Nikolaos Pantelaios

https://github.com/npantelaios https://scholar.google.com/citations?user=kdp9jEEAAAAJ

Applied Scientist with a PhD and 8+ years in Natural Language Processing, Code Generation and Security, combining extensive research and industry experience. Developed scalable models at Meta, saving $2.48 million annually. Published in top-tier conferences like ACM CCS and Usenix Security Symposium. Expertise in code generation, security analysis, cloud computing, and multi-modal AI. As a Software & Machine Learning Engineer, spearheaded complex research projects and met tight deadlines, with a strong focus on applying theoretical foundations of neural networks to practical, real-world scenarios.

EDUCATION

PhD, North Carolina State University, Computer Science Aug 2024 (Expected)
Research Interests: Code Generation, Natural Language Processing, Systems, Security, Privacy

BSc & MSc, National Technical University of Athens, Computer Engineering 2018
Thesis: Personality Traits Recognition from Speech using Autoencoders

EXPERIENCE

Meta New York City, NY
Applied Scientist Intern May 2022 - Aug 2022

- Implemented scalable Word2vec models, saving $2.48 million annually in CPU and memory infrastructure costs, enhancing efficiency in large-scale ML and cloud operations
- Established end-to-end ML pipelines and extracted key metrics, improving automated data labeling accuracy by 2% across 1 trillion Facebook and Instagram database columns

ByteDance Mountain View, CA
Security Engineer Intern May 2021 - Aug 2021

- Classified 1 billion failed SSL/TLS TikTok certificates by security severity from untrusted sources, enhancing data privacy and security protocols
- Evaluated 2 petabytes of logs and SSL certificates to uncover new trends in failed SSL categories, improving infrastructure resilience and security measures

ByteDance Mountain View, CA
Security Engineer Intern May 2020 - Aug 2020

- Devised a comparison system to detect thousands of compromised accounts and bots in the TikTok user base, strengthening platform security and user privacy
- Examined 50 million accounts using Hive and Hadoop, optimizing database management and cloud infrastructure with Python and Golang

North Carolina State University Raleigh, NC
Research Assistant Aug 2018 - Aug 2024

- Led the authorship of four first-author publications, including two in prestigious top-tier conferences (ACM CCS, Usenix Security Symposium), with reproducible code and significant real-world impact
- Conducted research in ML, NLP, Code Generation, Security, Privacy, Networking, Infrastructure, multi-modal AI, utilizing Cloud Computing (AWS, Colab). Ensured reproducibility and availability in Python, C++, JavaScript, C, and Golang, while managing & coordinating research groups from 3 universities

Intelen Athens, Greece
Applied Scientist Jan 2016 - Jan 2018

- Created energy consumption prediction tools using Python, AWS, and machine learning models (LSTM, RNN), applying time-series prediction algorithms (XGBoost), resulting in a 20% processing efficiency

· Analyzed client data with Spark & Hive to optimize usage patterns, enhancing reliability for 500 clients
· Managed data integration & API design with Flask, PostgreSQL, boosting backend performance by 15%

PROJECTS

LLM JavaScript Deobfuscator (Python, JavaScript) 2023 - 2024

· Achieved state-of-the-art performance in code deobfuscation using LLMs and malicious JavaScript
· Fine-tuned Llama-2-70B, Deepseek-LLM-67B and Gemma-7B LLMs using LoRA and PEFT

Forced Execution Browser for Evasion Detection (JavaScript, C++) 2022 - 2023

· Enhanced Chromium's code execution by 11%, detecting 28 malicious evasion categories in Node.js
· Flagged malicious code in 110 Chrome extensions, impacting over 2 million users, unsupervised learning

Malicious JavaScript Generator using Transformers (Python, C, JavaScript) 2021 - 2022

· Coordinated a group of 4 people to generate malicious JavaScript sequences using Transformers & PyTorch
· Detected malicious code snippets in-the-wild in C & JavaScript using a combination of fuzzing techniques

Personality Prediction using Speech Recognition (Python, Autoncoders) 2016 - 2018

· Engineered and optimized speech recognition models using autoencoders and convolutional neural networks, leveraging speech data from thousands of unique speakers
· Architected and deployed tens of personality prediction models achieving state-of-the-art results using TensorFlow and Keras on multi-GPU systems

PUBLICATIONS

Nikolaos Pantelaios, Trevor Dunlap, Greg Tystahl, Brad Reaves, William Enck, and Alexandros Kapravelos, LLM JavaScript Deobfuscator. 2024

Nikolaos Pantelaios, and Alexandros Kapravelos, FV8: A Forced Execution JavaScript Engine for Detecting Evasive Techniques. In proceedings of the Usenix Security Symposium, 2024

Nikolaos Pantelaios, Nick Nikiforakis, and Alexandros Kapravelos. Manifest V3 Unveiled: Navigating the New Era of Browser Extensions. arXiv preprint arXiv:2404.08310, 2024

Nikolaos Pantelaios, Nick Nikiforakis, and Alexandros Kapravelos. You've Changed: Detecting Malicious Browser Extensions through their Update Deltas. In Proceedings, ACM, CCS, 2020

HACKPACK

Hacking Group (HackPack) 2018 - Present

· Championed the HackPack Community for 6 years, focusing on educational hacking courses in ML and security. Engaged in tens of CTF competitions, coding web and binary exploits (XSS, DoS, ROP)
· Engineered custom exploits for HackPack CTF and for learning purposes, including Netcat, SQL injection, and reverse engineering challenges, engaging 10,000 participants
· Formulated, structured, and authored 50 machine learning challenges, incorporating hacking techniques, prompt engineering, LLMs, organized the infrastructure and handled cloud computing resources

CERTIFICATES

IBM Machine Learning Professional Certificate: Completed IBM Certification in Machine Learning, Natural Language Processing, Speech Recognition and Computer Vision

SKILLS

Programming Languages: Python, C, C++, C#, JavaScript, Golang, R, Java

AI Tools: LLM, Transformers, PyTorch, Tensorflow, Keras, Theano, HuggingFace, Autoencoders

Technical: Cloud (AWS, Colab), PostgreSQL, MongoDB, Hive, Hadoop, Linux, Docker, Kubernetes