégé for ontology design, LaTeX for scholarly writing, and Docker/Git for collaborative development.

References

Prof. Alex Thomo, University of Victoria, Computer Science, BC, Canada, thomo@uvic.ca

Prof. Venkatesh Srinivasan, Santa Clara University, Mathematics and Computer Science, CA, USA, vsrinivasan4@scu.edu

Prof. Yun Lu, University of Victoria, Computer Science, BC, Canada, yunlu@uvic.ca