

Shomir Wilson

Associate Professor of Human-Centered Computing and Social Informatics
Pennsylvania State University
shomir@psu.edu — <https://shomir.net>

HIGHLIGHTS

I am an academic **leader**:

- I recently served as the Interim Program Coordinator for our college’s B.S. in Data Sciences, and I will be the inaugural Program Coordinator for our new B.S. in Information Technology Ethics and Compliance.
- I serve on our University Faculty Senate and the university’s Faculty Affairs Committee.
- In Fall 2023 – Spring 2024 I chaired our college’s Awards Committee.

I am a **mentor** of students and faculty:

- I maintain a collection of advice (<https://shomir.net/advice>) for undergraduates, graduate students, postdocs, and faculty. Toward faculty mentorship, these materials include my narrative statement for tenure and promotion to associate professor, a guide of experiential advice for pretenure faculty, and a glossary of jargon that university faculty use.
- My undergraduate research advisees have first-authored papers accepted to major conferences, and my graduate research advisees have received best student paper awards.

I am a **researcher** in natural language processing (NLP) and digital privacy:

- I am the lead PI for the PrivaSeer Project, funded by a \$1.2M grant from the NSF, with four PIs across two universities and a nonprofit advocacy group.
- I received the NSF CAREER Award (\$556K) for an interdisciplinary project at the intersection of NLP and law.
- My lab’s publications have received awards or special recognitions from venues including EMNLP, WebConf, CSCW, PETS, and DocEng.

I am a **speaker** at events and in news media:

- My talks include keynotes at SEPLN and PrivateNLP and the HCOMP Encore Track.
- My media appearances include NBC News Now, Scripps Media, Inside Higher Ed, Digital Trends, and live television and radio interviews.

I am a **teacher** of a wide variety of undergraduate and graduate courses:

- My teaching portfolio includes undergraduate first-year seminars, senior-level required courses and electives, and graduate courses.
- In peer reviews of teaching, colleagues describe my teaching as highly interactive, and a consultant from our university’s center for teaching excellence described my style as “enthusiastic” and “welcoming”.

TABLE OF CONTENTS

Experience - p. 2	Media Appearances - p. 5	Papers - p. 11
Education - p. 3	Service - p. 7	Invited Activities - p. 18
Grants and Sponsorship - p. 3	Teaching - p. 9	Miscellany - p. 20
Awards and Honors - p. 4	Student Mentorship - p. 10	

External links in this PDF are [blue](#), and internal links are [red](#).

EXPERIENCE

Primary Affiliations

Pennsylvania State University College of Information Sciences and Technology (University Park, PA)	7/2024 –	Associate Professor with tenure Director, Human Language Technologies Lab Graduate Faculty, College of IST Graduate Faculty, Social Data Analytics Affiliate, Center for Socially Responsible Artificial Intelligence Affiliate, Social Science Research Institute
	8/2018 – 6/2024	Assistant Professor
University of Cincinnati EECS Department (Cincinnati, OH)	8/2016 – 8/2018	Assistant Professor Director, Human Language Technologies Lab Member, Institute for Analytics Innovation
Carnegie Mellon University School of Computer Science (Pittsburgh, PA)	8/2015 – 8/2016	Project Scientist <i>Supervisor: Norman Sadeh</i>
	1/2015 – 5/2015	Lecturer (Natural Language Processing)
	8/2014 – 7/2015	NSF International Research Fellow <i>Supervisor: Alan W Black</i>
	9/2011 – 7/2013	Postdoctoral Fellow <i>Supervisor: Norman Sadeh</i>
University of Maryland Computer Science Department (College Park, MD)	6/2006 – 5/2011	Graduate Research Assistant <i>Supervisor: Donald Perlis</i>
	8/2005 – 5/2006	Graduate Teaching Assistant

Visiting Affiliations

University of Edinburgh School of Informatics (Edinburgh, United Kingdom)	8/2013 – 7/2014	NSF International Research Fellow <i>Host: Jon Oberlander</i>
Nat'l Univ. of Singapore School of Computing (Kent Ridge, Singapore)	6/2010 – 8/2010	NSF EAPSI Fellow <i>Host: Min-Yen Kan</i>
Macquarie University School of Computing (Sydney, Australia)	6/2009 – 8/2009	NSF EAPSI Fellow <i>Host: Robert Dale</i>

EDUCATION

University of Maryland	2011 Ph.D. , Computer Science; 2008 M.S. , Computer Science
Virginia Tech	2005 B.S. , Computer Science, B.S. , Mathematics, B.A. , Philosophy
U. of Mary Washington	1999 part-time, non-degree seeking

GRANTS AND SPONSORSHIP

Financial Support External to the University

National Science Foundation: PI on #2237574 “CAREER: Large-Scale Exploration and Interpretation of Consumer-Oriented Legal Documents”. **\$556,397 to Penn State** for 2023-08-01 – 2028-07-31.

National Science Foundation: Co-PI on #2247723 “SaTC: CORE: Small: Toward Privacy Equity through Contextual Understanding of Self-Disclosure”. Other senior personnel: Sarah Rajtmajer (PI). **\$599,851 to Penn State** for 2023-06-01 – 2026-05-31.

National Science Foundation: Lead PI on #2105736 “Collaborative Research: SaTC: CORE: Medium: A Large-Scale, Longitudinal Resource to Advance Technical and Legal Understanding of Textual Privacy Information”. Other senior personnel: Lee Giles (Co-PI), Florian Schaub (PI at University of Michigan), Gabriela Zafir-Fortuna (PI at Future of Privacy Forum). **\$699,744 to Penn State and \$1,198,820 overall + \$21,000 REU supplement to Penn State** for 2021-07-01 – 2024-06-30.

National Institutes of Health: Co-I (subawardee) on #1R01MD015064-01A1 “Primed to (re)act: Can changes in procedural language prevent adverse events between police and minority male youth?”. Other senior personnel: Margaret Beale Spencer (PI at University of Chicago), Christopher Graziul (PI at University of Chicago), Karen Livescu (Co-I at Toyota Technological Institute at Chicago), Lisa Thureau (Co-I at Strategies for Youth). **\$228,054 to Penn State** for 2021-01-27 – 2023-12-30.

National Science Foundation: PI on #1914444 “SaTC: CORE: Medium: Collaborative: Automatically Answering People’s Privacy Questions”. Other senior personnel: Norman Sadeh (Lead PI at Carnegie Mellon University), Joel Reidenberg (PI at Fordham University), Tom Norton (Senior Personnel at Fordham University). **\$437,436 + \$14,000 REU supplement to Penn State** for 2019-07-15 – 2023-12-31.

National Science Foundation: Subawardee on #1330596 “TWC SBE: Option: Frontier: Collaborative: Towards Effective Web Privacy Notice and Choice: A Multi-Disciplinary Prospective”. Other senior personnel: Norman Sadeh (Lead PI at Carnegie Mellon University), et al. **\$76,249 to University of Cincinnati** for 2016-09-01 – 2018-07-31; **\$13,792 to Penn State** for 2018-09-01 – 2019-07-31.

National Science Foundation: PI on #1159236 “IRFP: Metalanguage Identification for Interactive Language Technologies”. **\$98,100** fellowship for 2013-08-01 – 2015-03-31.

National Science Foundation: PI on #1015666 “EAPSI: Parsing Metalanguage and the Use-Mention Distinction”. **\$5,000 + travel support** fellowship for 2010-06-13 – 2010-08-06.

National Science Foundation: PI on #0914091 “EAPSI: Distinguishing Use and Mention in Natural Language”. **\$5,000 + travel support** fellowship for 2009-06-22 – 2009-08-16.

Financial Support Internal to the University

Penn State Center for Socially Responsible Artificial Intelligence: Co-PI on “Understanding the Prevalence of Drinking Water Service Disruption through Large-Scale Analysis of News Articles and Social Media”. Other senior personnel: Christine Kirchhoff (Lead PI). **\$25,000** for 2022.

Penn State Center for Social Data Analytics: PI on “Exploring the Effects of Socioeconomic Status on Privacy Behaviors in an Online Social Network”. Other senior personnel: Sarah Rajtmajer (Co-PI). **\$25,000** for 2020.

Penn State Center for Cybersecurity Research and Education: PI on “Automatic Detection of Malicious Emails Using Discourse Analysis”. Other senior personnel: Susan Strauss (Co-PI). **\$7,500** for 2019 – 2021.

University of Cincinnati: PI on College of Engineering and Applied Sciences Faculty Development Grant. **\$2,350** for 2017.

Non-Financial Support

Pennsylvania Space Grant Consortium: Scholarship support to place undergraduate researchers in my lab for 2019 – 2022.

Penn State College of IST: Lead PI on “Web-Scale Search and Analysis of Privacy Policies”. Other senior personnel: Lee Giles (Co-PI). 2 years GRA support for 2019-08-26 – 2021-08-20.

Ohio Supercomputer Center: 9,000 Resource Units for research and teaching for 2016 – 2018.

AWARDS AND HONORS

Publication Awards and Recognitions

Diversity, Equity, and Inclusion Recognition at CSCW 2024.

Best Teaching Material Award at TeachNLP 2024.

Outstanding Paper Award at EMNLP 2023.

Best Student Paper Award (for student authors with me as co-author) at PETS 2023.

Best Student Paper Award (for my advisees with me as co-author) at DocEng 2023.

Best Short Paper Award at TrustNLP 2023.

Best Paper Finalist at WebConf 2016.

Other Honors

Future CRA Leaders Program (inaugural cohort), 2024.

Graduation Speaker, Virginia Tech Department of Philosophy, May 18, 2019.

University of Maryland International Conference Student Support Award, 2011.

University of Maryland Block Grant Fellowship, 2005–2007.

Virginia Tech Outstanding Senior in Computer Science, 2005.

Virginia Tech William H. Williams Senior Prize in Philosophy, 2004.

MEDIA APPEARANCES

National and International Media

Tech Brew: Quoted in “‘AI election’ roundup: Deepfakes and AI memes are taking off”, 2024-09-12. [\[article\]](#)

Scripps Media: Recorded video interview for story “States look to legislation to regulate AI political ads next election”, 2023-12-01. Featured in affiliated TV stations’ news broadcasts around the country and on Scripps’ social media accounts. [\[video\]](#)

Inside Higher Ed: Quoted in “Going Phishing on Campus”, 2023-07-18. [\[article\]](#)

Wealth Management: Quoted in “AI Will Heighten Cybersecurity Risks for RIAs”, 2023-07-12. [\[article\]](#)

NBC News NOW: Live video streaming interview about the potential impact of ChatGPT on online dating, NBC News NOW, 2023-03-17. [\[video\]](#)

VentureBeat: Quoted in “How bias creeps into the AI designed to detect toxicity”, 2021-12-09. [\[article\]](#)

Digital Trends: Quoted in “Will Google ever lose its throne as king of search? Here are its main contenders”, 2021-08-15. [\[article\]](#)

Local and Regional Media

Morning Wave on 90.5 FM BeFM (Busan, South Korea): Live interview during segment “DeepSeek Shakes Up AI World”, 2025-02-02.

Let’s Talk on WFXS 98.7 FM The Fox (State College, PA): Live radio interview about artificial intelligence, 2025-01-17.

Charlotte News Observer (Charlotte, NC): Quoted in “‘Over the top’ ad features a fake Mark Robinson. What to know about AI in political ads.”, 2024-10-10 [\[article\]](#); picked up by Yahoo News. [\[article\]](#)

Let’s Talk on WFXS 98.7 FM The Fox (State College, PA): Live radio interview about artificial intelligence, 2023-10-23.

The Daily Item (Sunbury, PA): Quoted in “Deepfake concerns growing along with AI technology”, 2023-09-30 [\[article\]](#); picked up by Yahoo News. [\[article\]](#)

WPMT-TV (York, PA): Recorded video interview for TV news segment “AI deepfakes becoming more lifelike, more misleading”, 2023-08-07. [\[video and article\]](#)

WPMT-TV (York, PA): Recorded video interview for TV news segment “How AI is changing the industry and what it means for Pa. businesses”, 2023-08-06. [\[video and article\]](#)

WPMT-TV (York, PA): Recorded video interview for TV news segment “How the AI revolution could impact education”, WPMT, 2023-07-21. [\[video and article\]](#)

WTAJ-TV (Altoona, PA): Recorded video interview for TV news segment “Artificial intelligence programs are causing concern for educators”, 2023-05-17. [\[video and article\]](#)

Centre County Report (State College, PA): Recorded video interview for online video about artificial intelligence being used by swatting perpetrators, 2023-04-21. [\[video\]](#)

WCPO-TV (Cincinnati, OH): Recorded video interview for TV news segment about FaceID privacy, 2017-11-14.

University Media

Penn State News: “IST announces new information technology ethics and compliance major”, 2025-05-05. [\[article\]](#)

Penn State News: “NCEMS working groups to answer molecular and cellular bioscience questions”, 2025-04-11. [\[article\]](#)

Penn State News: “Penn State congratulates Fall 2024 Emerging Academic Leaders on program completion”, 2025-03-27. [\[article\]](#)

Penn State News: “Q&A: DeepSeek AI assistant and the future of AI”, 2025-01-30. [\[article\]](#)

Penn State News: “IST faculty leads international privacy research event at German castle”, 2025-01-27. [\[article\]](#)

Penn State News: “New course helps students discover how computers make sense of human language”, 2024-12-10. [\[article\]](#)

Penn State News: “Ask an expert: AI and disinformation in the 2024 presidential election”, 2024-09-03. [\[article\]](#)

Penn State Research Magazine: “Faculty Highlights”, 2023. [\[article\]](#)

Penn State News: “Trained AI models exhibit learned disability bias, IST researchers say”, 2023-11-30. [\[article\]](#)

Penn State News: “Most websites do not publish privacy policies, researchers say”, 2023-10-25. [\[article\]](#)

iConnect Magazine: “Understanding the fine print”, 2023 Winter. [\[article\]](#)

Penn State News: “Positive triggering method reduces nationality bias in large text generators”, 2023-04-25. [\[article\]](#)

Penn State News: “What is ChatGPT and what can it be used for?”, 2023-04-21. [\[article\]](#)

Penn State News: “IST assistant professor Shomir Wilson receives NSF CAREER Award”, 2023-04-21. [\[article\]](#)

Penn State News: “Email scammers tailor methods to target universities”, 2023-04-20. [\[article\]](#)

Penn State News: “Center for Socially Responsible AI awards seed funding to 6 projects”, 2023-01-27. [\[article\]](#)

Penn State News: “Researchers propose methods for automatic detection of doxing”, 2022-12-09. [\[article\]](#)

Penn State News: “AI language models show bias against people with disabilities, study finds”, 2022-10-13. [\[article\]](#)

iConnect Magazine: “Studying Adverse Police Interactions”, 2021 Summer. [\[article\]](#)

Penn State News: “Women and lower-education users more likely to tweet personal information”, 2021-07-07. [\[article\]](#)

ICDS News: “\$1.2 million NSF grant to create search engine for online privacy research”, 2021-06-24. [\[article\]](#)

Penn State News: “Study of police language aims to find patterns that may lead to tragic outcomes”, 2021-05-25. [\[article\]](#)

Penn State News: “What if opting out of data collection were easy?”, 2021-01-14. [\[article\]](#)

Penn State News: “IST students and faculty ‘build together’ at Grace Hopper Celebration”, 2020-10-22. [\[article\]](#)

iConnect Magazine: “Simplifying Privacy”, 2019 Fall. [\[article\]](#)

Penn State News: “Research Aims to Automatically Answer User Questions on Online Privacy Policies”, 2019-07-23. [\[article\]](#)

Penn State News: “College of IST Awards Eight Seed Grants for Research Projects”, 2019-06-13. [\[article\]](#)

Penn State News: “Penn State’s Leadership in Artificial Intelligence Research”, 2019-04-18. [\[article\]](#)

Penn State News: “Outside of IST: Faculty and Staff Making a Difference”, 2018-10-17. [\[article\]](#)

Other Media

National Science Foundation: News from the Field: “Carnegie Mellon Researchers Create an AI to Help Us Make Sense of Privacy Policies”, 2018-03-01. [\[article\]](#)

National Science Foundation: SEE Innovation Research Highlight: “Refining a Computer’s Understanding of Language”, 2012.

SERVICE

Service items with future dates are confirmed commitments.

Peer Review

Editorial Board, *Journal of Intelligent Information Systems*, 2018 - present.

Standing Reviewer, *Computational Linguistics*, 2020 - present.

Standing Reviewer, ACL Rolling Review (centralized peer review for major NLP conferences, 2021 - present).

Area Chair: COLING 2025.

Program Committee and/or Reviewer: COLM 2025, PrivateNLP 2024, COLM 2024, AAAI 2024, IJCNLP-AAACL 2023, ACL 2023, EACL 2023, EMNLP 2022, PrivateNLP 2022, ACL 2022, AAAI 2022, MISQ (2021), PrivateNLP 2021, NAACL 2021, EMNLP 2020, IMWUT 2020, ACL 2020, PrivateNLP 2020, AAAI 2019, EMNLP 2019, NAACL 2019, LREC TA-COS Workshop 2018, ICDCIT 2017, IEEE CIC PiCSoc Workshop 2016.

Grant Proposal Review for the Research Community: Swiss National Science Foundation (2025, 2020, 2019), Marsden Fund of the Royal Society of New Zealand (2023), National Science Foundation (2018, 2017, 2016), Portuguese Foundation of Science and Technology (FCT) (2012).

Grant Proposal Review for the Academic Unit and University: Institute for Cyber Science (2023), Center for Socially Responsible Artificial Intelligence (2023, 2022), College of IST Seed Grants (2023, 2021, 2020), Center for Social Data Analytics (2020).

Research Community

Lead Organizer, Dagstuhl Seminar 25021: Grand Challenges for Research on Privacy Documents, 2025-01-(05-10).

Organizing Committee Member, the Sixth Workshop on Privacy in Natural Language Processing (PrivateNLP), at NAACL 2025.

Co-Lead, NSF SaTC PI Meeting Breakout Session (with Sarah Rajtmajer): “Privacy, Policy, and Social Factors”, 2024-09-04.

Lead Organizer, Birds-of-a-Feather Session: “NLP on Legal Text”, EACL 2023.

Lead Organizer, NSF SaTC PI Meeting Breakout Session (with Athina Markopoulou): “Privacy, Policy, and People”, 2022.

Student Research Workshop Mentor, NAACL 2022.

Co-Lead for Group Mentoring Session, ACL 2020.

Student Mentorship Program Participant, ACL 2019.

Lead Organizer, NSF SaTC PI Meeting Breakout Session (with Norman Sadeh): “Making Privacy Understandable for Internet Users”, 2019.

Lead Organizer, The AAAI Spring Symposium on Privacy-Enhancing Artificial Intelligence and Language Technologies, 2019. [\[hyperlink\]](#)

Session Chair, Social Applications II, EMNLP 2018.

Lead Organizer, The AAAI Fall Symposium on Privacy and Language Technologies, 2016. [\[hyperlink\]](#)

Session Chair, Logic Programming and Knowledge Engineering at AAAI 2008.

Department or College

Interim Program Coordinator, Data Sciences B.S. Program, Spring 2025.

College of IST Graduate Advisory Council, August 2024 - present.

Co-Organizer for IST-EECS Natural Language Processing Colloquium Series, Fall 2019 - present.

Privacy and Security Area Review Committee, 2024 Fall.

Faculty Mentor, October 2024 - present.

College of IST Awards Committee: Chair for Fall 2023 - Spring 2024, Member for Fall 2022 - Spring 2023.

Selection Committee for IST's Westin Scholars Award: Founder in 2021; Chair 2021 - 2022.

Course Committee Chair for Integration of Security and Privacy (SRA 472), 2021 - 2024.

Course Committee Chair for First-Year Seminar for Data Science Majors (PSU 17), 2021 - 2024.

Faculty Annual Review Committee, 2023.

College of IST Security/Privacy Faculty Hiring Committee: 2022-2023, 2019 - 2020, 2018 - 2019.

Committee Member for IST PhD Qualifying Exams, 2022, 2021 (committee chair), 2019.

College of IST Graduate Recruitment Committee, Fall 2020 - Spring 2022.

College Representative to the NSF Workshop on Departmental Plans for Broadening Participation in Computer Science, 2019 and 2022.

College of IST Committee to Assemble Proposal for an Undergraduate Degree in Artificial Intelligence and Informatics, 2019.

IST Senior Exit Survey Committee, 2019.

College of Engineering and Applied Sciences eLearning Task Force, 2018.

Member of the EECS Graduate Admissions Committee, 2017 - 2018.

Faculty Marshal for the College of Engineering and Applied Sciences: May 2018, May 2017.

Graduate Student Representative to the Computer Science Department Council, 2008 - 2009.

University

University Faculty Senate, Fall 2022 - Spring 2026.

Faculty Affairs Committee (via Faculty Senate), 2024 Fall - present.

Interviewer for the Penn State Millennium Scholars Program: 2025, 2024, 2023, 2022, 2021, 2020, 2019.

Panel on Undergraduate Research, Spark Program, 2024-11-06.

University Planning Committee (via Faculty Senate), 2022 Fall - 2024 Fall.

Steering Committee, Center for Socially Responsible Artificial Intelligence, Fall 2022 - present.

Advisory Board for DACTR (<https://dactrgroup.com/>) in the Nittany AI Challenge, 2020.

Participant in the Penn State - University of Auckland Research Collaboration Development Workshop, 2019.

Outreach

Pro bono consulting about artificial intelligence, for policy aides from the US Senate Special Committee on Aging, 2023 October - November.

Grand Awards Judge, International Science and Engineering Fair: 2021, 2018, 2015, and 2012.

Judge, Pittsburgh Regional Science and Engineering Fair, 2016.

Judge, Pittsburgh Regional FIRST Lego League, 2013 and 2011.

TEACHING

Courses for Credit

Understanding Artificial Intelligence for Human Language (DS 402): 2025 Spring. Pennsylvania State University.

Computational Foundations of Informatics (IST 510): 2025 Spring, 2023 Spring, 2022 Spring. Pennsylvania State University.

Natural Language Processing Meets the World (IST 197G): 2024 Fall. Pennsylvania State University.

Integration of Privacy and Security (SRA 472): 2023 Fall, 2022 Fall, 2022 Spring, 2021 Spring. Pennsylvania State University.

First-Year Seminar for Data Science Majors (PSU 17): 2023 Spring, 2021 Fall, 2020 Fall. Pennsylvania State University.

Honors Introduction to Information, People, and Technology (IST 110H): 2020 Fall, 2019 Fall. Pennsylvania State University.

Natural Language Processing for Sentiment, Semantics, and Discourse (IST 597): 2020 Spring, 2019 Spring. Pennsylvania State University.

Advanced Topics in Natural Language Processing (CS 7052): 2018 Spring, 2017 Spring. University of Cincinnati.

Natural Language Processing (CS 5134/6034; 11-411/11-611): 2017 Fall at University of Cincinnati; 2015 Spring at Carnegie Mellon University (co-taught with Chris Dyer and Alan W Black).

Non-Credit Courses

Natural Language Processing Methods and Applications: Workshop for faculty from Ming Chi University of Taiwan, 2018-07-(11-13). University of Cincinnati.

Ethics of Artificial Intelligence: Guest lectures over 2018-03-(11-15), 2018. Future University (Cairo, Egypt).

STUDENT MENTORSHIP

Research Advising and Mentorship

Advisees are listed with their graduation placements, when available.

Ph.D. Advisees: Pranav Venkit (2025, Research Scientist at Salesforce), Mukund Srinath (2024, Machine Learning Scientist at Expedia), Younes Karimi (2024, Senior Machine Learning Engineer at Lingopal)

M.S. Advisees: Sonu Gupta (2022, Data Scientist at Decision Spot), Duo Pan (2021, Associate Application Developer at ADP), Soundarya Sundareswara (2021, Software Engineer at Apple), Abhijith Athreya (2018, Chief Engineer at Samsung R&D), Baradwaj Aryasomayajula (2018, Software Developer at Incedo)

Undergraduate Honors Thesis Advisees: Jack Krehely (2025, Analyst at Deloitte), Kathryn Frankenberg (2022, Privacy Engineer at Google)

Non-Thesis Undergraduate Research Advisees: Avi Bewtra, Grace Ciambrone, Minh Doan, Miranda Goodman, Matthew Ihlenfeld, Samantha Kenny, Juan Larenas, Ananya Nijhawan, Nora O’Toole, Ellen Poplavska, Benjamin Siri, Josephine Soddano, Eesha Srivatsavaya, Kaitlyn Wassel

Thesis Committee Memberships

Ph.D. Thesis Committee Memberships: Michael Kranzlein (2024), Chase Bloch (2023), Kaixuan Zhang (2021), Maha Aljohani (2018)

M.S. Thesis Committee Memberships: Xi Lu (2020), Raphael Rodriguez (2020), Sampurna Ravindranathan (2018), Abhro Mondal (2017)

PAPERS

Papers with future publication dates are accepted to appear in their respective venues. Papers that received awards or other special recognition are marked with 🏆.

Peer-Reviewed Conference Proceedings

An Audit on the Perspectives and Challenges of Hallucinations in NLP. Pranav Narayanan Venkit, Tatiana Chakravorti, Vipul Gupta, Heidi Biggs, Mukund Srinath, Koustava Goswami, Sarah Rajtmajer, Shomir Wilson. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2024.

Incorporating Taxonomic Reasoning and Regulatory Knowledge into Automated Privacy Question Answering. Abhilasha Ravichander, Ian Yang, Rex Chen, Shomir Wilson, Thomas Norton, and Norman Sadeh. In *Proceedings of the 25th International Conference on Web Information Systems Engineering (WISE)*, 2024.

Do Generative AI Models Output Harm while Representing Non-Western Cultures: Evidence from A Community-Centered Approach. Sourojit Ghosh, Pranav Narayanan Venkit, Sanjana Gautam, Shomir Wilson and Aylin Caliskan. In *Proceedings of the 2024 AAAI/ACM Conference on AI, Ethics, and Society (AIES)*, 2024.

CALM: A Multi-task Benchmark for Comprehensive Assessment of Language Model Bias. Vipul Gupta, Pranav Narayanan Venkit, Hugo Laurençon, Shomir Wilson, and Rebecca J. Passonneau. In *The 1st Conference on Language Modeling (COLM)*, 2024.

Creation and Analysis of an International Corpus of Privacy Laws. Sonu Gupta, Geetika Gopi, Harish Balaji, Ellen Poplavska, Nora O’Toole, Siddhant Arora, Thomas Norton, Norman Sadeh, and Shomir Wilson. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING)*, 2024.

Understanding How to Inform Blind and Low-Vision Users about Data Privacy through Privacy Question Answering Assistants. Yuanyuan Feng, Abhilasha Ravichander, Yaxing Yao, Shikun Zhang, Rex Chen, Shomir Wilson, and Norman Sadeh. In *Proceedings of the 33rd USENIX Security Symposium (USENIX)*, 2024.

🏆 The Sentiment Problem: A Critical Survey towards Deconstructing Sentiment Analysis. Pranav Narayanan Venkit, Mukund Srinath, Sanjana Gautam, Saranya Venkatraman, Vipul Gupta, Rebecca J. Passonneau, and Shomir Wilson. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023. **Outstanding Paper Award.**

🏆 Privacy Lost and Found: An Investigation at Scale of Web Privacy Policy Availability. Mukund Srinath, Soundarya Nurani Sundareswara, Pranav Venkit, C. Lee Giles, and Shomir Wilson. In *Proceedings of the 23rd ACM Symposium on Document Engineering (DocEng)*, 2023. **Best Student Paper Award.**

Privacy Now or Never: Large-Scale Extraction and Analysis of Dates in Privacy Policy Text. Mukund Srinath, Lee Matheson, Pranav Venkit, Gabriela Zafir-Fortuna, Florian Schaub, C. Lee Giles and Shomir Wilson. In *Proceedings of the 23rd ACM Symposium on Document Engineering (DocEng)*, 2023.

Unmasking Nationality Bias: A Study of Human Perception of Nationalities in AI-Generated Articles. Pranav Narayanan Venkit, Sanjana Gautam, Ruchi Panchanadikar, Ting-Hao Huang, and Shomir Wilson. In *Proceedings of the Sixth AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES)*, 2023.

Creation and Analysis of a Corpus of Scam Emails Targeting Universities. Grace Ciambrone and Shomir Wilson. In *Companion Proceedings of the ACM Web Conference (WebConf)*, 2023.

Nationality Bias in Text Generation. Pranav Narayanan Venkit, Sanjana Gautam, Ruchi Panchanadikar, Ting-Hao Kenneth Huang, and Shomir Wilson. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 2023.

A Study of Implicit Language Model Bias against People with Disabilities. Pranav Venkit, Mukund Srinath, and Shomir Wilson. In *Proceedings of the 29th International Conference on Computational Linguistics (COLING)*, 2022.

STAPI: An Automatic Scraper for Extracting Iterative Title-Text Structure from Web Documents. Nan Zhang, Shomir Wilson, and Prasenjit Mitra. In *Proceedings of the 13th Language Resources and Evaluation Conference (LREC)*, 2022.

A Tale of Two Regulatory Regimes: Creation and Analysis of a Bilingual Privacy Policy Corpus. Siddhant Arora, Henry Hosseini, Christine Utz, Vinayshekhar Bannihatti Kumar, Tristan O. Dhellemmes, Abhilasha Ravichander, Peter Story, Jasmine Mangat, Rex Chen, Martin Degeling, Thomas Norton, Thomas Hupperich, Shomir Wilson, and Norman Sadeh. In *Proceedings of the 13th Language Resources and Evaluation Conference (LREC)*, 2022.

Automated Detection of Doxing on Twitter. Younes Karimi, Anna Squicciarini, and Shomir Wilson. In *Proceedings of the 25th ACM Conference on Computer-Supported Cooperative Work And Social Computing (CSCW)*, 2022.

Privacy at Scale: Introducing the PrivaSeer Corpus of Web Privacy Policies. Mukund Srinath, Shomir Wilson, and C. Lee Giles. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2021.

Breaking Down Walls of Text: How Can NLP Benefit Consumer Privacy? Abhilasha Ravichander, Alan W Black, Thomas Norton, Shomir Wilson, and Norman Sadeh. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2021.

PrivaSeer: A Privacy Policy Search Engine. Mukund Srinath, Soundarya Nurani Sundareswara, C. Lee Giles, and Shomir Wilson. In *Proceedings of the 21st International Conference on Web Engineering (ICWE)*, 2021.

A Large-Scale Exploration of Terms of Service Documents on the Web. Soundarya Nurani Sundareswara, Mukund Srinath, Shomir Wilson and C. Lee Giles. In *Proceedings of the 21st ACM Symposium on Document Engineering (DocEng)*, 2021.

From Prescription to Description: Mapping the GDPR to a Privacy Policy Corpus Annotation Scheme. Ellen Poplavska, Thomas B. Norton, Shomir Wilson, and Norman Sadeh. In *Proceedings of the 33rd International Conference on Legal Knowledge and Information Systems (JURIX)*, December 9-11, 2020.

Finding a Choice in a Haystack: Automatic Extraction of Opt-Out Statements from Privacy Policy Text. Vinayshekhar Bannihatti Kumar, Roger Iyengar, Namita Nisal, Yuanyuan Feng, Hana Habib, Peter Story, Sushain Cherivirala, Margaret Hagan, Lorrie Cranor, Shomir Wilson, Florian Schaub and Norman Sadeh. In *Proceedings of The Web Conference (WWW)*, April 20-24, 2020.

Question Answering for Privacy Policies: Combining Computational and Legal Perspectives. Abhilasha Ravichander, Alan Black, Shomir Wilson, Thomas Norton, and Norman Sadeh. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, November 3-7, 2019.

Vaccine: Obfuscating Access Pattern Against File-Injection Attacks. Hao Liu, Boyang Wang, Nan Niu, Shomir Wilson, and Xuetao Wei. In *Proceedings of the IEEE Conference on Communications and Network Security (CNS)*, June 10-12, 2019.

Supervised and Unsupervised Methods for Robust Separation of Section Titles and Prose Text in Web Documents. Abhijith Athreya Mysore Gopinath, Shomir Wilson, and Norman Sadeh. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Brussels, Belgium, November 2018.

Identifying the provision of choices in privacy policy text. Kanthashree Sathyendra, Shomir Wilson, Florian Schaub, Norman Sadeh, and Sebastian Zimmeck. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Copenhagen, Denmark, September 2017.

Automated analysis of privacy requirements for mobile apps. Sebastian Zimmeck, Ziqi Wang, Lieyong Zou, Roger Iyengar, Bin Liu, Florian Schaub, Shomir Wilson, Norman Sadeh, Steven M. Bellovin, and Joel Reidenberg. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, San Diego, California, March 2017.

The creation and analysis of a website privacy policy corpus. Shomir Wilson, Florian Schaub, Aswarth Abhilash Dara, Frederick Liu, Sushain Cherivirala, Pedro Giovanni Leon, Mads Schaarup Andersen, Sebastian Zimmeck, Kanthashree Mysore Sathyendra, N. Cameron Russell, Thomas B. Norton, Eduard Hovy, Joel Reidenberg, and Norman Sadeh. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, Berlin, Germany, August 2016.

🏆 Crowdsourcing annotations of websites' privacy policies: Can it really work? Shomir Wilson, Florian Schaub, Rohan Ramanath, Norman Sadeh, Fei Liu, Noah Smith and Frederick Liu. In *Proceedings of the 25th International World Wide Web Conference (WWW)*, Montreal, Canada, April 2016. **Best Paper Finalist.**

This table is different: A WordNet-based approach to identifying references to document entities. Shomir Wilson, Alan W Black, and Jon Oberlander. In *Proceedings of the 8th International Global WordNet Conference (GWC)*, Bucharest, Romania, January 2016.

Identifying relevant text fragments to help crowdsource privacy policy annotations. Rohan Ramanath, Florian Schaub, Shomir Wilson, Fei Liu, Norman Sadeh, and Noah Smith. In *Proceedings of the Second AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*, works-in-progress track, Pittsburgh, PA, November 2014.

Determiner-established deixis to communicative artifacts in pedagogical text. Shomir Wilson and Jon Oberlander. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, Baltimore, MD, June 2014.

Toward automatic processing of English metalanguage. Shomir Wilson. In *Proceedings of the 6th International Joint Conference on Natural Language Processing (IJCNLP)*, Nagoya, Japan, October 2013.

Privacy manipulation and acclimation in a location sharing application. Shomir Wilson, Justin Cranshaw, Norman Sadeh, Alessandro Acquisti, Lorrie Cranor, Jay Springfield, Sae Young Jeong, and Arun Balasubramanian. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, Zurich, Switzerland, September 2013.

Tweets are forever: A large-scale quantitative analysis of deleted tweets. Hazim Almuhiemi, Shomir Wilson, Bin Liu, Norman Sadeh, and Alessandro Acquisti. In *Proceedings of the 2013 ACM Conference on Computer Supported Cooperative Work (CSCW)*, San Antonio, TX, February 2013.

The creation of a corpus of English metalanguage. Shomir Wilson. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL)*, Jeju, South Korea, July 2012.

Application of MCL in a dialog agent. Darsana Josyula, Scott Fults, Michael L. Anderson, Shomir Wilson, and Don Perlis. In *Papers from the Third Language and Technology Conference (LTC)*, Poznań, Poland, October 2007.

Peer-Reviewed Hybrid Conference-Journal Articles

🔴 Race and Privacy in Broadcast Police Communications. Pranav Venkit, Christopher Graziul, Miranda Goodman, Samantha Kenny, and Shomir Wilson. In *Proceedings of the ACM on Human-Computer Interaction*, 2024. **Diversity, Equity, and Inclusion Recognition**. Conference: **CSCW**.

🔴 Researchers' Experiences in Analyzing Privacy Policies: Challenges and Opportunities. Abraham Mhaidli, Selin Fidan, An Doan, Gina Herakovic, Mukund Srinath, Lee Matheson, Shomir Wilson, and Florian Schaub. In *Proceedings on Privacy Enhancing Technologies (PoPETs)* 2023(4), 2023. **Best Student Paper Award**. Conference: **PETS**.

Digital Inequality Through the Lens of Self-Disclosure. Jooyoung Lee, Sarah Rajtmajer, Eesha Srivatsavaya, and Shomir Wilson. In *Proceedings on Privacy Enhancing Technologies (PoPETs)* 2021(3), 2021. Conference: **PETS**.

In search of the use-mention distinction and its impact on language processing tasks. Shomir Wilson. In *The International Journal of Computational Linguistics and Applications* 2(1-2), pp. 139-154, 2011. Conference: **CICLing**.

Peer-Reviewed Journal Articles

Online Self-Disclosure, Social Support, and User Engagement During the COVID-19 Pandemic. Jooyoung Lee, Sarah Rajtmajer, Eesha Srivatsavaya, and Shomir Wilson. *ACM Transactions on Social Computing (TSC)*, 2023.

Analyzing privacy policies at scale: From crowdsourcing to automated annotations. Shomir Wilson, Florian Schaub, Frederick Liu, Kanthashree Mysore Sathyendra, Sebastian Zimmeck,

Rohan Ramanath, Fei Liu, Norman Sadeh, and Noah Smith. In *ACM Transactions on the Web (TWEB)* 13(1), December 2018.

PrivOnto: A semantic framework for the analysis of privacy policies. Alessandro Oltramari, Dhivya Piraviperumal, Florian Schaub, Shomir Wilson, Sushain Cherivirala, Thomas B. Norton, N. Cameron Russell, Peter Story, Joel Reidenberg, and Norman Sadeh. In *Semantic Web Journal (SWJ)*, May 2017.

Nudges for privacy and security: Understanding and assisting users' choices online. Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, Yang Wang, and Shomir Wilson. In *ACM Computing Surveys (CSUR)* 50(3), August 2017.

Book Chapters

Professor in a Strange Land: An Expat Metaphor for Classroom Discussions. Shomir Wilson. In John M. Carroll (ed.), *Innovative Practices in Teaching Information Sciences and Technology: Further Experience Reports and Reflections*. Springer, 2024.

Nudges (and Deceptive Patterns) for Privacy: Six Years Later. Alessandro Acquisti, Idris Adjerid, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Yang Wang, and Shomir Wilson. In Sabine Trepte, Philipp Masur (Ed.), *The Routledge Handbook of Privacy and Social Media*. Taylor & Francis, 2023.

A bridge from the use-mention distinction to natural language processing. Shomir Wilson. In Saka, P., Johnson, M. (Ed.), *The Semantics and Pragmatics of Quotation*. Springer, 2017.

The metacognitive loop and reasoning about anomalies. Matthew Schmill, Michael L. Anderson, Scott Fults, Darsana Josyula, Tim Oates, Donald Perlis, Hamid Haidarian Shahri, Shomir Wilson, and Dean Wright. In Cox, M., Raja, A. (Ed.), *Metareasoning: Thinking About Thinking*. MIT Press, MA, 2010.

Magazine Articles

Reports on the 2019 AAAI Spring Symposium Series (Privacy-Enhancing Artificial Intelligence and Language Technologies). Shomir Wilson, et al. **AI Magazine** 40:3, Fall 2019.

Reports on the 2016 AAAI Fall Symposium Series (Privacy and Language Technologies). Patrícia Alves-Oliveira, Richard G. Freedman, Dan Grollman, Laura Herlant, Laura Humphrey, Fei Liu, Ross Mead, Frank Stein, Tom Williams, and Shomir Wilson. **AI Magazine** 38:2, Summer 2017.

A self-help guide for autonomous systems. Michael L. Anderson, Scott Fults, Darsana P. Josyula, Tim Oates, Don Perlis, Matthew D. Schmill, Shomir Wilson, and Dean Wright. **AI Magazine** 29:2, Summer 2008.

Peer-Reviewed Conference Poster Abstracts

The Role of Metalanguage in Prompt Injection Attacks. Lineesha Kamana and Shomir Wilson. In the Mid-Atlantic Student Colloquium on Speech, Language and Learning (**MASC-SLL**), 2025.

Questioning Privacy: Contrasting User Questions with Questions Answered by Privacy Policies. Tianyang Zhao, Younes Karimi, Thomas Norton, Norman Sadeh, Shomir Wilson, and Kami Vaniea. In the Mid-Atlantic Student Colloquium on Speech, Language and Learning (**MASC-SLL**), 2025.

Comparing Scam Emails and Email User Education at Universities. Duo Pan, Ellen Poplavska, Nora O'Toole, and Shomir Wilson. In *The Seventeenth Symposium on Usable Privacy and Security (unpublished work)* (**SOUPS**), 2021.

A Multilingual Comparison of Email Scams. Duo Pan, Ellen Poplavska, Yichen Yu, Susan Strauss, and Shomir Wilson. In *The Sixteenth Symposium on Usable Privacy and Security (unpublished work)* (**SOUPS**), 2020.

Automatic Title Generation to Improve the Readability of Privacy Policies. Abhijith Athreya Mysore Gopinath, Vinayshekhar Bannihatti Kumar, Shomir Wilson, Norman Sadeh. In *The Sixteenth Symposium on Usable Privacy and Security (unpublished work)* (**SOUPS**), 2020.

Privacy Not Found: A Study of the Availability of Privacy Policies on the Web. Soundarya Nurani Sundareswara, Shomir Wilson, Mukund Srinath, C. Lee Giles. In *The Sixteenth Symposium on Usable Privacy and Security (unpublished work)* (**SOUPS**), 2020.

Question Answering for Privacy Policies: Combining Computational and Legal Perspectives. Abhilasha Ravichander, Alan Black, Shomir Wilson, Thomas Norton, and Norman Sadeh. In *The Sixteenth Symposium on Usable Privacy and Security (published work)* (**SOUPS**), 2020.

Increasing the salience of data use opt-outs online. Namita Nisal, Sushain K. Cherivirala, Kanthashree M. Sathyendra, Margaret Hagan, Florian Schaub, Shomir Wilson, Lorrie Faith Cranor, and Norman Sadeh. In *The Thirteenth Symposium on Usable Privacy and Security (unpublished work)* (**SOUPS**), Santa Clara, CA, 2017.

Mobile app privacy compliance: Automated technology to help regulators, app stores and developers. Sebastian Zimmeck, Lieyong Zou, Bin Liu, Shomir Wilson, Steven M. Bellovin, Ziqi Wang, Roger Iyengar, Florian Schaub, Norman Sadeh, and Joel Reidenberg. In *The Thirteenth Symposium on Usable Privacy and Security (published work)* (**SOUPS**), Santa Clara, CA, 2017.

Visualization and interactive exploration of data practices in privacy policies. Sushain K. Cherivirala, Florian Schaub, Mads Schaarup Andersen, Shomir Wilson, Norman Sadeh, and Joel R. Reidenberg. In *The Twelfth Symposium on Usable Privacy and Security (unpublished work)* (**SOUPS**), Denver, CO, 2016.

Towards usable privacy policies: Semi-automatically extracting data practices from websites' privacy policies. Norman Sadeh, Alessandro Acquisti, Travis Breaux, Lorrie Cranor, Aleecia McDonald, Joel Reidenberg, Noah Smith, Fei Liu, N. Cameron Russell, Florian Schaub, Shomir Wilson, James Graves, Pedro Leon, Rohan Ramanath, and Ashwini Rao. In *The Tenth Symposium on Usable Privacy and Security (published work)* (**SOUPS**), Palo Alto, CA, 2014.

Peer-Reviewed Workshop Proceedings

📌 Documenting the Unwritten Curriculum of Student Research. Shomir Wilson. In *Proceedings of the Sixth Workshop on Teaching NLP (TeachNLP)*, 2024. **Best Teaching Material Award.**

Sociodemographic Bias in Language Models: A Survey and Forward Path. Vipul Gupta, Pranav Narayanan Venkit, Shomir Wilson, Rebecca J. Passonneau. In *Proceedings of the 5th Workshop on Gender Bias in Natural Language Processing (GeBNLP)*, 2024.

📌 Automated Ableism: An Exploration of Explicit Disability Biases in AIaaS Sentiment and Toxicity Analysis Models. Pranav Narayanan Venkit, Mukund Srinath, and Shomir Wilson. In *Proceedings of the Third Workshop on Trustworthy Natural Language Processing (TrustNLP)*, 2023. **Best Short Paper Award.**

Effects of Online Self-Disclosure on Receiving Social Support During the COVID-19 Pandemic. Jooyoung Lee, Sarah Rajtmajer, Eesha Srivatsavaya, and Shomir Wilson. In *1st Workshop on NLP for Positive Impact* (unpublished papers) at ACL, 2021.

Demystifying Privacy Policies with Language Technologies: Progress and Challenges. Shomir Wilson, Florian Schaub, Aswarth Dara, Sushain K. Cherivirala, Sebastian Zimmeck, Mads Schaarup Andersen, Pedro Giovanni Leon, Eduard Hovy, and Norman Sadeh. In *Proceedings of the Workshop on Text Analytics for Cybersecurity and Online Safety (TA-COS) at LREC*, Portoroz, Solvenia, May 2016.

Distinguishing use and mention in natural language. Shomir Wilson. In *Proceedings of the NAACL HLT Student Research Workshop*, Los Angeles, CA, June 2010.

The role of metacognition in robust AI systems. Matt Schmill, Tim Oates, Michael L. Anderson, Darsana Josyula, Don Perlis, Shomir Wilson, and Scott Fults. In *Papers from the Workshop on Metareasoning at the 23rd AAAI Conference on Artificial Intelligence*, Chicago, IL, July 2008.

Ontologies for reasoning about failures in AI systems. Michael L. Anderson, Scott Fults, Darsana Josyula, Tim Oates, Don Perlis, Matt Schmill, and Shomir Wilson. In *Proceedings of the First International Workshop on Metareasoning in Agent-Based Systems*, Honolulu, HI, 2007.

Other Peer-Reviewed Papers

Automated Detection and Analysis of Data Practices Using A Real-World Corpus. Mukund Srinath, Pranav Venkit, Maria Badillo, Florian Schaub, C. Lee Giles, and Shomir Wilson. In *Findings of the Association for Computational Linguistics: ACL 2024*.

Analyzing vocabulary intersections of expert annotations and topic models for data practices in privacy policies. Frederick Liu, Shomir Wilson, Florian Schaub and Norman Sadeh. In *Proceedings of the AAAI Fall Symposium on Privacy and Language Technologies*, Arlington, VA, November 2016.

Automatic extraction of opt-out choices from privacy policies. Kanthashree Mysore Sathyendra, Florian Schaub, Shomir Wilson and Norman Sadeh. In *Proceedings of the AAAI Fall Symposium on Privacy and Language Technologies*, Arlington, VA, November 2016.

Analyzing and predicting privacy law compliance of mobile apps. Sebastian Zimmeck, Ziqi Wang, Lieyong Zou, Roger Iyengar, Bin Liu, Florian Schaub, Shomir Wilson, Norman Sadeh, Steven M. Bellovin and Joel Reidenberg. In *Proceedings of the AAAI Fall Symposium on Privacy and Language Technologies*, Arlington, VA, November 2016.

The Metacognitive Loop: An architecture for building robust intelligent systems. Hamid Haidarian, Wikum Dinalankara, Scott Fults, Shomir Wilson, Don Perlis, Matt Schmill, Tim Oates, Darsana Josyula, and Michael Anderson. In *Proceedings of the AAAI Fall Symposium on Commonsense Knowledge*, Arlington, VA, November 2010.

Toward domain-neutral human-level metacognition. Michael L. Anderson, Matt Schmill, Tim Oates, Don Perlis, Darsana Josyula, Dean Wright, and Shomir Wilson. In *Proceedings of the 2007 AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning*, Palo Alto, CA, March 2007.

Dissertation

A Computational Theory of the Use-Mention Distinction in Natural Language. Shomir Wilson. University of Maryland, 2011.

Written Testimony

Contribution to Hearing 118-179, “Modern Scams: How Scammers Are Using Artificial Intelligence & How We Can Fight Back”. United States Senate Special Committee on Aging, November 2023.

Reports

Case studies and user interaction. Zinaida Benenson, Abdullah Elbi, Zekeriya Erkin, Natasha Fernandes, Simone Fischer-Hübner, Ivan Habernal, Els Kindt, Anna Leschanowsky, Pierre Lison, Christina Lohr, Emily Mower Provost, Jo Pierson, David Stevens, Francisco Teixeira, and Shomir Wilson. In *Privacy in Speech and Language Technology (Dagstuhl Seminar 22342)*, 2023.

Towards Automatic Classification of Privacy Policy Text. Frederick Liu, Shomir Wilson, Peter Story, Sebastian Zimmeck, and Norman Sadeh. Technical Report CMU-LTI-17-010, Carnegie Mellon University, 2017.

The Usable Privacy Policy Project: Combining crowdsourcing, machine learning and natural language processing to semi-automatically answer those privacy questions users care about. Norman Sadeh, Alessandro Acquisti, Travis Breaux, Lorrie Cranor, Aleecia McDonald, Joel Reidenberg, Noah Smith, Fei Liu, N. Cameron Russell, Florian Schaub, and Shomir Wilson. Technical Report CMU-ISR-13-119, Carnegie Mellon University, 2013.

Automatic categorization of privacy policies: A pilot study. Waleed Ammar, Shomir Wilson, Norman Sadeh, and Noah A. Smith. Technical Report CMU-LTI-12-019 / CMU-ISR-12-114, Carnegie Mellon University, 2012.

Other Papers

A ‘Sourceful’ Twist: Emoji Prediction Based on Sentiment, Hashtags and Application Source. Pranav Venkit, Zeba Karishma, Chi-Yang Hsu, Rahul Katiki, Kenneth Huang, Shomir Wilson, and Patrick Dudas. arXiv:2103.07833, 2021.

An active logic approach to Moore’s Paradox. Shomir Wilson. Scholarly paper for M.S. in Computer Science, 2008.

Evaluation of functional-linkage networks applied to protein annotation. Shomir Wilson. Honors thesis for B.S. in Computer Science, 2005.

Wittgenstein takes the Turing Test. Shomir Wilson. Selected for presentation at The First Undergraduate Philosophy Conference at Northwestern University, 2005.

The pathological liar: An exclusionary approach to self-referential contradictions in natural language. Shomir Wilson. *Aporia* 14(2), 2004.

Construction of a crystal graph simulation engine. Shomir Wilson. Honors thesis for B.S. in Mathematics, 2004.

INVITED ACTIVITIES

Invited Talks for Conferences and Similar Events

Keynote at Carnegie Bosch Institute Symposium on Responsible and Sustainable AI: “Case Studies of Sociodemographic Biases by Artificial Intelligence Models”, Carnegie Mellon University, 2025-03-11.

Postdoctoral Research Symposium Keynote: “Success, Failure, and Meaning as a Postdoc”. Penn State, 2024-12-05.

SEPLN 2024 Keynote: “Natural Language Processing for Privacy Empowerment and Fairness”. Valladolid, Spain, 2024-09-26.

PrivateNLP 2024 Keynote: “Understanding Privacy Through the Lens of NLP”. Bangkok, Thailand, 2024-08-16.

HCOMP 2016 Encore Track: “Crowdsourcing Annotations of Websites’ Privacy Policies: Can It Really Work?”. Austin, Texas, 2016-11-01.

Invited Talks for University Seminars

Brandeis University, Computer Science Department: “Natural Language Processing for Privacy and Social Good”, 2023-10-26.

Auburn University, AI@AU Forum: “Natural Language Processing for Privacy, Empowerment, and Social Good”, 2023-09-22.

Pennsylvania State University, Center for Language Science, Language Meets Technology Speaker Series: “Sociodemographic Biases in Natural Language Processing: Two Case Studies”, 2023-02-10.

Carnegie Mellon University, Privacy Seminar: “Natural Language Processing for Privacy and Social Good”, 2023-01-31.

University of Maine, Privacy Engineering - Regulatory Compliance Lab: “Natural Language Processing for Privacy and Beyond”, 2022-04-21.

Dalhousie University, Computer Science Seminar: “Human Language Technologies for Understanding Online Privacy”, 2018-12-05.

Pennsylvania State University, College of Information Sciences and Technology, Data Sciences Seminar: “Human Language Technologies for Understanding Online Privacy”, 2018-11-27.

Northern Kentucky University, College of Informatics: “Natural Language Processing for the Privacy of Internet Users”, 2017-03-24.

Cincinnati Children’s Hospital Medical Center, Division of Biomedical Informatics, Hutton Lecture Series: “Text Analysis to Support the Privacy of Internet Users”, 2017-02-17.

University of Dayton, Computer Science Department Colloquium Series: “Natural Language Processing to Support Internet Privacy”, 2017-02-03.

Georgetown University, Computer Science Colloquium Series: “Text Analysis to Support the Privacy of Internet Users”, 2016-11-21.

National University of Singapore, CHIME Text Seminar: “Introspective Users and Introspective Text: Some Recent Results”, 2016-01-05.

Cambridge University, NLIP Seminar Series: “Identifying Deixis to Communicative Artifacts in Text”, 2014-05-09.

University of Edinburgh, ILCC/HCRC Seminar Series: “An Empirical Approach to Metalanguage”, 2013-09-06.

Carnegie Mellon University, CL+NLP Lunch: “A Computational Approach to Metalanguage and the Use-Mention Distinction”, 2013-04-23.

Macquarie University, Centre for Language Technology Seminar Series: “Distinguishing Use and Mention in Natural Language”, 2009.

Other Invited Activities

Academic attendee hosted by Salesforce at Dreamforce 2024. San Francisco, CA, 2024-09-(17-18).

Panelist at “AI | ALL Summit: Future of Equitable and Inclusive AI”. Bangkok, Thailand, 2024-08-16.

Panelist for “Navigating the Promises and Pitfalls of Generative AI”, Penn State Institute for Computational and Data Sciences Symposium, 2023-10-19.

Participant in CRA LEVEL UP Workshop (event to gather best practices for inclusive undergraduate computing education), 2023-08-(07-08).

Participant in NSF SaTC Vision 2.0 Workshop, 2023-03-(08-09).

Panelist for “Tenure-Track Interview Panel Discussion for Postdocs”, organized by Penn State’s Office of the Senior Vice President for Research, 2023-01-20.

Participant in Dagstuhl Seminar “Privacy in Speech and Language Technology”, 2022-08-(21-26).

Panelist for Forums on “Meta-Level Control” and “Models of Self”, during the Metareasoning Workshop at the Twenty-Third AAAI Conference on Artificial Intelligence, 2008.

MISCELLANY

Penn State Emerging Academic Leaders Program (PSEAL), 2024.

Computing Research Association Future Leaders Program, 2024.

Professional Memberships: AAAI, ACL, ACM.

AED/CPR training, 2023, 2021, 2019.

QPR training (for suicide intervention), 2020.

Safer People, Safer Places training (for LGBT+ inclusiveness), 2018.

Opioid Overdose Prevention training, 2017.

Bias Busters training (for implicit bias identification and intervention), 2015.

Licensed amateur radio operator, technician class.