

# SHAHRIAR SHAYESTEH

**Research Scientist Intern | Responsible AI, NLP, AI Agents | Summer 2027**

State College, PA | +1 814-769-6022 | shahriar.shayesteh92@gmail.com | Website | Google Scholar | GitHub

## Summary

Ph.D. candidate studying how AI systems decide what information gets through. Audits disclosure and omission when LLMs and AI agents speak for users or explain complex information.

## Technical Skills

**Programming:** Python, SQL, Java, C++

**ML and NLP:** PyTorch, TensorFlow, scikit-learn, Hugging Face, pandas, NumPy

**LLM Systems:** LLM evaluation, tool-calling agents, LangChain, AutoGen, Model Context Protocol (MCP)

**Research Methods:** Experiment design, human annotation, topic modeling, fairness and privacy analysis

## Research Experience

**Graduate Research Assistant** | Human Language Technologies Lab, Penn State | *Aug 2023 – Present*

- Published a diagnostic audit of 2,344 third-party API and action schemas, identifying interface conditions that may invite unnecessary disclosure by tool-using LLM agents.
- Designed ongoing controlled evaluations for How Tool Schemas Shape LLM Agent Disclosure, varying schema structure, guidance, and declared purpose while holding the task and backend fixed.
- Built Easier to Read, Still Easy to Miss, a source-grounded audit of wireless-contract explanations using 24 topics and 3,899 verified provisions; targeted to ACM FAccT 2027.
- Analyzed 59,590 matched policy pairs for Privacy, More or Less (under revision for PoPETs 2027); contributed to PrivaSeer's 3,967,487-policy corpus and SoAC/SoACer's 195,495-website corpus.

**Research Intern** | Department of Canadian Heritage | *Feb 2022 – Apr 2022*

- Applied topic modeling and sentiment analysis to qualitative data on challenges faced by Canadian artists and translated results into policy-relevant findings.

**Graduate Research Assistant** | NLP Laboratory, University of Ottawa | *Jan 2021 – Jun 2023*

- Developed NDA-GAN, a semi-supervised text-classification model using negative data augmentation; improved overall accuracy and confidence on out-of-distribution samples relative to GAN-BERT.

## Selected Publications

- S. Shayesteh and S. Wilson. "From Conventional Web Privacy to Agentic Disclosure: How Tool Schemas May Invite LLM Oversharing." PrivateNLP @ ACL, 2026.
- S. Shayesteh et al. "SoAC and SoACer: A Sector-Based Corpus and LLM-Based Framework for Sectoral Website Classification." ACM DocEng, 2025.
- S. Wilson et al. "The PrivaSeer Project: Large-Scale Resources for Analysis of Privacy Policy Text." USENIX SOUPS, 2025.
- S. Shayesteh and D. Inkpen. "Generative Adversarial Learning with Negative Data Augmentation for Semi-Supervised Text Classification." FLAIRS, 2022.

## Education

**Pennsylvania State University** | Ph.D. in Informatics, GPA 3.93/4.0 | Aug 2023 – Expected Fall 2027

**University of Ottawa** | M.Sc. in Computer Science, GPA 4.0/4.0 | Sep 2020 – Jun 2023

**University of British Columbia** | Non-degree CS and Statistics coursework | Sep 2017 – May 2020

## Talks & Training

Guest lecture, IST 574 (scheduled Oct 2026); PrivateNLP @ ACL presenter (Jul 2026); Mila Responsible AI & Human Rights Summer School (Jun 2023)