Shubham Malaviya

Pune, Maharashtra   smalaviya311@gmail.com   LinkedIn   Google Scholar

ABOUT

AI & Cybersecurity Researcher with 6+ years of experience building AI systems in Federated Learning, Data Quality, and Threat Intelligence. Published research at top venues (ACM CCS, SIGMOD, SAC, CoLLAs) with multiple patents as lead inventor. Skilled at translating cutting-edge AI research into actionable insights and currently driving innovation in software supply chain security.

EXPERIENCE (6.25 YEARS)

Scientist at Tata Consultancy Services Aug 2019 - Present

- Designed a **Federated Learning–based distributed training system** to enable privacy-preserving multi-client learning, **preventing raw data exposure** and **reducing communication** costs by **over 8×**.
- **Led user-behavior security research** to model impulsive click patterns, enabling context-aware phishing detection and **early identification of high-risk user behaviors**.
- Built an **automated plaintext credential detection system**, proactively identifying exposed passwords and passphrases to **reduce security risk** and **enforce enterprise policy compliance**.
- Identified and **corrected data-quality issues** in cybersecurity dataset, **improving the reliability of threat intelligence signals**.
- Identified and **scoped high-impact AI-for-cybersecurity use cases**, assessing feasibility, risk, and scalability to **deliver roadmap-aligned technical proposals.**
- Authored **technical POVs and concept evaluations** to support long-term strategic planning.
- Presented research progress and future development plans in annual technical **reviews with an external academic advisor**, incorporating feedback to **validate technical direction and guide next-phase work.**
- **Led hiring** for the research group by screening resumes and **conducting 25+ AI/Cybersecurity interviews**.

Paper review committee member, ACM/SIGAPP SAC 2023 - Present

PROJECTS

User Behavior Analytics on Impulsive Clicks | Chromium, Python | ACM CCS 2025

- Designed a **session-level user behavior modeling** framework to quantify impulsive clicking from browser telemetry, showing that **user impulsivity is situational and context-dependent**.
- Engineered **behavioral and contextual security signals** (e.g., impulsivity score, engagement, domain familiarity, entropy) to **differentiate risky vs. habitual behavior** using real-world Chrome data of 20 users.

Annotation Errors in Cybersecurity | Transformers, PyTorch | ACM CCS 2024

- Curated a Cyber Threat Intelligence (CTI) dataset, LADDER, **fixing up to 17% test-label errors** in NER and relation detection to **restore evaluation reliability** and expose flaws in widely used datasets.
- **Quantified the impact of label noise**, showing that test-set errors can cause **>40% F1 degradation** and **misleading model rankings**, directly affecting security decision-making.
- Assessed **LLMs (Phi-3, Gemini)** readiness for CTI tasks under varying supervision levels.

Toward Scalable Annotation Error Detection (AED) | vLLM, PyTorch | HILDA-SIGMOD 2025

- Demonstrated that **label noise** causes up to **90% degradation** in AED performance when training Transformer models (DistilBERT, BERT, RoBERTa) with standard cross-entropy loss.
- Improved AED robustness using LogitClip regularization and ϵ-softmax loss, **achieving 20–45% performance gains**.
- Reduced manual annotation review cost by ~99%, enabling **cost-effective, scalable data quality pipelines**.

Privacy Preserving AI | Python, TensorFlow, Flower | ACM SAC 2022

- Designed FedFAME, a **data-augmentation–free federated** semi-supervised learning framework combining **model contrastive learning** and knowledge distillation, enabling effective training when **clients hold only unlabeled**, non-IID data.
- Validated FedFAME across vision and NLP datasets (CIFAR-10, Fashion-MNIST, 20NewsGroups, 5Abstracts-Group), achieving state-of-the-art accuracy **without client-side data augmentation**.
- Demonstrated **robustness to heterogeneous data** distributions using Dirichlet-based non-IID splits.

Reducing Communication Overhead in FL | PyTorch, Transformers, OpenDelta, Flower | CoLLAs 2023

- **Reduced federated communication cost** for BERT-Base by **>8× (420 MB → <50 MB)** using **PEFT** (parameter efficient finetuning) methods while preserving model quality.
- Evaluated benchmarked **PEFT techniques (Prefix-Tuning, Adapters, BitFit, LoRA)** on **BERT** across six GLUE tasks under non-IID client distributions.
- Showed **task-level transferability** of PEFT parameters enabling **zero-shot federated learning**.

Classify Enterprise Documents Based on the Content Sensitivity. | Python, TensorFlow, NLTK 2021

- Developed a text classification system for **sensitive data detection**, leveraging SOTA deep learning models to **enhance organizational data security**.
- Conducted comparative analysis of **LDA, BERT, and Hierarchical Attention Networks (HAN)** on the **WikiLeaks** dataset, identifying model strengths and trade-offs to optimize accuracy.

SKILLS

Cybersecurity	Data privacy, Phishing Detection, Secure ML, User Behavior Analytics, Threat Intelligence
AI/ML	NLP, LLM, Data-centric AI, Gen AI, Prompt eng, Agentic AI, PEFT, Robust Training
Distributed ML	Federated Learning, Non-IID Data, Communication-Efficient Training
Tech Stack	Python, PyTorch, Transformers, TensorFlow, vLLM, Flower
Soft Skills	Research Design & Validation, Critical Thinking, Technical Writing, Continuous Learning

ACHIEVEMENTS

- **TCS citation & Intellectual Property (IP) creation Award** for AI-driven cybersecurity innovation
- **Review Committee Member**, ACM/SIGAPP SAC, since 2023
- **Lead Inventor** on Multiple Patent Filings across geographies.
- **Publications at Top-Tier Venues**: ACM CCS, SIGMOD, SAC, CoLLAs
- **Service & Commitment Award** for sustained technical and organizational contributions
- **Google Cloud Skill Badge:** Generative AI Fundamentals & Vertex AI Prompt Design
- **SONY Certificate of Appreciation** for selection among Top Exhibited Photographs (2022).

EDUCATION

M.Tech – Dhirubhai Ambani Institute of Information and Communication Technology, Gujarat 2019
Specialization in Machine Learning (CGPA: 8.45)

B.E., I.T – L.D. College of Engineering, Gujarat (CGPA: 8.41) 2017

ADDITIONAL PROJECTS & PUBLICATIONS

- Passphrase Detection (Patent)
- Password Detection (Patent)
- PEFT in Cybersecurity (Paper)
- Influence-based Defense (Paper)
- Pdf Annotation Extraction