

Shahriar Shayesteh

PhD Candidate, Pennsylvania State University, USA
shahriar.shayesteh92@gmail.com

RESEARCH FOCUS

I study how privacy norms shift from policy-governed websites to **agentic AI systems**, where agents use tool-calling, APIs, and multi-agent workflows to construct and transmit personal information to web services at runtime.

SELECTED PUBLICATIONS

Generative Learning with Negative Data Augmentation for Semi-Supervised Text Classification.

Shahriar Shayesteh, Diana Inkpen. *FLAIRS-35*, 2022.

SoACer: Sector-Based Corpus & LLM Framework for Sectoral Website Classification.

Shahriar Shayesteh, Mukund Srinath, Lee Matheson, Lu Xian, Sinjoy Saha, Lee Giles, Shomir Wilson. *ACM DocEng*, 2025.

The PrivaSeer Project: Large-Scale Resources for Analysis of Privacy Policy Text.

Shomir Wilson, Florian Schaub, Lee Matheson, Shahriar Shayesteh, Lu Xian. *USENIX SOUPS*, 2025.

From Conventional Web Privacy to Agentic Disclosure: How Tool Schemas May Invite LLM Oversharing.

Shahriar Shayesteh, Shomir Wilson. *PrivateNLP Workshop at ACL*, 2026.

Research and Internship Experience

- **Graduate Research Assistant** Fall 2023 – Present
The Human Language Technologies Lab, Pennsylvania State University
 - **Project: PrivaSeer** Fall 2023 – 2025
A large-scale platform for collecting, indexing, and analyzing privacy policy text.
 - **Action:** Designed and implemented sector-aware analysis over 3M privacy policies, including extraction of data types, purposes, and third-party sharing practices.
 - **Outcome:** Enabled cross-sector comparison of privacy practices and supported empirical research on how organizations state and vary data handling commitments.
 - **Project: LLM-COLDs** 2025 – Present
A research project studying how AI systems advise users on consumer-oriented legal documents.
 - **Action:** Designed an analysis framework to evaluate when LLM-generated contract advice may be misleading, incomplete, or harmful to consumers.
 - **Outcome:** Supports research on trustworthy AI assistance for legal-document understanding and consumer decision-making.

Research Intern Feb 2022 – Apr 2022
Department of Canadian Heritage

- **Problem:** Understanding systemic challenges faced by artists using unstructured qualitative data.
- **Action:** Applied NLP methods (topic modeling, sentiment analysis) to analyze large text corpora; translated findings into policy-relevant insights.
- **Result:** Informed cultural policy design and strategic planning for government stakeholders.

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

- **Problem:** Social bias and reliability concerns in semi-supervised text classification.
- **Action:** Investigated fairness-aware training methods; implemented a GAN-based approach with negative data augmentation.
- **Result:** Achieved +3% performance improvement over strong baselines; work formed the basis of a Master's thesis.

RAI Summer School at Mila Jun 2023
Mila Institute – Responsible AI and Human Rights

- **Focus:** Ethical, legal, and governance challenges in AI systems.
- **Outcome:** Engaged in interdisciplinary discussions on accountability, fairness, and human-centric AI design.

Education

Pennsylvania State University
Ph.D. in Informatics

Fall 2023 – Fall 2027 (Expected)
GPA: 3.93/4.0

University of Ottawa
M.Sc. in Computer Science

Jan 2021 – Jun 2023
GPA: 4.0/4.0

University of British Columbia
Non-Degree Program, Computer Science and Statistics

Sep 2017 – May 2020
GPA: 3.73/4.0

Ongoing Research Projects

Privacy, More or Less: Large-Scale Sectoral Comparison of Privacy Policies Between Industries. (*Ongoing*)
Constructed a longitudinal corpus of 59K privacy policies to analyze sector-level norms, deviations, and mismatches between stated policies and observed practices.

LLM-COLDS: AI Advice on Consumer-Oriented Legal Documents. (*Ongoing*)
Studying how LLM-generated advice on consumer-oriented contracts can be misleading, incomplete, or harmful, and developing evaluation methods for trustworthy legal-document assistance.

References

Prof. Shomir Wilson
Associate Professor, IST, Pennsylvania State University
Email: shomirwilson@psu.edu