Pujan Paudel

pujanpaudel.com ppaudel@bu.edu 725-724-0359 codepujan

Education

- Ph.D. in Computer Engineering, Cumulative GPA: 3.93 *Expected May, 2025*
  Boston University (BU) *Boston, MA*
  Research Area: Content Moderation, Natural language Processing (NLP), Machine Learning (ML), Information Retrieval (IR), Cybersecurity
  Advisor: Dr. Gianluca Stringhini

- Bachelor of Science (Honours) in Computer Science, Cumulative GPA: 3.75 *May 2020*
  The University of Southern Mississippi (USM) *Hattiesburg, MS*
  Honours Thesis Title: *The Blind Spot of Twitter Bot Moderation*
  Advisor: Dr. Andrew Sung

Work Experience

Boston University *Boston, MA*

- Ph.D. Researcher at Security Lab (SeclaBU) *September 2020 - Present*
  - Developed a SOTA *Contrastive Textual Deviation* approach leveraging FLAN-T5 LLM for unsupervised stance detection, reducing false positives (20 to 2.1%) of automated soft moderation [1].
  - Developed an end-to-end reverse image search system for million-scale social media images using perceptual hashing and Milvus, improving visual soft moderation (by 13 times) on Twitter [2].
  - Implemented a *Learning To Rank* (LTR) based keyword extraction system to automate soft moderation in a dataset of 7M tweets, improving textual soft moderation (by 20 times) on Twitter [3].
  - Currently developing a real-time scanner to proactively discover fraudulent e-commerce websites by using streams of data from social media activity, search engines, and indicators of domain trustworthiness.
  - Currently experimenting with multi-modal embeddings (vision-language architectures) to address the challenges of cross-topic and cross-platform contextual content moderation.

- Graduate Teaching Fellow for Introduction to Software Engineering Course *September 2021 - May 2022*
  - Developed assignments, exams, and lab sessions for course topics such as Assembly programming and Object Oriented Programming in C++.
  - Mentored students in developing portfolio projects using multiple technologies (e.g. android development, game development, and web apps).

The University of Southern Mississippi *Hattiesburg, MS*

- Research Assistant *Jan 2017 - May 2020*
  - Conducted research in Twitter social bots using tools from topic modeling, network science, and information diffusion [8,9].
  - Built conversational agents tool using Amazon Alexa to assist the USM Psychology department in studying the feasibility of voice assistant technology in retention experiments of single-digit mathematical calculations for cognitively weaker children.

Technical Skills

- **Programming:** Python, C/C++, Javascript, Java, C#, Bash Scripting, R
- **Machine Learning:** Scikit-learn, PyTorch, Pandas, WandB, OpenCV, SciPy, NetworkX
- **Technologies:** ElasticSearch, Lucene, AWS, Docker, Flask, Spark, React.js, Django, Node.js, Github, AWS SageMaker, Google Cloud Platform
- **Database:** Milvus, Cassandra, MongoDB, SQL, PostgreSQL
- **Concepts:** Web crawling, Data mining, Exploratory Data Analysis, Generative AI, API, Transformers, Computer Vision, Large Language Models, Topic Modeling, Vector database, Big Data, Cloud Computing

Publications

[1] **P. Paudel**, M.H. Saeed, R. Auger, C. Wells and G. Stringhini, "Enabling Contextual Soft Moderation on Social Media through Contrastive Textual Deviation," 33rd Usenix Security Symposium, Philadelphia, PA, USA, 2024.

[2] **P. Paudel**, C. Ling, J. Blackburn and G. Stringhini, "PixelMod: Improving Soft Moderation of Visual Misleading Information on Twitter," 33rd Usenix Security Symposium, Philadelphia PA, USA, 2024.

[3] **P. Paudel**, J. Blackburn, E. De Cristofaro, S. Zannettou and G. Stringhini, "Lambretta: Learning To Rank For Twitter Soft Moderation," 2023 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 2023.

[4] N. Toraif, N. Gondal, **P. Paudel** and A. Frisellaa, "From colorblind to systemic racism: Emergence of a rhetorical shift in higher education discourse in response to the murder of George Floyd," PLoS one 18.8 (2023): e0289545.

[5] M. Singhal, C. Ling, **P. Paudel**, P. Thota, N. Kumarswamy, G. Stringhini, and S.Nilizadeh "SoK: Content Moderation in Social Media, from Guidelines to Enforcement, and Research to Practice," 2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P), 2023.

[6] **P. Paudel**, J. Blackburn, E. De Cristofaro, S. Zannettou and G. Stringhini, "A longitudinal study of the Gettr social network," International Workshop on Cyber Social Threats, 2022.

[7] **P. Paudel**, J. Blackburn, E. De Cristofaro, S. Zannettou and G. Stringhini, "Soros, child sacrifices, and 5G: understanding the spread of conspiracy theories on web communities," arXiv preprint, 2021.

[8] **P. Paudel**, TT. Nguyen, A. Hatua and AH. Sung, "How the tables have turned: Studying the new wave of social bots on Twitter using complex network analysis techniques," 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Vancouver, Canada, 2019.

[9] **P. Paudel**, TT. Nguyen, A. Hatua and AH. Sung, "User Level Multi-feed Weighted Topic Embeddings for Studying Network Interaction in Twitter," Big Data–BigData 2019: 8th International Congress, San Diego, CA, USA, 2019.

Additional Projects

- Profiling climate change misinformation on Reddit *January 2023 - May 2023*
  - ➪ Identified and measured the longitudinal evolution of climate skepticism claims across subreddits using Structural Topic Modeling (STM)
- ARPA Earmarks Analysis *March 2022 - May 2022*
  - ➪ Built a custom Named Entity Recognition (NER) model using Spacy to automatically infer policy buckets from ARPA amendment language to further analyze any disproportionate distribution of earmarked funding
- Scaling Remote Sensing Data Processing With Ray *March 2022 - May 2022*
  - ➪ Setup OpenTelemetry and Jaeger in Mass Open Cloud (MOC) for distributed profiling and finding bottlenecks on a NASA-JPL remote sensing application, proposing a new parallelization scheme with 3x speedup
- Cyberwarfare: Longitudinal Trends and Effects on Foreign Policy *May 2021 - July 2022*
  - ➪ Crawled, compiled, and curated a dataset of state-sponsored cyber attacks from three different data sources to analyze how cyber-severity of future attacks changes as an effect of policy actions between rival countries

Awards and Honors

- Pardee Center Graduate Summer Fellowship, **BU** *2021*
- Distinguished ECE PhD Fellowship, **BU** *2020*
- Runner Up, Undergraduate Research Symposium, **USM** *2019*
- Best Innovation Application, **CalHacks 2016** *2016*
- Best IBM Watson Hack, **HackRice 2016** *2016*