

Tarini Saka

Postdoctoral Researcher, Human-Centered Security and Privacy
Max Planck Institute for Security and Privacy (MPI-SP)

CONTACT INFORMATION

E-mail tarinisaka.work@gmail.com
Website <https://tarinisaka.github.io/>
LinkedIn www.linkedin.com/in/tarini-saka/
Google Scholar [Tarini Saka](#)

RESEARCH EXPERIENCE

- Postdoctoral Researcher** *Human-Centered Security and Privacy, Max Planck Institute for Security and Privacy*
Jan. 2026 – Present
- Critical examination of how generative AI amplifies traditional user scams through enhanced social engineering and automation.
 - Preparing a grant proposal to the Volkswagen Foundation (Security and Technology programme) on emerging AI-enabled threat vectors and user protection.
- Postdoctoral Researcher** *Cluster of Excellence CASA, Ruhr University Bochum*
Jan. 2025 – Dec. 2025
- Investigated how language barriers affect email processing and phishing detection through lab-based user studies to inform the design of more inclusive mail interfaces for multilingual users.
 - Investigated adoption and governance practices of LLM tools in organisations to identify risks, gaps, and expected practices.
 - Actively contributed to the research community through conference and workshop reviewing, workshop organisation, and service roles.
 - Supervised and mentored students across seminar, BSc, and MSc levels, providing guidance on research design, academic writing, and critical evaluation throughout their projects.
- Doctoral Researcher** *Artificial Intelligence and Applications Institute, The University of Edinburgh*
Apr. 2021 – Nov. 2024
- Systematized phishing research, identifying limitations of traditional text representations and gaps in contextual NLP feature usage.
 - Designed the Phishing Codebook to derive a novel contextual feature set and created a 503-email high-reliability annotated dataset for downstream phishing mitigation tasks.
 - Developed PhishCoder using fine-tuned transformer-based language models for multi-task information extraction via text classification and question answering.
 - Applied ensemble clustering, text similarity, and graph-based clustering techniques to identify phishing campaigns from contextual feature representations.
 - Collaborated with IT security professionals to design a first-of-its-kind phishing profiling and user-guidance tool for real-world mitigation.

UNIVERSITY EDUCATION

Apr. 2021 – Nov. 2024

Doctor of Philosophy,

Artificial Intelligence and Applications Institute,
School of Informatics, The University of Edinburgh,
Thesis: A Collaborative Human-AI Approach to Mitigate Phishing Attacks

Aug. 2015 – June 2020

Master's and Bachelor's of Science in Mathematics,

School of Mathematics,
Indian Institute of Science Education and Research – Thiruvananthapuram
CGPA: 8.52/10 (UK First-Class equivalent)

TEACHING EXPERIENCE

Teaching Assistant

Modules: Human-Computer Interaction; Usable Security and Privacy
University of Edinburgh

Assisted professors with classroom management and preparation of hybrid sessions; contributed to curriculum design and development of coursework and grading rubrics; evaluated student performance and provided actionable feedback to enhance learning outcomes.

Laboratory Demonstrator

Modules: Machine Learning Practical; Foundations of Natural Language Processing; Informatics 2D - Reasoning and Agents
University of Edinburgh

Delivered hands-on guidance in practical sessions; clarified complex concepts; mentored students in applying theoretical knowledge to real-world problems; provided targeted support to strengthen technical proficiency.

Examination Marker

Modules: Human-Computer Interaction; Usable Security and Privacy; Discrete Mathematics and Probability
University of Edinburgh

Assessed coursework and examinations with precision; ensured fair and consistent grading aligned with established rubrics; provided constructive feedback to students; worked with instructors to refine evaluation standards.

STUDENT SUPERVISION

Ruhr University Bochum

"Privacy Assessment of Speech-to-Text Mobile Applications: Combining Technical and Policy Analysis." by Maike Bendel, 2025. [Master thesis]

"Examining the Security of AI Email Assistants against Prompt Injection Attacks." by Julia Leckebusch, 2025. [Bachelor thesis]

"Responsible LLM Deployment: Understanding Training, Guidelines, and Employee Readiness in Organizations." by Sandy Hong, 2025. [Master thesis]

"Empowering Users in the AI Era: Communicating Security and Privacy Risks of Generative AI Tools" by Aleksandr Popok, 2025. [Seminar Paper]

"Creating A Framework for Evaluating Speech-to-Text Mobile Applications" by Maike Bendel, 2025. [Seminar Paper]

University of Edinburgh

"Using Large Language Models (LLM) Based Techniques to Mitigate Social Engineering Attacks." by Kalliopi Vakali, on-going, 2025. [Doctoral thesis]

"Impact of Real-Time Tailored AI-Guidance for Phishing on Users." by Kalliopi Vakali, 2025. [Master thesis]

"Evaluating the Impact of an Automated Email Analysis System on Phishing Susceptibility" by Sean Strain, 2023. [Master thesis]

PROFESSIONAL ACTIVITIES

Academic (Sub-) Reviewer

Computers & Security, ACM CCS 2026, ICML (Ethics reviewer) 2026, SOUPS 2026, NDSS 2026, DIMVA 2025–2026, ACSAC 2025, WOOT 2025, CHI-LBW 2024–2025

Conference Organization

- 2026 PC Member and Organizer of the *1st Workshop on Designing Human-AI Teaming for Security (HATS)* (co-located with SOUPS 2026).
- 2026 Publication Chair of the *1st Security and Privacy for Asian Internet Communities Workshop (SPAIC)* (co-located with AsiaCCS 2026).
- 2026 Poster Co-Chair of the *23rd Conference on Detection of Intrusions and Malware & Vulnerability Assessment (DIMVA '26)*.
- 2025 Poster Chair of the *22nd Conference on Detection of Intrusions and Malware & Vulnerability Assessment (DIMVA '25)*, Graz University of Technology, Austria.

Other Academic Service

- 2025–Present Founding Committee Member, *International Social Club of the Horst Görtz Institute (HGI)*, Ruhr University Bochum. Established a structured support network for international students and researchers; coordinated with administrative, financial, and international offices; and led needs assessment and outreach.
<https://hgi.rub.de/international-social-club>

AWARDS & SCHOLARSHIPS

- 2025 Received the Reviewer Award at the Annual Computer Security Applications Conference (ACSAC) for high-quality peer review.
- 2025 Selected for a *HIIT Postdoctoral Fellowship in Computer Science*, a fully funded postdoctoral fellow position for up to three years at the Helsinki Institute for Information Technology (HIIT); Offer declined due to timing.
<https://www.hiit.fi/hiit-postdoctoral-fellow-positions-in-computer-science>
- 2025 Received the Honorable Mention Award at ACM CHI conference on Human Factors in Computing Systems for our paper "*Judging Phishing Under Uncertainty: How Do Users Handle Inaccurate Automated Advice?*" (First Author). Awarded to the top 5% of submissions.
<https://programs.sigchi.org/chi/2025/program/content/189664>
- 2025 Finalist for 2025 SICSA Best PhD Dissertation Award for my PhD Dissertation. Awarded by Scottish Informatics and Computer Science Alliance (SICSA).
<https://sicsaconf.org/>
- 2024 Contributor on Excellence Cluster CASA Fundamental Research Project, *Improving IT Security and Privacy for Deaf and Hard-of-Hearing People* Funded for a one-year Postdoctoral Fellowship through the grant.
- 2021–2024 School of Informatics PhD Scholarship
Awarded by the University of Edinburgh's School of Informatics to support doctoral research in Automated Reasoning, Agents, Data Intensive Research, and Knowledge Management. Covered full tuition fees (£70,500), additional programme costs (£3,000), and provided a stipend of £15,285 per year (with annual increments) for 3.5 years.
<https://informatics.ed.ac.uk/>
- 2015–2020 INSPIRE Scholarship for Higher Education (SHE)
Awarded by the Government of India's Department of Science and Technology to meritorious students in the top 1% of the Class XII examination conducted by the Central Education Board in India. Awarded a five-year stipend to pursue an M.Sc. in Mathematics.
<https://online-inspire.gov.in/>

TECHNICAL SKILLS

Languages	Python (NumPy, Pandas, Scikit-learn), C, SQL, R
Mathematics and Statistics	Linear Algebra, Real Analysis, Probability Theory, Bayesian Inference, Hypothesis Testing, Statistical Modeling
Machine Learning	Supervised and Unsupervised Learning, Feature Engineering, Ensemble Methods (Random Forests, Gradient Boosting), Model Evaluation (Cross-Validation, ROC/AUC), PyTorch, TensorFlow
Deep Learning	Neural Network Architectures (CNNs, RNNs, Transformers), Transfer Learning, Hyperparameter Optimization, Regularization Techniques, GPU Acceleration (CUDA)
Natural Language Processing	Transformer Models (BERT, GPT, T5), Fine-Tuning, Tokenization Strategies, Sequence-to-Sequence Models, Named Entity Recognition, Language Modeling, Hugging Face Transformers
Human-Computer Interaction	User-Centered Design, Usability Testing, Heuristic Evaluation, Qualitative Research (Interviews, Thematic Analysis), Quantitative Research (Survey Design with Qualtrics/Forms, Hypothesis Testing, p-value Analysis, Statistical Validation)

List of Publications

- [P1] **Tarini Saka**, Kalliopi Vakali, Conor O'Shea, Nadin Kokciyan, Kami Vaniea. 2026. *Poster: Toward Task-Aware Anonymization of Phishing Emails for Research Data Sharing*. Under review in the 19th ACM Workshop on Artificial Intelligence and Security (AISec'26). Association for Computing Machinery.
- [P2] Kalliopi Vakali, **Tarini Saka**, Kami Vaniea, Nadin Kokciyan. 2026. *An Analysis of Factors Affecting LLM-based Phishing Detection*. Under review in the 19th ACM Workshop on Artificial Intelligence and Security (AISec'26). Association for Computing Machinery.
- [P3] Nicole Meng-Schneider, **Tarini Saka**, Adam Jenkins, Maria K Wolters, Kami Vaniea. 2026. *"Well, it just is not right for them to use it" - A Survey of Perceived Comfort with Sharing Smart Speakers*. Under review in the ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW).
- [P4] **Tarini Saka**. 2026. *Keeping Humans in Practice: Skill Atrophy as a Design Constraint for Human-AI Teaming in Cybersecurity*. To appear in the 2026 Workshop on Designing Human-AI Teaming for Security (HATS) at SOUPS.
- [P5] **Tarini Saka**, Sandy Hong, Veelasha Moonsamy. 2026. *When LLM Governance Falls to Employees: Informal Practices of Responsible AI Use at Work*. To appear in the AAAI Conference on AI, Ethics, and Society (AIES 2026).
- [P6] Sangavi Shanthakumar, **Tarini Saka**, Jens Christian Opdenbusch, Markus Schöps, Marco Gutfleisch. 2026. *An Empirical Comparison Study of Text-Based and Game-Based Security Policy Training*. To appear in the Mensch und Computer (MuC) conference. 2026.
- [P7] **Tarini Saka**, Kami Vaniea, and Nadin Kökciyan. 2026. *AI-Enhanced Email Security: A Novel Pipeline for Phishing Campaign Detection and Profiling*. To appear in ACM Transactions on Internet Technology (TOIT) Special Issue on Human-AI Collaboration in Security Operations Centres. Manuscript ID TOIT-2025-0128.
- [P8] Agata Stanczyk, Mindy Tran, **Tarini Saka**, Yixin Zou, and Veelasha Moonsamy. 2026. *Revealed or Reinforced: How Assistive Technologies Shape the Experience with Dark Patterns for Blind and Low-Vision Users*. In Proceedings of the ACM Designing Interactive Systems Conference (DIS). 2026.
- [P9] **Tarini Saka**, Nadin Kokciyan, Kami Vaniea. 2025. *SoK: Grouping Spam and Phishing Email Threats for Smarter Security*. IEEE Access, 13, 54938–54959. <https://doi.org/10.1109/ACCESS.2025.3555157>.

- [P10] **Tarini Saka**, Kalliopi Vakali, Adam Jenkins, Nadin Kokciyan, Kami Vania. 2025. [Judging Phishing Under Uncertainty: How Do Users Handle Inaccurate Automated Advice?](#) In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25). Association for Computing Machinery, New York, USA. [ Honourable Mention for top 5% Submissions.]
- [P11] **Tarini Saka**, Kami Vania, Nadin Kökciyan. 2024. [PhishCoder: Efficient Extraction of Contextual Information from Phishing Emails](#). In: Computer Security. ESORICS 2024 International Workshops. ESORICS 2024. Lecture Notes in Computer Science, vol 15264. Springer, Cham.
- [P12] **Tarini Saka**, Rachiyta Jain, Kami Vania, Nadin Kökciyan. 2024. [Phishing Codebook: A Structured Framework for the Characterization of Phishing Emails](#). arXiv preprint arXiv:2408.08967.
- [P13] **Tarini Saka**, Kami Vania, and Nadin Kökciyan. 2022. [Context-Based Clustering to Mitigate Phishing Attacks](#). In Proceedings of the 15th ACM Workshop on Artificial Intelligence and Security (AISec'22). Association for Computing Machinery, New York, NY, USA, 115–126.

PhD thesis

- [T1] **Tarini Saka**. [A Collaborative Human-AI Approach to Mitigate Phishing Attacks](#). The University of Edinburgh, Edinburgh, United Kingdom. 2024.